Rapid Evolution of Immune Proteins in Social Insects

By: Lumi Viljakainen, Jay D. Evans, Martin Hasselmann, Olav Rueppell, Salim Tingek, and Pekka Pamilo

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

Made available courtesy of Oxford University Press: http://www.oup.com/

*****Reprinted with permission. No further reproduction is authorized without written permission from Oxford University Press. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document.*****

Abstract:
The existence of behavioral traits connected to defense against pathogens manifests the importance of pathogens in the evolution of social insects. However, very little is known about how pathogen pressure has affected the molecular evolution of genes involved in their innate immune system. We have studied the sequence evolution of several immune genes in ants and honeybees. The results show high rates of evolution in both ants and honeybees as measured by the ratio of amino acid changes to silent nucleotide changes, the ratio being clearly higher than in *Drosophila* immune genes or in nonimmunity genes of bees. This conforms to our expectations based on high pathogen pressure in social insects. The codon-based likelihood method found clear evidence of positive selection only in one ant gene, even though positive selection has earlier been found in both ant and termite immune genes. There is now indication that selection on the amino acid composition of the immune-related genes has been an important part in the fight against pathogens by social insects. However, we cannot distinguish in all the cases whether the high observed d_N/d_S ratio results from positive selection within a restricted part of the studied genes or from relaxation of purifying selection associated with effective measures of behaviorally based colony-level defenses.
Key Words: ants, honeybees, molecular evolution, positive selection, immune system

Article:
During the course of evolution, social insects (ants, termites, social bees, and wasps) have faced the major challenge of managing the high risk of pathogen transmission due to close social contact of many individuals in large colonies (Schmid-Hempel 1998). Even though little studied outside the honeybee, *Apis mellifera*, numerous pathogenic viruses, bacteria, protozoans, and fungi have been described in social insects (Schmid-Hempel 1998). Social insects have evolved many colony-level defenses including hygienic behavior (Rothenbuhler 1964), mixing of antimicrobial compounds into nest material (Christe et al. 2003; Chapuisat et al. 2007), and increasing genetic diversity within colonies by multiple mating by the queen (e.g., Schmid-Hempel and Crozier 1999; Tarpy 2003; Hughes and Boomsma 2004; Seeley and Tarpy 2007). In certain taxa, immune defenses also include anatomical changes, such as the metapleural gland in ants (Beattie et al. 1986). The existence of these traits manifests the importance of pathogens in shaping the evolution of social insects.

As it is evident that pathogens have had a major impact on the evolution of social insect behavioral traits, it is startling that the genome of the honeybee *A. mellifera* was reported to contain substantially fewer of the innate immune genes when compared with the mosquito *Anopheles* and the fruit fly *Drosophila* (Honeybee Genome Sequencing Consortium 2006). The main components of the immune pathways are conserved, and the difference in the gene number stems from a smaller number of members in gene families covering all the stages of the pathways (recognition proteins, intracellular signaling molecules, and effector proteins; Evans et al. 2006). Diversification following gene duplication could be an important mechanism of improving the performance of the immune system, especially because insects lack adaptive immunity (such as somatic recombination of immune receptor genes and clonal expansion of activated lymphocytes), and thus the extent of memory and specificity characteristic of vertebrates. The relative paucity of immune gene repertoire in the

honeybee and the evolutionary success of bees and other social insects, despite high pathogen pressure, thus raise the question on the relative importance of physiological immunity and social immunity.

The molecular basis of the insect immune system has been studied in great detail in the fruit fly *Drosophila melanogaster*. Insect innate immune responses are in essence similar among insect taxa (Gillespie et al. 1997). Upon infection, innate defense mechanisms are activated comprising cellular and humoral responses. Phagocytosis by special blood cells, melanization, and encapsulation of large invaders constitute the cellular response. Humoral responses involve chains of events that begin in the hemolymph. The production of antimicrobial peptides is initiated via two distinct signaling pathways, Toll and Imd (reviewed by Ferrandon et al. 2007). Both pathways are triggered by the recognition of bacteria and fungi, Toll mainly by Gram-positive bacteria and Imd mainly by Gram-negative bacteria. There are two types of recognition proteins: PGRPs (peptidoglycan recognition proteins) and GNBPs (Gram-negative bacteria-binding proteins). The *Drosophila* genome encodes several PGRPs (Werner et al. 2000) of which some have enzymatic activity (called catalytic PGRPs, these do not function in bacterial recognition), whereas others have lost that function and operate solely in recognizing microbes (Mellroth et al. 2003; Bischoff et al. 2006). The recognition phase is followed by a complex signaling cascade culminating in the expression of genes encoding antimicrobial peptides primarily in the fat body tissue (Hoffmann 2003).

In a recent study, Sackton et al. (2007) identified 245 immune-related genes in *D. melanogaster* and showed in a comparison of six species of the *melanogaster* group that positive selection seems to have driven the evolution of many of them, and has been especially strong in recognition proteins. Signs of positive selection have also been found in *Drosophila* signaling pathway molecules (Schlenke and Begun 2003; Jiggins and Kim 2007). On the contrary, no evidence of positive selection has been found in antimicrobial peptide genes in *Drosophila* (Lazzaro and Clark 2003; Jiggins and Kim 2005; Sackton et al. 2007). In *Anopheles*, studies of five immune genes, including two antimicrobial peptides, point to the absence of positive selection and even to strong purifying selection in antibacterial peptides (Simard et al. 2007; Parmakelis et al. 2008).

The molecular evolution of immune genes is especially intriguing in social insects for two lines of arguments that result in opposing predictions. First, living in crowded thermoregulated colonies in which individuals are often closely related, creates a setting for efficient pathogen growth and transmission. This leads to the expectation that positive selection caused by high pathogen pressure is common in the immune genes of social insects. Second, although the key components of immune pathways are conserved in honeybees, the genome does not contain as many members in immune-related gene families as *Drosophila* and *Anopheles* do. In this context, it has been suggested that honeybees do not rely on individual immune defenses, but instead, the colony-level defenses play a major role in disease suppression (Evans et al. 2006). This could reduce the selective pressure caused by pathogens on the innate immunity.

Support for the first argument, that positive selection is common in immune genes of social insects, is provided in studies of termites. Bulmer and Crozier (2004) found positive selection in one copy of an antifungal peptide termicin that has been duplicated several times. These authors suggested that duplication and divergence of termicin loci may be associated with changes in the fungal pathogen spectrum accompanying speciation. Bulmer and Crozier (2006) have also investigated the evolution of two pathogen recognition proteins (GNBPs) and a transcription factor Relish that is part of the intracellular signaling cascade of insect immunity, and found positive selection in all the three genes. The selection pressure seemed to be highest in Relish, and based on this and the finding of positively selected sites in the regions important in its activation, the authors suggested that a host–pathogen arms race is driving positive selection in Relish (Bulmer and Crozier 2006).

Our aim here is to test to what extent social insect immune systems have been affected by selection. For this aim, we estimate rates of nucleotide substitutions in several immune genes in the honeybees (nine genes in five species) and ants (four genes in 12 species). The immune genes were chosen along the well-characterized Toll and Imd pathways of the insect innate immune system, and they represent the three stages of the pathways: recognition of pathogens, intracellular signaling, and production of effector proteins. Possible role of selection is

inferred on the basis of the overall ratio of nonsynonymous-to-synonymous substitutions, a comparison with the substitution rates in nonimmunity genes and estimates on substitutions within codons. Similarly to an earlier study on the evolution of an antimicrobial peptide defensin in ants (Viljakainen and Pamilo 2008), the present results show that the immune genes studied have a high ratio of d_N/d_S, much higher than in *Drosophila*, even though direct evidence for positively selected sites is scarce.

Materials and Methods

Immune Gene Identification

The ant immune genes used in this study were identified from *Myrmica ruginodis* using degenerate primers in 3' rapid amplification of cDNA ends (RACE), except for defensin, for which a primer Def-3 described by Viljakainen and Pamilo (2005) was used. Several degenerate primers were designed for 14 immune genes along Toll and Imd pathways based on conserved protein blocks between *A. mellifera*, *D. melanogaster*, and protein sequences deduced from the *Solenopsis invicta* expressed sequence tag (EST) database (Fourmidable Ant Sequence Database) using the program CODEHOP (Rose et al. 199). The immune system of *M. ruginodis* ants collected from Kiiminki, Finland, was activated by pricking the abdomen with a needle. After 18 h, total RNA was isolated from a pool of 20 individuals using TōTALLY RNA Kit (Ambion Inc., Austin, TX), and from that, cDNA synthesized with SuperScriptII Reverse Transcriptase (Invitrogen, Carlsbad, CA). The degenerate primers were used in 3'RACE (First Choice RLM-RACE Kit, Ambion), and the obtained polymerase chain reaction (PCR) products were used in 1.5% agarose gel electrophoresis. PCR products were purified from the agarose gel with the MinElute Gel Extraction Kit (Qiagen, Hilden, Germany) and cloned with the TOPO TA Cloning Kit (Invitrogen). Bacterial colonies grown overnight were suspended in 50 µl of sterile water and screened using PCR. The PCR products were purified with the MinElute PCR Purification Kit (Qiagen), and sequenced on both strands using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), and run on the ABI 3730 automated sequencer. Specific primers were designed on the basis of the 3'RACE products. These primers were used in 5'RACE (First Choice RLM-RACE Kit, Ambion) following the same procedures as described for 3'RACE above. The 3' and 5'RACE products were aligned and combined to a single sequence that was used in a BlastX search to check that the correct gene had been amplified. Functional domains were identified with the Conserved Domain Database of the National Center for Biotechnology Information and Prosite (Hulo et al. 2008), and signal peptides with SignalP 3.0 (Bendtsen et al. 2004). The degenerate primers and the gene-specific RACE-primers are available from the authors upon request.

Table 1
Gene-Specific PCR and Sequencing Primers for *Myrmica*

Locus and Primers	Primer Sequence (5′–3′)
PGRP	
MyrPGRP-PCR-F	GTACATCGTGACGACGACAACATT
MyrPGRP-PCR-R	TTGCAGTTTACGATCGCTCTGA
MyrPGRP-SEQ-1F	CATCATTTCTAGAGCGCAATGG
MyrPGRP-SEQ-2F	GAAAGGGTGGTCCGACATTG
MyrPGRP-SEQ-1R	CGAATGATGAATGATCACGTATG
MyrPGRP-SEQ-2R	ATCCGCGACCCTCGTAAAC
GNBP	
MyrGNBP-PCR-F	GTTACGCATTAAGCCGATCCTG
MyrGNBP-PCR-R	GCGGATACTGTTTCGTGGAGAC
MyrGNBP-SEQ-F	GCCTAAGAGACCATCGAGCA
MyrGNBP-SEQ-R	GCCAAACGATCCGTCTACAG
Abaecin	
Myr-abaecin-F	ATAAGTTGATCATATCATCCGTGA
Myr-abaecin-R	TCATTTTCCAATACACGACATC
Defensin	
MyrDef-F	GCGAACGTTCAGCATTCCTT
MyrDef-R	CAAAATTAACCAGCCCATGACTCTTA

Sequencing of Immune Genes

For the ant immune genes, PCR and sequencing primers were designed based on the 3' and 5'RACE products using the program Primer3Plus (Untergasser et al. 2007). The gene-specific primers are listed in table 1. Each gene was amplified from 10 *Myrmica* species, and if possible, from the outgroup *Manica rubida* (table 2) and sequenced as described above. Additionally, for each identified *Myrmica* immune gene, the corresponding

cDNA sequence of *S. invicta* was obtained from the *Solenopsis* EST database (Fourmidable Ant Sequence Database) by using Blast searches.

Table 2
Ant Samples and Sequenced Genes

Sample Code	Species	Location	*Defensin*	*Abaecin*	PGRP	GNBP
MY1	*Myrmica ruginodis*	Kiiminki, Finland	x	x	x	x
TV532	*Myrmica lobicornis*	Hanko, Finland	x	x	x	x
MY2	*Myrmica scabrinodis*	Kiiminki, Finland	x	x	x	x
MY7	*Myrmica sulcinodis*	Kiiminki, Finland		x	x	x
Pors36	*Myrmica rubra*	Hanko, Finland	x	x	x	x
TV529	*Myrmica sabuleti*	Hanko, Finland	x	x	x	x
Mrug1	*Myrmica rugulosa*	Helsinki, Finland	x	x	x	x
Mdetri24	*Myrmica detritinodis*	Canada	x	x	x	x
Mbrevi29	*Myrmica brevispinosis*	Canada	x	x	x	x
Malas31	*Myrmica alaskensis*	Canada		x	x	
ManicaE8	*Manica rubida*	Switzerland		x		x
			8	11	10	10

Honeybee specimens of the species *Apis cerana*, *Apis koschevnikovi*, *Apis dorsata*, and *Apis andreniformis* were collected in March 2007 at the Agricultural Research Station in Tenom (Sabah, Borneo, Malaysia) and directly stored in RNAlater (Ambion). Multiple individuals (10–50) from a single nest were collected per species. The head, thorax, and abdomen were separated and stored at –80 °C after transport until RNA extraction and cDNA synthesis. RNA was extracted from the head, thorax, and abdomen using the TōTALLY RNA Kit (Ambion). cDNA was synthesized with SuperScriptII reverse transcriptase from head and thorax (Invitrogen). PCR primers for several immune genes along Toll and Imd pathways were designed based on *A. mellifera* genomic sequences and if these did not amplify the target in some of the species, new primers were designed based on the already obtained products in other species (table 3). The obtained PCR products were sequenced as described above.

Table 3
Gene-Specific Primers for *Apis*

Locus and Primers	Primer Sequence (5′–3′)	Partial 5′ End (bp)[a]	Partial 3′ End (bp)[a]
PGRP-SA			
PGRP-SA-F	TCTCTGTGAACGAAGATGACCAA	49	44
PGRP-SA-R	CGAATTCACCCGATTTTTCTCA		
PGRP-SC2			
PGRP-SC2-F	GCTGAGCGGAGATGAAAATTG	97	
PGRP-SC2-R	TTCAGGCCAATTTTGAATCTGT		
GNBP1-1			
GNBP4-F	GGATTTATATACAAGTGAATGAAACAATG	3	62
GNBP4-R	CCAAATTCTTTGCCATCCACAA		
GNBP1-2			
GNBP9-F	GTCCAACGTGTCGTATTAAGAGAAT	96	159
GNBP9-R	ATCCGACATTTCTCCAAGGTT		
Cactus			
cact-F	CACGAATCGAATATCGAAATCAAA	60	54
cact-R	TGGATCAATGAGGGCTTTTACC		
Imd			
Imd_34F	CACATGTTAACGACCGATGC	243	66
Imd_774R	ATTTTTGCGAACGGATACCA		
Imd_183F	GAGCGATGCAATACCTGTCA		
Imd_781R	ACATCAAATTTTTGCGAACG		
Relish			
relish_52F	GCAGTGTTGAAGGAGCTGAA	162	138
relish_1250R	TGAACACATTTCGTTTGTTGTTT		
relish_168F	GGACGCTTTTCAGAATTGGA		
relish_749R	TGGATCAGCTCCAAGCATAA		
relish_666F	TCTGGAAAAACTCCATTGCAT		
relish_1298R	CCAATTTATCTCTGAATCTTCATCC		
Hymenoptaecin			
Hymen-F	CGATCAGCTCTACAGACAAGAGCA		56
Hymen-R	GTGAAAGCCTGCCACCAGGAC		
Defensin			
def1-F	TATTGTCGGCCTTCTCTTCATGG	52	
def1-R	TTTTTCATTTCTTAATCATTTCGCGTTT		

[a] The obtained gene sequences lacked some base pairs either from the 5′ end, the 3′ end, or both as indicated.

Sequence data from 11 nuclear and 2 mitochondrial honeybee nonimmunity genes were used as a comparison

for the immune genes. A fraction of the coding region (471 bp) of elongation factor EF-1α was amplified from the honeybee samples with primers forward: 5'-TGTGGAAATTCGAAACGTCA-3' and reverse: 5'-CGGAGAGCCTTGTCTGTAGG-3'. Sequence data representing the five species used in this study were retrieved from the GenBank for a nonimmunity gene,1,4,5-trisphosphate receptor (*itpr*) (accession numbers DQ468666 , DQ468659 , DQ468664 , DQ468661 , and DQ468657) and the mitochondrial genes for NADH dehydrogenase subunit 2 (*nad2*) (AY712666 , AY712672, AY712673, AY712677, and AY712681) and cytochrome oxidase subunit II (*cox2*) (AY588417 , AY587544, AY588415, AY587546, and AY587542). In addition, sequences of the following nuclear genes of *A. mellifera*, *A. cerana*, *A. dorsata*, and *Apis florea* were retrieved from the GenBank: rudimentary (*CAD*) (EU184806 –EU184808 and XM_393888), Ca^{2+}/calmodulin-dependent protein kinase (*CamKII*) (EU184793–EU184795 and AB013287), glycerol kinase (*gyk*) (EU184763–EU184765 and XM_392782), white (*w*) (EU184851–EU184853 and XM_001122252), sodium–potassium ATPase (*NaK*) (EU184748 –EU184750 and XM_623142), RNA polymerase II (*polII*) (EU184731–EU184733 and XM_623278), wingless (*wg*) (EU184714–EU184716 and XM_396946), long-wavelength rhodopsin (*LWRh*) (EU184838 , EU184839 , AY267162 , and AF091732), and arginine kinase (*ArgK*) (EU184831 , EU184832, AY267178, and EF032397).

Statistical Analyses

All the alignments were done with ClustalX version 2.0.8 (Thompson et al. 1997) and adjusted manually with GeneDoc (Nicholas et al. 1997) using the help of the deduced protein alignments. Synonymous and nonsynonymous substitution rate estimates (d_S and d_N) from the pairwise coding sequences for each gene were estimated with MEGA version 4 (Kumar et al. 2004) using the method of Pamilo and Bianchi (1993) and Li (1993).

No published phylogeny exists for the 10 investigated *Myrmica* species and therefore the phylogenetic tree of the *Myrmica* ants using *M. rubida* as an outgroup was reconstructed using partitioned data of the four identified immune genes including introns (*PGRP-SC2*, *GNBP1*, *defensin*, and, *abaecin*) with MrBayes version 3.1 (Ronquist and Huelsenbeck 2003). A nucleotide substitution model was chosen for each gene separately with the programs MrModeltest version 2.2 (Nylander 2004) and PAUP* version 4.0b10 (Swofford 2002). In MrBayes, four chains were run simultaneously for 1,000,000 generations with trees sampled every 100 generations. The sampled log likelihood values were plotted against the generations to verify stationarity of the chains, and based on that, 40% of the trees were excluded as burn-in. The stationarity in time of the nucleotide frequencies in the sequences was ascertained using the program Tree-Puzzle (Schmidt et al. 2002).

The nucleotide substitution rates in different ant and honeybee lineages were compared using a relative rate test in the program RRTree (Robinson-Rechavi and Huchon 2000). This method is phylogeny based, and for honeybees, the phylogeny reconstructed by Raffiudin and Crozier (2007) was used and for ants the phylogeny generated with MrBayes.

In order to detect positive selection, a maximum likelihood method based on a codon substitution model (Goldman and Yang 1994) implemented in a program codeml in the PAML 4 package was used (Yang 2007). The data were fitted to four likelihood models, M1, M2, M7, and M8. Model M1 is a neutral model which divides the codon sites into two categories, one having conserved sites with $\omega_0 = 0$ ($\omega = d_N/d_S$ = nonsynonymous–synonymous rate ratio) and the other involving neutral sites with $\omega_1 = 1$ (Nielsen and Yang 1998). Model M2 allows an additional category of sites with ω_2 estimated from the data and thus involving positively selected sites. Models M7 and M8 assume ω to follow a beta distribution with the shape parameters estimated in the interval (0, 1). The model M8 includes one additional category to account for positively selected sites (Yang et al. 2000). In these models, d_S is constant among sites, whereas d_N is variable (Nielsen and Yang 1998). The analyses were run twice, first with initial $\omega > 1$ and then initial $\omega < 1$ to avoid the model M8 being trapped at a local optimum (Anisimova et al. 2002). A likelihood ratio test was used to determine the model that best fits the data in two comparisons: M1 versus M2 and M7 versus M8. Swanson et al. (2003) have noted that if the beta distribution does not describe well the true distribution of ω, comparison of models M7 and M8 may result in false positives. Therefore, if M7–M8 comparison indicated positive selection, we applied a

comparison of M8A versus M8, as described in Swanson et al. (2003). Here, the model M8A is a modified null model assuming ω to follow a mixture between a beta distribution and a point mass at $\omega = 1$. Positively selected sites were identified with a Bayesian approach implemented in the codeml program using a posterior probability cutoff value of 0.95. For both ant and honeybee phylogenies, the total tree lengths in synonymous substitutions per synonymous sites were estimated using model M0 in codeml.

Correction for multiple testing in relative rate tests and PAML analyses was done by using the false discovery rate method implemented in the program QVALUE (Storey and Tibshirani 2003).

Results

Ant Immune Genes

We successfully isolated four immune genes from *M. ruginodis*: a *PGRP*; a *GNBP*; and two antimicrobial peptide genes (*abaecin* and *defensin*). The gene sequences identified spanned the entire coding regions in all but one gene, *GNBP*, which lacked approximately 45% of the coding region based on comparison with *A. mellifera GNBP1-2* sequence. The function of *GNBP1*, *abaecin*, and *defensin* is known from studies of other insects, and as the ant sequences shared high levels of amino acid identity with the known proteins (43–59% according to BlastX searches), they most likely have the same function, though experimental tests are required to confirm this. PGRPs constitute a highly diversified gene family, and thus, we tried to determine whether the ant PGRP is an ortholog to any of the known proteins. We generated a phylogeny using Bayesian inference including the deduced ant protein sequence and all the known different PGRPs from *A. mellifera*, *D. melanogaster*, and *Anopheles gambiae* (total number of sequences included was 30). The ant PGRP clustered with PGRP-S2 from *A. mellifera* (GB19301), exclusively.

The four ant immune genes were sequenced from the genomic DNA of 8–10 *Myrmica* species and if possible, also from *M. rubida* (table 2, GenBank accession numbers FJ546118 [GenBank] –FJ546127 [GenBank] , FJ546137 [GenBank] –FJ546165 [GenBank]). The sequences covered the entire coding region only in abaecin, whereas *PGRP* sequences lacked 49 bp and *defensin* 32 bp of the 5' end of the coding regions and *GNBP* 27 bp from the start of the identified region. The ant phylogenetic tree generated by using the information from all four genes including intron sequences is shown in figure 1. The outgroup *Manica* is the closest relative to *Myrmica* ants (Astruc et al. 2004; Brady et al. 2006).

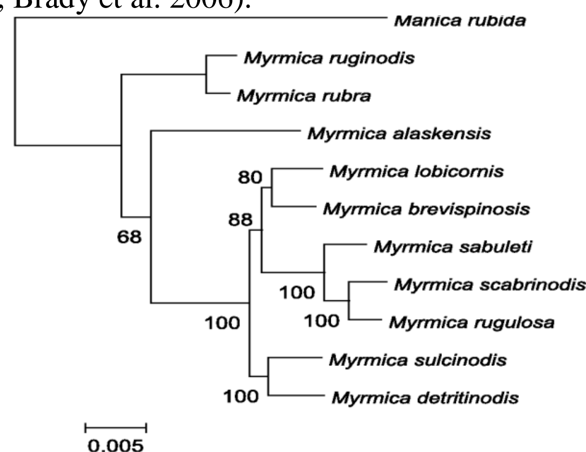

FIG. 1.—MrBayes phylogeny based on partitioned data of genomic sequences of four immune genes (*PGRP*, *GNBP*, *abaecin*, and *defensin* from *Myrmica* ants with posterior probability values shown in the nodes Scale bar is nucleotide substitutions per site.

Two sets of sequence data were used for the test of positive selection. The first data set included all the *Myrmica* sequences as well as the *Manica* sequences in *abaecin* and *GNBP*. The second data set additionally involved a sequence from a more distantly related *myrmicine* species, *S. invicta*, because including a more diverged sequence increases the amount of data and the number of nucleotide substitutions and thus improves the power of the test. Both data sets were analyzed using the entire length of the sequenced genes as well as using only the coding parts of the functional amidase domain of PGRP, and the mature peptides of *abaecin* and *defensin*. The sequenced part of *GNBP* was within the functional glyco-hydrolase domain. The tests of M1 versus M2 and M7 versus M8 gave essentially the same results, and thus only those of M7 versus M8 are shown (table 4). Positive

selection was indicated in *PGRP* both in the *Myrmica–Manica* data set and in the data involving *Solenopsis*, and these results were confirmed with a comparison of models M8A and M8 (table 4). Amino acid sites 1 P, 17 R, 44 I, 95 K, and 162 S were suggested to have been under selection, with the site 44 I being the only one consistently indicated by the different models.

Table 4
Likelihood Ratio Test to Detect Positive Selection with PAML in Ants[a]

Locus	N^b	N^c	Likelihood Ratio	df	P Value	q Value[d]	Global d_N/d_S	Tree Length
PGRP	168	10	16.488	2	2.63×10^{-4}	0.002	0.522	0.111
Amidase domain	144	10	15.781	2	3.74×10^{-4}	0.002	0.564	
GNBP	240	10	2.002	2	0.37	0.33	0.500	0.146
Abaecin	52	11	5.923	2	0.05	0.09	0.400	0.121
Mature region	33	11	3.960	2	0.14	0.18	0.300	
Defensin	77	8	0.250	2	0.88	0.72	0.500	0.053
Mature region	40	8	0	2	1	0.76	0.500	
Results including *Solenopsis invicta*								
PGRP	168	11	7.990	2	0.02	0.05	0.669	0.500
Amidase domain	144	11	9.294	2	0.01	0.04	0.815	
GNBP	240	11	6.221	2	0.05	0.09	0.557	0.619
Abaecin	52	12	3.181	2	0.21	0.21	0.400	0.353
Mature region	33	12	2.997	2	0.22	0.21	0.300	
Defensin	77	9	5.095	2	0.08	0.12	0.506	0.974
Mature region	40	9	2.992	2	0.22	0.21	0.400	

[a] Results of comparison of models M7 versus M8 are shown.
[b] Number of analyzed codons, excluding those containing gaps.
[c] Number of taxa.
[d] False discovery rate, q values ≤ 0.05 are considered significant indicating that model M8 is preferred.
[e] Total tree length in synonymous substitutions per synonymous sites obtained from model M0.

Pairwise synonymous and nonsynonymous substitution rate estimates among the *Myrmica* species gave an average d_S value of 0.027 and d_N value of 0.014 calculated over all the four immune genes (a total of 538 codons). The individual pairwise values had a large error estimate because of a small overall number of substitutions. The divergence between the *Myrmica* species and the outgroup *Manica* was calculated as an average of the pairwise distances in abaecin and GNBP (in total, 292 codons) resulting in $d_S = 0.075$ and $d_N = 0.038$. In the *Myrmica–Manica* data set, the global d_N/d_S estimates obtained from PAML varied from 0.300 in the mature part of *abaecin* to 0.564 in the amidase domain of *PGRP* (table 4). The median of the d_N/d_S ratios for the whole genes was 0.500. The relative rate test indicated, after the correction for multiple testing, a significantly higher nonsynonymous substitution rate in *defensin* of *Myrmica lobicornis*, *Myrmica brevispinosis*, and *Myrmica detritinodis* compared with *M. ruginodis* and *Myrmica rubra*.

Honeybee Immune Genes

The following list of eight immune genes was amplified and sequenced from cDNA of the honeybee species *A. mellifera*, *A. cerana*, *A. koschevnikovi*, *A. dorsata*, and *A. andreniformis*: *PGRP-S2*, *PGRP-SA*, and *GNBP1* in linkage group 4 (hereafter denoted as *GNBP1-1*), *GNBP1* in linkage group 9 (hereafter denoted as *GNBP1-2*), *cactus*, *Imd*, *Relish*, and *hymenoptaecin*. One nonimmunity gene (EF-1α) was amplified from all the species listed above. Additionally, *defensin* was amplified for all species mentioned except for *A. koschevnikovi* due to problems in amplification. The sequenced regions did not cover the entire coding parts of the genes but lacked a few codons in either the 3' or 5' end, or both (table 3). The sequences are in the GenBank with accession numbers FJ546093 [GenBank] –FJ546117 [GenBank] , FJ546128 [GenBank] –FJ546136 [GenBank] , and FJ546166 [GenBank] –FJ546175. PGRP-SA and the GNBPs are recognition proteins, cactus, IMD, and Relish are intracellular signaling proteins and hymenoptaecin and defensin are antimicrobial peptides. The honeybee *PGRP-S2* has been shown to be upregulated after immune challenge (Evans et al. 2006), but the exact function has not been verified.

In the honeybee, the tests for the detection of selection were done for the entire sequenced parts of the genes as well as for the functional domains or the mature peptides in the case of the antimicrobial peptides. The results concerning positive selection were the same for the whole genes and the functional regions and also using the different model comparisons. Evidence of positive selection was found in *GNBP1-2* (table 5). However, this gene contains a region with several closely located indels bringing uncertainty in the alignment. Notably, when this gene was reanalyzed excluding the region containing indels (42 bp), positive selection was not detected

(table 5). As the short region including indels had a major effect on the results, all the analyses of *GNBP1-2* were based on alignment from which the 42-bp indel region was removed.

Table 5
Likelihood Ratio Test to Detect Positive Selection with PAML in Honeybees[a]

Locus	N[b]	N[c]	Likelihood Ratio	df	P Value	q Value[d]	Global d_N/d_S	Tree length[e]
Immune genes								
PGRP-SA	145	5	2.903	2	0.23	1	0.203	0.415
Amidase domain	135	5	1.850	2	0.40	1	0.200	
PGRP-S2	162	5	0.513	2	0.77	1	0.239	0.288
Amidase domain	140	5	0	2	1	1	0.210	
GNBP1-1	287	5	1.518	2	0.47	1	0.200	0.276
Beta-GRP domain	265	5	0.141	2	0.93	1	0.200	
GNBP1-2	390	5	47.778	2	4.2×10^{-11}	1.4×10^{-9}	0.867	
GNBP1-2[f]	385	5	2.092	2	0.35	1	0.257	0.373
Beta-GRP domain	251	5	1.874	2	0.39	1	0.241	
Cactus	347	5	3.291	2	0.19	1	0.294	0.270
Ankyrin repeat region	117	5	0	2	1	1	0.230	
Relish	509	5	3.617	2	0.16	1	0.194	0.348
Ankyrin repeat region	285	5	0.832	2	0.66	1	0.208	
IPT domain	55	5	0	2	1	1	0.065	
Imd	155	5	0.084	2	0.96	1	0.318	0.260
Death domain	52	5	1.335	2	0.51	1	0.225	
Hymenoptaecin	110	5	0	2	1	1	0.299	0.475
Mature region	91	5	0	2	1	1	0.304	
Defensin	78	4	3.973	2	0.14	1	0.300	0.298
Mature region	52	4	0	2	1	1	0.272	
Nonimmunity genes								
EF1-alpha	157	5	0	2	1	1	0	
Itpr[g]	321	5	0	2	1	1	0.022	
nad2[g]	126	5	2.067	2	0.36	1	0.009	
cox2[g]	225	5	0	2	1	1	0.010	
CAD[g]	193	4	0	2	1	1	0.012	
CamKII[g]	92	4	0	2	1	1	0.076	
gyk[g]	139	4	0	2	1	1	0.036	
NaK[g]	472	4	0	2	1	1	0.001	
polII[g]	282	4	0	2	1	1	0	
LWRh[g]	151	4	0.150	2	0.93	1	0.092	
w[g]	55	4	0.038	2	0.98	1	0.111	
wg[g]	129	4	0	2	1	1	0	
ArgK[g]	156	4	2.674	2	0.26	1	0.100	

[a] Results of comparison of models M7 versus M8 are shown.
[b] Number of analyzed codons, excluding those containing gaps.
[c] Number of taxa.
[d] False discovery rate, q values ≤ 0.05 are considered significant indicating that model M8 is preferred.
[e] Total tree length in synonymous substitutions per synonymous sites obtained from model M0.
[f] Excluding the region containing indels.
[g] Sequences are from the GenBank.

Pairwise sequence differences between different honeybee species showed in the immune genes the mean d_S ranging from 0.052 (*A. cerana* and *A. koschevnikovi*) to 0.096 (*A. cerana* and *A. dorsata*) averaged over all the nine genes from *A. mellifera*, *A. cerana*, *A. koschevnikovi*, and *A. dorsata*. The mean d_N ranged from 0.016 to 0.035. The distance of *A. andreniformis*, which was used as an outgroup in the relative rate test, to the other *Apis* species was $d_S = 0.101$ and $d_N = 0.046$.

The nonimmunity genes could be compared from four or five species. The mean distance between species pairs over the 11 nuclear genes had the ranges of $d_{S\ nuclear} = 0.075–0.093$ when the more distant species *A. andreniformis* and *A. florea* were not included. The mean distance of these latter two to the other *Apis* species were $d_{Snuclear} = 0.109$ and $d_{Nnuclear} = 0.004$. The values of d_S were very similar to those obtained from the immunity genes. For the mitochondrial genes, the respective mean values were $d_{SmtDNA} = 0.691$ and $d_{NmtDNA} = 0.107$ excluding *A. andreniformis* and *A. florea*, and $d_{SmtDNA} = 0.812$ and $d_{NmtDNA} = 0.123$ between these two and the rest of the *Apis* bees.

The global d_N/d_S ratios obtained from the PAML analyses varied for the immune genes from 0.194 in *Relish* to 0.318 in *Imd* with a median of 0.269 (table 5). The global d_N/d_S ratios, estimated separately for the functional regions of the genes, were approximately the same or slightly lower than the estimate for the whole genes (table 5). The 13 nonimmunity genes showed significantly lower d_N/d_S ratios ranging from 0 to 0.111. All the ratios for the nonimmunity genes were smaller than any of those in the immune genes, the difference being statistically significant ($P < 0.001$). Using *A. andreniformis* as an outgroup and after the correction for multiple testing, the relative rate test indicated that the rate of nonsynonymous substitutions differed significantly in *GNBP1-2* in the

comparisons between *A. dorsata* and *A. koschevnikovi* ($Q = 0.003$), *A. dorsata* and *A. cerana* ($Q = 0.003$), and *A. dorsata* and *A. mellifera* ($Q = 0.005$) in such a way that *A. dorsata* had a lower value in all the comparisons.

Discussion

Evolutionary Rates in the Immune Genes of Social Insects

Our study included nine honeybee immune genes (*PGRP-S2*, *PGRP-SA*, two *GNBPs*, *cactus*, *Imd*, *Relish*, *hymenoptaecin*, and *defensin*) and four ant genes (*PGRP*, *GNBP*, *abaecin*, and *defensin*) which very likely function in immune defense. The results showed high evolutionary rates at the protein level (as indicated by the d_N/d_S ratio) in both ants and honeybees. The estimates of d_N/d_S ratios were very uniform over the genes with no gene showing a clearly higher or lower ratio. The d_N/d_S ratio found in the immune genes can be compared with ratios in nonimmunity genes in these same taxa and with ratios found in the immune genes in *Drosophila*. In honeybees the median of global d_N/d_S ratios in the immune genes was significantly higher than the median of d_N/d_S ratios in the four nonimmunity genes. The average rate of synonymous substitutions in the pairwise comparisons was very similar in the immune genes and the nuclear nonimmunity genes, indicating that this difference was caused by a higher rate of nonsynonymous substitutions. Unfortunately, we do not have sequences of nonimmunity genes from the ant species included in this study to carry out a similar comparison.

The median of d_N/d_S ratios in both honeybees (0.269) and ants (0.500) were considerably higher than the median of d_N/d_S ratios in immune-related genes in *Drosophila* ($d_N/d_S = 0.080$) (Sackton et al. 2007). The *Drosophila* estimate is based on more than 200 genes, and the genes may vary substantially in their selective pressures. However, if we take into account only the *Drosophila* genes that are involved in our study (*PGRP-S2*, *PGRP-SA*, *GNBP1*, *cactus*, *Relish*, *Imd*, and *defensin*), the median is 0.110, which is significantly lower than in both honeybees and ants (Mann–Whitney *U*-test, $P= 0.001$ for honeybee–*Drosophila* and $P = 0.016$ for ant–*Drosophila* comparison). These results indicate that the immune-related proteins have evolved at a faster rate in social insects compared with *Drosophila*.

One hypothesis explaining elevated d_N/d_S ratios in social insects is based on their putatively small effective population sizes, which could allow fixation of slightly harmful mutations (Bromham and Leys 2005). However, Bromham and Leys found no consistent patterns supporting this prediction, even though the d_N/d_S ratio was slightly increased in some advanced social species. In our study, the d_N/d_S ratios in bee immune genes were significantly higher than in the nonimmunity genes suggesting that selection rather than low N_e could explain the rapid evolution of immune genes in bees. A small effective population size would increase the rate of slightly harmful amino acid substitutions in all genes and elevate the d_N/d_S ratios. Although the immune genes in the honeybees had significantly higher mean d_N/d_S than the *D. melanogaster* group flies, the same did not apply for nonimmunity genes. We estimated the d_N/d_S ratios for the same set of nonimmunity genes by using the *Drosophila* sequence data available from FlyBase and Genbank. The d_N/d_S ratio of the 13 nonimmunity genes used in our comparison (table 5) was similar in the honeybees (mean 0.036, median 0.012) and the flies (mean 0.045, median 0.034) (the ratios for the two mitochondrial genes were intermediate in both taxa). Therefore, it seems likely that the high d_N/d_S ratios in the immune genes of bees and ants reflect changes in selection. Interestingly, there are two very different possibilities. First, if social immunity (Cremer et al. 2007) is efficient, purifying selection on innate immunity could be relaxed and amino acid changes could be allowed. Second, if the pathogen pressure is especially severe in social insects, we would expect positive selection driving amino acid replacements. The latter hypothesis should lead to repeated amino acid replacements at selected sites, that is, $d_N/d_S > 1$ at those sites. The possibilities to detect this (and to distinguish between the hypotheses) are limited if the total tree length is short, if positive selection has affected only specific phylogenetic lineages, or if many sites are selected and selection per site is thus diluted. We will discuss our findings in the light of these alternatives.

The median of the d_N/d_S ratios in the ant immune genes was almost twice that of the value in honeybees, and this difference was significant (Mann–Whitney *U*-test, $P = 0.005$). This suggests that the immune proteins in ants have been evolving faster than those of the honeybee. A possible explanation for this could be that the pathogen pressure is higher in ants, a scenario that would support the hypothesis of positive selection on the

immune genes. Although our data are insufficient to prove this is the case, it would be consistent with the differences between ant and bee life histories. For example, Boomsma et al. (2005) propose that bees could have a lower risk of diseases compared with ants due to the difference in their diets—ants are omnivorous whereas bees forage on relatively hygienic nectar and pollen. Moreover, honeybee nests are constructed on trees and shrubs whereas ant nests are in soil and probably less hygienic, for example they are exposed to decaying organic material and many nest visitors that can transmit diseases.

Sequences of four antimicrobial effector genes (*abaecin*, *apidaecin*, *hymenoptaecin*, *and defensin*) are available also from the bumble bee Bombus ignitus (Choi et al. 2008). A comparison of these sequences with the honeybee genes from *A. mellifera* show d_N/d_S ratios ranging from 0.176 to 0.375. The sequence differences between *Apis* and *Bombus* are an order of magnitude larger than within the genus *Apis*, with d_S estimates ranging from 0.690 to 1.045. Saturation of synonymous sites would affect the results by somewhat elevating the d_N/d_S ratios. The estimated d_N/d_S ratios from the *Apis–Bombus* comparison fell within the range shown in the comparisons of different *Apis* species, and even though they may be slightly overestimated, they support the view that the immune-related proteins have also evolved rapidly in social bees.

Positive Selection in the Immune Genes of Ants and Honeybees

In order to examine the role of positive selection in the evolution of immune genes, a codon-based likelihood method was applied. Evidence of positive selection was found only in the ant PGRP. It should be noted that although we did not find positive selection in *defensin*, an earlier study covering formicine ants has indicated positive selection in *defensin* (Viljakainen and Pamilo 2008). In the honeybee, we detected positive selection initially in GNBP1-2, but after excluding a 42-bp region, which was difficult to align reliably due to several indels, no indication of positive selection remained. Thus, even though the d_N/d_S ratios are high in ants and bees, there is no particularly strong support for the notion that positive selection is typical in their immune genes. Detection of positive selection with PAML is reliant on repeated amino acid changes at single sites and sufficient amount of changes are required to accumulate for the test to be significant. It should be noted that in ants and honeybees, the average tree lengths for the immune genes based on d_S were 0.108 and 0.334, respectively (tables 4 and 5). In *Drosophila*, the corresponding average tree length was 1.611 (Sackton et al. 2007). Thus, the paucity of detectable positive selection in the immune genes of ants and honeybees could be due to short overall tree lengths.

Support for lineage-specific selective pressure in immune genes has been found in *Drosophila* in which positive selection was shown in Relish in the *D. melanogaster* lineage, but not in the other species of the *D. melanogaster* group (Sackton et al. 2007). Also, in *Nasutitermes* termites, selective pressure appears to vary among lineages in two GNBPs and in Relish (Bulmer and Crozier 2006). As we did not have any prior expectation for finding positive selection in certain lineages, we did not test for branch-specific selection with PAML. Selection pressure may vary among lineages of both honeybees and ants, and this was hinted by the findings of variable rates of nonsynonymous substitutions in *GNBP1-2* of honeybees and in defensin of ants. Interestingly, defensin has been found to be under positive selection in a study containing several ant taxa (Viljakainen and Pamilo 2008). In that study, the overall significance of the selection test depended strongly on a significantly higher rate of nonsynonymous substitutions in a couple of phylogenetic lineages. Variable rates of nonsynonymous substitutions, but a lack of evidence of positive selection in defensin in the present study combined with the results of Viljakainen and Pamilo, suggest that defensin in ants may be involved in lineage-specific host–pathogen arms race.

A single gene, the ant *PGRP*, was found to have signs of positive selection in the present study. The ant and honeybee PGRPs formed a cluster in a phylogenetic analysis, and they share 52 % amino acid similarity. The *A. mellifera* PGRP-S2 has been shown to be upregulated after immune challenge and most likely also the ant PGRP functions in immune system. However, as none of the known PGRPs were included in the ant and honeybee PGRP cluster in the phylogeny, the function of these proteins remains unclear. PGRP-SC2 has been shown in *Drosophila* to function as a regulator of the Imd pathway preventing overresponse by degrading microbial peptidoglycan (Bischoff et al. 2006). It is worth noting that the ant and honeybee PGRPs both contain

the five amino acid residues reported to be crucial to enzymatic activity (Mellroth et al. 2003), which suggests, but does not prove, that the function could be the same. It is noteworthy that the five specific sites are also conserved in *Drosophila* PGRP-SB1/2 (Mellroth et al. 2003), which are not able to downregulate the immune response (Bischoff et al. 2006). As the function of the ant PGRP has not been verified, it is too early to say anything about the causes of positive selection on this gene.

The Importance of Colony-Level Defenses in Reducing Pathogen Pressure

Immune defense has been considered as a key factor in determining the success of a social insect colony (Baer and Schmid-Hempel 2006). The diverse repertoire of social defenses reviewed by Cremer et al. (2007) may act as a buffering mechanism and reduce selection pressure on the immune genes. There are several ways in which social insects can defend themselves. Behavioral traits have been selected for in order to avoid spreading of infections. However, some of the results concerning the avoidance behavior are contradictory as studies have shown that the level of trophallaxis (food exchange between workers) may decrease (Aubert and Richard 2008) or increase (de Souza et al. 2008) in infected worker ants. Antimicrobial secretions are also important preventive mechanism against infections (Stow et al. 2007). Colony-level defenses have been implicated when explaining the small number of immune genes in the honeybee genome, as compared with *Drosophila* (Evans et al. 2006), agreeing with the view of social defenses decreasing pathogen pressure. However, the rapidly diverging gene duplicates of the termite termicins (Bulmer and Crozier 2004) and positive selection in several genes in termites (Bulmer and Crozier 2006) and ants (Viljakainen and Pamilo 2008, this study) indicate that social defense mechanisms have their limitations and the evolution of the physiological defense mechanisms has also played an important role.

One aspect of social defense is the level of genetic variation within colonies. Strong genotypic host–parasite interactions have been clearly demonstrated in bumblebees (Schmid-Hempel and Reber Funk 2004). Such diversity refers more to balancing than directional selection. Balancing selection maintains high polymorphism within populations and data on intraspecific variation would be needed in order to evaluate its importance. Colony-level resistance has been shown to increase by genetic diversity resulting from polyandry (Hughes and Boomsma 2004) and mixing of genetic lines (Reber et al. 2008) in ants, and from outbreeding in termites (Calleri et al. 2006). Such results do not, however, specify whether genetic diversity affects through behavioral mechanisms or whether diversity of the immune-related genes is involved. Genetic variation in the immune response has been detected in bees (Decanini et al. 2007; Wilfert et al. 2007), but this stems, at least in some cases, from variation in the expression levels of genes (Decanini et al. 2007).

Selection for general efficiency (higher expression levels) rather than host–pathogen arms races has been suggested to be a major determinant in the evolution of antimicrobial effector molecules (Sackton et al. 2007). Infection experiments have shown that several antimicrobial peptides (abaecin, apidaecin, hymenoptaecin, and defensin) are expressed simultaneously in infected bees (Choi et al. 2008). Furthermore, the expression levels of abaecin and hymenoptaecin have also been shown to increase in connection with dealation of young ant queens (Tian et al. 2004), indicating that the time of nest founding, the only solitary life history stage in the life of ants, may be the most sensitive period in the ant colony life cycle. Lastly, high immune response levels have been shown to decrease the future productivity of the colonies, indicating a colony-level cost (Evans and Pettis 2005). Arguably the expression levels are important in the immune defense.

However, as mentioned above, there is evidence of genetic variation in immune responses and strain-specific effects, suggesting selection on the physiological mechanisms. The present results show similar d_N/d_S ratios in all the genes from different parts of the immune defense pathways, and positive selection in immune genes has been detected in both ants and termites (Bulmer and Crozier 2004; Viljakainen and Pamilo 2008). Based on these results, we suggest that selection on the amino acid composition of the immune-related genes has been an important part in the fight against pathogens by social insects.

In the future, it would be important to carry out a comparative study focusing on both substitution rates and sequence diversity within species, and including a larger number of immune genes from species that represent

different levels of sociality and different phylogenetic lineages in order to test the hypotheses concerning the relative importance of physiological and behavioral adaptations. Such effects also depend on the relative importance of defense by the queens and by the workers.

References

Anisimova M, Bielawski JP, Yang Z. Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol Biol Evol (2002) 19:950–958.

Astruc C, Julien JF, Errard C, Lenoir A. Phylogeny of ants (Formicidae) based on morphology and DNA sequence data. Mol Phylogenet Evol (2004) 31:880–893.

Aubert A, Richard FJ. Social management of LPS-induced inflammation in *Formica polyctena* ants. Brain Behav Immun (2008) 22:833–837.

Baer B, Schmid-Hempel P. Phenotypic variation in male and worker encapsulation response in the bumblebee *Bombus terrestris*. Ecol Entomol (2006) 31:591–596.

Beattie AJ, Turnbull CL, Hough T, Knox RB. Antibiotic production: a possible function for the metapleural glands of ants (Hymenoptera: Formicidae). Ann Entomol Soc Am (1986) 79:448–450.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol (2004) 340:783–795.

Bischoff V, Vignal C, Duvic B, Boneca IG, Hoffmann JA, Royet J. Downregulation of the *Drosophila* immune response by peptidoglycan-recognition proteins SC1 and SC2. PLoS Pathog (2006) 2:e14.

Boomsma JJ, Schmid-Hempel P, Hughes WOH. Life histories and parasite pressure across the major groups of social insects. In: Insect evolutionary ecology—Fellowes MDE, Holloway GJ, Rolff J, eds. (2005) Wallingford (United Kingdom): CABI. 139–175.

Brady SG, Schultz TR, Fisher BL, Ward PS. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc Natl Acad Sci USA (2006) 103:18172–18177.

Bromham L, Leys R. Sociality and the rate of molecular evolution. Mol Biol Evol (2005) 22:1393–1402.

Bulmer MS, Crozier RH. Duplication and diversifying selection among termite antifungal peptides. Mol Biol Evol (2004) 21:2256–2264.[

Bulmer MS, Crozier RH. Variation in positive selection in termite GNBPs and Relish. Mol Biol Evol (2006) 23:317–326.

Calleri DV, McGrail Reid E, Rosengaus RB, Vargo EL, Traniello JF. Inbreeding and disease resistance in a social insect: effects of heterozygosity on immunocompetence in the termite *Zootermopsis angusticollis*. Proc R Soc Lond B (2006) 273:2633–2640.

Chapuisat M, Oppliger A, Magliano P, Christe P. Wood ants use resin to protect themselves against pathogens. Pro R Soc Lond B (2007) 274:2013–2017.

Choi YS, Choo YM, Lee KS, Yoon HJ, Kim I, Je YH, Sohn HD, Jin BR. Cloning and expression profiling of four antibacterial peptide genes from the bumblebee *Bombus ignitus*. Comp Biochem Physiol Biochem B Mol Biol (2008) 150:141–146.

Christe P, Oppliger A, Bancalà F, Castella G, Chapuisat M. Evidence for collective medication in ants. Ecol Lett (2003) 6:19.

Cremer S, Armitage SA, Schmid-Hempel P. Social immunity. Curr Biol (2007) 17:R693–702.

de Souza DJ, Van Vlaenderen J, Moret Y, Lenoir A. Immune response affects ant trophallactic behaviour. J Insect Physiol (2008) 54:828–832.

Decanini LI, Collins AM, Evans JD. Variation and heritability in immune gene expression by diseased honeybees. J Hered (2007) 98:195–201.

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

Evans JD, Pettis SP. Colony-level impacts of immune responsiveness in honey bees, *Apis mellifera*. Evolution (2005) 59:2270–2274.

Ferrandon D, Imler JL, Hetru C, Hoffmann JA. The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. Nat Rev Immunol (2007) 7:862–874.

Fourmidable ant sequence database. [cited 2008 Feb]. Available from: http://fourmidable.unil.ch/.

Gillespie JP, Kanost MR, Trenczek T. Biological mediators of insect immunity. Annu Rev Entomol (1997) 4269:611–643.
Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol (1994) 11:725–736.
Hoffmann JA. The immune response of *Drosophila*. Nature (2003) 426:33–38.
Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature (2006) 443:931–949.
Hughes WO, Boomsma JJ. Genetic diversity and disease resistance in leaf-cutting ant societies. Evolution (2004) 58:1251–1260.
Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJ. The 20 years of PROSITE. Nucleic Acids Res (2008) 36:D245–D249.
Jiggins FM, Kim KW. The evolution of antifungal peptides in *Drosophila*. Genetics (2005) 171:1847–1859.
Jiggins FM, Kim KW. A screen for immunity genes evolving under positive selection in *Drosophila*. J Evol Biol (2007) 20:965–970.
Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform (2004) 5:150–163.
Lazzaro BP, Clark AG. Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. Mol Biol Evol (2003) 20:914–923.
Li WH. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol (1993) 36:96–99.
Mellroth P, Karlsson J, Steiner H. A scavenger function for a *Drosophila* peptidoglycan recognition protein. J Biol Chem (2003) 278:7059–7064.
Nicholas KB, Nicholas HB Jr, Deerfield DW II. GeneDoc: Analysis and visualization of genetic variation, EMBNEW.NEWS (1997) 4:14, 2 [cited 2009 May 1]. Available from: http://www.psc.edu/biomed/genedoc.
Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics (1998) 148:929–936.
Nylander JAA. MrModeltest 2.2. Program distributed by the author (2004) Evolutionary Biology Centre, Uppsala University.
Pamilo P, Bianchi NO. Evolution of the *zfx* and *zfy* genes: rates and interdependence between the genes. Mol Biol Evol (1993) 10:271–281.
Parmakelis A, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nkondjio C, Simard F, Caccone A, Powell JR. The molecular evolution of four anti-malarial immune genes in the *Anopheles gambiae* species complex. BMC Evol Biol (2008) 8:79.
Raffiudin R, Crozier RH. Phylogenetic analysis of honey bee behavioral evolution. Mol Phylogenet Evol (2007) 43:543–552.
Reber A, Castella G, Christe P, Chapuisat M. Experimentally increased group diversity improves disease resistance in an ant species. Ecol Lett (2008) 11:682–689.
Robinson-Rechavi M, Huchon D. RRTree: relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics (2000) 16:296–297/
Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics (2003) 19:1572–1574.
Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res (1998) 26:1628–1635.
Rothenbuhler WC. Behavior genetics of nest cleaning in honey bees. IV. Responses of F1 and backcross generations to disease-killed brood. Am Zool (1964) 4:111–123.
Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet (2007) 39:1461–1468.
Schlenke TA, Begun DJ. Natural selection drives *Drosophila* immune system evolution. Genetics (2003) 164:1471–1480.
Schmid-Hempel P. Parasites in social insects (1998) Princeton (NJ): Princeton University Press.

Schmid-Hempel P, Crozier RH. Polyandry versus polygyny versus parasites. Philos Trans R Soc Lond B (1999) 354:507–515.
Schmid-Hempel P, Reber Funk C. The distribution of genotypes of the trypanosome parasite, *crithidia bombi*, in populations of its host, *Bombus terrestris*. Parasitology (2004) 129:147–158.
Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics (2002) 18:502–504.
Seeley TD, Tarpy DR. Queen promiscuity lowers disease within honeybee colonies. Proc R Soc Lond B (2007) 274:67–72.
Simard F, Licht M, Besansky NJ, Lehmann T. Polymorphism at the *defensin* gene in the *Anopheles gambiae* complex: testing different selection hypotheses. Infect Genet Evol (2007) 7:285–292.
Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci USA (2003) 100:9440–9445.
Stow A, Briscoe D, Gillings M, Holley M, Smith S, Leys R, Silberbauer T, Turnbull C, Beattie A. Antimicrobial defences increase with sociality in bees. Biol Lett (2007) 3:422–424.
Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. Mol Biol Evol (2003) 20:18–20.
Swofford DL. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4 (2002) Sunderland (MA): Sinauer Associates.
Tarpy DR. Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth. Proc R Soc Lond B (2003) 270:99–103.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res (1997) 25:4876–4882.
Tian H, Vinson SB, Coates CJ. Differential gene expression between alate and dealate queens in the red imported fire ant, *Solenopsis invicta* buren (hymenoptera: Formicidae). Insect Biochem Mol Biol (2004) 34:937–949.
Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res (2007) 35:W71–W4.
Viljakainen L, Pamilo P. Identification and molecular characterization of *defensin* gene from the ant *Formica aquilonia*. Insect Mol Biol (2005) 14:335–338.
Viljakainen L, Pamilo P. Selection on an antimicrobial peptide defensin in ants. J Mol Evol (2008) 67:643–652.
Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. Proc Natl Acad Sci USA (2000) 97:13772–13777.
Wilfert L, Gadau J, Baer B, Schmid-Hempel P. Natural variation in the genetic architecture of a host-parasite interaction in the bumblebee *Bombus terrestris*. Mol Ecol (2007) 16:1327–1339.
Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol (2007) 24:1586–1591.
Yang Z, Nielsen R, Goldman N, Pedersen A-K. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics (2000) 155:431–449.