

A screen for immunity genes evolving under positive selection in *Drosophila*

F. M. JIGGINS & K. W. KIM

Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK

Keywords:

Drosophila;
immunity;
persephone;
serine protease;
serpin;
spirit.

Abstract

Genes involved in the immune system tend to have higher rates of adaptive evolution than other genes in the genome, probably because they are coevolving with pathogens. We have screened a sample of *Drosophila* genes to identify those evolving under positive selection. First, we identified rapidly evolving immunity genes by comparing 140 loci in *Drosophila erecta* and *D. yakuba*. Secondly, we resequenced 23 of the fastest evolving genes from the independent species pair *D. melanogaster* and *D. simulans*, and identified those under positive selection using a McDonald–Kreitman test. There was strong evidence of adaptive evolution in two serine proteases (*persephone* and *spirit*) and a homolog of the *Anopheles* serpin *SRPN6*, and weaker evidence in another serine protease and the death domain protein *dFADD*. These results add to mounting evidence that immune signalling pathway molecules often evolve rapidly, possibly because they are sites of host–parasite coevolution.

Introduction

Natural selection often drives rapid evolutionary change in immune system genes, suggesting that the selection pressures acting on these genes are continually changing. This is thought to result from pathogens evolving to evade the host immune response, which in turn selects for host counter adaptations. In one of the most comprehensive studies, it was recently shown that *Drosophila melanogaster* immunity genes evolve faster than nonimmunity genes (Schlenke & Begun, 2003). Furthermore, this difference was attributable to natural selection fixing selectively advantageous amino acid changing mutations. In this and other data sets, it has been found that many of the genes in immune signalling pathways are often subject to positive natural selection (Begun & Whitley, 2000; Schlenke & Begun, 2003). These include *Toll* and *necrotic* in the Toll pathway, and *Imd*, *Dredd* and *Relish* in the Imd pathway. It has been suggested that these signalling molecules may be targeted by pathogens that suppress the fly immune response,

which in turn selects for mutations in the fly genes that prevent this suppression (Begun & Whitley, 2000). There are also other classes of positively selected proteins, including the antiviral RNAi system and proteins that bind to pathogens during the cellular immune response (Lazzaro, 2005; Jiggins & Kim, 2006; Obbard *et al.*, 2006).

It is likely that immunity genes tend to evolve faster than other genes because they are adapting to novel pathogen challenges. Therefore, changes in the sequence of positively selected proteins have presumably altered the susceptibility of flies to infection. If this is the case, then the pattern of molecular evolution can be used to make predictions about the function of immune molecules. We expect that positively selected proteins will tend to be both functionally important in the immune system, and play an important role in adapting to novel pathogens or pathogen genotypes. Although other factors may drive the rapid evolution of immune-related proteins, screening for positively selected genes provides a rapid way to identify candidate genes within large genomic data sets.

In this study, we have screened a large set of genes that have been linked to the immune response for those under positive selection. We deliberately chose to include many poorly characterized genes whose involvement in the immune response is uncertain. In this way, we hoped

Correspondence: Francis M. Jiggins, Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Rd, Edinburgh EH9 3JT, UK. Tel.: (0) 131 650 6564; fax: (0) 131 650 5476; e-mail: francis.jiggins@ed.ac.uk

to identify those genes that merit further functional characterization. We began with a list of genes that have been linked to the immune response. Some of these genes were known from experimental studies, whereas others are homologous to known immune genes or are simply known to be strongly upregulated following an immune challenge. We then compared the sequence of these genes in *D. yakuba* and *D. erecta*, and narrowed the list down to just those that evolve particularly fast. To ensure that our test for positive selection is independent of the criteria used to draw up this shortlist, we then switched to a closely related but phylogenetically independent pair of *Drosophila* species (*D. simulans* and *D. melanogaster*). We sequenced a small number of alleles from both of these species, and used a McDonald–Kreitman test to identify those that are positively selected. This strategy of sequencing a small number of alleles in both species rather than a larger number from just one species should maximize the total number of polymorphisms detected for a given sequencing effort.

Methods

Identifying rapidly evolving immunity genes

We based our preliminary analysis on a list of 321 immunity-related genes in *D. melanogaster* compiled by Bruno LeMaitre (http://www.cnrs-gif.fr/cgm/immunity/drosophila_immunity_genes.html; Khush & Lemaitre, 2000). This list includes 156 genes that either have a known immune function or share sequence homology with known immunity genes. The remaining genes were identified in microarray studies of immune-challenged flies as having similar patterns of expression to known immunity genes (classified as *Drosophila* immune-related genes or DIRGs by De Gregorio *et al.*, 2001). Therefore, the list included many loci for which there is only tentative evidence for their involvement in the immune response.

To narrow down this list to those genes that evolve rapidly we compared their predicted sequences in the *D. yakuba* and *D. erecta* genomes (Release 1.0 of the Eisen Group Annotations for *D. yakuba* and *D. erecta*; D. Pollard and V. Iyer, pers. comm.). At the time of the analysis only 140 of the genes had been annotated in both species, and some of these annotations were incomplete (on average 3.6% of the sequence was missing, assuming the *D. yakuba* and *D. erecta* coding sequences are the same length as in *D. melanogaster*). Given the incomplete nature of the data set, this analysis was simply used to identify a rapidly evolving subset of the genes and no analysis of the data is presented. The predicted coding sequence was aligned using CLUSTALW (Thompson *et al.*, 1994 and the percentage divergence at nonsynonymous nucleotide sites (K_a) estimated using the method of Nei & Gojombori (1986). The 17 genes that had the highest K_a estimates but had not been included in previous evolu-

tionary studies were selected for resequencing (*dif*, *nec*, *spatzle*, *metchnikowin* and *dSr-CIII* were excluded as they had been studied before). An additional six of the 34 genes with the highest K_a estimates were also resequenced as there was additional evidence to implicate them in the immune response (experimental evidence, homology to immune genes or predicted domains that are likely to be immunity related). Of the 23 genes selected for resequencing, two (CG9928 and CG5765) were incompletely annotated in either *D. yakuba* or *D. erecta* and the alignment contained only just over half of the coding sequence. In CG5765 the missing region was highly repetitive and difficult to align between species, and we therefore did not resequence the missing region. However, the missing region of CG9928 was included in the resequencing described next.

Sequencing

Multiple alleles of these 23 genes were resequenced from *D. simulans* and *D. melanogaster*. The *D. simulans* stocks had been collected in Nairobi, Kenya, as isofemale lines (Dean & Ballard, 2004) and were inbred by sib mating for nine generations. The *D. melanogaster* stocks were collected as isofemale lines in 2002 by Bill Ballard and Sylvain Charlat in Franceville, Gabon. The appropriate chromosomes were made homozygous by standard crosses to SM1 or TM6 balancer stocks. We aimed to sequence four alleles of each gene from both species. In some cases we were unable to amplify the gene from all specimens, resulting in smaller samples. In the *D. simulans* sequences four sequences were excluded because they were heterozygous (one from each of these genes: CG5765, CG16743, CG10680 and CG9649). The region that was resequenced was on average 12% shorter than the region analysed in the preliminary screen. This was because some regions were excluded either due to the poor quality of the sequence, or to reduce the cost of the sequencing (e.g. short exons). The *D. melanogaster* genome sequence was included in the McDonald–Kreitman test analysis. We also used larger samples for one of the genes that we had previously sequenced for other reasons. Signal peptides were predicted using the program SIGNALP v3.0 (Bendtsen *et al.*, 2004). The clip domain and predicted site of proteolytic cleavage of serine proteases was taken from Ross *et al.* (2003). Data were submitted to GenBank under the accession numbers AM412815–AM412998 and only the coding sequence was analysed.

Results

We tested whether the rapid evolution of our genes results from positive selection by comparing polymorphism and divergence at synonymous and nonsynonymous sites. Under the neutral model, the ratio of synonymous : nonsynonymous polymorphic sites will be the

same as the ratio of synonymous : nonsynonymous interspecific differences. The McDonald–Kreitman test simply compares these two ratios in a 2×2 contingency table (McDonald & Kreitman, 1991). Five of our 23 candidate immunity genes showed some evidence of adaptive evolution (Table 1). In three of the genes this was highly significant, and in two further genes, this was on the borderline of significance (Table 1). The effect of multiple testing can be controlled for by setting a false discovery rate, which is the proportion of significant results that are type I errors (Benjamini & Hochberg, 1995). Using a false discovery rate of 5%, only the three genes with the smallest *P* values in Table 1 remain significant.

The excess of fixed nonsynonymous differences can be used to estimate the proportion of amino acid changes between *D. melanogaster* and *D. simulans* that have been fixed by natural selection (Smith & Eyre-Walker, 2002). In the three most significant cases of adaptive evolution, 79–84% of the differences were adaptive (Table 1). In the other two ‘borderline’ genes, an estimated 68% were fixed by selection. If the data are summed across the remaining 18 genes, there is a significant excess of fixed nonsynonymous differences. However, in the *D. melanogaster* data set, the estimated proportion of amino acid substitutions that are adaptive in these genes is 36%, which is below the genome average (Smith & Eyre-Walker, 2002). Therefore, there is little to suggest that the rate of evolution in these

genes has been accelerated because of host–parasite coevolution.

It is striking that most, and perhaps all, of these five proteins are involved in the signalling and control of immune responses (Table 2). Three of the five genes were serine proteases, one was a serine protease inhibitor (serpin) and one was a death domain protein (Table 2). The serine proteases *spirit* and *persephone* are both components of the Toll signalling pathway, and the death domain protein *dFADD* is part of the IMD pathway. The serpin *CG7219* is less well characterized, but it is known to be upregulated by the Toll pathway following immune challenge, and it is a homolog of a serpin in mosquitoes that is involved in defence against *Plasmodium*. The third serine protease is also upregulated by the Toll pathway, but little is known about its function.

The protein domains that are under positive selection may provide clues as to the selective forces acting on these molecules. Therefore, we have analysed the different domains of the three serine proteases separately. All three genes contain a predicted signal peptide, indicating that they are secreted proteins. There is no evidence of positive selection on these signal peptides, which is consistent with the fact that they are cleaved from the protein during secretion and are therefore unlikely to interact with pathogens (Table 3). The protein is secreted as an inactive zymogen, which is then activated when the NH_2 -terminal prodomain is cleaved from the COOH-terminal catalytic domain (Ross *et al.*, 2003). Two of the

Table 1 Genes studied.

| CG number | Length (bp) | Name | Function | Reference |
|-----------|-------------|--------------------|---|----------------------------------|
| CG9928 | 285 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG2056 | 1173 | <i>spirit</i> | Serine protease in toll pathway | Kambris <i>et al.</i> (2006) |
| CG6367 | 1145 | <i>persephone</i> | Serine protease in toll pathway | Ligoxygakis <i>et al.</i> (2002) |
| CG7219 | 1310 | | Serpin immune upregulated by Toll pathway and homolog of <i>Anopheles</i> gene SRPN6 involved in immunity against <i>Plasmodium</i> | Abraham <i>et al.</i> (2005) |
| CG9631 | 1086 | | Serine protease immune upregulated by Toll pathway | De Gregorio <i>et al.</i> (2002) |
| CG12297 | 429 | <i>dFADD (BG4)</i> | Death domain protein in IMD pathway | Leulier <i>et al.</i> (2002) |
| CG5765 | 390 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG5773 | 381 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG10912 | 633 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG6687 | 1278 | | Serpin, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG15066 | 345 | <i>IM23</i> | Peptide immune upregulated by Toll pathway | De Gregorio <i>et al.</i> (2002) |
| CG6467 | 650 | <i>Jonah 65Aiv</i> | Serine protease immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG10680 | 576 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG9434 | 765 | <i>Fst</i> | Immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG18067 | 666 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG2217 | 1053 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG9080 | 309 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG14957 | 288 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG9120 | 426 | <i>lysX</i> | Lysozyme | |
| CG9649 | 1343 | | Serine protease immune upregulated by Toll pathway | De Gregorio <i>et al.</i> (2002) |
| CG15293 | 861 | | Immune induced and found in haemolymph clots | Scherfer <i>et al.</i> (2004) |
| CG16772 | 846 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |
| CG16743 | 525 | | Uncharacterized, immune upregulated | De Gregorio <i>et al.</i> (2001) |

Table 2 McDonald–Kreitman tests on immunity genes.

| Name | Number alleles | | Fixed differences | | <i>D. melanogaster</i> polymorphisms | | | | <i>D. simulans</i> polymorphisms | | | | Both species polymorphisms | |
|--------------------|----------------|-----|-------------------|----|--------------------------------------|----|-------------|----------|----------------------------------|----|-------------|----------|----------------------------|----------|
| | sim | mel | Syn | NS | Syn | NS | % adaptive* | <i>P</i> | Syn | NS | % adaptive* | <i>P</i> | % adaptive* | <i>P</i> |
| CG9928 | 4 | 5 | 6 | 5 | 2 | 0 | 100 | n.s. | 7 | 1 | 83 | n.s. | 87 | n.s. |
| <i>spirit</i> | 4 | 5 | 24 | 20 | 20 | 2 | 88 | 0.005 | 3 | 1 | 60 | n.s. | 84 | 0.004 |
| <i>persephone</i> | 4 | 11 | 30 | 16 | 30 | 3 | 81 | 0.009 | 15 | 2 | 75 | n.s. | 79 | 0.006 |
| CG7219 | 4 | 5 | 17 | 17 | 19 | 11 | 42 | n.s. | 50 | 3 | 94 | < 0.0001 | 79 | 0.0004 |
| CG9631 | 4 | 5 | 20 | 22 | 11 | 4 | 67 | n.s. | 9 | 3 | 70 | n.s. | 68 | 0.04 |
| <i>DFADD (BG4)</i> | 5 | 5 | 7 | 16 | 10 | 6 | 74 | 0.059 | 9 | 8 | 61 | n.s. | 68 | 0.06 |
| CG5765 | 4 | 4 | 6 | 9 | 5 | 2 | 73 | n.s. | 4 | 3 | 50 | n.s. | 63 | n.s. |
| CG5773 | 4 | 5 | 2 | 2 | 5 | 1 | 80 | n.s. | 3 | 2 | 33 | n.s. | 63 | n.s. |
| CG10912 | 3 | 4 | 20 | 28 | 2 | 5 | 0 | n.s. | 10 | 2 | 86 | 0.02 | 58 | n.s. |
| CG6687 | 4 | 5 | 33 | 42 | 2 | 2 | 21 | n.s. | 16 | 10 | 51 | n.s. | 48 | n.s. |
| <i>IM23</i> | 4 | 5 | 9 | 4 | 10 | 3 | 33 | n.s. | 2 | 1 | 0 | n.s. | 25 | n.s. |
| <i>Jonah 65Aiv</i> | 4 | 5 | 9 | 11 | 2 | 3 | 0 | n.s. | 1 | 0 | 100 | n.s. | 18 | n.s. |
| CG10680 | 4 | 4 | 12 | 13 | 3 | 1 | 69 | n.s. | 6 | 7 | 0 | n.s. | 18 | n.s. |
| <i>Fst</i> | 4 | 5 | 18 | 14 | 5 | 8 | 0 | n.s. | 31 | 18 | 25 | n.s. | 7 | n.s. |
| CG18067 | 4 | 5 | 15 | 10 | 9 | 3 | 50 | n.s. | 11 | 10 | 0 | n.s. | 3 | n.s. |
| CG2217 | 3 | 5 | 22 | 12 | 16 | 5 | 43 | n.s. | 12 | 10 | 0 | n.s. | 2 | n.s. |
| CG9080 | 4 | 5 | 6 | 5 | 1 | 1 | 0 | n.s. | 2 | 2 | 0 | n.s. | 0 | n.s. |
| CG14957 | 4 | 5 | 5 | 9 | 0 | 1 | – | n.s. | 1 | 3 | 0 | n.s. | 0 | n.s. |
| <i>lysX</i> | 4 | 5 | 8 | 2 | 7 | 4 | 0 | n.s. | 9 | 9 | 0 | n.s. | 0 | n.s. |
| CG9649 | 3 | 5 | 24 | 15 | 10 | 11 | 0 | n.s. | 16 | 7 | 30 | n.s. | 0 | n.s. |
| CG15293 | 4 | 5 | 26 | 28 | 2 | 6 | 0 | n.s. | 9 | 7 | 28 | n.s. | 0 | n.s. |
| CG16772 | 4 | 5 | 17 | 8 | 9 | 10 | 0 | n.s. | 25 | 12 | 0 | n.s. | 0 | n.s. |
| CG16743 | 4 | 4 | 10 | 12 | 0 | 7 | – | n.s. | 7 | 2 | 76 | n.s. | 0 | n.s. |

*See Smith & Eyre-Walker (2002).

mel, *D. melanogaster*; sim, *D. simulans*; syn, synonymous; NS, nonsynonymous; *P*, Fisher exact test *P*-value.

| Region | Fixed differences | | Polymorphisms | | % adaptive | <i>P</i> |
|--|-------------------|----|---------------|----|------------|----------|
| | Syn | NS | Syn | NS | | |
| Signal peptide | 4 | 9 | 2 | 4 | 11 | n.s. |
| Clip domain | 6 | 2 | 7 | 2 | 14 | n.s. |
| COOH-terminal catalytic domain | 48 | 19 | 64 | 4 | 84 | 0.0005 |
| NH ₂ -terminal prodomain (excl. Clip) | 16 | 28 | 15 | 5 | 81 | 0.007 |

Table 3 McDonald–Kreitman tests on different domains of the serine proteases. The data summed across both species and the genes *spirit*, *persephone* and CG9631.

serine proteases also contain a clip domain, which may be involved in protein–protein interactions (Ross *et al.*, 2003). Analysing the three domains separately, there is no evidence of positive selection on the clip domain (Table 3). However, the rest of the prodomain and the catalytic domain have similar rates of adaptive evolution (Table 3). Therefore, selection is probably acting on the inactive zymogen, and not just the active protease.

Discussion

Many immunity genes, including key signalling molecules, have proved to be under positive selection. The aim of this study was to screen a set of candidate immunity genes for those evolving adaptively. This allows us to predict which molecules may be coevolving

with pathogens, and which are likely to be functionally important in the immune system.

Our first striking result was that four of the five genes showing some evidence of positive selection have experimental evidence to support their role in the immune system of either flies or mosquitoes (Table 2). In contrast, none of the remaining 18 genes had such evidence. This not only supports the contention that positive selection is associated with immunity genes, but also suggests that many of the remaining genes may not have important roles in the immune system. Many of our candidate genes had been identified in microarray studies of immune-challenged flies. The fact that they show a different pattern of molecular evolution to the known immune genes suggests that they are either not important in the immune system, or they are

components of the immune system that are not under positive selection (perhaps because they do not coevolve with pathogens).

All the genes included in this study had a high level of nonsynonymous divergence between *D. yakuba* and *D. erecta*. There are three possible reasons why most of these genes did not prove to be positively selected in *D. melanogaster* or *D. simulans*. First, these genes may be evolving rapidly because of low selective constraints. Indeed, the proportion of polymorphisms that were nonsynonymous was highest in the genes that we did not detect as being positively selected, which is consistent with there being weak selective constraints on these genes (data not shown). Secondly, different genes may be under positive selection in different species of *Drosophila*. Thirdly, in very large data sets some genes will have a large number of nonsynonymous substitutions due to chance. Whatever the cause, it is clear that the K_a alone is a poor predictor of which genes are evolving adaptively. Although we chose to select genes on the basis of nonsynonymous divergence (K_a), using the ratio nonsynonymous : synonymous substitutions (K_a/K_s) would have made little difference to the list of genes we studied. Furthermore, the small number of synonymous sites in short genes means that this statistic can have a high variance. Therefore, comparisons of pairs of genomes are likely to be of limited value in identifying adaptively evolving genes.

Three of the positively selected genes that we identified are well-characterized components of immune-signalling pathways (Tables 1 and 2). The remaining two, a serine protease and a serpin, may also play a role in the control or modulation of the immune system. This corroborates a pattern described before, where signalling molecules are one of the most rapidly evolving components of the immune system (Begun & Whitley, 2000; Schlenke & Begun, 2003). It has been proposed that this is because of signalling pathways being targeted by parasite molecules that suppress the immune response (Begun & Whitley, 2000). Therefore, selection will favour mutations in the signalling molecules that prevent this suppression. This hypothesis is supported by the growing number of pathogen molecules that suppress host immune-signalling pathways (Begun & Whitley, 2000; Thoetkiattikul *et al.*, 2005). Further clues as to the nature of these selection pressures can come from looking at the protein domains that are evolving adaptively. The prodomain (excluding the clip region) and the catalytic domain of the three serine proteases show similar rates of adaptive evolution. This suggests that selection may be acting before the inactive zymogen is cleaved to produce an active protease. It is plausible that the selection pressures result from parasite molecules that block activation of the zymogen. However, the hypothesis that parasite immune suppressors drive the rapid evolution of immune-signalling pathways still lacks direct experimental support, and other explanations exist. For example, if selection

favours continual changes in the pattern of immune gene expression, this could result in adaptive evolution of signalling molecules.

These results add to a growing list of invertebrate immunity genes that are under directional selection. These genes all have elevated rates of amino acid substitutions between species. However, to date there is no evidence for balancing selection maintaining ancient polymorphisms as has been observed in some vertebrate immunity genes. Therefore, although theoretical models often predict that host–parasite interactions can result in frequency-dependent selection, there has been little evidence that this has resulted in the long-term maintenance of polymorphisms within invertebrate immunity genes. Instead, it is possible that coevolution between invertebrates and their parasites may commonly involve the recurrent fixation of new mutations.

Acknowledgments

The work was funded by the Wellcome Trust. Dan Pollard and Venky Iyer kindly supplied unpublished genome annotations.

References

- Abraham, E.G., Pinto, S.B., Ghosh, A., Vanlandingham, D.L., Budd, A., Higgs, S., Kafatos, F.C., Jacobs-Lorena, M. & Michel, K. 2005. An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 16327–16332.
- Begun, D.J. & Whitley, P. 2000. Adaptive evolution of relish, a *Drosophila* NF- κ B/I κ B protein. *Genetics* **154**: 1231–1238.
- Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**: 783–795.
- Benjamini, Y. & Hochberg, Y. 1995. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodological* **57**: 289–300.
- De Gregorio, E., Spellman, P.T., Rubin, G.M. & Lemaitre, B. 2001. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl. Acad. Sci. U.S.A.* **98**: 12590–12595.
- De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M. & Lemaitre, B. 2002. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J.* **21**: 2568–2579.
- Dean, M.D. & Ballard, J.W.O. 2004. Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Mol. Phylogenet. Evol.* **32**: 998–1009.
- Jiggins, F.M. & Kim, K.-W. 2006. Contrasting evolutionary patterns in *Drosophila* immune receptors. *J. Mol. Evol.* **63**: 769–780.
- Kambris, Z., Brun, S., Jang, I.-H., Nam, H.-J., Romeo, Y., Takahashi, K., Lee, W.-J., Ueda, R. & Lemaitre, B. 2006. *Drosophila* immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation. *Curr. Biol.* **16**: 808–813.

- Khush, R.S. & Lemaitre, B. 2000. Genes that fight infection – what the *Drosophila* genome says about animal immunity. *Trends Genet.* **16**: 442–449.
- Lazzaro, B.P. 2005. Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics* **169**: 2023–2034.
- Leulier, F., Vidal, S., Saigo, K., Ueda, R. & Lemaitre, B. 2002. Inducible expression of double-stranded RNA reveals a role for dFADD in the regulation of the antibacterial response in *Drosophila* adults. *Curr. Biol.* **12**: 996–1000.
- Ligoxygakis, P., Pelte, N., Hoffmann, J.A. & Reichhart, J.-M. 2002. Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* **297**: 114–116.
- McDonald, J.H. & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**: 652–654.
- Nei, M. & Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- Obbard, D.J., Jiggins, F.M., Halligan, D.L. & Little, T.J. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* **16**: 580.
- Ross, J., Jiang, H., Kanost, M.R. & Wang, Y. 2003. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene* **304**: 117–131.
- Scherfer, C., Karlsson, C., Loseva, O., Bidla, G., Goto, A., Havemann, J., Dushay, M.S. & Theopold, U. 2004. Isolation and characterization of hemolymph clotting factors in *Drosophila melanogaster* by a pullout method. *Curr. Biol.* **14**: 625–629.
- Schlenke, T.A. & Begun, D.J. 2003. Natural selection drives *Drosophila* immune system evolution. *Genetics* **164**: 1471–1480.
- Smith, N.G.C. & Eyre-Walker, A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022.
- Thoetkiattikul, H., Beck, M.H. & Strand, M.R. 2005. Inhibitor kappa B-like proteins from a polydnavirus inhibit NF-kappa B activation and suppress the insect immune response. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 11426–11431.
- Thompson, J.O., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673–4680.

Received 18 September 2006; revised 1 December 2006; accepted 16 December 2006