Comparative transcriptome analysis for immune response against

fungal infection in *Drosophila virilis.*

Yosuke Seto

Department of Biological Sciences

Tokyo Metropolitan University

2013

CONTENTS

GENERAL INTRODUCTION

A major goal in evolutionary biology is to uncover the mechanisms of phenotypic evolution. It is considered that many phenotypic characters have been evolved to adapt to various environmental conditions by natural selection (Darwin 1859). So far, molecular evolutionary biologists have tried to understand the adaptive evolution at molecular level using theoretical and experimental analyses (Hughes and Nei 1988, Nielsen and Yang 1998, Suzuki and Gojobori 1999, Smith and Eyre-Walker 2002, Yokoyama et al., 2008). In these studies, it was postulated that changes of nucleotide or protein sequence causing functionary alteration of the gene are the most reflect the adaptive evolution. A representative example of the theoretical studies is that focused on evolution of immune systems. Hughes and Nei (1988) revealed that the rate of nonsynonymous substitution (dN) was significantly higher than the rate of synonymous substitution (dS) in antigen recognition sites (ARS) of major histocompatibility complex (MHC) genes by comparing protein-coding region of MHC genes among mammals, whereas dN for non-ARS region was significantly lower than dS. From this results, they claimed that the ARS have been evolved under positive selection and the highly polymorphism of ARS have been maintained by overdominant selection, based on the prediction from the neutral theory of molecular evolution (Kimura 1983). Similarly, many immune-related genes involving in innate immune system are shown to be rapidly evolving compared to non-immune-related genes (Schlenke and Begun, 2003, Sackton et al., 2007, Obbard et al., 2009). These observations were explained by coevolutionary interactions between hosts and pathogens, so-called "arms races" (Dawkins and Krebs 1979). Under this conception, compared to non-immune-related genes, immune-related genes are expected to be rapidly evolving or to have elevated polymorphism to maintain various alleles to cover ever-changing pathogens (Schlenke and Begun, 2003, Sackton

et al., 2007, Obbard et al., 2009). These studies were focused on to detect evolutionary traces from the primary structure of DNA or protein sequences.

However, Wilson, Maxson and Sarich (1974) proposed that phenotypic evolution have more arisen from changes of gene regulatory system than from changes of protein function. Consistently, in recent years, it has been revealed that changes in gene expression pattern play an important role in phenotypic evolution, e.g., novelty of pigmentation pattern on *Drosophila* wings generated by changing spatial expression pattern of *yellow* (Gompel et al., 2005), changes in butterfly eyespots on the wings by changes in *Distal-less* expression pattern (Beldade, Brakefield and Long 2002) and changes in beak morphology in Darwin's Finches generated by gene expression changes of *BMP4* (Abzhanov et al., 2004). In *Drosophila* immune system, similar situation was also reported. Sackton and Clark (2009) found that the expression patterns of antimicrobial peptide (AMP) genes against bacterial infection by septic injury were different between two *Drosophila* species, *Drosophila melanogaster* and *D. virilis*. Although they suggested that this difference in the immune-response was due to different ecological traits of the two species, they did not clarify the relationship between the phenotype and the gene expression pattern.

D. melanogaster feeds on fermented or rotting fruits, which mainly harbor Baker's yeast, *Saccharomyces cerevisiae*, whereas *D. virilis* feeds on slime flux and decaying bark of trees, on which a variety of yeasts and filamentous fungi thrive (Carson 1971, Throckmorton 1975, Weber, Davoli and Anke 2006, Weber 2006). From this difference in the natural habitat, *D. virilis* is supposed to have a higher risk of the infection by a variety of fungi. However, according to 12 *Drosophila* species genomes analysis, *D. virilis* does not have the antifungal peptide, Drosomycin (Sackton et al., 2007), which is known to be an essential AMP in antifungal immune system of *D. melanogaster* (Lemaitre et al., 1996, Tzou, Reichhart and

Lemaitre 2002). This raises the question about the immune mechanism contributed to the antifungal resistance of *D. virilis*, which is thought to be an important factor for understanding the adaptive evolution of *D. virilis* to its habitat in moldy environment. To answer this question, I investigated the immune gene response to the fungal infection of *D. virilis* to clarify what immune system of *D. virilis* has evolved to defend against fungal infection.

My comparative transcriptome analysis revealed that many immune-related genes, such as AMP genes and immune-induced molecule (IM) genes, showed extensively different expression pattern between *D. melanogaster* and *D. virilis* in response to the infection of *Penicillium* fungus. Furthermore, I found a possibility that unknown immune-related genes have been recruited in antifungal immune system of *D. virilis* during its evolution. This *D. virilis*-specific immune gene response may contribute to the observed high resistance to the fungal infection. My results provide an important example for understanding the mechanism of phenotypic evolution by gene expression changes proposed by Wilson, Maxson and Sarich (1974).

Literature cited

Abzhanov A., M. Protas, B. R. Grant, P. R. Grant and C. J. Tabin, "Bmp4 and Morphological Variation of Beaks in Darwin's Finches," Science, vol. 305, no. 5689, pp. 1462-1465, 2004.

Beldade P, P. M. Brakefield and A. D. Long, "Contribution of Distal-less to quantitative variation in butterfly eyespots," *Nature*, vol. 415, no. 6869, pp. 315-318, 2002.

Carson H. L., "The ecology of *Drosophila* breeding sites", *Harold L. Lyon Arboretum Lecture Number Two*, University of Hawaii, pp. 1-28, 1971.

Darwin C., "On the origin of species by means of natural selection,"John Murray, London, 1859.

Dawkins R. and J. R. Krebs, "Arms races between and within species," *Proceedings of the Royal Society of London B Biological Sciences*, vol. 205, no. 1161, pp. 489-511, 1979.

Gompel N., B. Prud'homme, P. J. Wittkopp, V. A. Kassner and S. B. Carroll, "Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*," Nature, vol. 433, no. 7025, pp. 481-487, 2005.

Hughes, A. L., and M. Nei, "Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection," *Nature*, vol. **335**, no. 6186, pp. 167-170, 1988.

Janeway C. A. Jr, P. Travers, M. Walport, et al., "Immunobiology: The Immune System in Health and Disease," 5th edition, New York, Garland Science, 2001.

Kimura M., "The neutral theory of molecular evolution," Cambridge University Press, Cambridge, 1983

Lemaitre B., E. Nicolas, L. Michaut, J. M. Reichhart and J. A. Hoffmann, "The dorsoventral regulatory gene cassette Spatzle/Toll/cactus controls the potent antifungal response in

Drosophila adults," *Cell*, vol. 86, no. 6, pp. 973-983, 1996.

Nielsen, R. and Z. Yang, "Likelihood models for detecting positive selected amino acid sites and applications to the HIV-1 envelope gene," *Genetics*, vol. 148, no. 3, pp. 929-936, 1998.

Obbard, D. J., J. J. Welch, K. W. Kim and F. M. Jiggins, "Quantifying Adaptive Evolution in the Drosophila Immune System," *PLoS Genetics*, vol. 5, no. 10, Article ID e1000698, 2009.

Sackton T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark and A. G. Clark, "Dynamic evolution of the innate immune system in *Drosophila*," *Nature Genetics*, vol. 39, no. 12, pp 1461-1468, 2007.

Sackton T. B. and A. G. Clark, "Comparative profiling of the transcriptional response to infection in two species of *Drosophila* by short-read cDNA sequencing," *BMC Genomics*, vol. 10, article 259, 2009.

Schlenke T. A. and D. J. Begun, "Natural Selection Drives *Drosophila* Immune System Evolution," *Genetics*, vol. 164, no. 4, pp. 1471-1480, 2003.

Smith N. G. C and A. Eyre-Walker, "Adaptive protein evolution in Drosophila," *Nature*, vol. 415, no. 6875, pp. 1022–24, 2002.

Suzuki, Y. and T. Gojobori, "A method for detecting positive selection at single amino acid sites," *Molecular Biology and Evolution*, vol. 16, no. 10, pp. 1315–1328, 1999.

Throckmorton L. H., "The phylogeny, ecology and geography of *Drosophila*," in *Invertebrates of Genetic Interest*, R. C. King (ed), Plenum Press, New York, pp. 421-469, 1975.

Tzou P., J. M. Reichhart and B. Lemaitre, "Constitutive expression of a single antimicrobial peptide can restore wild-type resistance to infection in immuno-deficient *Drosophila* mutants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2152-2157, 2002.

Weber R. W. S., P. Davoli and H. Anke, "A microbial consortium involving the astaxanthin

producer Xanthophyllomyces dendrorhous on freshly cut birch stumps in Germany", *Mycologist*, Vol. 20, no. 2, pp. 57-61, 2006.

Weber R. W. S., "On the ecology of fungal consortia of spring sap-flows", *Mycologist*, Vol. 20, no. 4, pp. 140-143, 2006.

Wilson A. C., L. R. Maxson and V. M. Sarich, "Two Types of Molecular Evolution. Evidence from Studies of Interspecific Hybridization," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 71, no. 7, pp. 2843-2847, 1974.

Yokoyama S., T. Tada, H. Zhang and L. Britt, "Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 36, pp. 13480-13485. 2008

ABSTRACT

The innate immune system of *Drosophila* is activated by ingestion of microorganisms. *D. melanogaster* breeds on fruits fermented by *Saccharomyces cerevisiae*, whereas *D. virilis* breeds on slime flux and decaying bark of tree housing a variety of bacteria, yeasts and molds. In this study, it is shown that *D. virilis* has a higher resistance to oral infection of a species of filamentous fungi belonging to the genus *Penicillium* compared to *D. melanogaster.* In response to the fungal infection, a transcriptome profile of immune-related genes was considerably different between *D. melanogaster* and *D. virilis*: the genes encoding antifungal peptides, Drosomycin and Metchnikowin, were highly expressed in *D. melanogaster* whereas the genes encoding Diptericin and Defensin were highly expressed in *D. virilis*. On the other hand, the immune-induced molecule (IM) genes showed contrary expression patterns between the two species: they were induced by the fungal infection in *D. melanogaster* but tended to be suppressed in *D. virilis*. Our transcriptome analysis also showed newly predicted immune-related genes in *D. virilis*. These results suggest that the innate immune system has been extensively differentiated during the evolution of these *Drosophila* species.

ABSTRACT

The innate immune system of *Drosophila* is activated by ingestion of microorganisms. *D. melanogaster* breeds on fruits fermented by *Saccharomyces cerevisiae*, whereas *D. virilis* breeds on slime flux and decaying bark of tree housing a variety of bacteria, yeasts and molds. In this study, it is shown that *D. virilis* has a higher resistance to oral infection of a species of filamentous fungi belonging to the genus *Penicillium* compared to *D. melanogaster*. In response to the fungal infection, a transcriptome profile of immune-related genes was considerably different between *D. melanogaster* and *D. virilis*: the genes encoding antifungal peptides, Drosomycin and Metchnikowin, were highly expressed in *D. melanogaster* whereas the genes encoding Diptericin and Defensin were highly expressed in *D. virilis*. On the other hand, the immune-induced molecule (IM) genes showed contrary expression patterns between the two species: they were induced by the fungal infection in *D. melanogaster* but tended to be suppressed in *D. virilis*. Our transcriptome analysis also showed newly predicted immune-related genes in *D. virilis*. These results suggest that the innate immune system has been extensively differentiated during the evolution of these *Drosophila* species.

1. INTRODUCTION

In natural environments, *Drosophila* species feed and breed on fermenting fruits, slime fluxes on decaying parts of tree, etc., where biochemical processes of bacteria and fungi are extremely active (Carson 1971, Throckmorton 1975, Markow and O'Grady 2007). Therefore, *Drosophila* species are exposed to a huge number of microorganisms throughout their developmental stages. Feeding on decaying or fermented materials results in the ingestion of a wide variety of microorganisms in their digestive organs. Recent studies on larval immune response of *D. melanogaster* to oral infection of bacteria and fungi showed that the fat body mediated systemic immune response including antimicrobial peptide (AMP) production was triggered by infections of gram-negative bacterial species such as *Pseudomonas entomophila* and *Erwinia carotovora carotovora* 15 (Ecc15) and of a dimorphic fungal species, *Candida albicans* (Basset et al., 2000, Vodovaret al., 2005, Glittenberg et al., 2011).

In the expression of AMP genes, two major signaling pathways, Toll and Imd pathways, play a critical role. The Toll pathway is especially important in immune response to infection of fungi and gram-positive bacteria (Lemaitre et al., 1996, Rutschmann, Kilinc and Ferrandon 2002). After beta-(1.3)-glucans and Lys-type peptidoglycans, which are components of cell wall of fungi and gram-positive bacteria, are recognized by the gram-negative bacteria binding protein 3 (GNBP3) and peptidoglycan-recognition protein-SA (PGRP-SA), the Toll pathway is triggered by cleavage and binding of the ligand, Spatzle, to lead to degradation of Cactus, an inhibitor of NF-kappaB like transcription factor. The degradation allows NF-kappaB (NF-kB) like transcription factor, Dif and Dorsal, to translocate into nucleus and activate the transcription of a set of target genes. On the other hand, the Imd pathway has a key function in immune response to infection of gram-negative bacteria. After DAP-type peptide glycan, which is a component of cell wall of gram-negative bacteria, is recognized by peptidoglycan-recognition

protein-LC (PGRP-LC), a transcription factor, Relish, is phosphorylated and cleaved into the active form. As the result, expressions of a group of target genes are triggered (Ferrandon et al., 2007, Lemaitre and Hoffmann 2007) (Figure 1). In addition to these two pathways, JAK/STAT and JNK pathways are also important for immune response to infection of microorganisms in *Drosophila* (Boutros, Agaisse and Perrimon 2002, Agaisse and Perrimon 2004, Delaney et al., 2006, Lemaitre and Hoffmann 2007). The JAK/STAT signaling pathway mainly regulates phagocytosis, hemolymph coagulation and melanization (Agaisse and Perrimon 2004).

AMPs are cationic small secretory peptides that exhibit a wide range of activities against bacteria, fungi and/or viruses, playing an essential role in the innate immune system of *Drosophila* (Lemaitre and Hoffmann 2007). To date, seven AMP families, i.e., Attacin, Cecropin, Defensin, Diptericin, Drosocin, Drosomycin and Metchnikowin, have been identified in *Drosophila melanogaster* (Lemaitre and Hoffmann 2007). According to Sackton et al. (2007), it was indicated by their sequence analysis of the 12 *Drosophila* genomes that only the species belonging to the *melanogaster* species group of the subgenus *Sophophora* had Drosomycin genes. Drosomycin is known to be a major antifungal peptide (Fehlbaum et al., 1994, Lemaitre et al., 1996, Tzou, Reichhart and Lemaitre 2002). This suggests that antifungal immune response varies among different *Drosophila* species and attacks from different bacteria and/or fungi might have produced different immune responses in *Drosophila*. Therefore, it is hypothesized that the differences in the environmental factors caused the difference in the immune system.

For instance, *D. virilis* feeds and breeds on slime flux and decaying bark of trees, which are infected by various bacteria, yeasts and molds. Indeed, many yeasts other than *Saccharomyces cerevisiae* and filamentous fungi, such as *Xanthophyllomyces dendrorhous*, *Cryptococcus* spp., *Fusarium* spp., etc., have been isolated from slime flux and decaying wood (Weber, Davoli and

Anke 2006, Weber 2006), whereas *S. cerevisiae* solely ferments various fruits, which *D. melanogaster* thrives on (Carson 1971, Throckmorton 1975, Markow and O'Grady 2007). From this difference in the microbial community in host materials of *D. virilis* and *D. melanogaster*, it is conceivable that *D. virilis* is exposed to a wider variety of fungi and therefore *D. virilis* has a higher resistance to fungi compared to *D. melanogaster*. To test this hypothesis, I examined the immune response of *D. virilis* and *D. melanogaster* to a fungus species belonging to the genus *Penicillium*. Since *Penicillium* species are commonly found in both slime flux and rotting fruits (Coates and Johnson 1997, Peterson, Bayer and Wicklow 2004), both *D. virilis* and *D. melanogaster* likely have high risk of *Penicillium* infection throughout their developmental stages. To measure resistance of *D. virilis* and *D. melanogaster* to the fungal infection, adult flies of these species were reared on the culture medium that *Penicillium* fungi grew. The results showed that *D. virilis* adult flies survived more than two times longer than *D. melanogaster* flies (Figure 2), suggesting that *D. virilis* has a higher resistance to *Penicillium* infection. This higher antifungal activity without having Drosomycin motivated us to investigate the immune system of *D. virilis*.

In this study, to clarify the immune mechanism responsible for the higher antifungal resistance of *D. virilis*, larval immune response to the fungal infection between *D. virilis* and *D. melanogaster* were compared by means of comparative transcriptome analyses. Using a Roche 454 GS Junior sequencer, I examined the transcriptome of fat body and salivary gland of 3rd-inster larvae with and without infection of a *Penicillium* species. Genes showing different expression pattern in response to the fungal infection between *D. virilis* and *D. melanogaster* were extracted and compared. These genes included those encoding AMPs and 'immune-induced molecule (IM)'. Extensive differences were observed in the expression pattern of already known AMP and IM genes between *D. virilis* and *D. melanogaster*.

10

Additionally, two potential AMP genes were newly identified from function-unidentified genes. Furthermore, three novel putative immune-related genes were identified: the products of them had a homology to an IM, Ras-like GTP binding protein Rho1 involved in many signaling pathways and Ficolin-2 binding to a cell wall component of bacteria and fungi, respectively.

2. MATERIALS and METHODS

2.1 Measurement of antifungal resistance

Twenty to twenty five adult flies 1-day after eclosion were reared at 25 ºC on a cornmeal-malt medium (50 g cornmeal, 50 g malt powder, 40 g dried brewer's yeast, 50 g sucrose, 5 ml propionic acid and 5 g agar in 1 liter water) with and without *Penicillium* fungi. The medium containing *Penicillium* fungi was prepared by inoculating a small amount of spores of a *Penicillium* species (identified by its nucleotide sequence of 18S RNA gene) onto the cornmeal-malt medium and incubated at 20 ºC for a week or more until the surface was completely covered by the growing fungi. After the flies were transferred onto the medium with or without fungi, the number of flies alive was counted every day. To measure the resistance to the infection of the *Penicillium* species, the 50% lethal time (LT50) was estimated by the generalized linear method implemented in R version 2.15.2 software (R Development Core Team 2008). These processes were independently replicated three times.

2.2 Induction of gene expression by fungal-infection

A small amount of *Penicillium*'s spores were inoculated and cultured on a sabouraud dextrose agar (SDA) medium (10 g peptone, 40 g Dextrose and 15 g agar in 1 liter water) at 20 °C for several days until the fungi grew to cover the surface of medium. To prepare the fungus infected larvae, twenty 3rd-instar larvae of *D. virilis* or *D. melanogaster* were reared on the fungus-covered SDA medium for 12 hours at 20 °C. The induction of AMP genes is usually

detected in three hours after the infection and continued at least 24 hours at 25 °C (Vodovar et al., 2005, Glittenberg et al., 2011). However, I reared the larvae at 20 °C to postpone their pupation. The response to the fungal infection was confirmed by the raised expression level of the Metchnikowin gene (*Mtk*) (known antifungal AMP gene) measured by RT-PCR and only the induction confirmed samples were used for the transcriptome sequencing described in the next section. As the control, the naïve larvae were prepared by rearing with the same condition on fungus-free SDA medium.

2.3 Transcriptome sequencing

I analyzed transcriptome of larval fat body and salivary grand. This is because all AMPs were shown to be expressed in fat body and a major antifungal AMP, Drosomycin, was highly expressed in larval salivary gland in *D. melanogaster* (Tzou, De Gregorio and Lemaitre 2002). Larval fat bodies and salivary glands dissected from twenty fungus infected or naïve 3rd-instar larvae were pooled and the total RNA was extracted from these fat bodies and salivary glands by acid-guanidium phenol-chloroform (AGPC) method (Chomczynski and Sacchi 1987). Then, mRNA was isolated by using Dynabeads mRNA purification kit (Invitrogen) according to the supplier's instruction. The complementary DNA (cDNA) library was constructed according to the Roche GS Junior cDNA rapid library preparation protocol with a modification to keep short molecules expected for AMP genes. The double-stranded cDNA was synthesized by using cDNA synthesis system (Roche Diagnostics) with random hexamer primers. The resultant cDNA was purified by using AMPure XP kit (Agencourt) and the end-polished cDNA fragments were ligated with the FAM-labeled RL adaptor included in Lib-L GS FLX Titanium Rapid

Library Preparation kit (Roche Diagnostics). The adaptor-ligated cDNA was then purified by using Agencourt AMPure XP system and finally eluted in 50 μl TE buffer. The cDNA solution was then concentrated by extracting with the equal volume of 2-butanol twice and subsequently with diethyl ether to remove the residual 2-butanol. Instead of the sizing procedure described in the standard protocol, I conducted 2% agarose-gel electrophoresis, excised the gel section containing 200 bp to 1 kb DNA fragments and extracted the cDNA using High Pure PCR Clean-up kit (Roche diagnostics). The quality and quantity of the cDNA was evaluated by using QuantiFluor™-P Handheld Fluorometer (Promega) and Agilent 2100 Bioanalyzer High Sensitivity DNA kit (Agilent Technologies). The pyrosequencing was conducted by using a 454 GS junior sequencer after the emulsion PCR according to manufacturer's instructions (Roche diagnostics).

2.4 Gene prediction for pyrosequencing reads

All the sequence reads obtained from a 454 GS Junior sequencer were filtered by the shotgun full processing of GS Run Processor application with the default setting. The filtered pyrosequencing reads of *D. melanogaster* and of *D. virilis* were queried to the complete mitochondrial genome sequence of *D. melanogaster* (Flybase genome database release 5.46, ftp://ftp.flybase.net/genomes/) and that of *D. virilis* (NCBI; gi 190710421), respectively, by using the stand-alone BLAST 2.2.25+ software (Altschul et al., 1990, Camacho et al., 2009) to remove the reads derived from mitochondrial genes. The reads that did not hit the mitochondrial genome sequence were then queried to *D. melanogaster* ribosomal RNA (rRNA) sequences (NCBI; gi 158246) to remove the reads from rRNA. To identify the gene, from which each read

14

derived, each read was queried against the Flybase *D. virilis* database release 1.2 or *D. melanogaster* database release 5.46 downloaded from Flybase FTP site (ftp://ftp.flybase.net/genomes/), depending on which species it was derived from. Using the stand-alone BLAST 2.2.25+ software, I first queried against the CDS database and the reads that did not hit were subsequently queried against gene and transcript databases (Figure 3a). Finally, the reads that did not hit any target were used for further analyses to search for novel immune -related genes as explained later in the section 2.6

For the genes identified in the *D. virilis* genome, most of them have different names from their orthologues in the *D. melanogaster* genome. In this study, however, I used the gene names of *D. melanogaster* for both species for the ease of comparison between species. The correspondence of gene ID between the two species was according to the 12 *Drosophila* genome analyses (ftp://ftp.flybase.net/genomes/12_species_analysis/clark_eisen/homology/) (*Drosophila* 12 Genomes Consortium 2007). For genes that have multiple IDs corresponding to multiple copies in either or both species, one-to-one correspondence of homologue between the two species was determined by TBLASTN search with the translated protein sequence of *D. virilis* gene as the query against the *D. melanogaster* CDS database. Whether a gene is immune-related or not was determined by referring to the list of *Drosophila* immune-related genes (Sackton et al., 2007). (Figure 3b)

The *D. virilis* genes of unknown function, which did not have homologue in the *D. melanogaster* genome, were further BLAST searched for their homologues in other organisms' genomes (http://blast.ncbi.nlm.nih.gov/) (Altschul et al., 1990). In this homology search, only the genes, for which the number of reads was significantly different between fungus infected and naïve larvae, were used. For the genes that did not hit any homologue in any organism (*D. virilis*-specific genes), their functions were predicted by using domain and motif search

programs available in NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) and Pfam (http://pfam.sanger.ac.uk/) (Figure 3b). When any conserved domain or motif was not predicted, the presence of signal peptide was predicted by using SignalP (v4.0) (Petersen et al., 2011) and ProP (v1.0) (Duckert, Brunak and Blom 2004) programs as a criterion to consider the possibility of antimicrobial peptide. For the candidates with putative signal peptide, the molecular weight, net charge and structural features were computed by using JEMBOSS (v1.5) program (Carver and Bleasby 2003). Finally, from the amino acid sequence of putative mature peptide after removal of the putative signal peptide, the possibility of antimicrobial peptide was examined by AMP prediction web programs, AntiBP2 (Lata, Mishra and Raghava 2009), CAMP (Thomas et al., 2010) and AMPA (Torrent, Nogués and Boix 2009).

2.5 Estimation of gene expression level

To estimate the expression level of each gene, the total number of reads to hit the gene in the BLAST search was counted (Figure 3b). To calibrate the difference in transcript length among different genes, the number of reads counted was then standardized to be the number of reads per site per million reads (RPSM) as follows.

$$RPSM = \text{Number of reads} / \text{Total number of reads} / \text{Transcript length} \times 1{,}000{,}000$$

I further normalized RPSM to take the difference in total gene expression level between the samples into account and computed Trimmed Mean of M values (TMM) (Robinson and

Oshlack 2010), using TCC package implemented in R version 2.15.2 software (R Development Core Team 2008, Sun et al., 2013). For each gene, the TMM for the fungus infected larvae was compared to that for the control naïve larvae to quantify the extent of gene expression change in terms of the induction coefficient (IC) as follows.

$$IC = TMM \text{ of the infected larvae} / TMM \text{ of the naïve larvae}$$

To test a statistical significance of the induction, the difference in the number of actual reads was compared between the fungus infected and naïve larvae. In this test, ribosomal protein L32 (*RpL32*) and glyceraldehyde 3 phosphate dehydrogenase (GAPDH) genes were used as endogenous control genes. Although Actin was also a well-known endogenous control gene, Actin was reported to play an important role in phagocytosis against fungi in *Drosophila* S2 cell (Stroschein-Stevenson et al., 2006) and that the expression of an actin gene (*Act42A*) of *D. melanogaster* 3rd-instar larvae was induced by *Saccharomyces cerevisiae* contained in the culture medium (Gershman et al., 2007). Indeed, the expression of *D. melanogaster Act42A* was not detected in the control naïve larvae but in the fungus infected larvae (the number of reads was 6 and TMM = 0.0619). Therefore, only *RpL32* and GAPDH genes were used as the endogenous control genes in this study. Since the homogeneity of the numbers of reads for the two genes between the fungus infected and the naïve larvae was statistically supported ($P = 0.14$ in *D. virilis* and $P = 0.51$ in *D. melanogaster* by Fisher's exact test, Supplementary Table 1), the total number of reads derived from the two genes was used as the number of reads for the endogenous control genes. Finally, the difference in the number of reads between the fungus infected larvae and the naïve larvae was tested on the 2x2 contingency table with the numbers for the endogenous control genes by Pearson's chi-square test or Fisher's exact test dependent

on whether the minimum number of reads was five or more or not.

2.6 Prediction of new immune-related genes in *D. virilis*

The pyrosequencing reads which were derived from the fungus infected *D. virilis* but not matched any known gene were subjected to predict a new gene (Figure 3c). These pyrosequencing reads were mapped to the *D. virilis* genome sequence by Newbler GS reference mapper software (Roche Diagnostics) with the default parameter setting designated for CDS sequences to obtain continuous transcript sequences. Since the median length (192 bp) of the obtained contigs was similar to that (230 bp) of 3'-UTR of *D. melanogaster* (Sackton and Clark 2009), many contigs might not include protein coding region at all. Therefore, for each contig, the corresponding genome sequence plus 250 bp each of its upstream and downstream flanking regions were extracted to build a query sequence to search for new gene. All the query sequences obtained were subjected to BLASTX search against Swissprot protein database downloaded from the Uniprot web site (http://www.uniprot.org/downloads) with the condition of e-value <= 1E-05. For the identified putative genes, the difference in the number of reads was statistically tested between the fungus infected and the naïve larvae in the same way as that for the known genes described above and if the number of reads was significantly different, then the gene ontology was analyzed by STRAP software (v1.1.0.0) (Bhatia et al., 2009).

2.7 Pyrosequencing and data analyses of oligo-capped full length cDNA

The 5'-end sequences of the new immune-related genes described in section 2.3 was determined by the BAP-TAP method (Maruyama and Sugano 1994, Suzuki et al., 1997) (Figure 4). Total RNA extraction and mRNA purification from twenty *Penicillium*-fungus infected 3rd-instar larvae of *D. virilis* were performed by the same way as described in section 2.3. The purified mRNA was treated by 2 U bacterial alkaline phosphatase (BAP) (Nippon Gene) in a 50 μl mixture containing 10 mM Tris-HCl and 0.1 mM $MgSO_4$ at 37 °C for 1 hour. After the reaction, BAP was removed by Phenol-Chloroform purification. The BAP treated mRNA was then treated with 45 U tobacco acid pyrophosphatase (TAP) (Nippon Gene) in a 50 μl mixture containing 5 mM Sodium Acetate (pH 5.5), 0.5 mM EDTA (pH 8.0), 1 mM 2-Mercaptoethanol at 37 °C for 1 hour and then TAP was removed by Phenol-Chloroform purification The BAP-TAP treated mRNA was ligated with 400 ng 5' RNA adaptor designed for 454 GS Junior sequencer (Table 1) by 40 U T4 RNA ligase in 50 μl mixture containing 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol (DTT), 1 mM ATP and 0.01 % bovine serum albumin (BSA) at 16 °C for 3 hours. The resultant oligo-capped mRNA was purified by Phenol-Chloroform extraction, and then treated with 2 U DNase I (Invitrogen) in 20 μl of 20 mM Tris-HCl (pH 8.4), 2 mM $MgCl_2$, 50 mM KCl at room temperature for 15 minutes. The reaction was terminated by adding 2 μl 25 mM EDTA and incubated at 65 °C for 10 minutes. The first-strand cDNA of the oligo-capped mRNA was synthesized in a 40 μl mixture containing the DNase I-treated oligo-capped mRNA, 50 mM Tris-HCl (pH 8.3 at room temperature), 1X First-Strand buffer (75 mM KCl, 3 mM $MgCl_2$), 2.5 μM oligo-$(dT)_{15}$ added 3' adaptor primer (Table 1), 0.5 mM dNTP mix, 5 mM DTT and 400 U SuperScript III Reverse Transcriptase (Invitrogen). The oligo-capped mRNA, the oligo-$(dT)_{15}$ + 3' adaptor primer and the dNTP mix was firstly mixed and incubated at 65 °C for 5 minutes. Then, the primer mixed reaction was placed on ice more than 2 minutes. Finally, the first-strand buffer and the reverse

transcriptase were added into the primer mixed reaction and incubated at 50 °C for 50 minutes. The reverse transcription reaction was terminated by placing at 70 °C for 15 minutes.

Double-stranded oligo-capped cDNA was synthesized by using 1.25 U Taq polymerase in a 50 μl mixture containing 1X PCR buffer (50 mM KCl, 10 mM Tris-HCl (pH 8.5), 2 mM $MgCl_2$ and 0.001 % gelatin), 0.2 mM dNTP mix, 250 nM FAM-labeled 5' adaptor primer and the first-strand cDNA as the template (Table 1). The reaction was conducted by incubation at 95 °C for 2 minutes followed by 20 cycles of 95 °C for 15 seconds and 68 °C for 2 minutes. Then, the 3' adaptor primer was added into the reaction to amplify the synthesized double-stranded oligo-capped cDNA by PCR. The PCR amplification was performed with 25 cycles of 95 °C for 15 seconds, 68 °C for 2 minutes, and a final extension at 72 °C for 5 minutes. To purify the PCR products, the 3-fold volume of Binding Buffer (5 M Guanidine Thiocyanate; 100 mM Tris-HCl (pH 6.6) and 10 μl silica particles (5 μm diameter) suspended in 0.01 N HCl were added into the PCR products. The mixture was incubated at room temperature for 5 minutes and centrifuged at 12,000 rpm for 1 minute. After the supernatant was removed, the precipitated silica particles were washed twice with Wash Buffer (10 mM Tris-HCl (pH 7.5), 100 mM NaCl : Ethanol = 1 : 4). Finally, the purified PCR products (oligo-capped cDNA library) were eluted in 50 μl TE buffer. Concentrating, sizing and 454 GS Junior sequencing of the PCR products were conducted by the same way as described in section 2.3 with 5' adaptor primer (Figure 4).

The obtained 5'-end-enrichd pyrosequencing reads were assembled by using Newbler GS De Novo Assembler software (Roche Diagnostics) with the default parameter setting designated for CDS sequences. The nucleotide sequence of each contig obtained in section 2.3 was queried against the assembled 5'-end-enriched sequences by using the stand-alone BLAST 2.2.25+ software (Camacho et al., 2009). In the obtained full length cDNA, the protein-coding region was predicted by using getorf program implemented in EMBOSS (Carver and Bleasby 2003).

20

Finally, the secondary structure of the obtained protein sequence was predicted by using Disulfind and Jpred programs (Ceroni et al., 2006, Cole, Barber and Barton 2008).

2.8 Real-time reverse transcriptase PCR (RT-PCR)

The total RNA was extracted by AGPC method (Chomczynski and Sacchi 1987) from pooled fat bodies, salivary glands and guts dissected from ten fungus infected or naïve 3rd-instar larvae. The first-strand cDNA was synthesized from 1 μg of total RNA by the same way as described in section 2.7 except that the reaction was conducted in a half volume of mixture and 2.5 μM oligo-(dT)$_{28}$ primer was used for the first-strand cDNA synthesis instead of the oligo-(dT)$_{15}$ + 3' adaptor primer. The same reaction without the reverse transcriptase was conducted to verify the absence of genomic DNA. Real-time RT-PCR was conducted by using StepOne PLUS real-time PCR system (Applied Biosystems). The amplification of the PCR product was detected by SYBR Green I (Camblex Bio Science, Rockland). Primers were designed for the amplicon size to be less than 150 bp. The primers were listed in Table 2. In this analysis, *RpL32* was used as the endogenous control to normalize gene expression level. An 20 μl PCR mixture was contained 1X PCR buffer (50 mM KCl, 10 mM Tris-HCl (pH 8.5), 2 mM MgCl$_2$ and 0.001 % gelatin), 0.2 mM dNTP Mix, 250 nM gene-specific forward and reverse primers, one 20 thousandth diluted SYBR Green I, and 0.67 U Taq polymerase. The real-time RT-PCR amplification was conducted under the condition of 95 °C for 15 seconds followed by 40 cycles of 95 °C for 15 seconds, 62 °C for 20 seconds and 72 °C for 20seconds. The melting-curve analysis was then performed under the condition of 95 °C for 15 seconds, 60 °C for 1 minutes and then slow heating at 0.3 °C per second up to 95 °C. The obtained gene expression levels

were compared by the comparative Ct analysis method (Livak 1997) between the fungal infected and naïve samples. The gene expression level was measured in three biological repetitions with two technical replications.

3. RESULTS

3.1 Difference in antifungal resistance between *D. virilis* and *D. melanogaster*

To compare antifungal resistance between *D. virilis* and *D. melanogaster*, adult flies of these species were reared on a culture medium harboring *Penicillium* fungi and their survival time was measured. The results showed that the *D. virilis* flies survived more than two times longer than the *D. melanogaster* flies did (Figure 2); the average 50% lethal times (LT50) of *D. virilis* and *D. melanogaster* flies were 6.04 days and 1.75 days, respectively, whereas their survival time on the normal culture medium without fungi was much longer (LT50 >> 10 days). This suggests that *D. virilis* has a higher resistance to the infection of the *Penicillium* species than *D. melanogaster* at the adult stage.

3.2 Summary of transcriptome analysis

Many AMP genes encode relatively short peptides less than 100 amino acids long. Therefore, to avoid the loss of sequences derived from such short transcripts, the 454 GS junior sequencing was adjusted for cDNA library containing cDNA fragments longer than 200 bp long, whereas the standard sizing procedure selects DNA fragments of 600 - 900 bp long on average by removing those shorter than 350 bp long to be less than 10%. This resulted in 109,106 reads with the average length of 226 bp and 119,533 reads with the average length of 217 bp from the fungus infected and the naïve (uninfected) *D. virilis* larvae, respectively (Table 3). On the other hand, 110,578 reads with the average length of 242 bp and 91,947 reads with the average length

of 219 bp were obtained from the fungus infected and the naïve (uninfected) *D. melanogaster* larvae, respectively (Table 3).

After removing the reads derived from mitochondrial genes and rRNA genes, the total numbers of the remaining reads were 77,558 and 90,836 for the fungus infected and naïve *D. virilis* larvae, respectively, and 65,670 and 48,474 for the fungus infected and naïve *D. melanogaster* larvae, respectively. They were thought to be derived from mRNA transcribed from nuclear protein-coding genes. For 55,358 and 62,110 out of the 77,558 and 90,836 reads, respectively, I found BLAST hits for 5,155 and 4,709 genes, respectively, in *D. virilis*, whereas for 63,555 and 46,536 out of the 65,670 and 48,474 reads, respectively, I found BLAST hits for 4,735 and 4,275 genes, respectively, in *D. melanogaster*. It is noteworthy that the numbers of the remaining reads for *D. virilis* were 22,200 (fungus infected) and 28,726 (naïve), which were more than ten times as many as the corresponding 2,115 (fungus infected) and 1,938 (naive) for *D. melanogaster* (Table 3).

3.3 Expression pattern of immune-related genes

According to Sackton et al. (2007), innate immune system is categorized into three functional classes, 'recognition', 'signaling' and 'effector.' In the *D. virilis* transcriptome analysis, 128 immune-related genes were detected, in which 23, 68 and 37 were assigned to recognition, signaling and effector classes, respectively (Table 4, Supplementary Table 2 and Figure 5). In the case of the *D. melanogaster* transcriptome, 129 immune-related genes were detected, in which 28, 62 and 39 genes were assigned to recognition, signaling and effector classes, respectively (Table 5, Supplementary Table 3, Figure 5). Among the immune-related

genes, many of recognition and signaling class genes expressed in the fungus infected larvae were present in both *D. virilis* and *D. melanogaster* (Figure 5). In the recognition class genes, *PGRP-SA*, *PGRP-LC*, *PGRP-LE* and *GNBP3* involved in Toll and Imd pathways were expressed in both species (Figure 1, Supplementary Tables 2 and 3). The expression of genes for nimrod and complement-like proteins called thioester-containing proteins (TEPs), which activate cellular immune response such as phagocytosis, were also detected in both species. Among the TEP genes, *TEPII* (IC = 5.359, *P* = 4.68E-22) and *TEPIV* (IC = 2.515, *P* = 8.24E-05) were significantly up-regulated in *D. melanogaster* (Table 5, Supplementary Table 3), whereas the expressions of their homologs in *D. virilis* were not induced by the fungal infection (Table 4, Supplementary Table 2). I also detected the genes for negative regulators of systematic immune response, such as *PGRP-SC1a*, *PGRP-SC2* and *PGRP-LB* (Mellroth, J. Karlsson and Steiner 2003, Bischoff et al., 2006, Zaidman-Remy et al., 2006, Paredes et al., 2011), as well as the genes for activators. Consistent with the expression of these recognition class genes, the expressions of signaling class genes, e.g., Myd88, Rel, STAT92E, hep, etc., involved in Toll, Imd, JNK and JAK/STAT pathways, were also detected in both species (see Tables 4 and 5, Supplementary Tables 2 and 3 for details).

3.4 Between-species differences in the expression pattern of effecter class genes

Since the effectors directly function against infected microbes, in this study I focus on the response of the effector class genes to the *Penicillium* infection to elucidate the differences in the antifungal resistance between *D. melanogaster* and *D. virilis*. In contrast to the shared expression pattern between the species observed in the recognition and signaling class genes,

substantial differences in the expression pattern were observed in the effector class genes.

AMPs are known to be a major effector that has a critical role in the innate immune system of *Drosophila* (Tzou, Reichhart and Lemaitre 2002). In *D. melanogaster*, 20 AMP genes belonging to seven AMP gene families have been found, whereas 15 AMP genes belonging to five AMP gene families have been identified in *D. virilis* (Drosocin and Drosomycin in *D. melanogaster* are missing in *D. virilis*) (Sackton et al., 2007). In both *D. virilis* and *D. melanogaster*, many AMP genes (11 of 15 in *D. virilis* and 14 of 20 in *D. melanogaster*) were expressed in the fungus infected larvae (Tables 4 and 5, Supplementary Tables 2 and 3). In *D. virilis*, genes encoding Diptericin (*GJ19916*, TMM = 3.812), Defensin (*GJ22479*, TMM = 2.445) and Cecropin (*Cec2B*, TMM = 1.604 and *Cec3*, TMM = 1.475) showed high TMM and Diptericin (*GJ19916*) was most highly expressed in the fungus infected larvae (Table 4). In contrast, the expression level of Metchnikowin (*GJ22469*), which was the only known antifungal peptide in *D. virilis*, was not so high (TMM = 0.660; Table 4). In contrast, Drosomycin (*Drs*) and Metchnikowin (*Mtk*), which were known as antifungal peptide genes, were most strongly expressed in the fungus infected *D. melanogaster* larvae (TMM = 23.817 and 23.719, respectively), followed by Diptericin (*Dpt*, TMM = 11.568), Attacin (*AttC*, TMM = 4.684) and Drosocin (*Dro*, TMM = 4.237) (Table 5). Among the Drosomycin gene family, only *Dro5* responded to the fungal infection, suggesting that *D. melanogaster* uses the specific Drosomycin gene copy against the *Penicillium* species. However, the expression level of *Dro5* was 100-fold lower than that of *Drs* (TMM = 0.276) (Table 5). These observations indicate substantial differences in the AMP usage between the species, i.e., against the fungal infection, Diptericin, Defensin and Cecropin were the three major AMPs in *D. virilis*, whereas Drosomycin and Metchnikowin were the two major AMPs in *D. melanogaster* (Figure 6).

Among other effector class genes, the immune-induced molecule (IM) genes showed

distinct expression pattern between the species. The IM genes are known as the genes induced by bacterial or fungal infection in *D. melanogaster*. However, their functions mostly have not been characterized. In this study, 10 IM genes were identified to be expressed in the fungus infected *D. melanogaster* larvae and five of them, *IM1*, *IM4*, *IM10*, *IM14* and *IM18* were significantly up-regulated by 2-fold or more (Table 5 and Supplementary Table 3). For most of the *D. melanogaster* IMs, their expressions tended to be induced by the fungal infection. On the other hand, five IM genes, *IM1* (*GJ19885*), *IM4* (*GJ18607*), *IM10* (*GJ21308*, *GJ21309*) and *IM23* (*GJ22454*), were identified to be expressed in *D. virilis,* but their expression tended to be down-regulated by the fungal infection (Table 4, Supplementary Table 2). Especially, the expressions of *IM1* (*GJ19885*), *IM4* (*GJ18607*) and *IM10* (*GJ21308*) were significantly reduced by the fungal infection by half or less (Table 4). These differences in the expression pattern may indicate that IMs play separate roles in the immune response to fungal infection in *D. melanogaster* and *D. virilis*.

3.5 Novel AMP genes in ~~the annotated~~ *D. virilis* ~~genes~~

Using the BLAST search against all the known *D. melanogaster* genes, I could not find the homologues for three *D. virilis* annotated genes significantly up-regulated by the fungal infection. They were *GJ10737* (IC = 2.503, *P* = 0.0037), *GJ11722* (IC = 3.198, *P* = 0.032) and *GJ18291* (IC = 3.909, *P* = 0.047). Additional queries to orthologue database (orthoDB: http://cegg.unige.ch/orthodb6) (Waterhouse et al., 2012) and the non-redundant gene database in the NCBI BLAST web server failed to find any known gene, suggesting that they were *D. virilis*-specific genes. Although I further searched for annotated domains and motifs in the

expected products of these genes using the domain and motif search programs on NCBI Conserved Domain Database and Pfam, no conserved domain or motif was predicted. However, using SignalP (v4.0) (Petersen et al., 2011), ProP (v1.0) (Duckert, Brunak and Blom 2004) and JEMBOSS (v1.5) (Carver and Bleasby 2003) programs, the expected products of *GJ10737* and *GJ18291* were predicted to be secretory peptides having propeptide sequences and positively charged mature peptide (Table 6). These features are commonly found in AMPs. Indeed, AMP prediction web programs, CAMP (Thomas et al., 2010) and AMPA (Torrent, Nogués and Boix 2009), predicted them to be AMPs, although another program, AntiBP2 (Lata, Mishra and Raghava 2009), did not (Table 6). These results suggested the possibility that *D. virilis* possesses unknown AMP genes functioning in its innate immune system.

3.6 Novel immune related genes in *D. virilis*

In our BLAST analysis described above, 22,200 and 28,726 pyrosequencing reads respectively from the fungal infected and naïve *D. virilis* larvae did not hit any known gene, whereas such reads were only 2,115 (infected) and 1,938 (naive) in *D. melanogaster* (Table 3). I hypothesized that this is because there were many unidentified genes in *D. virilis*. To examine whether or not these reads were derived from unidentified immune related genes, I assembled these reads by mapping each read onto the *D. virilis* genome sequence to make contigs. Then, I performed a BLASTX search against Swissprot protein database using each of these contigs as the query.

Out of the 22,200 reads, 21,488 (about 97%) were mapped onto the *D. virilis* genome sequence to be assembled to 3,269 contigs of the average length 237 bp in total (Figure 7). This

28

indicates that these reads were actually derived from transcripts of the *D. virilis* genome rather than possible contaminants and that there are unidentified transcription units potentially encoding polypeptide. Since most of the contigs were shorter than the median length of 3'-UTR of *D. melanogaster* genes, I extended each contig with 250 bp each of upstream and downstream genome sequences to make a query sequence subjected to the BLAST search against Swissprot protein database. As the result, I identified 620 putative genes in the 3,269 contigs. Among them, 27 putative genes showed a statistically significant difference in the number of reads between the fungus infected and naïve larvae. Three out of the 27 putative genes, *PG00034*, *PG01778* and *PG02420*, were assigned to potential immune-related genes for subsequent GO analysis (Supplementary Table 4). *PG00034* was homologous to *IM14* of *D. melanogaster*. Although the expression of *IM14* was significantly up-regulated in *D. melanogaster* (Tables 5 and 7), the expression of *PG00034* was significantly down-regulated by the fungal infection in *D. virilis*. *PG01778* was homologous to a Ras-like GTP-binding protein, *Rho1*, of *D. melanogaster*. This gene is known to play a role in regulating actin genes involved in phagocytosis (Hariharan et al., 1995, Magie et al., 1999, Greenberg and Grinstein 2002, Magie and Parkhurst 2005). The expression was observed only in the infected larvae in *D. virilis* and induced by the fungal infection (IC = 2.020) in the *D. melanogaster* larvae, indicating that this gene was up-regulated by the fungal infection in both species. *PG02420* was homologous to *Ficolin-2* that binds to the cell wall component of bacteria and fungi (Ma et al., 2004, Endo, Matsushita and Fujita 2007), and the expression of *PG02420* was significantly down-regulated in the infected *D. virilis* (IC = 0.208) (Table 7).

For the remaining 2,649 contigs, I did not find any homologue in Swissprot protein database. However, among the 2,649 contigs, the number of pyrosequencing reads was significantly different between the fungal infected and naïve larvae in 64 contigs and 26 of them

were considered to be up-regulated by the fungal infection. In order to predict the protein-coding region for these 26 contigs, I tried to determine the 5'-end sequence by the oligo-capping method adjusted for the use of 454 GS Junior sequencer. As the results, 50,573 reads with the average length of 190 bp were obtained from the infected *D. virilis* larvae, and these reads were assembled to construct continuous transcript sequences. Out of the 50,573 reads, 41,423 (about 82%) were assembled to 900 contigs (Table 8). Combining the 900 contigs with the assembled sequences in the section 2.3, I determined 5'-end sequence of 8 contigs. Two of them, *PG00667* and *PG01875*, were identified to be *GJ11849* and *GJ22451* by NCBI BLAST analysis, respectively. *GJ11849* was ribosomal protein L26 and *GJ22451* was a homologue of *IM3* of *D. melanogaster*. Interestingly, the expression of *GJ22451* was up-regulated (IC = 10.396, Table 9), whereas the expressions of all other IM gene homologues were tend to be down-regulated in the fungal-infected *D. virilis*. In this study, it was predicted that *GJ22451* encoded small secretary peptide having weak positive net charge (net charge = 0.5). In addition, *GJ22451* was predicted to have two beta-sheets and to be stabilized by a disulfide-bridge between two cysteine residues (Figure 8). Since these features are found in some AMPs, whether *GJ22451* could function as AMP was evaluated by using AMP prediction programs. As the result, AntiBP2 AMP prediction program predicted that *GJ22451* was similar to beta-Defensin of mammals (Table 9).

Since the remaining 6 contigs did not show similarity to any known genes, I predicted the protein-coding region using getorf program implemented in EMBOSS (Carver and Bleasby 2003). As the result, it was predicted that *PG01471* encoded a proline-rich and positively-charged secretory peptide, which is the features often observed in AMPs. Indeed, *PG01471* was predicted to be AMP by AntiBP2 AMP prediction program (Table 9). Particularly, C-terminal region of the *PG01471* was very similar to that of Metchnikowin (Figure 9). Since

30

any sequence showing homology to *PG01471* was not found in other *Drosophila* genome sequences excepting *D. mojavensis*, *PG01471* seemed be a lineage-specific AMP gene, which may contribute to different antifungal resistance between *D. virilis* and *D. melanogaster* (Figure 10).

3.7 Local expression of Defensin gene in response to the fungal infection

My transcriptome analysis indicated that *D. virilis* uses a specific Defensin in response to the fungal infection. However, while my transcriptome analysis particularly focused on systemic immune response to the fungal infection, local immune response is also important to defend from the infection (Tzou et al., 2000, Liehl et al., 2006). Therefore, I surveyed differences in expression responses of Metchnikowin and Defensin genes among fat body, salivary gland and gut of the fungal-infected *D. virilis* larvae by real-time RT-PCR. The analysis for fat body and gut was conducted three biological replications, whereas no replication was made for salivary gland, because the tissue is too small to extract enough amount of RNA. In the fat body, strong induction of Defensin gene (*GJ22479*) expression was observed in all replicates, consistent with the result of the transcriptome analysis. Contrary, the expression of Metchnikowin gene (*GJ22469*) was not always induced by the fungal infection (Figures 11). Additionally, expression level of *GJ22479* was tended to be higher than that of Metchnikowin (*GJ22469*) (Figure 11). In the salivary gland, the expressions of both Metchnikowin (*GJ22469*) and Defensin (*GJ22479*) genes were induced by the fungal infection (Figure 12). The expression of another Defensin gene (*GJ21126*) was observed in neither the fat body nor salivary gland (Figures 11and 12). Interestingly, *GJ21126* was locally expressed in the gut, and the expression

31

was tended to be up-regulated by the fungal infection, whereas the expression of *GJ22479* gene was tended to be down-regulated by the fungal infection in the gut (Figure 13). Nevertheless, the expression level of *GJ21126* was very low compared to that of *GJ22479* even in the gut (Figure 13). These results support the hypothesis that *D. virilis* uses mainly one of the Defensin genes, *GJ22479*, against the fungal infection.

4. DISCUSSION

In this study, I first clarified that the antifungal resistance against *Penicillium* fungal infection is higher in *D. virilis* than in *D. melanogaster*. In general, adult flies of most *Drosophila* species are attracted to, feed and breed upon a variety of fermenting substances such as fallen fruit and flowers, slime fluxes of forest trees, decaying bark of trees, mushrooms, etc. (Carson 1971). However, there are inter-species variations of the fermenting substances utilized by *Drosophila* species for feeding and breeding. For instance, *D. virilis* is known to feed on slime flux and decaying bark of tree harboring many yeasts and filamentous fungi, such as *Xanthophyllomyces dendrorhous*, *Cryptococcus* spp., *Fusarium* spp., etc. (Weber, Davoli and Anke 2006, Weber 2006), whereas *D. melanogaster* feeds on fermented fruits, which mainly harbor Baker's yeast, *Saccharomyces cerevisiae* (Carson 1971, Throckmorton 1975, Markow and O'Grady 2007). The *Penicillium* species is ubiquitously and abundantly found in natural environment, where *Drosophila* species live, and grow on both decaying woods and fruits (Coates and Johnson 1997, Peterson, Bayer and Wicklow 2004). Therefore, both *D. virilis* and *D. melanogaster* are likely to be infected by them in nature during their life time. According to the theory of evolutionary adaptation, the higher antifungal resistance of *D. virilis* observed in this study (Figure 2) is expected to reflect the result of higher risk of the infection in their living environments over the evolutionary time compared to *D. melanogaster*. This raises the question of the immune mechanism attributed to the higher antifungal resistance of *D. virilis*, and it is thought to be a key factor for understanding the adaptive evolution of *D. virilis* to its habitat in moldy environment. To answer this question, I compared the immune response to the fungal infection between *D. virilis* and *D. melanogaster* by analyzing their transcriptome extracted from larval salivary gland and fat body. Although the antifungal resistance was compared at the

adult stage, I focused on the transcriptome at the larval stage. Since the larvae live and feed on fermented substances in their habitat environment and cannot escape from the surrounding microbes as the adults fly away, the larvae are consistently infected by microbes. Therefore, I assume that the resistance at the larval stage is more important for their adaptation to environment. Unfortunately, it was difficult to measure the antifungal resistance at the larval stage since the larvae became pupae within several days and some larvae avoided immediate infection of fungi by digging the medium deeply. Accordingly, my interpretation in the following is on the basis of the assumption that the resistance at the adult stage correlates with the resistance at the larval stage.

My comparative transcriptome analysis revealed that the genes involved in all major signaling pathways for immune response, i.e., Toll, Imd, JAK/STAT and JNK, were triggered by the infection of the *Penicillium* species in both *D. virilis* and *D. melanogaster* (Tables 4 and 5, Supplementary Tables 2 and 3). These pathways regulate humoral and cellular immune responses, such as AMP production, phagocytosis, etc. (Lemaitre and Hoffmann 2007, Agaisse and Perrimon 2004, Kallio et al., 2005). Among the signaling pathways, the Toll pathway plays an essential role against fungal infection in *D. melanogaster* (Lemaitre et al., 1996, Lemaitre, Reichhart and Hoffmann 1997). The Toll pathway regulates expressions of two antifungal peptides, Drosomycin and Metchnikowin (De Gregorio et al., 2002). Consistent with this fact, the expression levels of Drosomycin and Metchnikowin genes were highest in the fungus infected *D. melanogaster* larvae (Table 5). The response of these AMP genes to the infection of an entomopathogenic fungus, *Beauvaria bassiana*, was highest in adult *D. melanogaster* as well (De Gregorio 2001, Irving et al., 2001). Interestingly, seven genes encoding Drosomycin have been found in *D. melanogaster* genome (*Drs*, *Drsl*, *Dro2*, *Dro3*, *Dro4*, *Dro5* and *Dro6*) (Sackton et al., 2007). Nevertheless, I found that only *Drs* and *Dro5* were induced by the fungal infection

in the *D. melanogaster* larvae (Table 5). This specificity of the expression pattern was consistent with the result of the microarray analysis by De Gregorio et al. (2001), suggesting that the specific genes, *Drs* and *Dro5*, are used against the fungal infection at both larva and adult stages. In contrast, any Drosomycin gene is absent in the *D. virilis* genome and the expression of the Metchnikowin gene (*GJ22469*) was not high (TMM = 0.660) compared to that of other AMP genes in the fungus infected *D. virilis* larvae (Table 4, Supplementary Table 2, Figure 6). This result was rather unexpected since Metchnikowin was the only known antifungal peptide in *D. virilis*, suggesting that Metchnikowin of *D. virilis* does not compensate for the lack of Drosomycin. Since the comparison of *D. melanogaster* and *D. virilis* genomes revealed that *Mtk* is present as a single copy gene in both species (Sackton et al., 2007), it is implausible that *D. virilis* has an additional copy of *Mtk* responsible for the observed higher antifungal resistance.

On the other hand, the genes encoding Diptericin (*GJ19916*), Defensin (*GJ22479*) and Cecropin (*Cec2B* and *Cec3*) were highly expressed (TMM = 3.812, TMM = 2.445, TMM = 1.604 and TMM = 1.475, respectively) in the fungus infected *D. virilis* larvae compared to other AMP genes (Table 4), suggesting a substantial difference in the AMP usage in response to the fungal infection between the two species and a possibility that Diptericin, Defensin and Cecropin have an antifungal function in *D. virilis*. The antifungal activity of Diptericin and Defensin against an ascomycete fungus, *Fusarium oxysporum*, has been reported, although they are not effective against other fungi (*Neurospora crassa*, *Beauvaria bassiana* and *Aspergillus fumigatus*) in *D. melanogaster* (Tzou, Reichhart and Lemaitre 2002). Comparing the Diptericin protein sequence of *D. virilis* to its orthologue in *D. melanogaster,* I found substantial amino acid differences (50-70%) (Figure 14). This may indicate the possibility that Diptericin of *D. virilis* has a different activity spectrum against fungi from that of *D. melanogaster*, although the main activity of the latter is not antifungal but antibacterial (Wicker et al., 1990). In contrast,

35

amino acid sequences of mature peptide from *Cec2B* and *Cec3* of *D. virilis* are almost identical (92.5-100%) to those of Cecropin of *D. melanogaster*, and the few amino acid substitutions observed are all conservative to maintain physicochemical properties of the peptide (Figure 15). Therefore, it is likely that the functions of Cecropin are conserved in the two species. A notable difference was observed in the Defensin gene. Defensin is known to be an AMP of main specificity to gram-positive bacteria in *D. melanogaster* (Dimarcq et al., 1994). However, the *Drosophila* Defensin is classified into Defensin_2 superfamily (Pfam: PF01097), which has antifungal activity in mosquito (*Anopheles gambiae*) and sand fly (*Phlebotomus duboscqi*) (Vizioli et al., 2001, Boulanger et al., 2004). *D. virilis* has two Defensin genes (*GJ21126* and *GJ22479*). The mature peptide sequence translated from *GJ21126* is closely related to the *D. melanogaster* Defensin gene as expected from their phylogenetic relationship of species, whereas the mature peptide sequence translated from *GJ22479* is more similar to those of *Anopheles gambiae* (AgaDef) and *Phlebotomus duboscqi* (PduDef), which have antifungal activity (Figures 16 and 17). In my transcriptome analysis for fat body and salivary gland, I detected the expression of *GJ22479* but not *GJ21126* in response to the *Penicillium* infection. This result was confirmed by real-time RT-PCR analysis. However, although the expression level was much lower than that of *GJ22479*, the expression of *GJ21126* was detected in gut (Figure 13). This observation suggests the possibility that the functions of the two Defensin genes have been differentiated through *D. virilis* evolution. A possible speculation based on these observations is that Defensin functions differently as an antifungal peptide in *D. virilis* from that in *D. melanogaster*. Since the expression of these three AMPs are under the regulation of the Imd pathway rather than the Toll pathway (Imler and Hoffmann 2000, De Gregorio et al., 2002), this result suggests that the Imd pathway plays an important role in the response to the fungal infection in *D. virilis*, in contrast to the fact that the Toll pathway is more important to

regulate the Drosomycin genes as the antifungal response in *D. melanogaster*. Alternatively, the Diptericin, Defensin and Cecropin genes may be under the Toll pathway regulation in *D. virilis*. To examine this possibility, I analyzed the upstream region of these genes to see differences in the binding sites of NF-kB-like transcription factors, DIF, Dorsal and Relish between *D. virilis* and *D. melanogaster*. In addition to these binding sites, I also compared the binding site of a GATA factor, Serpent, which regulates synergistically the expressions of AMP genes with the NF-kB-like transcription factors (Senger et al., 2004). Senger et al. (2004) discussed that the organizations of these transcription factor binding sites of AMP genes were related to whether the Toll or Imd pathway had main effect on their expression regulation. However, there was no clear difference in the number, position and direction of these binding sites, suggesting that the alternative possibility is not likely (Figures 18, 19 and 20, Appendices I and II).

A striking difference in the expression pattern was observed in the immune-induced molecule (IM) genes. The IM genes of *D. melanogaster* showed a similar expression pattern to that observed in the previous study conducted by De Gregorio et al. (2001). In this study, ten IM genes were expressed in the fungus infected *D. melanogaster* larvae and five of them, *IM1*, *IM4*, *IM10*, *IM14* and *IM18*, were significantly up-regulated by 2-fold or more and down-regulated gene was not observed (Table 5, Supplementary Table 3). Similar inductions of IM genes were observed in adult flies by the infection of *B. bassiana* (De Gregorio 2001). This suggests that the IM genes play a similar role in antifungal immunity in larvae and adults of *D. melanogaster* and against *Penicillium* and *Beauvaria* fungi, although the function of the IM genes has not been characterized. However, the IM genes showed contrary expression pattern in *D. virilis*: the expressions of five IM genes, *IM1* (*GJ19885*), *IM4* (*GJ18607*), *IM10* (*GJ21308*, *GJ21309*) and *IM23* (*GJ22454*), detected in *D. virilis* were rather down-regulated by the fungal infection (Figure 6). Indeed, three of them, *IM1* (*GJ19885*), *IM4* (*GJ18607*) and *IM10* (*GJ21308*),

showed statistically significant reductions (Table 4, Supplementary Table 2). This result suggests differences in the functions of IMs between *D. virilis* and *D. melanogaster*. In other words, the definition of immune-induced molecule (IM) holds true in *D. melanogaster* but not necessarily so in other *Drosophila* species. It can be speculated that *D. virilis* may have other immune-related genes that have the functions of IMs in *D. melanogaster*. Based on the comparative transcriptome analysis using bacterial-infected *D. melanogaster* and *D. virilis* flies, Sackton and Clark (2009) suggested that new components were recruited into the immune system of *D. virilis*. Therefore, my results as well as their observation motivated us to search for novel immune-related genes in *D. virilis*.

In our transcriptome analysis, I found that three *D. virilis*-specific genes were induced by the fungal infection and two of them, *GJ10737* and *GJ18291*, were predicted to encode novel AMPs (Table 6). This suggests that *D. virilis* has acquired lineage-specific AMPs against fungal infection through its evolution. Since no orthologous sequences of these genes were found in other *Drosophila* genomes either, these genes seemed to be recruited to the *D. virilis* genome de novo. In addition to the fraction of these genes of unknown function, I also predicted new *D. virilis* genes from the pyrosequencing reads that did not show any BLAST hit.

In our BLAST analyses of the pyrosequencing reads, approximately 30% of the reads from *D. virilis* did not hit any gene, whereas only 3-4% of the reads from *D. melanogaster* fell in the same situation (Table 3). This may suggest the possibility that many genes in the *D. virilis* genome have not been identified yet. Actually, I found 620 putative genes in 3,469 contigs and three of them, *PG00034*, *PG01778* and *PG02420*, were predicted to be immune-related genes with expression level significantly changed by the fungal infection. *PG00034* is homologous to *IM14* and *PG01778* is homologous to a Ras-like GTP-binding protein, *Rho1*, which regulates actin cytoskeletal organization (Hariharan et al., 1995, Magie et al., 1999) and is involved in

phagocytosis (Greenberg and Grinstein 2002, Magie and Parkhurst 2005) in *D. melanogaster* (Table 7). *PG02420* is homologous to *Ficolin-2* of *Bos taurus*. Ficolin binds to a cell wall component of bacteria and fungi and is involved in phagocytosis (Ma et al., 2004, Endo, Matsushita and Fujita 2007). Although the expression of the *IM14* was significantly up-regulated by the fungal infection in the *D. melanogaster* larvae, the expression of *PG00034* was significantly down-regulated as in the case of other homologues of IM genes in the *D. virilis* larvae. Similarly, the expression of *PG02420* was significantly down-regulated in the infected *D. virilis* larvae. On the other hand, the expression of *PG01778* was significantly up-regulated by the fungal infection in *D. virilis*. For the remaining 2,649 contigs, I could not find any homologue in Swissprot protein database. This seems partly because many of them are too short to find a homology to a known gene, domain or motif in the homology search (Figure 7). Further experimental determination of their full length sequence is necessary for a better prediction of novel protein coding genes. From this perspective, I tried to determine the 5'-end sequence of these contigs using the oligo-capping method adjusted to 454 GS Junior sequencer. As the result, I found two candidate genes, which encode potential AMP. One of them, *GJ22451*, is a homologue of *IM3* of *D. melanogaster*. The expression of *GJ22451* is exceptionally up-regulated in the fungal-infected *D. virilis*, whereas other IM genes tended to be down-regulated. *GJ22451* is predicted to be similar to mammalian beta-Defensin by AntiBP2 prediction program (Table 9). The other candidate gene, *PG01471*, was predicted to encode a Metchnikowin-like proline-rich secretory peptide (Table 9, Figure 9). Although *GJ22451* is present in the all 12 *Drosophila* species, homologue of *PG01471* was not found in *D. melanogaster* (Figure 10). This observation suggests that *PG01471* may contribute to the higher antifungal resistance of *D. virilis*. Antifungal activity of *PG01471* should be experimentally verified.

Our comparative transcriptome analysis revealed extensive differences in the immune response to the infection of *Penicillium* species between *D. virilis* and *D. melanogaster* at the transcriptome level. These results provide an important insight to the different role of immune system between ecologically diverged species. It is quite natural to consider that the observed differences resulted from evolutionary adaptation to their different habitat. This presumption should be further experimentally examined by the investigation of antimicrobial activities of AMPs, e.g., Diptericin and Defensin, to identify the component responsible for the higher antifungal resistance of *D. virilis*.

CONCLUSION

In general, *Drosophila* species feed and breed on fermenting fruits, slime fluxes on decaying parts of tree and so on, in which a variety of microbes are extremely active (Carson 1971, Throckmorton 1975, Markow and O'Grady 2007). Therefore, anti-microbes immune system is an essential trait for *Drosophila* species to survive. The evolution of the immune system is likely responsible for the diversity of *Drosophila* species adapting to a variety of microbial environments. In this study, a substantial difference in antifungal activity against a *Penicillium* species between two *Drosophila* species, *D. virilis* and *D. melanogaster* living in different environments, was demonstrated.

My comparative transcriptome analysis showed extensive differences in the expression pattern of immune-related genes, i.e., antimicrobial peptide (AMP) and the immune-induced molecule (IM) genes, in response to the *Penicillium* infection between *D. virilis* and *D. melanogaster*. Furthermore, I predicted novel immune-related genes responding to the fungal infection in *D. virilis*. These results indicate that the innate immune system has been substantially differentiated during the evolution of these *Drosophila* species. The extensive differences in the immune system may have been evolved as an adaptive response to microbial environments, which remains open to further investigations.

LITERATURE CITED

Agaisse H. and N. Perrimon, "The roles of JAK/STAT signaling in *Drosophila* immune responses," *Immunological Reviews*, vol. 198, no. 1, pp. 72-82, 2004.

Altschul S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403-410, 1990.

Basset A., R. Khush, A. Braun, L. Gardan, F. Boccard, J. A. Hoffmann and B. Lemaitre, "The phytopathogenic bacteria, *Erwinia carotovora*, infects *Drosophila* and activates an immune response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 7, pp. 3376-3381, 2000.

Bhatia V. N., D. H. Perlman, C. E. Costello and M. E. McComb, "Software tool for researching annotations of proteins: open-source protein annotation software with data visualization," *Analytical Chemistry*, vol. 81, no. 23, pp. 9819-9823, 2009.

Bischoff V., C. Vignal, B. Duvic, I. G. Boneca, J. A. Hoffmann and J. Royet, "Downregulation of the *Drosophila* Immune Response by Peptidoglycan-Recognition Proteins SC1 and SC2," *PLoS Pathogen*, vol. 2, no. 2, Article ID e14, 2006.

Boulanger N., C. Lowenberger, P. Volf, R. Ursic, L. Sigutova, L. Sabatier, M. Svobodova, S. M. Beverley, G. Spa¨th, R. Brun, B. Pesson and P. Bulet, "Characterization of a Defensin from the Sand Fly *Phlebotomus duboscqi* Induced by Challenge with Bacteria or the Protozoan Parasite *Leishmania major*," *Infection and Immunity*, vol. 72, no. 12, pp. 7140-7146, 2004.

Boutros M., H. Agaisse and N. Perrimon, "Sequential activation of signaling pathways during innate immune responses in Drosophila," *Developmental Cell*, vol. 3, no. 5, pp. 711-722, 2002.

Camacho C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden,

"BLAST+: architecture and applications", *BMC Bioinformatics*, vol.10, article 421, 2009.

Carson H. L., "The ecology of *Drosophila* breeding sites", *Harold L. Lyon Arboretum Lecture Number Two*, University of Hawaii, pp. 1-28, 1971.

Carver T. and A. Bleasby, "The design of Jemboss: a graphical user interface to EMBOSS," *Bioinformatics*, vol. 19, no. 14, pp. 1837-1843, 2003.

Ceroni A., A. Passerini, A. Vullo and P. Frasconi1,"DISULFIND: a disulfide bonding state and cysteine connectivity prediction server," *Nucleic Acids Research*, vol. 34, Web Server issue, pp. W177–W181, 2006.

Chomczynski P. and N. Sacchi, "Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction," *Analytical Biochemistry*, vol. 162, no. 1, pp. 156-159, 1987.

Coates L. and G. Johnson, "Postharvest diseases of fruit and vegetables," in *Plant Pathogens and Plant Diseases*, J. F. Brown and H. J. Ogle (ed.), Rockvale publications, pp 533-548, 1997.

Cole C., J. D. Barber and G. J. Barton, "The Jpred 3 secondary structure prediction server," *Nucleic Acids Research*, vol. 35, Web Server issue, pp. W197-201, 2008.

De Gregorio E., P. T. Spellman, G. M. Rubin and B. Lemaitre, "Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 22, pp. 12590-12595, 2001.

De Gregorio E., P. T. Spellman, P. Tzou, G. M. Rubin and B. Lemaitre, "The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*," *The EMBO Journal*, vol. 21, no. 11, pp. 2568-2579, 2002.

Delaney J. R., S. Stoven, H. Uvell, K. V. Anderson, Y. Engstrom and M. Mlodzik, "Cooperative

control of *Drosophila* immune responses by the JNK and NF- κ B signaling pathways," *The EMBO Journal*, vol. 25, no. 13, pp. 3068-3077, 2006.

Dimarcq J. L., D. Hoffmann, M. Meister, P. Bulet, R. Lanot, J. M. Reichhart and J. A. Hoffmann, "Characterization and transcriptional profiles of a *Drosophila* gene encoding an insect defensin. A study in insect immunity," *European Journal of Biochemistry*, vol. 221, no. 1, pp. 201-209, 1994.

Drosophila 12 Genomes Consortium, "Evolution of genes and genomes on the *Drosophila* phylogeny," *Nature*, vol. 450, no. 7167, pp. 203-218, 2007.

Duckert P., S. Brunak and N. Blom, "Prediction of proprotein convertase cleavage sites," *Protein Engineering, Design and Selection*, vol. 17, no. 1, pp. 107-112, 2004.

Endo Y., M. Matsushita and T. Fujita, "Role of ficolin in innate immunity and its molecular basis," *Immunobiology*, vol. 212, no. 4-5, pp. 371-379, 2007.

Fehlbaum P., P. Bulet, L. Michaut, M. Lagueux, W. F. Broekaert, C. Hetru and J. A. Hoffmann, "Insect immunity. Septic injury of *Drosophila* induces the synthesis of a potent antifungalpeptide with sequence homology to plant antifungal peptides," *The Journal of Biological Chemistry*, vol. 269, no. 52, pp. 33159-33163, 1994.

Ferrandon D., Jean-Luc Imler, C. Hetru and J. A. Hoffmann, "The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections," *Nature Reviews Immunology*, vol. 7, no. 11, pp. 862-874, 2007.

Gershman B., O. Puig, L. Hang, R. M. Peitzsch M. Tatar and R. S. Garofalo, "High-resolution dynamics of the transcriptional response to nutrition in *Drosophila*: a key role for dFOXO," Physiol Genomics vol. 29, no. 1, pp. 24-34, 2007.

Glittenberg T. M., I. Kounatidis, D. Christensen, M. Kostov, S. Kimber, I. Roberts and P. Ligoxygakis, "Pathogen and host factors are needed to provoke a systemic host response to

gastrointestinal infection of *Drosophila* larvae by *Candida albicans*," *Disease Models and Mechanisms*, vol. 4, no. 4, pp. 515-525, 2011.

Greenberg S. and S. Grinstein, "Phagocytosis and innate immunity," *Current Opinion in Immunology*, vol. 14, no. 1, pp. 136-145, 2002.

Hariharan I. K., K. Q. Hu, H. Asha, A. Quintanilla, R. M. Ezzell and J. Settleman, "Characterization of rho GTPase family homologues in *Drosophila melanogaster*: overexpressing Rho1 in retinal cells causes a late developmental defect," *The EMBO Journal,* vol. 14, no. 2, pp. 292-302, 1995.

Higgins D., J. Thompson, T. Gibson. J. D. Thompson, D. G. Higgins, T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673-4680, 1994.

Imler J. L. and J. A. Hoffmann, "Signaling mechanisms in the antimicrobial host defense of *Drosophila*", *Current Opinion in Microbiology*, vol.3, no. 1, pp. 16-22, 2000.

Irving P., L. Troxler, T. S. Heuer, M. Belvin, C. Kopczynski, J. M. Reichhart, J. A. Hoffmann and C. Hetru, "A genome-wide analysis of immune responses in *Drosophila*," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 98, no. 26, pp. 15119-15124, 2001.

Kallio J., A. Leinonen, J. Ulvila, S. Valanne, R. A. Ezekowitz and M. Ramet, "Functional analysis of immune response genes in *Drosophila* identifies JNK pathway as a regulator of antimicrobial peptide gene expression in S2 cells," *Microbes and Infection*, vol. 7, no. 5-6, pp.811-819, 2005.

Lata S., N. K. Mishra and G. P. S. Raghava, "AntiBP2: Improved version of antibacterial peptide prediction," *BMC Bioinformatics*, vol. 11, S19, 2009.

Lemaitre B., E. Nicolas, L. Michaut, J. M. Reichhart and J. A. Hoffmann, "The dorsoventral regulatory gene cassette Spatzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults," *Cell*, vol. 86, no. 6, pp. 973-983, 1996.

Lemaitre B., J. M. Reichhart and J. A. Hoffmann, "*Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. no. 26, 94, pp. 14614-14619, 1997.

Lemaitre B. and J. A. Hoffmann, "The host defense of *Drosophila melanogaster*," *Annual Review of Immunology*, vol. 25, pp. 697-743, 2007.

Liehl P., M. Blight, N. Vodovar, F. Boccard, and B. Lemaitre, "Prevalence of Local Immune Response against Oral Infection in a Drosophila/Pseudomonas Infection Model," PLoS Pathogen, vol. 2, no. 6, Article ID e56, 2006.

Livak K. J., "ABI PRISM 7700 Sequence detection system user bulletin #2: relative quantitation of gene expression (PN 4303859)," Applied Biosystems, 1997 (2001 updated)

Ma Y. G., M. Y. Cho, M. Zhao, J. W. Park, M. Matsushita, T. Fujita and B. L. Lee, "Human mannose-binding lectin and L-ficolin function as specific pattern recognition proteins in the lectin activation pathway of complement," *The Journal of Biological Chemistry*, vol. 279, no. 24, pp. 25307-25312, 2004.

Magie C. R., M. R. Meyer, M. S. Gorsuch and S. M. Parkhurst, "Mutations in the Rho1 small GTPase disrupt morphogenesis and segmentation during early *Drosophila* development," *Development*, vol. 126, no. 23, pp. 5353-5364, 1999.

Magie C. R. and S. M. Parkhurst, "Rho1 regulates signaling events required for proper *Drosophila* embryonic development," *Developmental biology*, vol. 278, no. 1, pp. 144-154,

2005.

Markow T. A. and P. M. O'Grady, "*Drosophila* biology in the genomic age," *Genetics*, vol. 177, No. 3, pp. 1269-1276, 2007.

Maruyama K. and S. Sugano, "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides," *Gene*, vol. 138, pp. 171-174, 1994.

Mellroth P., J. Karlsson and H. Steiner, "A scavenger function for a *Drosophila* peptidoglycan recognition protein," *The Journal of Biological Chemistry*, vol. 278, no. 9, pp. 7059-7064, 2003.

Paredes J. C., D. P. Welchman, M. Poidevin and B. Lemaitre, "Negative Regulation by Amidase PGRPs Shapes the *Drosophila* Antibacterial Response and Protects the Fly from Innocuous Infection", *immunity*, vol. 35, no. 5, pp. 770-779, 2011.

Petersen T. N., S. Brunak, G. Heijne and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, no. 10, pp. 785-786, 2011.

Peterson S. W., E. M. Bayer and D. T. Wicklow, "*Penicillium thiersii*, *Penicillium angulare* and *Penicillium decaturense*, new species isolated from wood-decay fungi in North America and their phylogenetic placement from multilocus DNA sequence analysis", *Mycologia*, vol. 96, no. 6, pp. 1280-1293, 2004.

R Development Core Team, "R: A language and environment for statistical computing", *R Foundation for Statistical Computing*, Vienna, Austria, 2008.

Robinson M. D. and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, no. 3, article R25, 2010.

Rutschmann S., A. Kilinc and D. Ferrandon, "Cutting edge: TheToll pathway is required for resistance to gram-positive bacterial infections in *Drosophila*," *The Journal of Immunology*,

vol. 168, no. 4, pp. 1542–1546, 2002.

Sackton T. B., B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark and A. G. Clark, "Dynamic evolution of the innate immune system in *Drosophila*," *Nature Genetics*, vol. 39, no. 12, pp 1461-1468, 2007.

Sackton T. B. and A. G. Clark, "Comparative profiling of the transcriptional response to infection in two species of *Drosophila* by short-read cDNA sequencing," *BMC Genomics*, vol. 10, article 259, 2009.

Senger K., G. W. Armstrong, W. J. Rowell, J. M. Kwan, M. Markstein and M. Levine, "Immunity regulatory DNAs share common organizational features in *Drosophila*," *Molecular Cell*, vol. 13, no. 1, pp. 19-32, 2004.

Stroschein-Stevenson S. L., E. Foley, P. H. O'Farrell and A. D. Johnson, "Identification of *Drosophila* Gene Products Required for Phagocytosis of *Candida albicans*," *PLoS Biology*, vol. 4, no. 1, Article ID e4, 2006.

Sun J., T. Nishiyama, K. Shimizu and K. Kadota, "TCC: an R package for comparing tag count data with robust normalization strategies," *BMC Bioinformatics*, vol.14, article 219, 2013.

Suzuki Y., K. Yoshitomo-Nakagawa, K. Maruyama, A. Suyama and S. Sugano, "Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library," *Gene*, vol. 200, pp.149-156, 1997.

Tamura K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods," *Molecular Biology and Evolution*, vol. 28, no. 10, pp. 2731-2739, 2011.

Thomas S., S. Karnik, R. S. Barai, V. K. Jayaraman and S. Idicula-Thomas, "CAMP: a useful resource for research on antimicrobial peptides," *Nucleic Acids Research*, vol. 38, Database

issue, 2010.

Throckmorton L. H., "The phylogeny, ecology and geography of *Drosophila*," in *Invertebrates of Genetic Interest*, R. C. King (ed), Plenum Press, New York, pp. 421-469, 1975.

Torrent M., V. M. Nogués and E. Boix, "A theoretical approach to spot active regions in antimicrobial proteins," *BMC Bioinformatics*, vol. 10, article 373, 2009.

Tzou P., S. Ohresser, D. Ferrandon, M. Capovilla, J. M. Reichhart, B. Lemaitre, J. A. Hoffmann and J. L. Imler, "Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia," *Immunity*, vol. 13, no. 5, pp. 737-748, 2000.

Tzou P., J. M. Reichhart and B. Lemaitre, "Constitutive expression of a single antimicrobial peptide can restore wild-type resistance to infection in immuno-deficient *Drosophila* mutants," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 4, pp. 2152-2157, 2002.

Tzou P., E. De Gregorio and B. Lemaitre, "How *Drosophila* combats microbial infection: a model to study innate immunity and host-pathogen interactions," *Current Opinion in Microbiology*, vol. 5, no. 1, pp. 102-110, 2002.

Vizioli J., A. M. Richman, S. Uttenweiler-Joseph, C. Blass and P. Bulet, "The defensin peptide of the malaria vector mosquito *Anopheles gambiae*: antimicrobial activities and expression in adult mosquitoes," *Insect Biochemistry and Molecular Biology* vol. 31, no. 3, pp. 241-248, 2001.

Vodovar N., M. Vinals, P. Liehl, A. Basset, J. Degrouard, P. Spellman, F. Boccard and B. Lemaitre, "*Drosophila* host defense after oral infection by an entomopathogenic *Pseudomonas* species," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 32, pp. 11414-11419, 2005.

Waterhouse R. M., F. Tegenfeldt, J. Li, E. M. Zdobnov and E. V. Kriventseva, "OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs," Nucleic Acids Research, vol. 41, Database issue, pp. D358-D365, 2012.

Weber R. W. S., P. Davoli and H. Anke, "A microbial consortium involving the astaxanthin producer Xanthophyllomyces dendrorhous on freshly cut birch stumps in Germany", *Mycologist*, Vol. 20, no. 2, pp. 57-61, 2006.

Weber R. W. S., "On the ecology of fungal consortia of spring sap-flows", *Mycologist*, Vol. 20, no. 4, pp. 140-143, 2006.

Wicker C., J. M. Reichhart, D. Hoffmann, D. Hultmark, C. Samakovlis and J. A. Hoffmann. "Insect immunity. Characterization of a *Drosophila* cDNA encoding a novel member of the diptericin family of immune peptides," *The Journal of Biological Chemistry*, vol. 265, no. 36, pp. 22493-22498, 1990.

Zaidman-Remy A., M. Herve, M. Poidevin, S. Pili-Floury, M. S. Kim, D. Blanot, B. H. Oh, R. Ueda, D. Mengin-Lecreulx and B. Lemaitre, "The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection," *Immunity*, vol. 24, no. 4, pp. 463-473, 2006.

Table 1: Sequences of oligo-capping adaptors and sequencing primers.

Sequence name	Sequence
5' RNA adaptor	CCAUCUCAUCCCUGCGUGUCUCCGACGACU
5'-end FAM primer	FAM-CCATCTCATCCCTGCGTGTCTCCGACGACT
5'- adaptor primer	CCATCTCATCCCTGCGTGTCTCCGACGACT
oligo-$(dT)_{15}$ + 3' adaptor primer	CCTATCCCCTGTGTGCCTTGGCAGTCGACT[T]15

Table 2: PCR primers used for real-time RT-PCR

Primer name	Sequence
Drosophila_Rp49_41F	AGCGCACCAAGCACTTCA
Drosophila_Rp49_178R	CGTAACCGATGTTGGGCA
Dvir_Mtk_54F	CCTGAGCCTGAACCTGTCG
Dvir_Mtk_146R	GGCTGATTGGGATTGAATGG
GJ21126_26F	TGCTACTAATCCTGGCTGTGAC
GJ21126_163R	AGGCTGTGTTCTTTACGTTCCA
GJ22479_29F	TGGCTCTACTGGTGTGCTTG
GJ22479_151R	ATGTGGCACGCTTCTGACG

Table 3: Summary statistics of 454 GS junior sequencing and BLAST analysis.

	D. virilis		*D. melanogaster*	
	Infected	Naïve	Infected	Naïve
Total no. of reads	109,106	119,533	110,578	91,947
Maximum length (bp)	715	667	710	580
Minimum length (bp)	40	40	40	40
Average Length (bp)	226	217	242	219
No. of mtDNA-derived reads	5,557	6,197	5,998	7,483
No. of rDNA-derived reads	25,991	22,500	38,910	35,990
No. of other reads	77,558	90,836	65,670	48,474
No. of BLAST hits	55,358	62,110	63,555	46,536
(No. of genes)	(5,155)	(4,709)	(4,735)	(4,275)
No. of unidentified reads	22,200	28,726	2,115	1,938

Table 4: Number of reads, trimmed mean of M value (TMM) and induction coefficient (IC) for recognition, signaling and effector class immune genes showing significant changes in expression level by fungal infection in *D. virilis*.

D. virilis gene	*D. melanogaster* homologue	Infected		Naive		IC	Functional Class	Notes
		No. of reads	TMM	No. of reads	TMM			
GJ20666	*CG13422*	6	0.153	0	0	Infinity	Recognition	Beta-glucan binding domain
GJ12160	*PGRP-SB1*	11	0.235	2	0.040	5.864	Recognition	PGRP domain
GJ18074	*nimB3*	2	0.067	12	0.376	0.178	Recognition	Nimrod-related
GJ12373	*msn*	9	0.024	1	0.002	9.595	Signaling	Kinase
GJ20603	*Pvr*	15	0.038	2	0.005	7.996	Signaling	Receptor
GJ19441	*SPE*	3	0.033	15	0.155	0.213	Signaling	Protease
GJ22479	*Def*	53	2.445	0	0	Infinity	Effector	Antimicrobial peptide
GJ21173	*AttC*	47	0.818	0	0	Infinity	Effector	Antimicrobial peptide
Cec2B	*CecA1 / CecA2*	25	1.604	0	0	Infinity	Effector	Antimicrobial peptide
Cec3	*CecC*	23	1.475	0	0	Infinity	Effector	Antimicrobial peptide
GJ22469	*Mtk*	9	0.660	0	0	Infinity	Effector	Antimicrobial peptide
GJ19916	*Dpt*	104	3.812	4	0.138	27.720	Effector	Antimicrobial peptide
GJ19917	*DptB*	39	1.120	3	0.081	13.860	Effector	Antimicrobial peptide
GJ20572	*AttA*	49	0.856	24	0.393	2.177	Effector	Antimicrobial peptide
GJ17981	*fon*	217	1.641	370	2.624	0.625	Effector	Coagulation
GJ18607	*IM4*	79	7.542	151	13.521	0.558	Effector	IM
GJ21308	*IM10*	23	0.350	51	0.727	0.481	Effector	IM
GJ19885	*IM1*	37	3.302	123	10.296	0.321	Effector	IM

Genes are sorted in order of induction coefficient at each functional class.

Table 5: Number of reads, trimmed mean of M value (TMM) and induction coefficient (IC) for recognition, signaling and effector class immune genes showing significant changes in expression level by fungal infection in *D. melanogaster*.

D. melanogaster gene	Infected		Naive		IC	Functional Class	Notes
	No. of reads	TMM	No. of reads	TMM			
PGRP-SB1	29	0.779	0	0	Infinity	Recognition	PGRP domain
PGRP-SC1b	11	0.288	0	0	Infinity	Recognition	Amidase degradation
PGRP-SB2	9	0.225	0	0	Infinity	Recognition	PGRP domain
Mcr	4	0.011	0	0	Infinity	Recognition	Tep
PGRP-SC2	20	0.603	3	0.102	5.891	Recognition	Amidase degradation
TepII	188	0.708	31	0.132	5.359	Recognition	Tep
nimC2	43	0.310	9	0.073	4.222	Recognition	Nimrod-related
GNBP3	15	0.164	4	0.049	3.313	Recognition	Beta-glucan binding domain
CG13422	17	0.569	5	0.189	3.004	Recognition	Beta-glucan binding domain
TepIV	37	0.131	13	0.052	2.515	Recognition	Tep
PGRP-SD	27	0.626	13	0.341	1.835	Recognition	PGRP domain
Rel	14	0.067	0	0	Infinity	Signaling	Transcription factor
aop	6	0.026	0	0	Infinity	Signaling	Transcription factor
brm	5	0.016	0	0	Infinity	Signaling	Transcription factor
Myd88	4	0.019	0	0	Infinity	Signaling	.
CG6361	15	0.185	1	0.014	13.254	Signaling	Protease
cact	11	0.081	1	0.008	9.720	Signaling	.
dom	8	0.085	1	0.012	7.069	Signaling	Transcription factor
Stat92E	11	0.050	3	0.016	3.240	Signaling	Transcription factor
srp	18	0.080	5	0.025	3.181	Signaling	Transcription factor
phl	32	0.135	9	0.043	3.142	Signaling	.
mask	10	0.012	3	0.004	2.945	Signaling	.
spirit	22	0.231	7	0.083	2.777	Signaling	Protease
CecC	35	1.521	0	0	Infinity	Effector	Antimicrobial peptide
CecA1	14	0.663	0	0	Infinity	Effector	Antimicrobial peptide
Def	11	0.461	0	0	Infinity	Effector	Antimicrobial peptide
CecB	7	0.288	0	0	Infinity	Effector	Antimicrobial peptide
dro5	6	0.276	0	0	Infinity	Effector	Antimicrobial peptide
AttC	252	4.684	2	0.042	111.333	Effector	Antimicrobial peptide
Dpt	343	11.568	24	0.916	12.628	Effector	Antimicrobial peptide

Table 5 continued

DptB	80	2.974	6	0.252	11.781	Effector	Antimicrobial peptide
Pu	79	0.687	7	0.069	9.972	Effector	Melanin synthesis cascade
TotC	10	0.311	1	0.035	8.836	Effector	Tot
IM18	62	1.403	8	0.205	6.848	Effector	IM
Mtk	380	23.719	52	3.673	6.457	Effector	Antimicrobial peptide
Dro	192	4.237	27	0.674	6.283	Effector	Antimicrobial peptide
yellow-f	23	0.277	6	0.082	3.387	Effector	Melanin synthesis cascade
IM14	68	5.101	19	1.613	3.162	Effector	IM
AttA	96	2.113	27	0.673	3.142	Effector	Antimicrobial peptide
IM4	56	2.194	16	0.709	3.093	Effector	IM
IM10	355	6.147	116	2.273	2.704	Effector	IM
IM1	247	11.541	82	4.336	2.662	Effector	IM
AttB	74	1.428	27	0.590	2.422	Effector	Antimicrobial peptide
IM2	139	6.250	62	3.155	1.981	Effector	IM
Tsf1	145	1.209	68	0.642	1.884	Effector	Iron binding
TotA	182	5.213	98	3.177	1.641	Effector	Tot
Drs	551	23.817	299	14.627	1.628	Effector	Antimicrobial peptide
Tig	22	0.053	12	0.033	1.620	Effector	Coagulation
IM3	330	18.401	188	11.864	1.551	Effector	IM

Genes are sorted in order of induction coefficient at each functional class.

Table 6: Trimmed mean of M value (TMM), induction coefficient (IC), number of amino acids of mature peptide, molecular weight, net charge and protein structural feature for putative antimicrobial peptide genes in *D. virilis* predicted by AMP prediction programs.

D. virilis gene	TMM	IC	Mature peptide size (aa)	Molecular weight (kDa)	Net Charge	Structural features	AMP prediction		
							AntiBP2	CAMP	AMPA
GJ10737	1.368	2.503	35	4.07	12	Arg+Val rich (51%)	-	+	+
GJ18291	0.316	3.909	61	6.70	25	Lys+Ser rich (46%)	-	+	+

Table 7: Number of reads and induction coefficient (IC) for putative immune-related genes in *D. virilis* and their homologues in *D. melanogaster*.

	D. virilis				*D. melanogaster*		
Putative gene	No. of reads		IC	Homologue	No. of reads		IC
	Infected	Naïve			Infected	Naïve	
PG00034	17 *	37	0.477	*IM14*	68 **	19	3.162
PG01778	7 *	0	infinity	*Rho1*	16 *	7	2.020
PG02420	2 *	10	0.208	-	-	-	-

Table 8: Summary statistics of 5'-end-riched 454 GS junior sequencing and de novo assemble.

454 GS Jr sequencing	No. of reads	50573
	Average length (bp)	190
	Longest read (bp)	599
	Shortest read (bp)	40
De novo assemble	No. of assembled reads	41423
	No. of contigs	900

Table 9: Trimmed mean of M value (TMM), induction coefficient (IC), identity gene of *D. virilis*, their homologues in *D. melanogaster* and AMP prediction for putative genes whose expression were significantly changed by fungal.infection in *D. virilis*.

Gene	Infected		Naïve		IC	Identity gene of *D. virilis*	*D. melanogaster* homologue	AMP prediction
	No. of reads	TMM	No. of reads	TMM				
PG00098	1131	13758.005	814	9532.570	1.443	-		-
PG00273	32	389.263	15	175.662	2.216	-		-
PG00667	17	206.796	5	58.507	3.535	*GJ11849*	ribosomal protein L26	-
PG01259	21	255.454	7	81.975	3.116	-		-
PG01471	9	109.480	1	11.701	9.356	-		+
PG01875	10	121.645	1	11.701	10.396	*GJ22451*	IM 3	+
PG02341	10	121.645	1	11.701	10.396	-		-
PG03038	46	559.565	22	257.637	2.172	-		-

Supplementary Table 1: Number of reads, trimmed mean of M value (TMM) and induction coefficient (IC) for endogenous control genes in *D. virilis* and *D. melanogaster.*

Organism	Symbol	Gene	Infected		Naïve		IC
			No. of reads	TMM	No. of reads	TMM	
D. virilis	RpL32	*RpL32*	38	1.156	31	0.884	1.307
	Gapdh	*GJ20812*	0	0	2	0.022	0
		GJ20492	27	0.333	36	0.416	0.800
D. melanogaster	RpL32	*RpL32*	71	2.216	87	3.073	0.721
	Gapdh	*Gapdh1*	27	0.351	40	0.589	0.597
		Gapdh2	11	0.134	21	0.289	0.463

Supplementary Table 2: Number of reads, trimmed mean of M value (TMM) and induction coefficient (IC) for recognition, signaling and effector class immune-related genes observed in *D. virilis*.

D. virilis gene	*D. melanogaster* homologue	Infected		Naïve		IC	Functional Class	Notes
		No of reads	TMM	No of reads	TMM			
GJ20666	CG13422	6 *	0.153	0	0	Infinity	Recognition	Beta-glucan binding domain
GJ18161	nimB5	5	0.070	0	0	Infinity	Recognition	Nimrod-related
GJ22101	Corin	1	0.003	0	0	Infinity	Recognition	Scavenger receptor
GJ11092	modSP	1	0.009	0	0	Infinity	Recognition	Scavenger receptor
GJ12160	PGRP-SB1	11 *	0.235	2	0.040	5.864	Recognition	PGRP domain
GJ20388	Mcr	4	0.009	1	0.002	4.265	Recognition	Tep
GJ13216	modSP	10	0.073	5	0.034	2.132	Recognition	Scavenger receptor
GJ19730	crq	2	0.017	1	0.008	2.132	Recognition	Scavenger receptor
GJ15950	PGRP-SA	10	0.208	7	0.137	1.523	Recognition	PGRP domain
GJ16225	TepII	28	0.082	26	0.071	1.148	Recognition	Tep
GJ13386	PGRP-LF	2	0.022	2	0.021	1.066	Recognition	PGRP domain
GJ13383	PGRP-LC	7	0.054	8	0.058	0.933	Recognition	PGRP domain
GJ18075	nimB2	29	0.298	36	0.347	0.859	Recognition	Nimrod-related
GJ18565	PGRP-LE	3	0.034	4	0.042	0.800	Recognition	PGRP domain
GJ18162	nimC2	11	0.063	15	0.081	0.78	Recognition	Nimrod-related
GJ17482	pes	2	0.015	3	0.021	0.71	Recognition	Scavenger receptor
GJ14102	GNBP2	5	0.149	8	0.223	0.666	Recognition	Beta-glucan binding domain
GJ13082	GNBP3	6	0.050	10	0.078	0.640	Recognition	Beta-glucan binding domain
GJ23926	modSP	4	0.026	7	0.042	0.609	Recognition	Scavenger receptor
GJ18229	TepIV	29	0.081	54	0.141	0.573	Recognition	Tep
GJ21683	CG30148	7	0.230	14	0.431	0.533	Recognition	Beta-glucan binding domain

Supplementary Table 2 continued

GJ18073	nimB4	1	0.009	3	0.025	0.355	Recognition	Nimrod-related
GJ18074	nimB3	2 *	0.067	12	0.376	0.178	Recognition	Nimrod-related
GJ11132	r2d2	3	0.039	1	0.012	3.198	Other	.
GJ20897	Dcr-2	4	0.010	5	0.011	0.853	Other	.
GJ15620	et	5	0.028	0	0	Infinity	Signaling	.
GJ18697	Ulp1	5	0.023	0	0	Infinity	Signaling	Protein modification
GJ21884	bsk	2	0.022	0	0	Infinity	Signaling	Kinase
GJ16598	phl	2	0.011	0	0	Infinity	Signaling	.
GJ20084	casp	1	0.034	0	0	Infinity	Signaling	Negative regulator of imd
GJ21215	grass	1	0.027	0	0	Infinity	Signaling	Protease
GJ21265	her	1	0.012	0	0	Infinity	Signaling	Transcription factor
GJ16702	Dredd	1	0.008	0	0	Infinity	Signaling	Protease
GJ16121	Traf4	1	0.008	0	0	Infinity	Signaling	.
GJ20901	Iap2	1	0.008	0	0	Infinity	Signaling	.
GJ14961	Su(var)2-10	1	0.007	0	0	Infinity	Signaling	.
GJ24520	ea	1	0.005	0	0	Infinity	Signaling	Protease
GJ23176	Ser	1	0.003	0	0	Infinity	Signaling	Ligand
GJ12373	msn	9 *	0.024	1	0.002	9.595	Signaling	Kinase
GJ20603	Pvr	15 **	0.038	2	0.005	7.996	Signaling	Receptor
GJ23773	Hel89B	5	0.011	1	0.002	5.331	Signaling	.
tub	tub	7	0.055	2	0.015	3.732	Signaling	.
GJ20044	dom	10	0.013	3	0.004	3.554	Signaling	Transcription factor
Ras1	Ras85D	3	0.065	1	0.020	3.198	Signaling	.
GJ20758	Nup214	3	0.007	1	0.002	3.198	Signaling	Nuclear transport

Supplementary Table 2 continued

GJ16165	*Pvf2*	5	0.055	2	0.020	2.665	Signaling	Ligand
GJ21092	*Myd88*	5	0.044	2	0.017	2.665	Signaling	.
GJ18227	*ref(2)P*	18	0.113	9	0.053	2.132	Signaling	.
GJ11038	*Mkk4*	4	0.038	2	0.018	2.132	Signaling	Kinase
GJ18559	*smt3*	2	0.080	1	0.038	2.132	Signaling	Protein modification
GJ10528	*mask*	16	0.016	9	0.008	1.895	Signaling	.
GJ12202	*Uev1A*	7	0.197	4	0.105	1.866	Signaling	Protein modification
GJ19711	*lwr*	5	0.128	3	0.072	1.777	Signaling	Protein modification
GJ24191	*Stat92E*	13	0.069	8	0.040	1.733	Signaling	Transcription factor
GJ17162	*cact*	8	0.059	5	0.035	1.706	Signaling	.
GJ19035	*ben*	3	0.081	2	0.051	1.599	Signaling	Protein modification
GJ16904	*hep*	4	0.012	3	0.009	1.422	Signaling	Kinase
GJ23481	*Rel*	7	0.030	6	0.024	1.244	Signaling	Transcription factor
GJ22783	*SPE*	7	0.071	7	0.067	1.066	Signaling	Protease
GJ10450	*cher*	6	0.024	6	0.022	1.066	Signaling	.
GJ16233	*Dif*	4	0.027	4	0.025	1.066	Signaling	Transcription factor
GJ10642	*Kay*	3	0.016	3	0.015	1.066	Signaling	Transcription factor
GJ21432	*POSH*	3	0.014	3	0.013	1.066	Signaling	.
GJ22384	*Jra*	2	0.028	2	0.026	1.066	Signaling	Transcription factor
GJ13209	*Rac1*	1	0.021	1	0.020	1.066	Signaling	.
GJ12360	*Rac2*	1	0.021	1	0.020	1.066	Signaling	.
GJ21078	*key*	1	0.010	1	0.010	1.066	Signaling	Kinase
GJ20083	*casp*	1	0.008	1	0.007	1.066	Signaling	Negative regulator of imd
GJ10286	*Socs36E*	1	0.006	1	0.006	1.066	Signaling	Negative regulator of JAKSTAT

GJ24058	*mbo*	1	0.006	1	0.005	1.066	Signaling	Nuclear transport
GJ15619	*dome*	1	0.003	1	0.003	1.066	Signaling	Receptor
GJ15330	*hop*	1	0.003	1	0.003	1.066	Signaling	Kinase
GJ20945	*18w*	1	0.003	1	0.003	1.066	Signaling	Toll-like
GJ10951	*Sp7*	35	0.371	40	0.398	0.933	Signaling	Phenoloxidase cascade
GJ20054	*Egfr*	5	0.014	6	0.016	0.888	Signaling	Receptor
GJ14938	*nec*	16	0.126	20	0.148	0.853	Signaling	Protease
GJ18893	*Pvf1*	3	0.038	4	0.047	0.800	Signaling	Ligand
GJ21412	*Spn27A*	5	0.046	7	0.061	0.762	Signaling	Phenoloxidase cascade
GJ20786	*Stam*	7	0.040	10	0.053	0.746	Signaling	.
GJ18599	*Ntf-2*	2	0.063	3	0.088	0.711	Signaling	Nuclear transport
GJ14460	*MP1*	14	0.145	22	0.213	0.678	Signaling	Phenoloxidase cascade
GJ23376	*srp*	10	0.033	16	0.050	0.666	Signaling	Transcription factor
GJ10449	*cher*	6	0.017	10	0.027	0.640	Signaling	.
GJ15197	*psh*	14	0.145	25	0.242	0.597	Signaling	Protease
GJ11780	*brm*	5	0.012	9	0.021	0.592	Signaling	Transcription factor
GJ23435	*ben*	1	0.027	2	0.051	0.533	Signaling	Protein modification
GJ18753	*slpr*	1	0.003	2	0.006	0.533	Signaling	.
GJ17781	*ush*	4	0.011	10	0.026	0.426	Signaling	Transcription factor
GJ21441	*smt3*	1	0.045	3	0.127	0.355	Signaling	Protein modification
GJ18140	*p38b*	1	0.011	3	0.032	0.355	Signaling	Kinase
aop	*aop*	1	0.005	3	0.015	0.355	Signaling	Transcription factor
GJ11480	*BG4*	1	0.018	4	0.067	0.267	Signaling	.
GJ19441	*SPE*	3 *	0.033	15	0.155	0.213	Signaling	Protease

Supplementary Table 2 continued

GJ22479	Def	53 **	2.445	0	0	Infinity	Effector	Antimicrobial peptide
GJ21173	AttC	47 **	0.818	0	0	Infinity	Effector	Antimicrobial peptide
Cec2B	CecA1 / CecA2	25 **	1.604	0	0	Infinity	Effector	Antimicrobial peptide
Cec3	CecC	23 **	1.475	0	0	Infinity	Effector	Antimicrobial peptide
GJ22469	Mtk	9 **	0.660	0	0	Infinity	Effector	Antimicrobial peptide
GJ22662	AttD	5	0.113	0	0	Infinity	Effector	Antimicrobial peptide
GJ13808	Tsf2	3	0.015	0	0	Infinity	Effector	Iron binding
GJ15754	Jafrac1	2	0.042	0	0	Infinity	Effector	Gut protection
Ddc	Ddc	2	0.016	0	0	Infinity	Effector	Melanin synthesis cascade
GJ21181	PO45	2	0.012	0	0	Infinity	Effector	Phenoloxidase
Cec1	CecC	1	0.064	0	0	Infinity	Effector	Antimicrobial peptide
GJ12891	Jafrac2	1	0.017	0	0	Infinity	Effector	Gut protection
GJ14144	Duox	1	0.003	0	0	Infinity	Effector	Gut protection
GJ19916	Dpt	104 **	3.812	4	0.138	27.720	Effector	Antimicrobial peptide
GJ19917	DptB	39 **	1.120	3	0.081	13.860	Effector	Antimicrobial peptide
GJ12507	LysD	5	0.146	1	0.027	5.331	Effector	Lysozyme, c-type
GJ20571	AttA	5	0.087	1	0.016	5.331	Effector	Antimicrobial peptide
GJ20572	AttA	49 *	0.856	24	0.393	2.177	Effector	Antimicrobial peptide
GJ20545	Pu	83	1.257	58	0.824	1.526	Effector	Melanin synthesis cascade
GJ10563	Irc	19	0.110	14	0.076	1.447	Effector	Gut protection
GJ11603	Catsup	6	0.056	6	0.053	1.066	Effector	Melanin synthesis cascade
GJ20669	Tsf3	1	0.006	1	0.005	1.066	Effector	Iron binding
GJ21009	CG6426	7	0.177	8	0.190	0.933	Effector	Lysozyme, I-type
GJ15168	Tig	47	0.087	55	0.095	0.911	Effector	Coagulation

Supplementary Table 2 continued

GJ10805	*yellow-f2*	2	0.018	3	0.025	0.711	Effector	Melanin synthesis cascade
GJ15366	*Tsf1*	111	0.713	172	1.036	0.688	Effector	Iron binding
GJ17981	*fon*	217 *	1.641	370	2.624	0.625	Effector	Coagulation
GJ21309	*IM10*	22	0.257	37	0.405	0.634	Effector	IM
GJ18607	*IM4*	79 *	7.542	151	13.521	0.558	Effector	IM
GJ21008	*CG6426*	2	0.051	4	0.095	0.533	Effector	Lysozyme, I-type
GJ21308	*IM10*	23 *	0.350	51	0.727	0.481	Effector	IM
GJ18065	*CG15293*	9	0.069	22	0.159	0.436	Effector	Coagulation
GJ13134	*CG14823*	2	0.036	5	0.085	0.426	Effector	Lysozyme, I-type; destabilase
GJ19885	*IM1*	37 **	3.302	123	10.296	0.321	Effector	IM
GJ18809	*CG16799*	1	0.025	5	0.117	0.213	Effector	Lysozyme, c-type
GJ22454	*IM23*	1	0.039	6	0.218	0.178	Effector	IM
GJ12894	*LysD*	1	0.029	7	0.187	0.152	Effector	Lysozyme, c-type

Genes are shown in order of induction coefficient at each functional class.

*, ** Significant difference from the number of reads for naïve larvae (* $P < 0.05$, ** $P < 0.01$).

Supplementary Table 3: Number of reads, trimmed mean of M value (TMM) and induction coefficient (IC) for recognition, signaling and effector class immune-related genes observed in *D. melanogaster*.

D. melanogaster gene	Infected		Naïve		IC	Functional Class	Notes
	No. of reads	TMM	No. of reads	TMM			
PGRP-SB1	29 **	0.779	0	0	Infinity	Recognition	PGRP domain
PGRP-SC1b	11 **	0.288	0	0	Infinity	Recognition	Amidase degradation
PGRP-SB2	9 **	0.225	0	0	Infinity	Recognition	PGRP domain
Mcr	4 *	0.011	0	0	Infinity	Recognition	Tep
GNBP2	3	0.035	0	0	Infinity	Recognition	Beta-glucan binding domain
Sr-CIV	2	0.024	0	0	Infinity	Recognition	Scavenger receptor
nimB1	1	0.013	0	0	Infinity	Recognition	Nimrod-related
PGRP-SC1a	1	0.030	0	0	Infinity	Recognition	Amidase degradation
PGRP-SC2	20 **	0.603	3	0.102	5.891	Recognition	Amidase degradation
TepII	188 **	0.708	31	0.132	5.359	Recognition	Tep
nimC2	43 **	0.310	9	0.073	4.222	Recognition	Nimrod-related
GNBP3	15 **	0.164	4	0.049	3.313	Recognition	Beta-glucan binding domain
CG13422	17 **	0.569	5	0.189	3.004	Recognition	Beta-glucan binding domain
TepIV	37 **	0.131	13	0.052	2.515	Recognition	Tep
crq	5	0.043	2	0.020	2.209	Recognition	Scavenger receptor
PGRP-SA	7	0.103	3	0.050	2.062	Recognition	PGRP domain
TepI	7	0.028	3	0.013	2.062	Recognition	Tep
PGRP-SD	27 **	0.626	13	0.341	1.835	Recognition	PGRP domain
modSP	2	0.014	2	0.016	0.884	Recognition	Scavenger receptor
PGRP-LC	1	0.009	1	0.010	0.884	Recognition	PGRP domain
PGRP-LE	1	0.009	1	0.010	0.884	Recognition	PGRP domain
pes	1	0.007	1	0.007	0.884	Recognition	Scavenger receptor
PGRP-LB	1	0.013	1	0.015	0.884	Recognition	Amidase degradation
nimB2	24	0.279	29	0.382	0.731	Recognition	Nimrod-related
nimB4	4	0.045	5	0.064	0.707	Recognition	Nimrod-related
nimB3	24	0.898	44	1.864	0.482	Recognition	Nimrod-related
CG30148	1	0.026	2	0.058	0.442	Recognition	Beta-glucan binding domain
emp	1	0.007	3	0.025	0.295	Recognition	Scavenger receptor
Dcr-2	3	0.009	2	0.007	1.325	Other	.
Rel	14 **	0.067	0	0	Infinity	Signaling	Transcription factor
aop	6 **	0.026	0	0	Infinity	Signaling	Transcription factor
brm	5 *	0.016	0	0	Infinity	Signaling	Transcription factor

Myd88	4 *	0.019	0	0	Infinity	Signaling	.
ush	3	0.008	0	0	Infinity	Signaling	Transcription factor
Mpk2	3	0.034	0	0	Infinity	Signaling	Kinase
ird5	3	0.020	0	0	Infinity	Signaling	Kinase
slpr	3	0.010	0	0	Infinity	Signaling	.
lwr	3	0.045	0	0	Infinity	Signaling	Protein modification
hop	3	0.010	0	0	Infinity	Signaling	Kinase
Ulp1	3	0.010	0	0	Infinity	Signaling	Protein modification
tamo	2	0.009	0	0	Infinity	Signaling	Nuclear transport
Traf4	2	0.015	0	0	Infinity	Signaling	.
Pvf2	1	0.008	0	0	Infinity	Signaling	Ligand
edl	1	0.011	0	0	Infinity	Signaling	.
Dsor1	1	0.009	0	0	Infinity	Signaling	Kinase
mbo	1	0.007	0	0	Infinity	Signaling	Nuclear transport
Rac2	1	0.011	0	0	Infinity	Signaling	.
CG6361	15 **	0.185	1	0.014	13.254	Signaling	Protease
cact	11 **	0.081	1	0.008	9.720	Signaling	.
dom	8 *	0.085	1	0.012	7.069	Signaling	Transcription factor
Rac1	4	0.037	1	0.010	3.534	Signaling	.
spz	4	0.033	1	0.010	3.534	Signaling	Ligand
Stat92E	11 *	0.050	3	0.016	3.240	Signaling	Transcription factor
srp	18 **	0.080	5	0.025	3.181	Signaling	Transcription factor
phl	32 **	0.135	9	0.043	3.142	Signaling	.
mask	10 *	0.012	3	0.004	2.945	Signaling	.
spirit	22 **	0.231	7	0.083	2.777	Signaling	Protease
hep	6	0.031	2	0.012	2.651	Signaling	Kinase
emb	6	0.022	2	0.008	2.651	Signaling	Nuclear transport
Pvr	8	0.026	3	0.011	2.356	Signaling	Receptor
Sp7	13	0.112	6	0.059	1.914	Signaling	Phenoloxidase cascade
ben	15	0.138	7	0.073	1.893	Signaling	Protein modification
Egfr	6	0.022	3	0.012	1.767	Signaling	Receptor
Stam	4	0.023	2	0.013	1.767	Signaling	.
Ntf-2	4	0.022	2	0.012	1.767	Signaling	Nuclear transport
Hel89B	2	0.005	1	0.003	1.767	Signaling	.
key	2	0.025	1	0.014	1.767	Signaling	Kinase
Mkk4	2	0.013	1	0.008	1.767	Signaling	Kinase
Dif	2	0.013	1	0.007	1.767	Signaling	Transcription factor

tub	2	0.017	1	0.009	1.767	Signaling	.
imd	2	0.025	1	0.014	1.767	Signaling	.
dome	3	0.010	2	0.008	1.325	Signaling	Receptor
grass	3	0.038	2	0.029	1.325	Signaling	Protease
nec	18	0.186	14	0.164	1.136	Signaling	Protease
ref(2)P	10	0.075	8	0.068	1.104	Signaling	.
SPE	7	0.093	6	0.090	1.031	Signaling	Protease
MP1	9	0.101	9	0.114	0.884	Signaling	Phenoloxidase cascade
Jra	3	0.033	3	0.037	0.884	Signaling	Transcription factor
msn	2	0.006	2	0.007	0.884	Signaling	Kinase
Iap2	1	0.008	1	0.009	0.884	Signaling	.
Dredd	1	0.008	1	0.0010	0.884	Signaling	Protease
ytr	1	0.008	1	0.009	0.884	Signaling	.
Traf-like	1	0.007	1	0.008	0.884	Signaling	.
kay	1	0.004	1	0.004	0.884	Signaling	Transcription factor
pnt	1	0.006	1	0.007	0.884	Signaling	Transcription factor
Uev1A	7	0.087	9	0.127	0.687	Signaling	Protein modification
psh	3	0.030	4	0.046	0.663	Signaling	Protease
Spn27A	2	0.015	3	0.025	0.589	Signaling	Phenoloxidase cascade
Su(var)2-10	2	0.016	3	0.026	0.589	Signaling	.
cher	4	0.019	7	0.038	0.505	Signaling	.
smt3	3	0.068	9	0.229	0.295	Signaling	Protein modification
CecC	35 **	1.521	0	0	Infinity	Effector	Antimicrobial peptide
CecA1	14 **	0.663	0	0	Infinity	Effector	Antimicrobial peptide
Def	11 **	0.461	0	0	Infinity	Effector	Antimicrobial peptide
CecB	7 **	0.288	0	0	Infinity	Effector	Antimicrobial peptide
dro5	6 **	0.276	0	0	Infinity	Effector	Antimicrobial peptide
Hml	2	0.003	0	0	Infinity	Effector	Coagulation
CG18107	2	0.118	0	0	Infinity	Effector	IM
TotB	2	0.063	0	0	Infinity	Effector	Tot
CecA2	2	0.094	0	0	Infinity	Effector	Antimicrobial peptide
Tsf3	1	0.007	0	0	Infinity	Effector	Iron binding
AttC	252 **	4.684	2	0.042	111.333	Effector	Antimicrobial peptide
Dpt	343 **	11.568	24	0.916	12.628	Effector	Antimicrobial peptide
DptB	80 **	2.974	6	0.252	11.781	Effector	Antimicrobial peptide
Pu	79 **	0.687	7	0.069	9.972	Effector	Melanin synthesis cascade

TotC	10 **	0.311	1	0.035	8.836	Effector	Tot
IM18	62 **	1.403	8	0.205	6.848	Effector	IM
Mtk	380 **	23.719	52	3.673	6.457	Effector	Antimicrobial peptide
Dro	192 **	4.237	27	0.674	6.283	Effector	Antimicrobial peptide
Jafrac1	5	0.083	1	0.019	4.418	Effector	Gut protection
yellow-f	23 **	0.277	6	0.082	3.387	Effector	Melanin synthesis cascade
IM14	68 **	5.101	19	1.613	3.162	Effector	IM
AttA	96 **	2.113	27	0.673	3.142	Effector	Antimicrobial peptide
IM4	56 **	2.194	16	0.709	3.093	Effector	IM
IM10	355 **	6.147	116	2.273	2.704	Effector	IM
IM1	247 **	11.541	82	4.336	2.662	Effector	IM
AttB	74 **	1.428	27	0.590	2.422	Effector	Antimicrobial peptide
IM2	139 **	6.250	62	3.155	1.981	Effector	IM
Tsf1	145 **	1.209	68	0.642	1.884	Effector	Iron binding
IM23	6	0.216	3	0.122	1.767	Effector	IM
TotA	182 **	5.213	98	3.177	1.641	Effector	Tot
Drs	551 **	23.817	299	14.627	1.628	Effector	Antimicrobial peptide
Tig	22 **	0.053	12	0.033	1.620	Effector	Coagulation
CG15293	9	0.126	5	0.079	1.590	Effector	Coagulation
IM3	330 **	18.401	188	11.864	1.551	Effector	IM
CG33470	9	0.167	6	0.126	1.325	Effector	IM
Irc	26	0.182	22	0.174	1.044	Effector	Gut protection
CG16799	7	0.164	10	0.265	0.619	Effector	Lysozyme, c-type
fon	80	0.696	121	1.192	0.584	Effector	Coagulation
Catsup	2	0.018	8	0.082	0.221	Effector	Melanin synthesis cascade

Genes are shown in order of induction coefficient at each functional class.

*, ** Significant difference from the number of reads for naïve larvae (* $P < 0.05$, ** $P < 0.01$).

Supplementary Table 4: Number of reads, induction coefficient (IC) and predicted function of the putative genes (PG) in *D. virilis*.

Putative gene	No. of reads		IC	Swissprot ID	Definition	Organism	Immune-related
	Infected	Naïve					
PG00034	17	37	0.48	P83869	Immune-induced peptide 14	*Drosophila melanogaster*	+
PG00098	1131	814	1.44	A7Y3K2	Putative membrane protein ycf1	*Ipomoea purpurea*	-
PG00273	32	15	2.22	P05389	60S acidic ribosomal protein P2	*Drosophila melanogaster*	-
PG00326	13	4	3.38	P29742	Clathrin heavy chain	*Drosophila melanogaster*	-
PG00604	11	1	11.44	P60892	Ribose-phosphate pyrophosphokinase 1	*Rattus norvegicus*	-
PG00667	17	5	3.53	P61255	60S ribosomal protein L26	*Mus musculus*	-
PG00683	17	54	0.33	P61210	ADP-ribosylation factor 1	*Locusta migratoria*	-
PG01080	36	86	0.43	O44390	Acyl-CoA Delta(11) desaturase	*Trichoplusia ni*	-
PG01083	12	0	Infinity	P54385	Glutamate dehydrogenase, mitochondrial	*Drosophila melanogaster*	-
PG01215	53	122	0.45	P07701	Salivary glue protein Sgs-5	*Drosophila melanogaster*	-
PG01259	21	7	3.12	P02553	Tubulin alpha chain (Fragment)	*Lytechinus pictus*	-
PG01319	15	0	Infinity	P23194	Uricase	*Drosophila virilis*	-
PG01327	8	1	8.32	Q7KN62	Transitional endoplasmic reticulum ATPase TER94	*Drosophila melanogaster*	-
PG01341	10	0	Infinity	P79398	Eukaryotic translation initiation factor 4 gamma 2	*Oryctolagus cuniculus*	-
PG01370	1	14	0.074	Q03168	Lysosomal aspartic protease	*Aedes aegypti*	-
PG01376	2	11	0.19	Q962Q6	40S ribosomal protein S24	*Spodoptera frugiperda*	-
PG01460	62	27	2.39	P31403	V-type proton ATPase 16 kDa proteolipid subunit	*Manduca sexta*	-
PG01494	17	4	4.42	P20007	Phosphoenolpyruvate carboxykinase [GTP]	*Drosophila melanogaster*	-
PG01518	1	8	0.13	Q8T8R1	CCHC-type zinc finger protein CG3800	*Drosophila melanogaster*	-
PG01778	7	0	Infinity	P48148	Ras-like GTP-binding protein Rho1	*Drosophila melanogaster*	+
PG01865	55	105	0.54	P54361	Ornithine decarboxylase antizyme	*Drosophila melanogaster*	-
PG01979	18	39	0.48	P02707	Hepatic lectin	*Gallus gallus*	-

PG01980	15	0	Infinity	P11997	Larval serum protein 1 gamma chain	*Drosophila melanogaster*	-
PG02420	2	10	0.21	Q5I2E5	Ficolin-2	*Bos taurus*	+
PG02437	10	30	0.35	Q9GPH3	Activating transcription factor of chaperone	*Bombyx mori*	-
PG03038	46	22	2.17	Q9NJH0	Elongation factor 1-gamma	*Drosophila melanogaster*	-
PG03151	6	0	Infinity	P60517	Gamma-aminobutyric acid receptor-associated protein	*Rattus norvegicus*	-
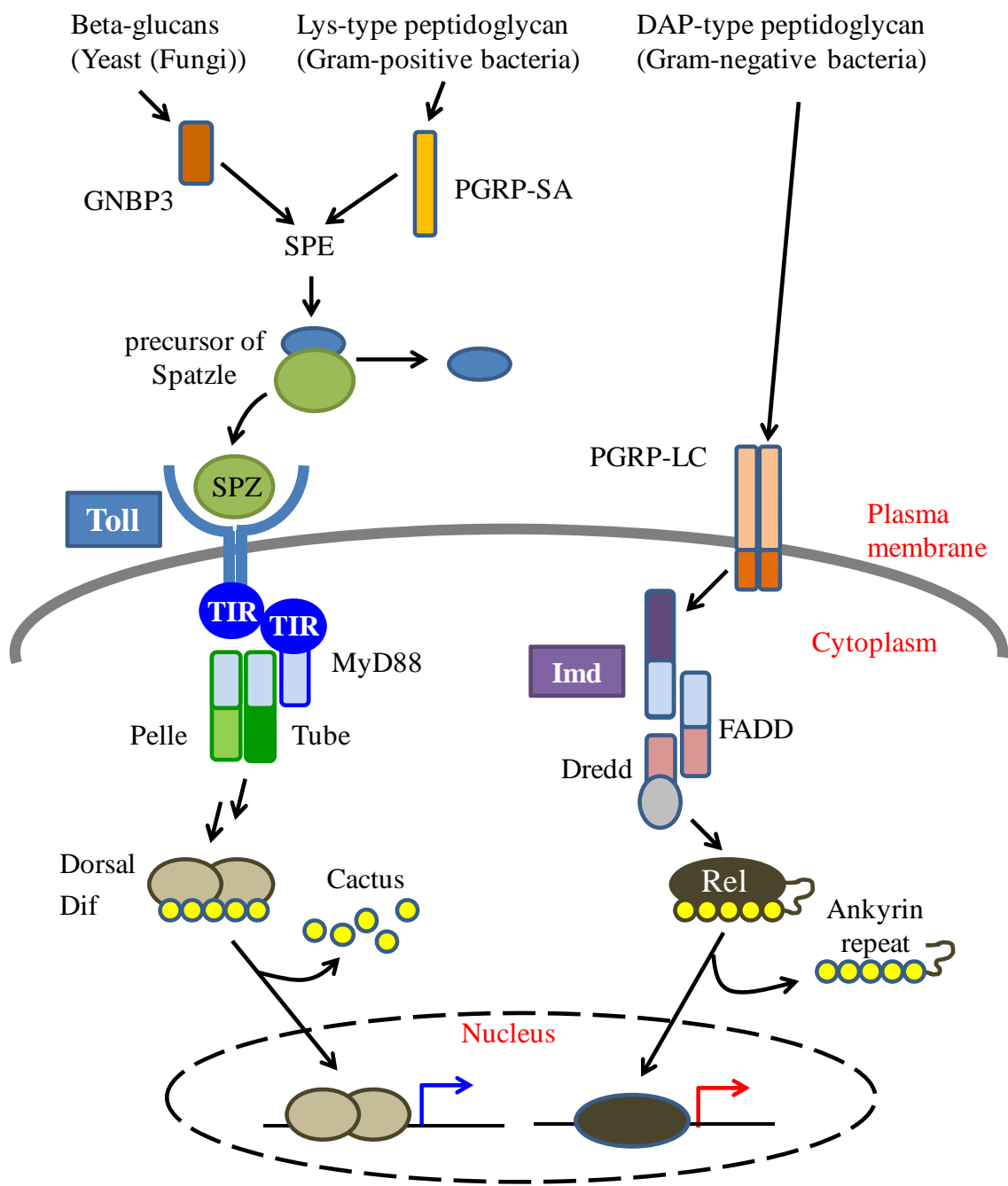
Figure 1: Diagram of Toll and Imd pathway in *Drosophila*.

The Toll pathway is mainly activated by infection of fungi and gram-positive bacteria, and the Imd pathway is largely activated by infection of gram-negative bacteria. The infection of fungi, gram-positive and gram-negative bacteria is sensed by pattern recognition receptors, such as gram-negative bacteria binding protein (GNBP) and peptidoglycan-recognition protein (PGRP). In Toll pathway, after beta-(1.3)-glucans and Lys-type peptidoglycans, which are a component of cell wall of fungi and gram-positive bacteria, are recognized by GNBP3 and PGRP-SA, spatzle processing enzyme (SPE) cleaves the precursor of Spatzle (SPZ). Binding of the SPZ to Toll receptor triggers conformational changes in the receptor and activates the Toll receptor. The activation of Toll receptor cause formation of MyD88-Tube-Pelle complex and the signal proceeds to the phosphorylation and degradation of the Cactus, which is an inhibitor of NF-kappaB (NF-kB) like transcription factor, Dorsal and Dif. Then the Dif and the Dorsal move into nucleus and activate the transcription of a set of target genes. In Imd pathway, after DAP-type peptide glycan, which is a component of cell wall of gram-negative bacteria, is recognized by PGRP-LC, the signal activates intracellular adaptor Imd. The Imd interacts with FADD, and the FADD activates caspase, Dredd. Ankyrin-repeat of Relish (Rel), is cleaved by Dredd, and the Relish is converged to active form. Then the Relish is translocated into nucleus and activate the transcription of a group of target genes.
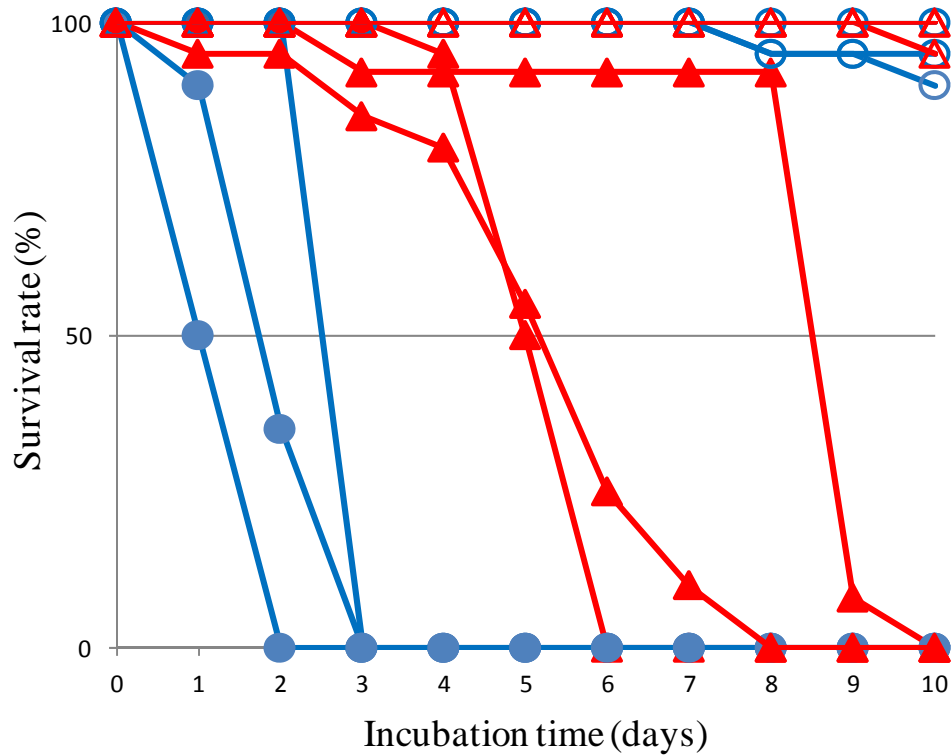
Figure 2: Survival curves of fungal infected and naïve *D. virilis* and *D. melanogaster*. Twenty to twenty five flies 1-day after eclosion were reared at 25 ºC on the culture medium covered by a *Penicillium* species (infected) or without fungus (naïve). The red lines with filled and open triangle data points indicate fungus-infected and naïve *D. virilis*, respectively, whereas the blue lines with filled and open circle data points indicate fungus-infected and naïve *D. melanogaster*, respectively. All measurements were independently replicated in three times. The number of lines indicates the experimental replications.
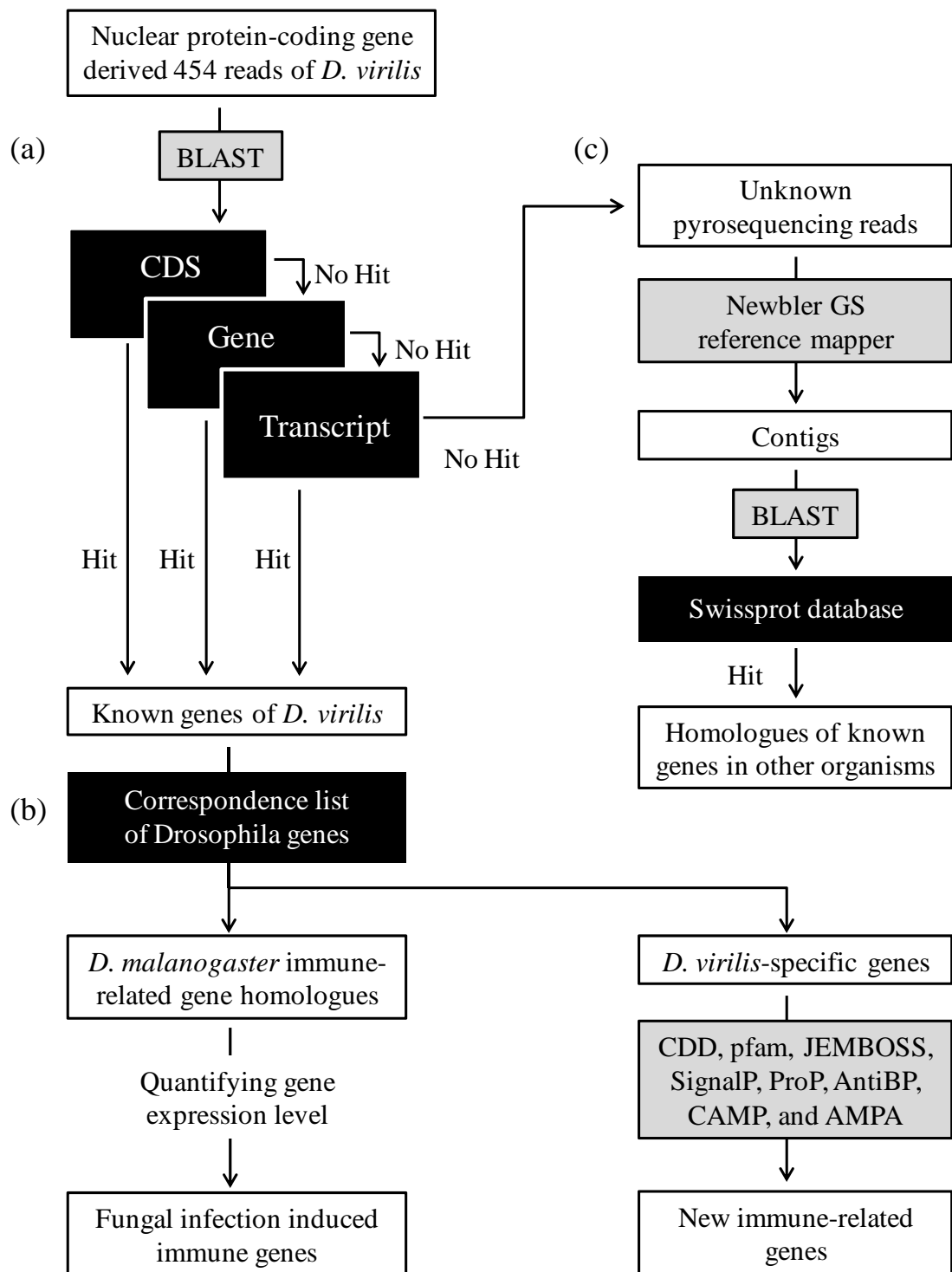
Figure 3: Workflow of data analyses for gene identification (a), gene expression (b) and prediction of immune related gene (c). Input data in an open box is processed by program(s) in the grey box on the following arrow with or without a database in the black box leading to its outcome in the open box.
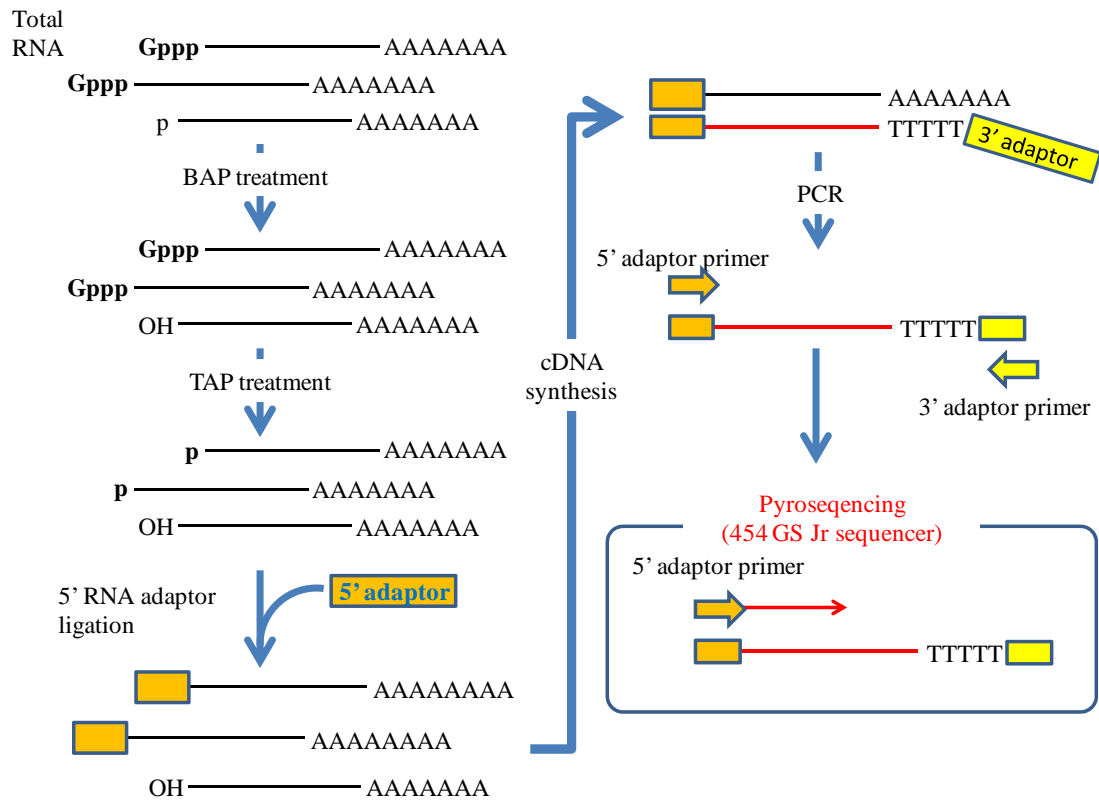
Figure 4: Workflow of construction of oligo-capping full-length cDNA library and 5'-end-enriched pyrosequencing by using 454 GS Jr sequencer.
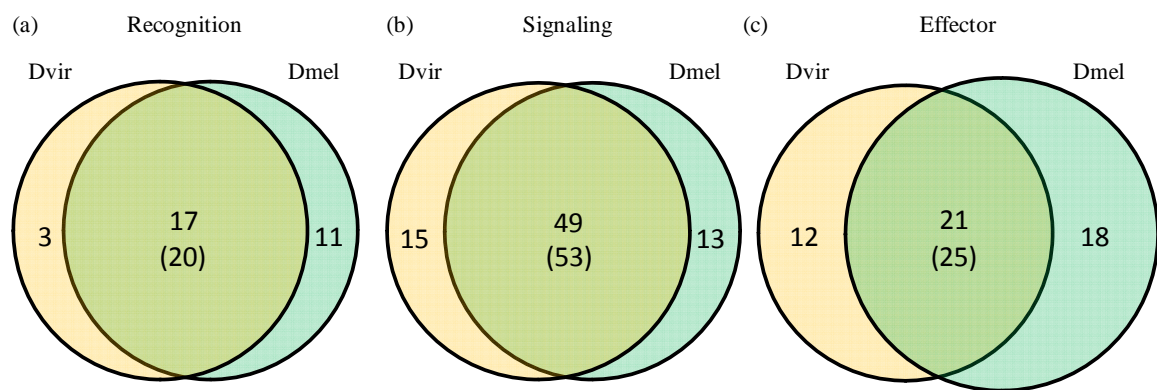
Figure 5: Venn diagrams that represent the numbers of expressed immune-related genes for recognition (a), signaling (b) and effectors (c) observed in the *Penicillium*-infected *D. virilis* (Dvir) and *D. melanogaster* (Dmel) larvae. The numbers in parentheses indicate the numbers of duplicated genes in *D. virilis*.

Dvir

- Dpt (GJ19916) **
- Def (GJ22479) **
- CecA1, CecA2 (Cec2B) **
- CecC (Cec3) **
- Pu (GJ20545)
- DptB (GJ19917) **
- AttA (GJ20572) *
- AttC (GJ21173) **
- Mtk (GJ22469) **
- LysD (GJ12507)
- AttD (GJ22662)
- Irc (GJ10563)
- AttA (GJ20571)
- CecC (Cec1)
- Catsup (GJ11603)
- Jafrac1 (GJ15754)
- Jafrac2 (GJ12891)
- Ddc (Ddc)
- Tsf2 (GJ13808)
- PO45 (GJ21181)
- Tsf3 (GJ20669)
- Duox (GJ14144)
- IM4 (GJ18607) *
- IM1 (GJ19885) **
- fon (GJ17981) *
- Tsf1 (GJ15366)
- IM10 (GJ21308)
- IM10 (GJ21309)
- CG6426 (GJ21009)
- Tig (GJ15168)
- CG15293 (GJ18065)
- CG6426 (GJ21008)
- IM23 (GJ22454)
- CG14823 (GJ13134)
- LysD (GJ12894)
- CG16799 (GJ18809)
- yellow-f2 (GJ10805)

Dmel

- Drs **
- Mtk **
- IM3 **
- Dpt **
- IM1 **
- IM2 **
- IM10 **
- TotA **
- IM14 **
- AttC **
- Dro **
- DptB **
- IM4 **
- AttA **
- CecC **
- AttB **
- IM18 **
- Tsf1 **
- Pu **
- CecA1 **
- Def **
- TotC **
- CecB **
- dro5 **
- yellow-f **
- IM23
- Irc
- CG33470
- CG15293
- CG18107
- CecA2
- Jafrac1
- TotB
- Tig *
- Tsf3
- Hml
- fon
- CG16799
- Catsup

Induction coefficient

	infinity
	2 <
	1 ~ 2
	0.5 ~ 1
	< 0.5

80

Figure 6: Summary of changes in gene expression level of the effector genes in the *Penicillium*-infected larvae.

The effector class genes are piled in order of the expression level in terms of trimmed mean of M values (TMM). Expressions of genes observed only in the *Penicillium*-infected larvae are displayed in red. Genes of the induction coefficient greater than 2.0, between 1.0 and 2.0, between 0.5 and 1.0 below 0.5 are displayed in dark orange, light orange, light blue and dark blue, respectively. The AMP genes and the IM genes homologous between *D. virilis* and *D. melanogaster* are connected to each other by red lines and blue lines, respectively. For each *D. virilis* gene, the gene name of its homologue in *D. melanogaster* is indicated and the gene name of *D. virilis* is indicated in parenthesis. Asterisks indicate a statistically significant difference in the number of reads observed between the infected and naïve larvae (* $P < 0.05$; ** $P < 0.01$).

Used reads	22,200
Mapped reads	21,488
Total no. of contig	3,269
Largest contig (bp)	2,633
Shortest contig (bp)	100
Average length of contig (bp)	237
Median (bp)	192
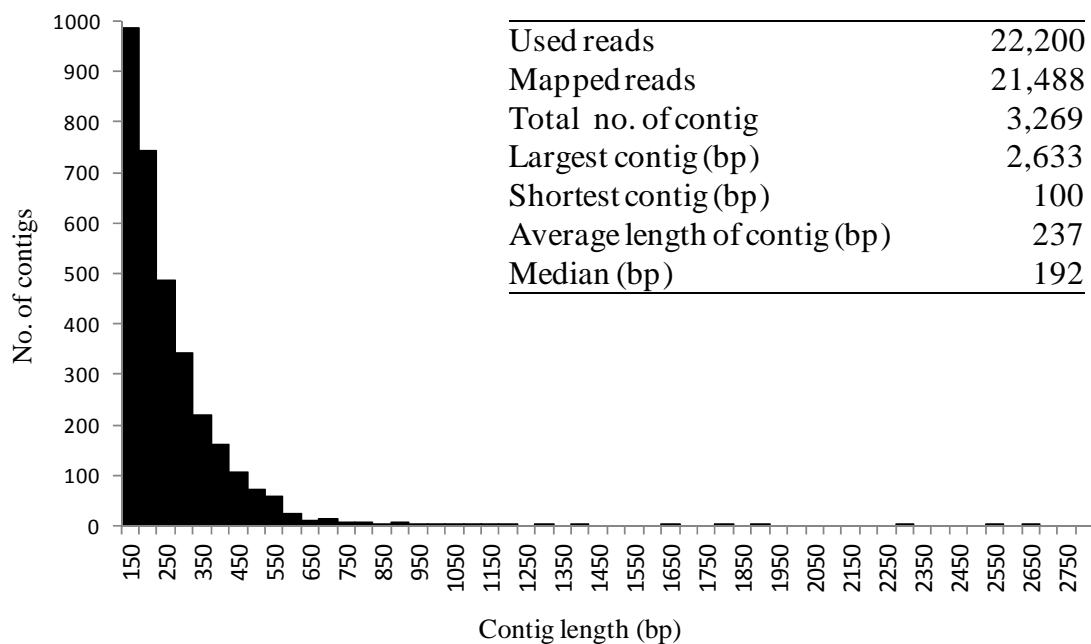
Figure 7: Distribution of sequence length (bp) of contigs constructed from the pyrosequencing reads of *D. virilis* that did not hit any annotated genes.

Signal peptide ← → Mature peptide

```
Dmel_CG15065   MKWMSLVFLCGLLAMAVASPLNPGNVIINGDCRHCNVRGG*
Dvir_GJ22451   MKWLSLAFVMALLALASANPLQPGHVIINGDCKVCNVRGD*
```
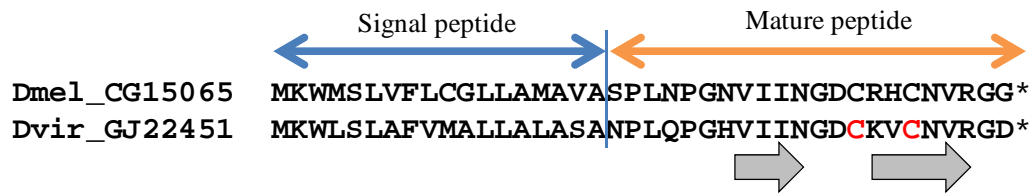
Figure 8: Amino acid sequences and structural features of *GJ22451* and its homologue of *D. melanogaster* (*CG15065*). *GJ22451* was predicted to have signal peptide at N-terminal region indicated by the blue doubleheaded arrow. The predicted mature peptide is indicated by the orange doubleheaded arrow. Additionally, *GJ22451* was predicted to have two beta-sheets indicated by gray arrows and to be stabilized by disulfide bridge between two cysteine residues in red.

```
Dmel (Mtk)               --HRHQGPI- --------FD TRPSPFNPNQ PR-PGPIY--
Dsim (GD11114)           --........- --------.. ..........  ..-......--
Dper (GL20065)           --RHR....- --------.. .........P ..GG-.F---
Dsec (GM21609)           --........- --------.. ..........  ..-......--
Dyak (GE11702)           --........- --------.. ..........  ..G.....--
Dere (GG20517)           --..QR...- --------.. ..........  ..-..R..--
Dana (GF13036)           --........- --------.. .........A .G-.S.P.RG
Dpse (GA24917)           --RHR....- --------.. .........P ..GG-.F---
Dwil (GK19217)           --RHR....- --------.. ..........  ..GG....--
Dgri (GH20235)           --S.R....- --------.. .........P ..GG-.YV--
Dmoj (GI20102)           --RHREDRNP --------.. .........P ..-...YV--
Dvir (GJ22469)           --RH.E.R.P --------.. ..........  ..-...YI--
Dvir_PG01471             NI...HL.PP PPGNQWNP.N FN......G. ..----Y---
Dmoj_PG01471_homologue   NL...PM.PP PPGNQWNP.. F.......G. ..----F---
```

Figure 9: Amino acid sequence alignment of mature peptide of Metchnikowin from *D. melanogaster* and its homologues in 11 *Drosophila* species and *PG01471* from *D. virilis* (indicated by bold face) and its homologue from *D. mojavensis.* Multiple alignment was constructed by CLUSTAL W (Higgins et al., 1994) program. The amino acid residues identical to the uppermost sequence are indicated by dot. The red boxes surround highly conserved regions. For each gene, abbreviated four-letter species code (Dmel: *D. melanogaster*, Dsec: *D. sechellia*, Dsim: *D. simulans*, Dere: *D. erecta*, Dyak: *D. yakuba*, Dana: *D. ananassae*, Dper: *D. persimilis*, Dpse: *D. pseudoobscura*, Dgri: *D. grimshawi*, Dmoj: *D. mojavensis*, Dvir: *D. virilis* and Dwil: *D. willistoni*) with the gene name in parenthesis is shown.

species	GJ10737	GJ18291	GJ22451	PG 01471
Dmel	-	-	◯	-
Dsim	-	-	◯	-
Dsec	-	-	◯	-
Dyak	-	-	◯	-
Dere	-	-	◯	-
Dana	-	-	◯	-
Dpse	-	-	◯	-
Dper	-	-	◯	-
Dwil	-	-	◯	-
Dmoj	-	△	◯	◯
Dvir	◯	◯	◯	◯
Dgri	-	-	◯	-
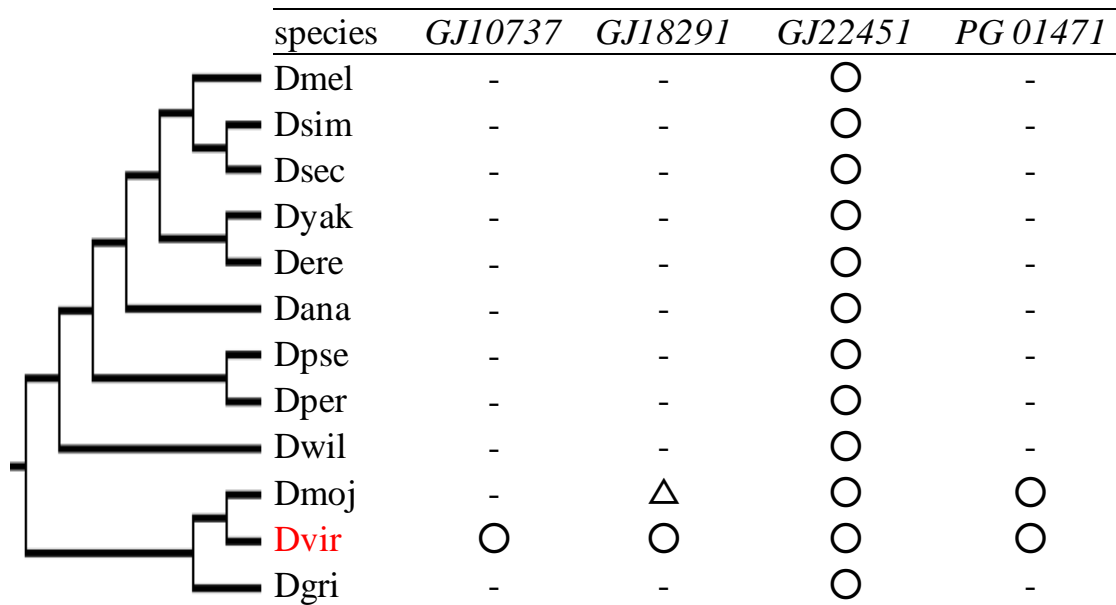
Figure 10: Distribution of the four predicted immune-related genes, which may function as AMP, among 12 *Drosophila* species. Presence of the gene is indicated by circle (identity is 50% or more) or triangle (identity is less than 50%). The phylogenetic relationship of the 12 *Drosophila* species is displayed at the right side.
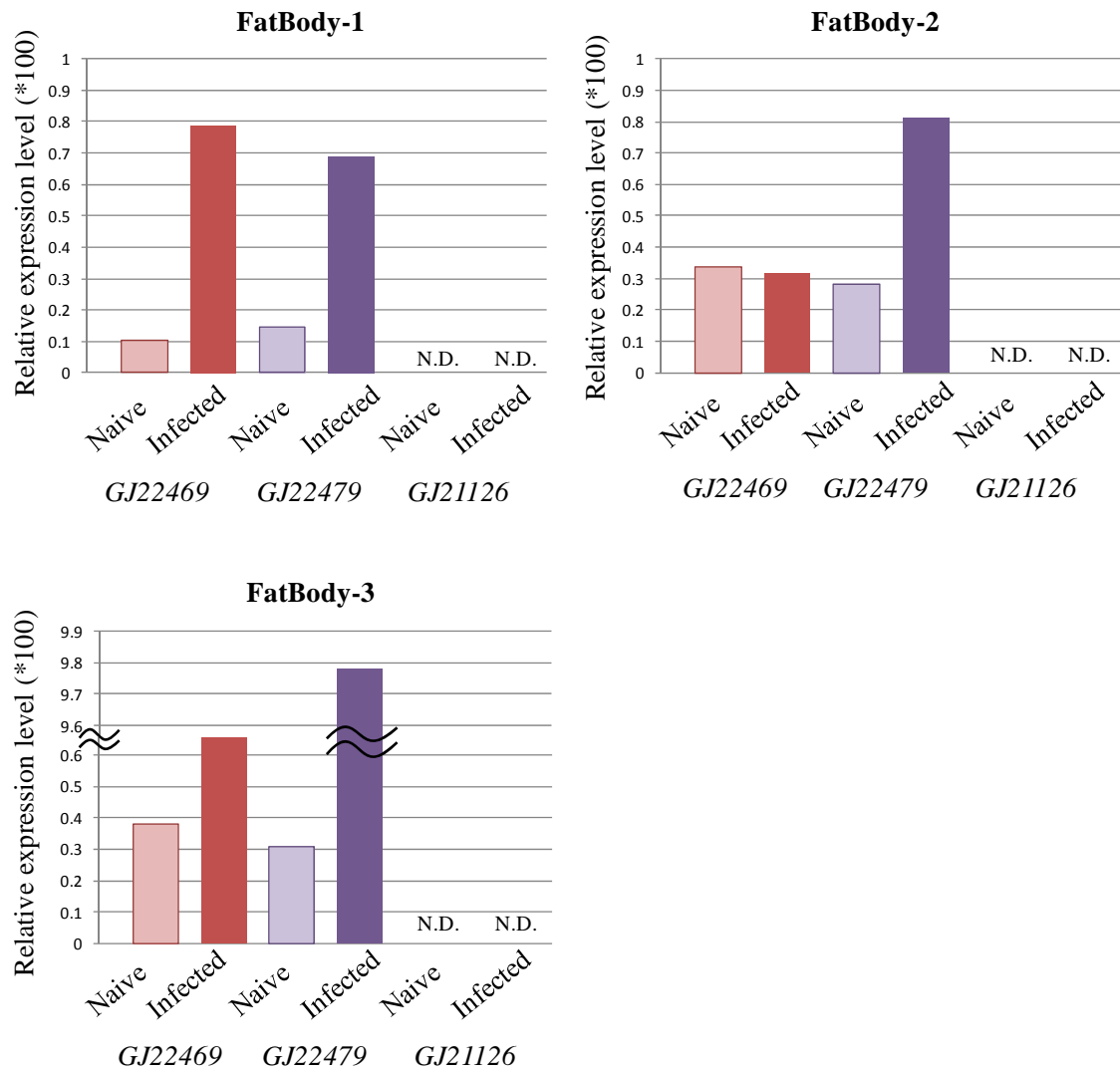
Figure 11: Three biological replicates (Fat body-1, -2 and -3) of comparison of gene expression level of Metchnikowin (*GJ22469*) and Defensin (*GJ22479* and *GJ21126*) in fat body of *D. virilis*. The gene expression level was normalized by the gene expression level of *RpL32*. N.D.; Not detected.
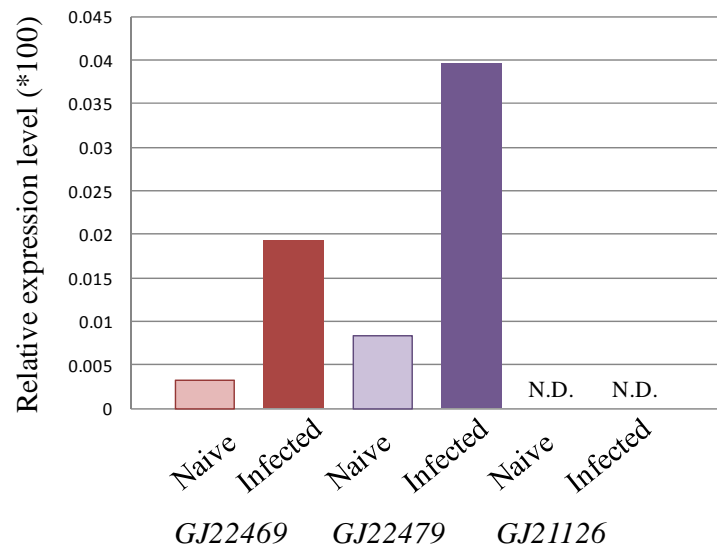
Figure 12: Gene expression response of Metchnikowin (*GJ22469*) and Defensin gene (*GJ22479* and *GJ21126*) to the fungal infection in salivary gland of *D. virilis*. The gene expression level was normalized by the gene expression level of *RpL32*. N.D.; Not detected.
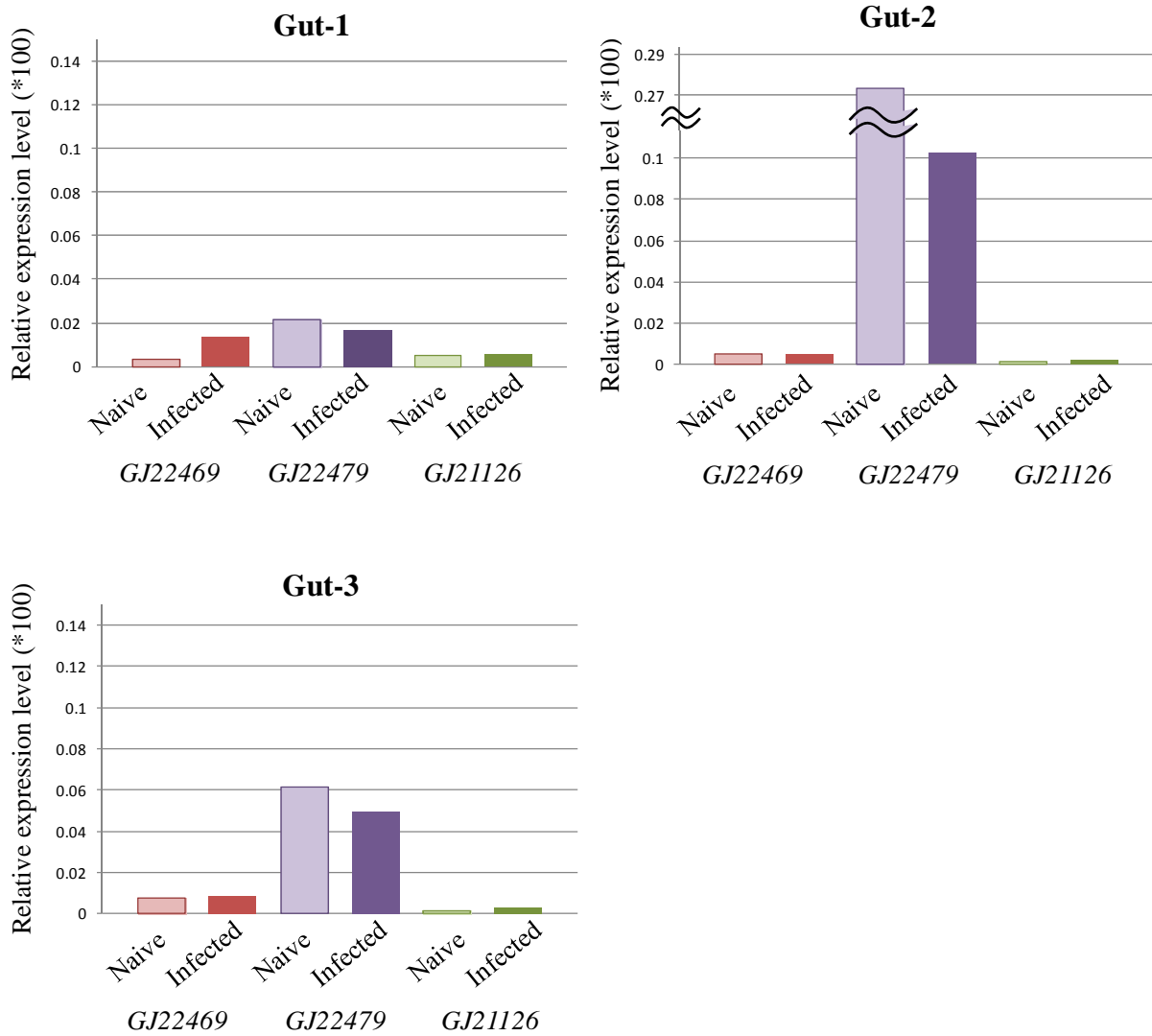
Figure 13: Three biological replicates (Gut-1, -2 and -3) of gene expression level of Metchnikowin (*GJ22469*) and Defensin (*GJ22479* and *GJ21126*) in gut of *D. virilis*. The expression level was normalized by that of *RpL32* expression level.

(a)

```
Dmel (Dpt)       DDMTMKP--- --TPPPQY-P LN----LQGG GGGQSGDGFG FAVQGHQKVW TSDNGRHEIG
Dvir (GJ19915)   -NPEE..KGD VW.ERQ.FN. P.EQRF.LD. .YNKDKS.KD VWA.AQVP.. ..E.K...FD
Dvir (GJ19916)   -NPEE..KGD VW.ERQ.FN. P.EQRF.LD. .YNKDKS.KD VWA.AQVP.. ..E.K...FD

Dmel (Dpt)       LNGGYGQHLG GPYGNSEPSW KVGSTYTYRF PNF
Dvir (GJ19915)   VI.K...... ..W....... GA.GN.RF.. ---
Dvir (GJ19916)   VI.K...... ..W....... GA.GN.RF.. ---
```

(b)

```
Dmel (DptB)      ---------- -----DPREI VNLQ---PEP -LAYAPNFDV P----LHRVR RQFQLNGGGG
Dvir (GJ19917)   LLTVDDEPAT QLVSAK..SL LS.RLMV.D. NKQL.E.Y.W APSEQVEQL. VPR...VQ..

Dmel (DptB)      GSPKQGFDLS LNGRAPVWQS PNGRHSFDAT GSYAQHLGGP YGNSRPQWGA GGVYTFRF
Dvir (GJ19917)   ...R...... V......... ......L... .Q.S...... ......N... .AQ.....
```

Figure 14: Amino acid sequence alignment of mature peptide of Diptericin from *D. melanogaster* and *D. virilis*. (a) alignment of Diptericin (Dpt) of *D. melagnoaster* (Dmel) and its homologues (*GJ19915* and *GJ19916*) of *D. virilis* (Dvir). (b) alignment of DiptericinB (DptB) of *D. melanogaster* and its homologue of *D. virilis* (*GJ19917*). The sequence alignment was constructed by CLUSTAL W (Higgins et al., 1994) program. The amino acid residues identical to the uppermost sequence are indicated by dot. Gaps are indicated by hyphens.

```
Dmel (CecA1)    GWLKKIGKKIERVGQHTRDATIQGLGIAQQAANVAATARG
Dmel (CecA2)    ........................................
Dmel (CecB)     ...R.L......I.......S..V................
Dmel (CecC)     .....L..R...I...........................
Dvir (Cec1)     ............I...........................
Dvir (Cec2A)    ........................................
Dvir (Cec2B)    ........................................
Dvir (Cec3)     ............I...........................
```

Figure 15: Amino acid sequence alignment of mature peptide of Cecropin from *D. melanogaster* (Dmel) and *D. virilis* (Dvir). The multiple alignment was constructed by CLUSTAL W (Higgins et al., 199) program. The amino acid residues identical to the uppermost sequence are indicated by dot.
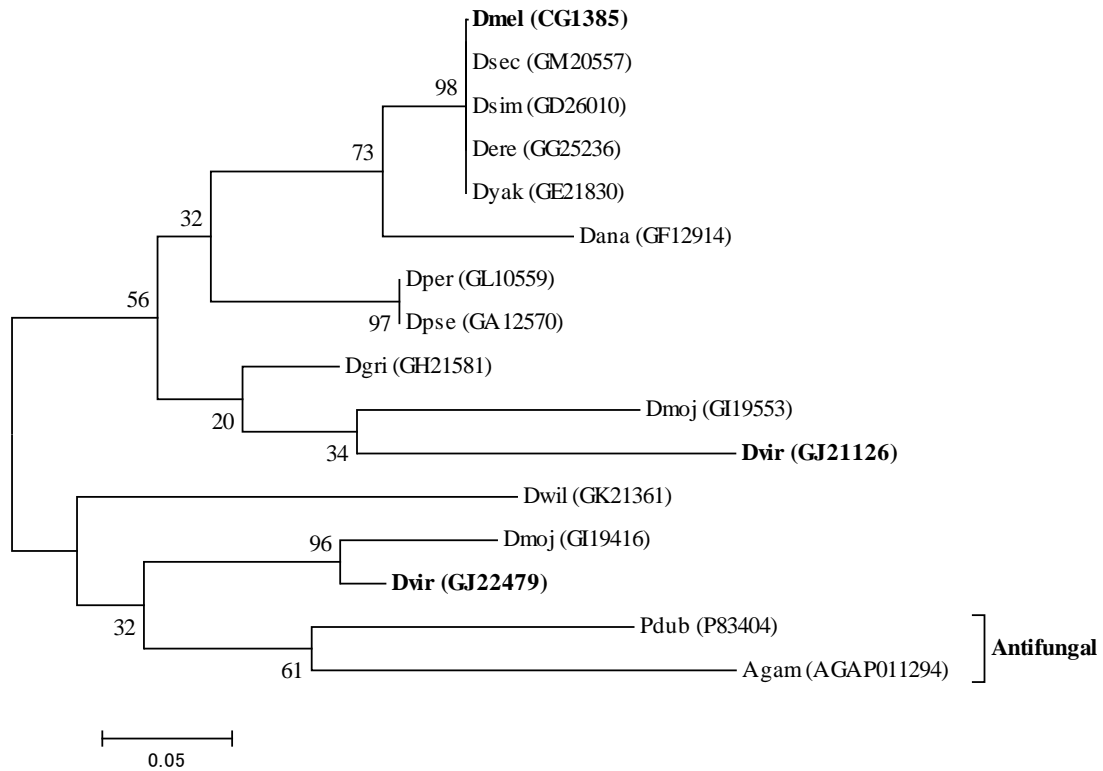
Figure 16; Neighbor-joining phylogenetic tree of *Drosophila* Defensin genes with antifungal Defensin genes of sand fly (*Phlebotomus duboscqi*) and mosquito (*Anopheles gambiae*). Amino acid sequences of the mature peptide were aligned by CLUSTAL W (Higgins et al., 1994) and the phylogenetic tree was reconstructed with the Poisson model by MEGA5 (Tamura et al., 2011). For each Defensin gene, abbreviated four-letter species code (Dmel: *Drosophila melanogaster*, Dsec: *D. sechellia*, Dsim: *D. simulans*, Dere: *D. erecta*, Dyak: *D. yakuba*, Dana: *D. ananassae*, Dper: *D. persimilis*, Dpse: *D. pseudoobscura*, Dgri: *D. grimshawi*, Dmoj: *D. mojavensis*, Dvir: *D. virilis*, Dwil: *D. willistoni*, Pdub: *Phlebotomus duboscqi* and Agam: *Anopheles gambiae*) with Gene ID or Uniprot ID in parenthesis is shown as an operational taxonomic unit. The Defensins genes of *D. melanogaster* and *D. virilis* were indicated by bold face. The number along each branch is the bootstrap value computed by 1,000 bootstrap replicates.

```
Dmel (CG1385)       ATCDLLSKWNWNHTACAGHCIAKGFKGGYCNDKAVCVCRN
Dvir (GJ21126)      .......F..VKN...VA..L.RRY......N..I....R
Dvir (GJ22479)      .......GF.V..S...A...GL.RS.............R
Pdub (P83404)       .......AFGVG.A...A...GH.YR.....S....T..R
Agam (AGAP011294)   .....A.GFGVGSSL..A....RRYR.....S........
```

Figure 17: Amino acid sequence alignment of mature peptide of Defensin from *Drosophila melanogaster* (Dmel), *D. virilis* (Dvir), *Phlebotomus duboscqi* (Pdub) and *Anopheles gambiae* (Agam). The multiple alignment was constructed by CLUSTAL W (Higgins et al., 1994) program. The amino acid residues identical to the uppermost sequence are indicated by dot.
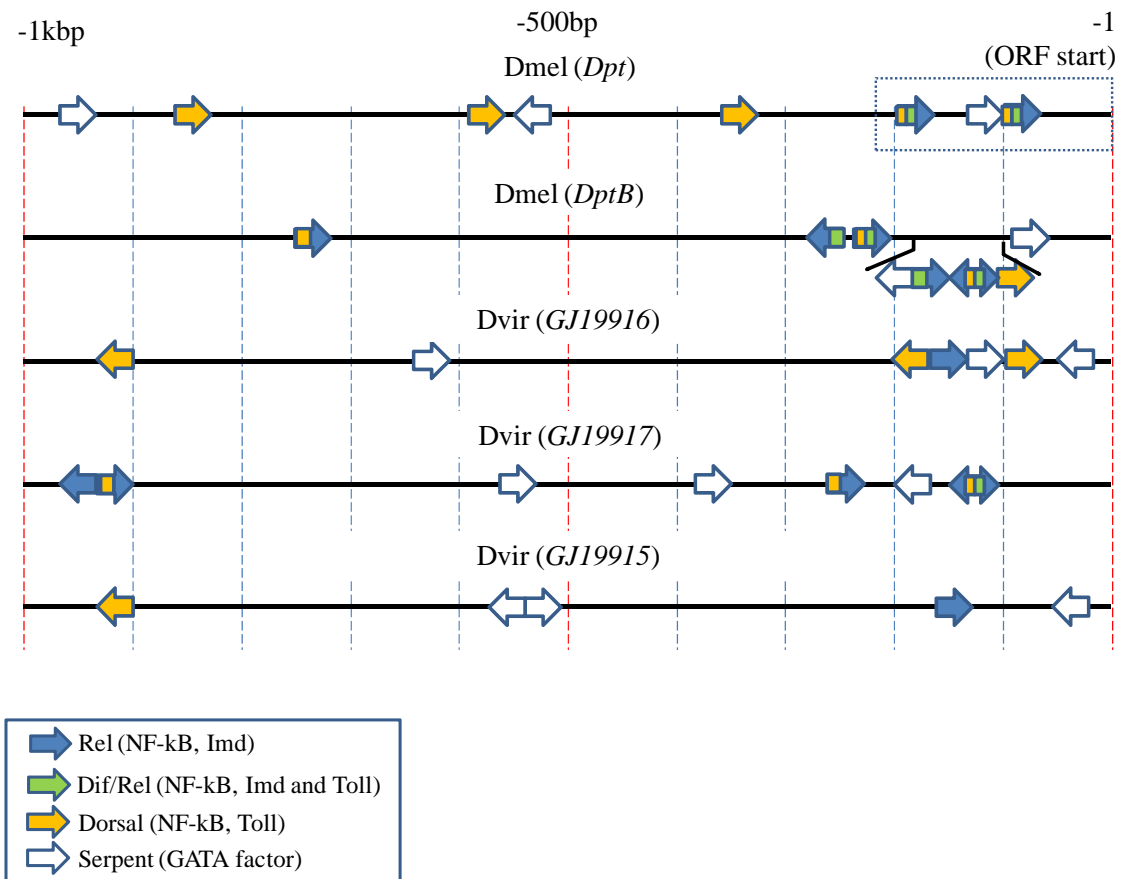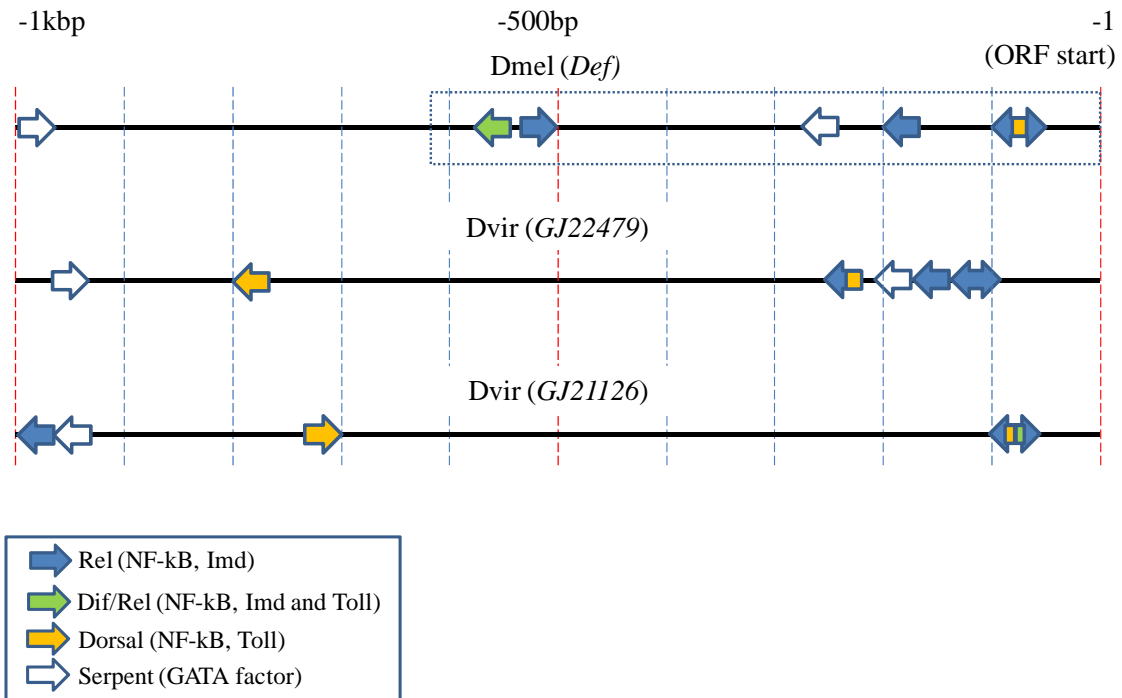
Figure 18: Organization of NF-kB-like transcription factor (Relish, Dif and Dorsal) and GATA transcription factor (Serpent) binding sites on 1 kb upstream region from initiation codon of Diptericin genes (*Dpt* and *DptB*) in *D. melanogaster* and those (*GJ19915*, *GJ19916* and *GJ19917*) in *D. virilis*. Among these Diptericin genes, expression of *GJ19915* was not observed in our transcriptome analysis. Binding sites of Relish, Dif/Relish heterodimer and Dorsal are designated by blue, green, and orange arrows, respectively, whereas white arrow represents Serpent binding site. The dashed box represents the segment experimentally tested for enhancer activity in Senger et al., 2004. From the sequence comparison, the sites on which multiple transcription factors could bind are represented by arrows with two or three colors. Doubleheaded arrows indicate palindromic binding sites.
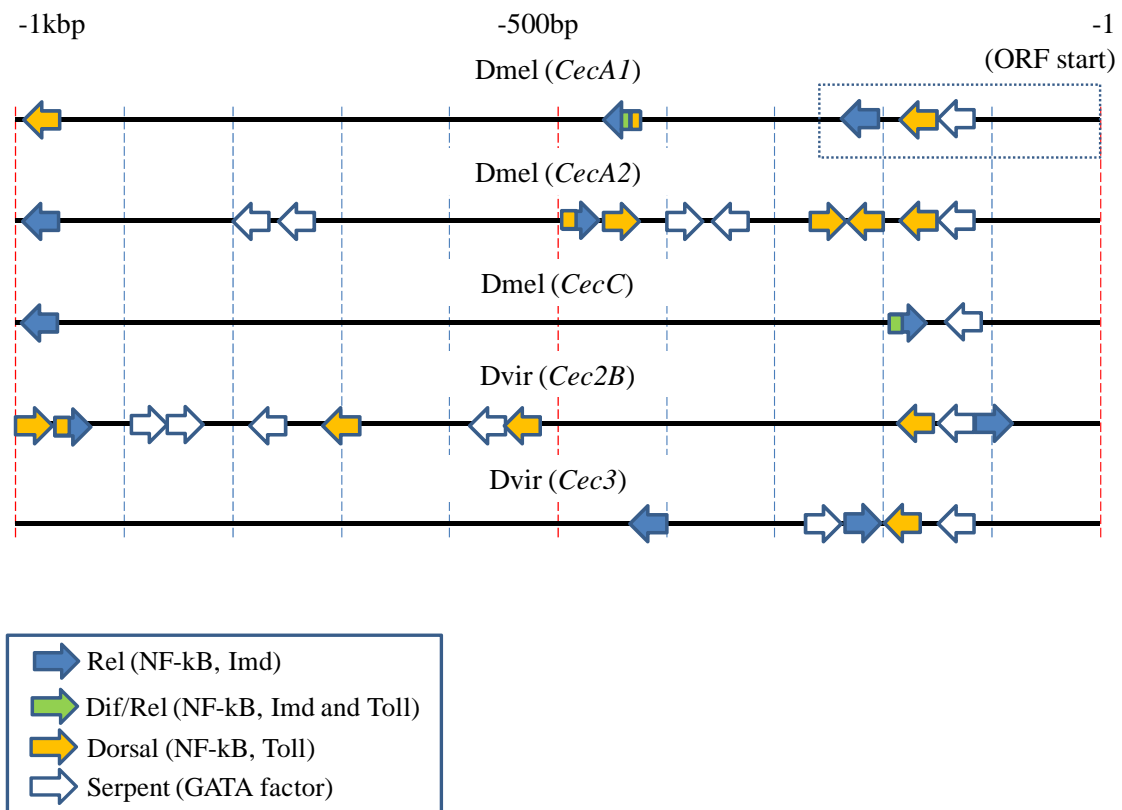
Figure 19: Organization of NF-kB-like transcription factors (Relish, Dif and Dorsal) and GATA transcription factor (Serpent) binding site on 1 kb upstream region from initiation codon of Defensin genes (Def) in *D. melanogaster* and those (*GJ21126* and *GJ22479*) in *D. virilis*. Among these Defensin genes, although expression of *GJ21126* was not observed in fat body and salivary gland, the expression was detected in gut. Binding sites of Relish, Dif/Relish heterodimer and Dorsal are designated by blue, green, and orange arrows, respectively, whereas white arrow represents Serpent binding site. The dashed box represents the segments experimentally tested for enhancer activity in Senger et al., 2004. From the sequence comparison, the sites on which multiple transcription factors could bind are represented by arrows with two or three colors. Doubleheaded arrows indicate palindromic binding sites.

Figure 20: Organization of NF-kB-like transcription factors (Relish, Dif and Dorsal) and GATA transcription factor (Serpent) binding site on 1 kb upstream region from initiation codon of Cecropin genes (*CecA1*, *CecA2* and *CecC*) in *D. melanogaster* (*CecA1*, *CecA2* and *CecC*) and those (*Cec2B* and *Cec3*) in *D. virilis*. Binding sites of Relish, Dif/Relish heterodimer and Dorsal are designated by blue, green, and yellow arrows, respectively, whereas white arrow represents Serpent binding site. The dashed box represents the segments experimentally tested for enhancer activity in Senger et al., 2004. From the sequence comparison, the sites on which multiple transcription factors could bind are represented by arrows with two or three colors. Doubleheaded arrows indicate palindromic binding sites.

Appendix I-1: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *Def* in *D. melanogaster*.

CAAGACAAGACAACCTGGT**CGATAACA(Serpent_fwd)**AAGGTAAACAGGCAACGGCCAGCC
AAGGAGCAGGGCAACTGAAAAAGCCTCGGTCGTGGGATTCGATCGGGATTGTCGGCTCAGCGCG
ACTGGGCTGAGAACCAGATGCAGATGCCGATACAGATATAGATACACGTACGCGCAGATACGGA
TTCAGATACAAGTACACCCGCCCCTGCCGCTGTATGCCCAACTAATCATTGTGTGATTCTTGTT
TGTTTATTTGCCCGGCATTATGAAGAGACTTTTCGGTAGAAATTATTTATTGTCGCATGTGTTT
ATGTATCCGTAACCGAGTATCTCAGTTGCTTGAGCCAACTGTGTAGCTGTGTAGCTGTGAGTAT
AGCCCTTAAAGTGGCACCCAATCGGTCAGTTAGCTAGAAATTCAGATGATTAAATA**TGGATTCC**
C(Dif/Rel)CTACATCAGCTAATTTCAACAGTTT**GGGAGTAAT(Rel)**AAAATCGAAATTGGA
TGCTACTAAAGGGCACATATTTACTTAGGCTTTTATCAACGTTGCATATATACAAATATCCTGC
ATATTTCGCAAACCAAAGATTCTTTCTCAAGTAAGGCCTAAACAATTTGAAATGGTTAATTTCG
TAGATGTTGCTTTTTACAATTAACTTGTCATGTGGAATATACTTTACTGCCTAAAATTTAAGGC
AGTTAAAATCCCTAGAAATGCAAATAACTTATTGCAGAAACGGGCTCTGTCGGCTGTATTTTGC
TCTTATCT(Serpent)ATGAAATATTGTCAATATTTTCCAGGCAAAGCACATGAAATAATGAT
CTAGACAACG**GTTTCTCCC(Rel)**ATTTGCAGTGAACTTAAAAATTAAAAACCCCCGAGACGTG
TCTTCCTGCACAGAAAAAGAGACAATGGGAAGGTAAGTCACCGGGT**GGGAGTCCC(Dorsl,Re**
l)TGGGCCGAATCGATCAGCCCGTCGCATTGCTATATAAGCTCGGCGAAACCACAATCTGCAAC
AACAGTATCTCTCCAGTTGTATTCCAAG<mark>ATGAAGTTCTTCGTTCTCGTGGCTATCGCTTTTGCT</mark>
<mark>CTGCTTGCTTGCGTGGCGCAGGCTCAGCCAGTTTCCGATGTGGATCCAATTCCAGAGGATCATG</mark>
<mark>TCCTGGTGCATGAGGATGCCCACCAGGAGGTGCTGCAGCATAGCCGCCAGAAGCGAGCCACATG</mark>
<mark>CGACCTACTCTCCAAGTGGAACTGGAACCACACCGCCTGCGCCGGCCACTGCATTGCCAAGGGG</mark>
<mark>TTCAAAGGCGGCTACTGCAACGACAAGGCCGTCTGCGTTTGCCGCAATTGA</mark>**(Def CDS)**

Appendix I-2: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *Dpt* in *D. melanogaster*.

GTATAATTTTGGTTATAACAAGTAACTTTAC**TGATAAGA(Serpent)**CTTGGATTCTCTTTAT
AATATATTTAACAGAGATGTATATGATGCAATTTATAGAATTTAAAAAATCTTAAGAAACTTAA
AATGTTGCTTCAAGATCCTGGTCATCATTGCCCCA**GGGAAATTC(Dorsal)**CGTCTTTTCCGG
TGGACCTTCACCCCTTGAACCAATTTCAAGCTCTATATAATCGGCAGATGACTGGGCTGTGACG
TGTCCCAGCTGTAAAGTAAATATGCATATATATTTTTTTTTTATATTTTTTTATTTTCGGATTAA
TTGTATATTTTTCCTTTGCCGCACTCACCCAGCAGACCAAAAAAACGGCCAATATTTTCATTAA
TATCGATTGTAACAACATTTCTCTCGGCTGCCTCGGCACTCAACTGCTGAT**GGGAAACTG(Dor
sal)**TTTTTACGTTTGCTCAACCTCTGCTTTTAATC**AATTATCA(Serpent)**CTTATTAAATA
TTTATATTTGTTTTTTTTTTTGTTTGCCTGCATACAAACATACATCGCTCTTTGTCTGTCGCCG
CGGAGAGGTTTTAAAATTAATCCGTGGAACTGGGAAAAGGATGAACTTTCGTTTATTTTCGAAG
GGAAATCATTAATGTTTTAATTGTTAATAATAAGCTGGGAGGTT**GGGATATTG(Dorsal)**TTC
TTAAGATACATATTTAAAAACTTCGTGGAATAAGAGGTTACAAATTTTATCATTTAATAAGTAT
TTAACCTCTTGTTTGTCAAATGAAAATAAGGTGTGAGTCCTCGTTTAAGAAAGATCCCCTGGTG
GTATTTGTTTTTGCATCG**GGGATTCCT(Rel/Dorsal)**TTTTTATGACCGGTAATCAATCTTG
GGTTCTAATTATGAGACAATAACCGCCGTAGGTATACTTTCTGAGT**AGATAAGG(Serpent)**T
GACATCG**GGGATTCCT(Rel/Dorsal)**TTTGGAAAGCGGCCTATAAAAGAGCATCGAAACTGC
AGCAAAGGTATCAGTCAGCATATTCCAGTTCTTCAATTGAGAACAACTGAGATGCAGTTCACCA
TTGCCGTCGCCTTACTTTGCTGCGCAATCGCTTCTACTTTGGCTTATCCGATGCCCGACGACAT
GACCATGAAGCCCACTCCACCACCGCAGTACCCACTCAATCTTCAGGGAGGCGGCGGTGGCCAG
AGCGGCGATGGTTTTGGCTTTGCAGTCCAGGGTCACCAGAAGGTGTGGACCAGCGACAATGGAC
GCCACGAGATTGGACTGAATGGAGGATATGGACAGCACTTGGGAGGACCATATGGCAACTCAGA
ACCGAGCTGGAAAGTGGGAAGCACCTACACCTACAGATTTCCGAATTTCTAA **(Dpt CDS)**

97

Appendix I-3: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *DptB* in *D. melanogaster*.

ACTCCGTTGGGTAATAACAAATTACGATGACAGGCGGTCTTAATGAAACCTGCCAAGACAAAAT
GTTTTTTGCTAATTGGATTTATTGCGTAAACTGTCGAGAGAACCTTCGATGGAAATTTATCATG
CGCAACATCCGCAGCCTAACCCGAAGATTAAAATATTACGCCATCCCCAACGAAAGCGAATAAG
ATTCGCTTCGAGTATTTGAAAAATGATTCGTGCAGAGGCTGTATTCTAAACTCTTTTACAACCA
AATTTG**GGGATTACC(Rel/Dorsal)**AAAGCTTTTCACATCATTAATTCCAGCTAAGAATGGC
CAAATTGAGTTGACCTTGTAAAATGATGAATCGATTTGGGAAACTAAATTCTGCTTGGGTCTGC
GGAACCTGTAAGGTAATTTTGCTTGGGAGGAAATCAAATTTTGGCGAAAGTAACATCCGATATT
CTTACTAATAGAAAACACGAAATTGGAGATGATTAAAAATAGAATATATGTTGAATGAAAATGT
ATTATATATGCATAAAACTCAGGTTACCAGAAAGAAAGTTGGTGTTAAAATGATAAAACATAGA
TTTGCATCTTAACTAGAATGCATAAAAGAAATATAACTCTTCGTATAATGCTCTTTTTTATTAG
AAAGAATTAATTTAGAAAGAATTAATGTTAGCCTAAGAATCTGAAATAGAATTAGCTTGTAGTA
GTTAATATACATATATTATTTTCACAGCTGCCACACATGA**ATGGATCCC(Rel,Dif/Rel)**AA
TCGAAATGCGAGTACCTCTACATTCATTGTTTCTAGAT**GGGATTCAC(Dorsal,Rel)**TTTGG
CATCTGCCTCTATTTGAGGAGCCTACACGTACC**TCTTATCA(Serpent**)ACGAAATGTG**GGGA
TCCAC(Rel,Dif/Rel)**TGGTGAACCACTGCACCGTCGTCTTGACA**GGGATTCCC(Rel,Dif
/Rel,Dorsal)**AATTG**GGGAATCTC(Dorsal)**ATCTGCGAGTAC**TGATAAGA(Serpent)**C
ACAGATCCGAGCTATATAAGACCTGAGTTCGACTCTAGCCGCATTCAGTTGACAAAGCCTAATC
AAATCAAAATGCATTTCACCGCTAGTCTTCTATTCATTGGACTGGCTTGTGCCTTCTCGAGTGC
CTGGGCTTATCCCTATCCTGATCCCCGAGAGATTGTGAATCTGCAGCCTGAACCACTGGCAGTA
AGTTTTTATGAAACTTTTCACTTGGAGCACAGTTTAATAGAGTATTTTTCTAGTATGCTCCCAA
TTTTGATGTGCCCCTGCACAGAGTGCGTCGCCAGTTCCAATTGAATGGCGGTGGCGGTGGAAGC
CCAAAGCAAGGATTCGATCTGAGCCTCAACGGACGTGCTCCCGTTTGGCAGAGCCCCAATGGAC
GCCACTCCTTCGATGCCACGGGATCGTATGCCCAGCACCTTGGTGGACCCTATGGCAACAGCCG
GCCTCAGTGGGGAGCCGGTGGAGTGTATACCTTCAGGTTTTAG **(DptB CDS)**

Appendix I-4: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *Mtk* in *D. melanogaster*.


```
TAGCGCCACGTTCAACCTCTTTTGCAGCCCCATTCTGCTGCGAGAAAACTAACAAAGTGCTCTA
ATCGAGCCAAGGGGCAATTTCTTGTGTTGCTGCAGCTGCACTTTGCACCTCCGCATCCGTGCAC
C**CAAAAACCC(Dorsal)**GCTTTCTAGATGTTCTCATCATGCACTGAAAAAGAATCCAAATTTT
TACAAGAAATAGTTTAAAATTAGGTAATGTGAAAGATATCGGCACACGGACAGGCCGAATTATC
TGTTGTAAAACTAGCTGCTCAGTTATTAAAAACATTTGTAGTTGCTGACGTTTCCATACAGAGA
CTAATTTTATTTTCACGGACAG**GGGTTTTCC(Dorsal)**GCTTTAATTGCTTCATTTTTGTTGC
TTTATTGCGTGTATATTGCCCCACAAAACAGATATAAATCATTCGCGCATATCGTAAATGTTGG
TAGAAAATGACAAACAAGAGAAAAAAGATGATTAAAAGCTTCAAGACAATCCTCTATAGGATCT
GATTAAATATGAATATTTTATTTATTAGTTTTCTTTCTGTGTACGGCTTAGAAGGCAGAAGCTG
CGAGGGGCGTAGGGCAGTGGGCGTGGCTCCGTGTTGACGCATGTTGACTATGCCTTTGAATGGC
TGCCGTGGTTGTCGGT**GGGTAATTT(Dorsal)**GCAATGCAGAAAAACCAACAGGGCGCTAAAA
AGGAGAGTGTTTTCGTGGGAGGTGGAGATGGTCACTGGGGGCAACATAAATATTCAGCGAGAAA
CGTCATATTTACATTTAGTCTAGGC**TGATAATC(Serpent)**CGGGACCGT**GGGAAGTCCC(Do
rsal,Rel)**CTTTGGGTGGTGCTGGC**TGGGTTCCC(Dif,Rel)**CTGGCCACAATC**GGTTATCT
(Serpent)**GCCCCCGGCTGACACTTGCCCGTCATTCATTCGGCT**GCTTATCG(Serpent)**CA
GAAGCTCAAATA**AAAAGTCCC(Dorsal)**CAATCTGCGACTCGTTTGTCTGGGACTGAGCTATA
AAAGCCTCACCATCTCAACGCTCAAAGCATCAATCAATTCCCGCCACCGAGCTAAGATGCAACT
TAATCTTGGAGCGATTTTTCTGGCCCTGCTGGGTGTGATGGCCACGGCTACATCAGTGCTGGCA
GAGCCTCATCGTCACCAGGGACCCATTTTCGATACGAGGCCGTCGCCCTTCAATCCTAACCAAC
CAAGACCGGGTCCAATTTATTAA **(Mtk CDS)**
```

Appendix I-5: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *CecA1* in *D. melanogaster*.

CATCAGTGTAAAATTC**GGAAAACCC(Dorsal)**AGCGATCTAGTTATGAAATACTTTGTGGTCC
TTGTCGTCCTGGCCCTCATTTTGGCCATCAGCGTGGGTCCTTCGGATGCAGTATTTATTGATAT
TCTTGACAAAGTGGTTTGTTTCTTCTTTAAACAATTGTAGTTTACAATGAAGCTTAAACATTTG
TATTTCTACAGGAAAACGCAATACACAATGCTGCTCAAGTGGGAATTGGCTTTGCTAAGCCCTT
TGAAAAATTGATCAATCCGAAGTAATTCTGCACTGCAATTTAATTAATGTATCGTTTAACGAAA
ATAAACACAAATTTTAAAATCTGAAAAACAACTAAGTTACTAACGCAAGACTTTTAGTTAAGTT
AGTTAATATAGACCGAGATGTATGTACATACATACCGCTTTCGCTTACAATAAAATGTTAAATA
AGTTTTCAGATTCGTACGTGCTCAGTAAACAATTATTTTTTATTGTCATTTAATGCCTATTGAA
TTTTTCAAACTTAATTTAGTGCCTTTAGTAAAATATTGTA**GTGATTCCC(Rel,Dorsal,**
andDif/Rel)CTCGAAAAATACCACAAATTGGATGCGTTTATGTAAATAAATTGCCCTTGAGT
GATAGAGTAAATTTGAATTTGACTGTCTTAGAAAGATAGAAAGAGATCAATTCAAAATGCCAAA
AGGATAGAGTTATTAAAGCTCTAATTCAAATTGGCCCAGAACCGTTTAAAGGATATTACAATTT
GTAATTTACATATTTGGATTATAGCATTGAAATCCCCG**ATTGTTCCC(Rel)**TAGATGTGCAGA
TGTGTGCTTGGAATCAGATCGGTTACCTTCAGTGTACTTTTCTCTGC**AAAAATCCC(Dorsal)**
CGTGCATG**CCTTATCT(Serpent)**GTCATTTTGTTTTTCAAGCTGGCTGTTCGCCTATAAAAG
CTCTCGCCTTTTGTATCGCAGTCATCAGTCGCTCAGACCTCACTGCAATATCAATATCTTTAGC
TTCTCCTAAGAAAAAATCAAGAAAATATCACCATGAACTTCTACAACATCTTCGTTTTCGTCGC
TCTCATTCTGGCCATCACCATTGGACAATCGGAAGCTGGGTGGCTGAAGAAAATTGGCAAGAAA
ATCGTAAGTTCTTCCATTTGAAATCTGTTAAGACGGAAACTAACTGACTAACTTCTTTTCGAAG
GAACGCGTTGGTCAGCACACTCGGGATGCCACAATCCAGGGACTGGGAATCGCTCAACAAGCCG
CCAATGTCGCCGCAACTGCCCGAGGTTGA **(CecA1 CDS)**

Appendix I-6: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *CecA2* in *D. melanogaster*.

ATTTATTCT**GTTGCTCCC(Rel)**TGTAAATAAAACAATTTTAAAAATTTAAAGAATTCTATTCA
AACTTTGTTTTTTAAAGAGTTGGAGAAAAGCGAACTCTTGAATTTATACACACATTTTAAATAC
ACTTAAGAGGCATTATTTATACAGGATATTACAAATCGCTTCTTTTCCGATTTGGAAAGGCCGA
GATTAT**GTCTTATCT(Serpent)**GTTGAAATATAATTCGTTTCACCTATAAAAGGACCAGTCT
TTTAGTTTA**AATTATCA(Serpent)**GTCGCTTGTCAAATACTGAAACAATTAGATTAATTTGT
GGATTTTATTTGTCCTCATCCTGACCACTTATTGGCCACAATTGGAAGCTGGCTTCGACGGGAC
ATTAGTAAGCTTAGTCATTTTAAAAGATTTCTTTGCATCTAACTATGATTCTAAATCCTCAGAA
GGACGTTGGTCTATACACCCTAAATGCTACCCTGCAAGTTGCTGAAGTCGCTTCGAAAGCAGCC
AATGTGGCAATCACTGCCAG**GGGATAAAC(Rel,Dorsal)**TTAAGTTA**GGGTATTAT(Dorsa
l)**TTATAAGAAATTAAATTAATAGATTTTATTTTATATATTTTTTGTATATTGTTATTCAAAC**T
GATAATG(Serpent)**TAATATACGCTTTTCAAACGATCATTCCAAATCAGTTGTGG**GCTTATC
G(Serpent)**CAAATGATTTCGTAGTGTTTTTATTTTGATTGATTCAAAGAAG**GGGTTTCCT(D
orsal)**CTCTGATTCTTAGTCTCCCGCATTGACGAGGTA**AAAAATCCC(Dorsal)**TATGCATA
TGAAATATGCAAATTT**AAAAATCCC(Dorsal)**CCAATCCGACAGGTTGGTTTTGATCGGTTTG
GATTCCTCTCGTGTACTTTTCAGCCAT**AAAAATCCC(Dorsal)**CTTTCGAG**CCTTATCA(Ser
pent)**GGCGCTGAACTTAAGCTGATTCGCCTATAAAAGCTCTCGGCGTTCCTGGTGCAATCAAC
AGTCGATCACTTTCCATTGCAACAGCAACATCAGAGCTATAGCTACTCTTGCAAAATCTAAAGT
CAAATAAAACCACCATGAACTTCTACAACATCTTCGTTTTCGTCGCTCTCATTCTGGCCATCAC
CATTGGACAATCGGAAGCTGGTTGGCTAAAGAAAATTGGCAAGAAAATCGTAAGTCCATTCTAT
TTGAAATTTGTTAAACCGGAAACTAACTAACTCCTTTTCATAGGAACGTGTTGGTCAGCACACT
CGCGACGCCACAATCCAGGGACTGGGAATCGCTCAACAGGCCGCCAATGTTGCAGCCACTGCTC
GAGGTTAA **(CecA2 CDS)**

101

Appendix I-7: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *CecC* in *D. melanogaster*.

ATCACTGTAATATCATTTAAGACTTGTCTGCCGA**ATGGCTCCC(Rel)**TCCAATTTGAGTAGTT
GAGCCGGCTTTTGCAACCAGTTGCGGCATTTACTGACTTACTTACGCTTCCCTGGACTCACCAA
CTAGATGCTTCGAAGTCTGAGAATGTGAATGAGGACGAGTCCTGGCGGTTATGACACAGGACTC
GCTGGCTCCTTTTGTGTCTCCGTCTCTGCCACTTGTAGCTGTCACTCAGCGGTTAATTCGAGCG
ATTTTATTTACATTTCGCAGAGGCCATGGAACCGGAAAGGAGCAGCAAGGAAGCGGAGTCAAGG
CCCAAGATGGATAAAGCGTTTTTACTTGTTAAGGAAATTAGTGGCATATCCTGACAGGGCGCCC
ATCTTCCTCGCAGCTTCGCATCCTAGATGCTCTATTCCTATATCCTTCCGCATTGTGTGTGTTT
TTGTGTGTGTGTTTTGCCTCCCTTGCTGATTTAATTCATTTGTTGTCTTTGCTGCGCGTCTATT
GTCTGCCCCTCGTTGTTGGTTTTTTATGGCTGAAATTAAAGTTACATTTTCGTGGTGGTTGATA
TTATGTGCCGATGTGCGCGAAAGCCTACATCCTGGGCCATCCCCCTCTCAAAAATCAAGTACAG
TTGCGTGATTGTGTCCTCTTTGTAAATCTAAATTTTATTTGAAAATATTGTTTAGAAGAAGTTA
GCTATTGCTTTTTGCACACATGAGAGCTAAGCGAAGAACGCTCCATTTTTACTAGCAGCTGCTC
AAACAGATTACCGAAGACAGTCTTCGTCTAACAAAGAAG**GGGATCCAC(Rel,Dif/Rel)**TGC
AGTCTTTCTCTTCTCGCTGCGAAAAGTTCCCCGTCGTCG**CCTTATCG(Serpent)**GCATCGCA
TTCTTCGCTATAAAAGCCGCCTGTGCCAGAAGTCCAGTCATCAGTCGCTCAGTTTCCACAGCAG
CTAAACAGCTAAATCGCAATCTATATATATATATATATACTAAGGAATTAAACCTAGAAAATTC
ACCATGAACTTCTACAAGATCTTCGTTTTCGTCGCCCTCATCCTGGCCATCAGCATTGGACAAT
CGGAAGCCGGTTGGCTGAAGAAACTTGGCAAGAGAATCGTAAGTTCAGCAACAAAATATATTAA
ATACTTGCAAATTTACTAATTTGTTTTATATTTACTTGCAAAGGAGCGCATTGGCCAGCACACC
CGGGATGCAACCATTCAAGGACTGGGAATTGCGCAACAGGCCGCCAATGTGGCAGCCACCGCCA
GAGGATGA **(CecC CDS)**

Appendix II-1: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ21126* in *D. virilis.*

TCCAAGTGCATGGCTATTTGCAGTATGCCCCGTTCCACAGATT**CAAAGTCCC(Dorsal)GCCTATC
G(Serpent)**TAGGGTATGTAAAAGTAGGGGAAGACGCCGTGCACATGCATGCAACACTTTTGGCCA
TTGGCATTGGCGCCAAATATGCGCACAATGGGCACCTGTTCAACAAATATAAGGGCCTTAGCATATA
ATTAACAATTTCAAATTAGTTGCTATATTTACGCGCTTAATTTCCTTGCCGCGCAACTCCGAGTAGC
AGGGATCCAGGCCAAATAC**GGGTTTTTC(Dorsal,Rel)**CATGTAAAAATCGGCAATTACAAGCCG
CACGGAGTAAATGCCGTCTATATCGGACATTTTCTGTTGTTTTGTGTGCAAATATAAACGTAGTATT
CATTAAAAACCAGCAACAACTTAATTTTTTGCAACTTGTTTACTTTTATGCGAATAACTGCCAGGGC
TGGCGAAACTCGTTGCGGCAACATTGTGAAATTCAATTTGCGACTCGTTACGAGTACTTGGTACTTG
CACATTCAAATAATTGCCGCCATTTAGGGCTAGCAACTAGGGCTGGCAATGCTGCACAAGTACAATG
CTGACTCGTGACAAGTACGACTACTCGTTCAAATTCAAACTTGTGGCGCCAATTAAACGTATCTATT
GTAAATTTAATGAATTCTGCATTTGTTATTAGATATATTATATTTATATGTTCTTTGCATTATATTT
ATATAAGAAATAGTTCATTGTAATTGATTCTGTGCTGTATTTTATGACATCCGGGGCCAAGCTTAGT
CTAGACTCTGCTATTGACTTTCCATTGTTGTCTTCTTTGAACAACTGCTCATCAACCGAGATGATTA
AAAGATCCCCCTGTACGTGCCCAGCCTGTGAATACGACTTGAAAGCAAAGCAGTGTCCACC**GGGAAT
CCC(Dorsal,Rel,Dif/Rel)**TTAGCTTGACAAAAGCGACACTTTCGCTGTATAAATACGTGGAG
CCATGCCGCTTGGCATCATCAGTTGCCAGACAAACTGCATTACAATGAAGTTGATCGTATGCCTGGG
CCTGCTACTAATCCTGGCTGTGACAAATGCGTTGACTTCTACGCATGAGGCGACAGCTCCAACGGGA
ATGGAATGGCATCCACGCCAGAAACGCGCCACCTGCGATCTGCTGAGCTTCTGGAACGTAAAGAACA
CAGCCTGTGTGGCCCACTGCCTGGCCAGGCGCTACAAGGGCGGCTACTGCAATAACAAGGCCATTTG
TGTTTGTCGGCGTTAA **(GJ21126 CDS)**

Appendix II-2: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ22479* in *D. virilis.*

CAGACAGTATGACGCCAAATTCAATACAACATCGT**CGATAACA(Serpent)**AAGGTAAACAGCCCA
TTCAAAAGGGGTTCCAGGCAGGGGCTCAAACAACAGGCAACCACTCGACTGGTGTCAATAATAATAA
TAGCCGCAGTCCACTCGGCCAACGATGAAGACGCATTTGGTAGCCCCGATGTAGCTGATGCGCAGAA
CAAAA**AAAAAACCC(Dorsal)**AAACCGAAACAACTACAGATACAGATACAATACGAATAGTTACAG
ATACAAAGATACAAATACACAAACATTACGCCCGCTTGTGTACTTAACTAATCATTTTGTGTTGCTT
CTCTGTTTGTTTTTTGTGAAGATACAATTCGGCTGAATTTATTTGTTGTATCTACGCATTTGTATCT
GTATCTGTAGCTGTAGCTGTAGCTGTAGCTGTATCTGTATCTGCATCGCTTCAATAGAATCGAATCT
GCGCACTTGCAAAACAATTAAAACATTTACGAGGAAATCGAGGCATTTGTTTGGCTATGGCCAAAAT
TTTGAAGCGAACAACTTGAACGCTAATCAAAACAATCTATGGCCAAATTAATTAGGCCAATTCTTGA
AAAGAATAATAATTCTTAAAAGATTTTATGCTTATGCATAAGGCGTTAAAGTGAAAAGTACAGTTGA
AACATTTGCTTATGTTGATAAAAAAAAAGGTTTGCTCAGATTATATTCTATATAGTCTATATACTAT
TTAGCTATATGTTATTTTTTATAAGACGGAGCTAACCTATTTGCCTGAATATC**AGATGTCCC(Rel)**
GTTTATATTCTATGTGGTCTATATATAAATTT**GGTTATCA(Serpent)**GCGCTCTATGAATAATAC
CTAGAACAGCGAA**ATACTTCCC(Rel)**CCCAG**GGGAACTCCC(Rel)**TTAGCTTGTTGCTGCTGAGA
CTCTGATGTTAAGCCTTGATAGAGCCAAGCCTGTGGCGTATAAAAGCCAGCTGTAGCTTCCGAGCAC
ATCATTCAGTGACAGTGACAGAGCAGTTCAAGCAAG<mark>ATGAAATTTACCATACTCCTAGGCGTTCTGG</mark>
<mark>CTCTACTGGTGTGCTTGGCCCAGGCACAACCTGTCGAACAGGATTCACTCGCTGAGCGGGAGCCCGG</mark>
<mark>GGCAGTTGAGCCCATGCCACAGGACTTGCACAGCCGTCAGAAGCGTGCCACATGCGATCTGCTGAGC</mark>
<mark>GGCTTCAATGTGAATCACTCGGCCTGTGCAGCTCACTGTATTGGCCTGGGCAGGAGTGGAGGCTATT</mark>
<mark>GTAATGATAAGGCTGTTTGCGTTTGTCGACGTTGA</mark> **(GJ22479 CDS)**

Appendix II-3: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ19915* in *D. virilis*.

ACAGCTGTGATATATATGTTCAAATATATAATATTTATAACAAATCAAGCTGGACAACATCTGTCAA
CT**AATTATCCC(Dorsal)**GATCCGTCCACTCGCCGCGTGCATCATCCCAGACAGGAAACCAGCTCT
TGGTGGGCGCATTTGCCCCATTCACGCCCAGTTCTGCATCATCGACTTGTGCCACGCCCAGCTGCAG
GTGGCAAACAGTTAGTCAGAGCTGCTAAAGTAGTTACATATATCAATTTATTACCCCGCACAGCAGA
GCCAATAATATGGCCATAATATTTATGTAAATGATTCGCAGCCGCATTTTGCACTCTGTCAAAGTTG
ACTCTCTCGACTGATGCACGTTGTGCGACTGTTTCGCTGTTTGTTTATAATATTTAAACGGGGAAAA
AAACACAGTGCACAATATTTATTTAATAATTTATTCATGCTACATGGG**CATTATCA(Serpent)**GC
AGC**TGATAAGC(Serpent)**CGCACAGACGAGACTCGTAAAAATTGAAATGAACACGAAATGCTGAA
AATGTTTTCACAGAAGCCACAAAATGCTCATAAATCAATGAATGCTAAATTAAAAAAATAAACAATA
ACACTTGCCATGTCGAATCTGTGGTCAGAGCAGAAATGTGTCATAACTATAGGCCAACTATGCCAGA
CAAATGTTTTTCCGCATGTTCAATAAAATATTTCAATAAATGTTCATATGAAAGAACAATAAAAGAA
AAATACGATTATATACGTACTCCATTGGGCAAGGCAATGCGCAAAGCAAGCAAAACAAAAAACATAA
TAAATGTTTAATTGAAGACAAACAAGTGCCAGCTCAGAAATACGCCAATTTCAACAAAAGGAATCAC
CCTG**GGGAAGCAC(Rel)**CCGGCAGATAAAGACAGAGCTGCGGCTCCACGGTGATTTTTTTGGTTCT
GTGCATCTTCGACTGCTGCGTGGTCTAGCTATAAAAGCTCGCTGTTCCGGCTAGATG**CATTATCA(S
erpent)**GTCAATCAGCTCCAGATTCACATACAACATGCAGGTTACGCTCATCGCTTTGCTCTGCTG
CATTGTCGGCTCCGCCCTGGCCATGCCCCAACCCAATCCCGAGGAGAAGCCGAAGGGAGATGTGTGG
ACTGAGCGCCAGCAGTTCAATCCGCCAAACGAGCAGCGTTTTCTTTTGGATGGCGGCTACAACAAGG
ACAAGAGCGGCAAGGATGTATGGGCCCAGGCTCAGGTGCCCGTTTGGACCAGCGAGAACAAACGTCA
CGAGTTCGATGTGATTGGCAAATATGGACAGCATCTGGGTGGACCCTGGGGCAACAGCGAGCCTTCA
TGGGGCGCGGGTGGCAACTACAGATTTCGATTTTAA **(GJ19915 CDS)**

Appendix II-4: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ19916* in *D. virilis*.

```
TTTGAAAGTTAGCCTGAGCCCTGGACAAAAAGATTCACACAACAGCTTCGGCGATTTTTCCACTCTT
GCGGATGGAATTTTCCAAAATTTCCC(Dorsal)TGTCTTAGTGATTCAGTCAGTCGGCCAGACTGA
CCAGTCGGATTGTTAGTTAATCAATCAGTCAATCAATCACTCAGTGTGTTTTTCAATCAGTTAGTAA
GTCAGTCGGTCAGTCAGTCAGTCAATCAGTCAGTCAGTCAATCAGTCAATCAGTCAGTCAATCAGTC
AGTCAATCAGTCAGTTAGACTGTCAGTCAGTAAGTTAGACAGACAGTTAGTCAGCCACTGTGTTAAT
AAGTCTGTCAGTCATTCAGTCTCTCAACCTGTCAACCAATCAGTCAGATAATC(Serpent)AGTCA
ATCAGTCAGTCAGTCAATCAGTAACGAATTCAGTCTGTCAATCAGTCCTTCAGTCAGTTAGTCTGTC
AGTCAGTAAGTTAGACAGACAGTTAGTCAGCCACTGTGTCAATAAGTCTGTCAGTCATTCAGTCTCT
AAACCTGTCAACAAATCAGTCAGAAACTCATTCAGTCAGTCAATCAGTCTGCCAGATGAAGAATATT
AAAGGTATTTGAAGATCTTTCTTTCCGTCGACAGTCTATTAAAAAGAATTACAAACCGTATGAAAGA
CAATCTGAGAGATACGTACTTACATTATATATTTTAATATACTCATCAGTATTGCTATAATAATGTG
TTTATATGTATGTATACTATAATACTTTGTTTGAAGTAATAATGAAGGATGAATATTAAATGTATAT
CATTTACTTTGTGCGGCCAAGCGACCTTAAGATGAAAATCCC(Dorsal)CCCAGTCATAAAAAAGG
AACCACCCAGGGGAAGTAC(Rel)TAAGCAGATAACA(Serpent)TAGAACAGCGAGCTGGCAGTG
GGTCATCT(Dorsal)CACCCATCGGGTGGTGCTCGTCTTGTCTGCCTATAAAAGCAATGTTCAAGC
TGGACGCATTATCA(Serpent)GTCAATCAGCTCCAGATTCACATACAACATGCAGGTTACACTCA
TCGCTTTGCTCTGCTGCATTGTCGGCTCCGCCCTGGCCATGCCCCAACCCAATCCCGAGGAGAAGCC
GAAGGGAGATGTGTGGACTGAGCGCCAGCAGTTCAATCCGCCAAACGAGCAGCGTTTTCTTTTGGAT
GGCGGCTACAACAAGGACAAGAGCGGCAAGGATGTATGGGCCCAGGCTCAGGTGCCCGTTTGGACCA
GCGAGAACAAACGTCACGAGTTCGATGTGATTGGCAAATATGGACAGCATCTGGGTGGACCCTGGGG
CAACAGCGAGCCTTCATGGGCGCGGGTGGCAACTACAGATTTAGATTTTAA (GJ19916 CDS)
```

Appendix II-5: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ19917* in *D. virilis.*

CGTTCCTATATCAAATTACTTCAAATTAAAATGAATCATTTTCGAAGCTCT**AGAACTCCC(Rel)**GAAAC**GGGAATTAC(Rel)**AAAATGATTTTTAGGCAGAAAAAATTTGCATACATTTTTCGTTGCCAATTTAAACTGACCTTGGTAAGTTGAGGAAGGGTTTGGCTTACACGCGCCACATATCTGGCATATAATTATGTGCTGTCTATGATTAAATTTGTTAAGCCCCTGTATTGGATATTATTGCTTTGTATGGCTTAAGAGTTGCGATATTTTTAGAGATAGATATAGATAGAGAGAGAGAGAGCGAGAGAGAGGGTGAAAAAGCGAGAGAATGAGAAAATGTATATGGAGATAGAGATAGAGAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGAGATAGA**AGATAGTG(Serpent)**AGAGTGAGAGTGAGAGTGAGAGAGAGAGAGTGAGAGAGAGACAGGGGTTTATATACATATGTCTAGAAAAAGTTTCCAATGTTGATGACCGCAGCTCGGAGATGGGAACAGAGTATCTTCTAGTATATAGGAATATCTT**AAGATAGTT(Serpent)**AAGATCAATTGATCTGGTTGATTAACTGAACAACATTGACATTTCTTTTATTTCAGGTATAACTGCAGCTGTTCTATATTTGTGTAGTGCCTAATTGAAGACCTTAGCATTGCTTGGCTCATCAGGCG**GGGATTTAC(Rel)**ATGCTGGTCTTTATCAACCTACCATATTATTTGCCAAGAGATCACGCA**TCTTATCA(Serpent)**GCTGCCTTCGCCGGGTATTCACTTGTGAAGCACAGTCTCCCATGCCAGACAATGTGCCACAA**GGGATTCCC(Rel,Dorsal)**TGCGACGGTAGAGCTTGTCTATGGGCCCCAAGCATATATAAGGCACATGGAGCCACTCATTCGTCATTTTAGTATCAACATCAACAATAGCAGCCACCACAGCTCAAAACCACAAG<mark>ATGCAACTTAAACTCAGTTTGCTGCTGCTCGTCCTCGGCGTCTGCGCCTGCGCCTGGGCCTATCCTAATCCGCTCCTGACAGTTGACGACGAGCCGGCCACACAATTGGTAAGTGCAAAGCCCAGAAGTTTGTTGAGCTTAAGACTTATGGTGCCCGATCCCAACAAGCAGCTGGCCGAGAACTACGATTGGGCGCCCAGTGAACAGGTGGAGCAGCTTCGCGTGCCTCGCCAACTGAATGTGCAGGGCGGCGGCAGTCCACGTCAGGGCTTCGACTTGAGCGTCAATGGACGCGCGCCCGTCTGGCAGAGTCCCAACGGTCGCCACTCGCTGGACGCGACGGGACAGTATTCCCAGCACCTAGGCGGACCCTACGGCAACAGTCGGCCCAATTGGGGCGCGGGCGCACAGTATACGTTCCGTTTCTAG</mark> **(GJ19917 CDS)**

Appendix II-6: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *GJ22469* in *D. virilis*.

TCTTGGCAGGTTTTTAATGTAGAGCTTTATGATAAATGATATGGTATTCGACTGT**TGATAAGA**TTTT
GTTATTATATTAGGCGCATG**CTTAAACCC(Dorsal)**GTTAAACACAAGTTTGAACAATTTATCTTT
AGCAAAAAAGAATATCATTCAGATTTTGTTCATATAAAAAGAATTTGGTGAACAAAACGTTGGTGAA
TATATCTAAGCTTTCGATCCCAAGGCAGCTATTTATAAGCTATAGTAGCTCGATCCTTAGAACAAAA
CTTATTTTTTACAAAAAATACATGTGTCATATACCTTCCATATAATACATATATCTTCTCTTACGCA
TCACACTTGCTATGATATAATTGAAATACGCTCTACAAGCAGGGTATGCATAAAACGGAAGTAACCG
CACAAATGGTAAAGTGCAT**GGGAATATT(Dorsal)**TACTGAAAATGTCAGTTGGCGCATTCAAAAG
GGCGCCAAAGATGAGGCTGCCTGCTGTAACATGGCCTTTGCCAGAGGAAAACTAACAAAGTGCTCTA
ATAGAGCCAAAGGGGGCTCTGGCAGTGGCAGCTGCAACAGCAGAAGTTCGCCTTAGCTAGACACGAT
GGCAACTGCTTGTTGCAGTGGGCGTGGCGTTGTGTTGACGCATGTTGCTGTTGCCTTTGAATGGCGT
CGTTGTTGCTAGTAATTTGCAATGCAAAAACAACGTAAAGTGGCTGCTGACTAGGCAGTCAACGTCG
GTTGATTCGGGGCGAAACCTGGGCATTTGGCTGCGACCTTGGCGCAAACTAAATATTCAACCAGCAC
ATTTCATATTTCGTTTTCTCAGTTTTTGTGCTCGGCT**GTGCATCCC(Rel)**CGCATTGAAT**AGTTAT**
C(Serpent)AGCGCAGCTGAGCTGACCCCGTCTGGCATTTAAA**GCTTATCT(Serpent)**GGCGAA
CAACAAAA**CAAAATCCC(Dorsal)**CCTGGTTGCCTGCGCTTTGCGTATAAAAACCCTGCCGCTTGG
GCGCTGTGGCATCAATCAACTAACAACTGACTTTTAGACAAG ATGCAACTCAATCTGCGTGGTCTGC
TGCTGCTGCCCCTGCTGCTGGTGCTGAGCCTGAGCCTGAACCTGTCGCTAACCGAGGCGCGACATCA
TGAGGGCCGCATACCGTTTGACACCCGACCATCGCCATTCAATCCCAATCAGCCCAGGCCCGGGCCG
TACATATAG **(GJ22469 CDS)**

Appendix II-7: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *Cec2B* in *D. virilis.*

CTGGTATCTGAACATCGCG**GGGAATTTT(Dorsal)**TCTTTGAGTCTCTCCGCAAATCGATCCACT**G**
GGACTCAC(Rel,Dorsal)AGCTTGCACCGCCAGAGCAGAGTAGAGCAATTCACGCAGCAAGCT**GA**
TAACT(Serpent)GGGAAGAAACGCT**GATAACG(Serpent)**CGCTATGGCAACGGAGACAGAGCA
ACAACACCTGGCTGACTTTTAAGTGTTTGCACACTAAACTGATCCGACATTGAAAATATTCA**TATTA**
TCA(Serpent)GCCCGATCGGGTCAGTGGCATATTCACACATATTGAATACAGTTGTAGTCGGAAA
TGGGGAAGCA**GAAAAACCC(Dorsal)**ACGTGAAGTTTCATTCATAAATTGATTTGTAGTAAAAATG
TTGTAGTTGTCTCTGTTTCATGGCAAGCATTTTTGGTGGTTACCCGACAAATCCACTTCATATGTCA
AT**CCTATCT(Serpent)**GTGGGGCCAATTGCGAGCGTGCG**GAAATACCC(Dorsal)**AACATTTTG
GCATTGCATCTTACATGGCAAAGCTTTGTTGTGAGGATTAGAATTTTTAAATTTTGAATCTAACTTG
GGGCTTCTGACCATATCTCGAAGCTTAATAGAACTAGATACGTGAAATTTTTGGAGTCTTCTTAACT
GCAGCCTGCTTAAGTAACTTTTTCTTACAATATAATCTTTTGAATAATTATTCAGTATTAATAGATT
TTATAGAATTTATTACATATACTATAACATCCGACGTTTATGTTTAAGTCGATCATATAAATATTGT
TTACGGCATATCTTAACATTCTCTGCACTTGCTTTTGCTTTGCTTGCTTTTGATCCGTGTACTTTTT
CGCCGAG**CAAAATCCC(Dorsal)**CTCGTTCATCCATCAATTGTTGAGTTTTTGCGCTGCGAG**GCTT**
ATCA(Serpent)GCATCG**GGGAGCTAT(Rel)**ATGAGCTATAAAAAAGGCACTGACTGGTCGTATC
AGTCATCAGTCGCTCAGCATCTGCACCCACAGCAACAGCTCAAAGCAAAGCTCCTAGTTTCCAAGGA
ATATCCAATACCTGAACTCAAT<mark>ATGAACTTCTACAAGGTCTTCATCTTCGTTGCGCTCATCCTGGCC
ATCAGCTTGGGTCAATCCGAGGCTGGTTGGTTGAAGAAGATTGGCAAGAAAATCGTAAGTCTTTCAA
CTACATCTTTAGTATGAGCAAT</mark>TACTTATGATAGATTTACCACTACAGG<mark>AACGCGTTGGCCAGCACA
CTCGGGATGCCACCATCCAGGGATTGGGCATTGCCCAGCAGGCTGCGAATGTGGCCGCCACGGCCAG
GGGCTAA</mark> **(Cec2B CDS)**

Appendix II-8: Distribution of NF-kB-like (Relish (Rel), Dorsal and Dif/Rel heterodimer) and GATA (Serpent) transcription factor binding sites on 1 kb upstream of *Cec3* in *D. virilis.*

```
AAATTAGTGGCATATCCTGACAATGGGCACATCCTGCTCCTGCTCCTAGGCTTGTGCTCGGCGTTGC
CTTCCTTGATTTAATTCATTTGTTGTCTTTGCTACGAGTCGCTTTTTGTTGCTTGTTTCTTTGTTTT
GGTGGTTGACAGTGTTTCAGTTGCCACGCCCACTAATTGTCGGCTTTGGACACGGGTACAGTTGACT
TCTCGATCCGCACTAGATTGCAATTTGCCCAGACCAAATATTATGCTTGCAAACAATTTCAAGCTAA
GCAAATGAAACGAGTTTCGGAATTGCCGCAGCCTGGTAACAAACTAATCGTTCTAAATTTTAATCAA
TAGTTATTTGTTAAATTCGGAATGGCTTGGCTTTGTATAAAAATGCCAACATCTACACACATCTGTG
CAAGCTACACTTTTAGCCTATTTATTACCAATTTCAACAAGACAATCATGTTTGTGCTTCTCATTCT
GGCCACCAGCTTGGGCCAAAGCGAGGCTGGTTGGTTGAAGAAGATTGGTTAGAGAAGTAAATATTAG
AAAATTTTAAAACAAGTATATTTGGTTTCTTTGTGCTACTAGCCATGCTCT**GTGGCTCCC(Rel)**AG
TTCTGCTACAGCCCCTGGCTAAAGACAGTGTTAAGTTAGGCTAAGTTTGTACATCATTTCATACTGT
CCTATATCCAATGCTCAACTTTACTCTGCCTAAAAAAAGTCTACTCTCGATATTGCAAACGAAATTT
**CGATAACC(Serpent)**TGCCTACCTTTTGTGTTTATTCTGCAAGCCCTGCTAGC**GGGAAGCAC(Re
l)**AAATTGTCTGTTCACAAA**CAAAATCCC(Dorsal)**CGAGCCCCAACGATCGATCTGTTTTGTCTT
TGCGCTGCGAG**GCTTATCA(Serpent)**CCGCCGGAGAGCTATATGAGCTATAAAAGAGGCACCCAC
TGTCGTATCAGTCATCAGTCGCTCAGCATCAGCAACAAAAGCAACACCTCAAATCCCAGCTCCTAGT
TTCCAAGGAATATTCAATACTAGAACTCAAT`ATGAACTTCTACAAGGTCTTCATCTTCGTTGCGCTC`
`ATCCTGGCCATCAGCTTGGGTCAATCCGAGGCTGGTTGGTTGAAGAAGATTGGCAAGAAAATCG`TAA
GTCTATTAACCATATTGTTTAGTATAAACAAATACTTATTAAAGATTTACACTTACAGG`AACGCATT`
`GGCCAGCACACTCGGGATGCCACCATCCAGGGACTGGGAATTGCCCAACAGGCTGCGAATGTGGCCG`
`CCACGGCCAGGGGCTAA` **(Cec3 CDS)**
```