Accepted Manuscript

Genes encoding cuticular proteins are components of the Nimrod gene cluster in *Drosophila*

Gyöngyi Cinege, János Zsámboki, Maite Vidal-Quadras, Anne Uv, Gábor Csordás, Viktor Honti, Erika Gábor, Zoltán Hegedűs, Gergely I.B. Varga, Attila L. Kovács, Gábor Juhász, Michael J. Williams, István Andó, Éva Kurucz

PII: S0965-1748(17)30082-6

DOI: 10.1016/j.ibmb.2017.06.006

Reference: IB 2958

To appear in: *Insect Biochemistry and Molecular Biology*

Received Date: 9 November 2016

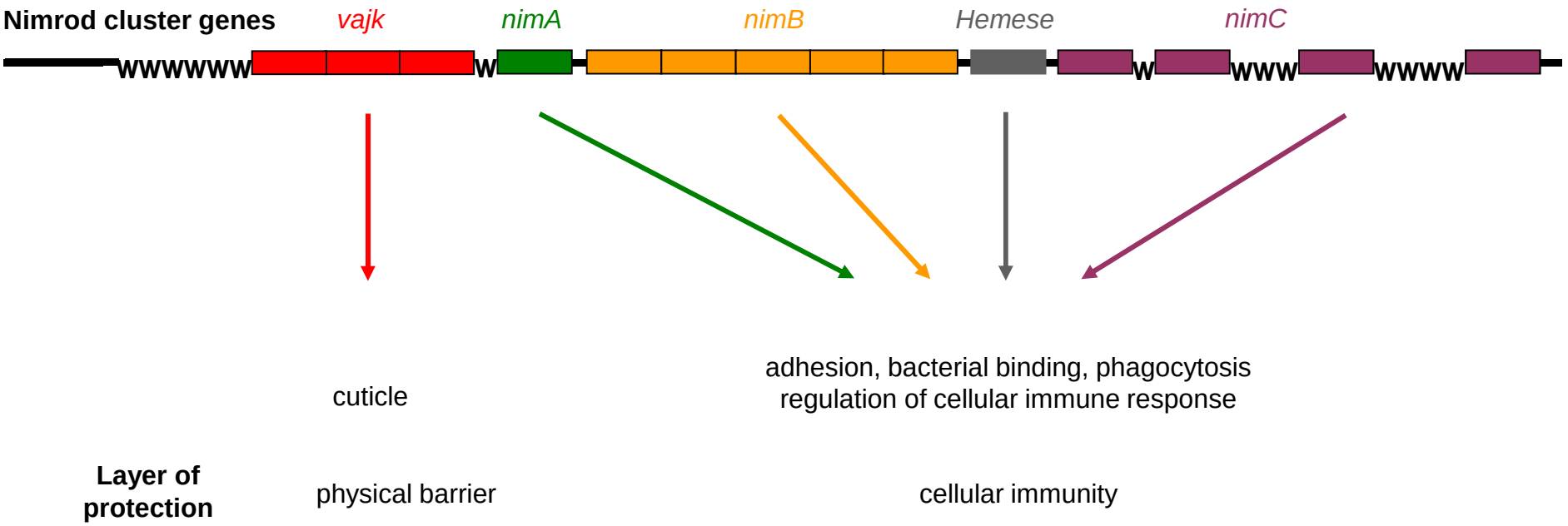
Revised Date: 4 May 2017

Accepted Date: 10 June 2017

Please cite this article as: Cinege, Gyö., Zsámboki, Já., Vidal-Quadras, M., Uv, A., Csordás, Gá., Honti, V., Gábor, E., Hegedűs, Zoltá., Varga, G.I.B., Kovács, A.L., Juhász, Gá., Williams, M.J., Andó, Istvá., Kurucz, É., Genes encoding cuticular proteins are components of the Nimrod gene cluster in *Drosophila*, *Insect Biochemistry and Molecular Biology* (2017), doi: 10.1016/j.ibmb.2017.06.006.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

Genes encoding cuticular proteins are components of the Nimrod gene cluster in *Drosophila*

Gyöngyi Cinege[1], János Zsámboki[1], Maite Vidal-Quadras[2], Anne Uv[2], Gábor Csordás[1], Viktor Honti[1], Erika Gábor[1], Zoltán Hegedűs[3], Gergely I.B. Varga[1], Attila L. Kovács[4], Gábor Juhász[4], Michael J. Williams[5], István Andó[1], Éva Kurucz[1*]

[*]Corresponding author:
E-mail address: kurucz.eva@brc.mta.hu (Éva Kurucz)
Phone number: 00-36-62-599 462

[1]Immunology Unit, Institute of Genetics, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary
H-6701, Szeged, P.O.Box 521
cinege.gyongyi@brc.mta.hu
latvanypekseg@gmail.com
csordas.gabor@brc.mta.hu
honti.viktor@brc.mta.hu
gabor.erika@brc.mta.hu
varga.gergelyistvan@brc.mta.hu
ando.istvan@brc.mta.hu

[2]Department of Medical Biochemistry and Cell Biology, Instititute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, SE-40530 Gothenburg, Sweden
anne.uv@medkem.gu.se
maite.vidal.quadras@gu.se

[3]Laboratory of Bioinformatics, Institute of Biophysics, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary
H-6701, Szeged, P.O.Box 521
hegedus.zoltan@brc.mta.hu

[4] Department of Anatomy, Cell and Developmental Biology, Eötvös Loránd University, Budapest, Hungary
juhasz.gabor@brc.mta.hu
alkova@caesar.elte.hu

[5]Department of Neuroscience, Functional Pharmacology,
Uppsala University, Uppsala, Sweden
Institutionen för neurovetenskap
BMC, Husargatan 3, Box 593
751 24 UPPSALA, Sweden
michael.williams@neuro.uu.se

ABSTRACT

The Nimrod gene cluster, located on the second chromosome of *Drosophila melanogaster*, is the largest synthenic unit of the *Drosophila* genome. Nimrod genes show blood cell specific expression and code for phagocytosis receptors that play a major role in fruit fly innate immune functions. We previously identified three homologous genes (*vajk-1*, *vajk-2* and *vajk-3*) located within the Nimrod cluster, which are unrelated to the Nimrod genes, but are homologous to a fourth gene (*vajk-4*) located outside the cluster. Here we show that, unlike the Nimrod candidates, the Vajk proteins are expressed in cuticular structures of the late embryo and the late pupa, indicating that they contribute to cuticular barrier functions.

Keywords: gene cluster, Nimrod, innate immunity, *Drosophila*, cuticle development, trachea

1. Introduction

The protection of organisms from pathogens requires the concerted action of several layers of defense, including mechanical and immune components. The first line of defense against infection comprises a physical barrier, the skin, the mucosa and chemical barriers that stop microbes and parasites from entering the body. If the pathogens are able to get past the physical barrier, a constitutive second line of defense attacks the pathogens, which involves proteolytic cascades in the body fluid, as well as phagocytosis mediated by a group of immune system cells.

The third line of defense is comprised of inducible antimicrobial peptides and effector cells, differentiating after immune induction, involved in separation of large particles. These three layers of defense are the major components of the innate immune system, a phylogenetically conserved module, based on the same principles in all multicellular organisms, including insects and mammals.

The Nimrod gene cluster, on the *Drosophila melanogaster* second chromosome, is a major component of the innate immune response (Kurucz et al., 2007; Kurant et al., 2008; Zsámboki et al., 2013). The distribution of genes in this cluster is similar over distantly related arthropod species, indicating that the cluster remained predominantly intact during the last 300-350 million years (Somogyi et al., 2010). The expression pattern of several genes located here is blood cell specific, and the encoded proteins play a central role in the immune response. All Nim proteins contain a characteristic EGF-repeat, the Nim-domain (Kurucz et al., 2007; Somogyi et al., 2008). NimA, NimB1, NimB2 and NimC1 proteins possess bacterial binding capacity and are important for the phagocytosis of bacteria (Kurucz et al., 2007; Zsámboki et al., 2013), the NimC4 (SIMU) protein has role in glial phagocytosis of apoptotic neurons (Elliott and Ravichandran 2008; Kurant et al., 2008; Shklyar et al., 2013). Another Nim family member, Hemese, was shown to play a modulatory role in hemocyte activation or recruitment (Kurucz et al., 2003).

Besides the *Hemese* and *Nim* genes, the cluster also contains candidates with yet unknown function. The *vajk-1*, *vajk-2* and *vajk-3* genes are located in the large intron of *ance-3,* and are transcribed in the opposite direction. The *ance* genes were suggested to be involved in immune defense, as in *Locusta migratoria* hemocytes, their expression is significantly enhanced after stimulation by bacterial lipopolysaccharides (Macours et al., 2003; Duressa et al., 2016) or following a *Wolbachia* infection in the fruitfly (Xi et al., 2008). The *vajk-4* gene, a paralogue of the *vajk-1*, *vajk-2* and *vajk-3* genes, is located outside of the cluster, but still on the second chromosome. The *vajk* genes were previously suggested to play roles in the innate immune defense (Irving et al., 2005; Somogyi et al., 2010). We have therefore carried out a molecular and immunological analysis of the encoded proteins and reveal that they exhibit characteristics of cuticular proteins and therefore are essential components of the cuticle, the mechanical barrier facing the challenges by pathogen.

2. Materials and methods

94

95 *2.1. Drosophila stocks and culturing*

96

97 The following fruit fly stocks were used: wild type *Oregon R*, w^{1118}, *w;actin-GAL4/CyO,GFP*
98 (4414), *w;Da-GAL4* (55849), w; *alphaTub67C-GAL4* (7063), *w;UAS-GFP* (4775) from
99 Bloomington Drosophila Stock Center, *l(3)mbn-1/TM6B, Tb* (Konrad et al., 1994), *w;btl-GAL4*
100 (Shiga et al., 1996), *Flip-FRTy,w,UAS-FLP;Act5C-FRT-STOP-FRT-GAL4,UAS-GFP* lineage
101 tracing line (Honti et al., 2010). *UAS-RNAi* transgenic stocks were obtained from the Vienna
102 *Drosophila* RNAi Center (http://stockcenter.vdrc.at) and from the National Institute of Genetics
103 RNAi Center (NIG-fly), Kyoto (www.shigen.nig.ac.jp/fly/nigfly/). For the analysis of the *vajk-1*
104 gene the V102445/KK, V38024, V38026 and 16886-R3 for the *vajk-4* gene the 36642 and the
105 15514 *UAS-RNAi* transgenic stocks were used. Strains were kept at 25°C on standard yeast-
106 cornmeal food. Age-synchronized *Drosophila* populations were used in the experiments. The
107 early pupae represents 1:1 mixture of 24 and 48 h, and the late pupae 1:1 mixture of 72 and 96 h
108 pupae harvested after puparium formation.

109

110 2.2. *Isolation of RNA and synthesis of the cDNA*

111

112 The RNA was isolated using TRIzol Reagent (Invitrogen). The contaminating DNA was
113 eliminated by DNase I treatment (Fermentas): 1 µg RNA was treated with 1 u DNase I and then
114 reverse transcribed using RevertAid First Strand cDNA Synthesis Kit (Fermentas).

115

116 *2.3. Reverse transcribed polymerase chain reaction (RT-PCR)*

117

118 In the RT-PCR 1 µl undiluted cDNA was used as template and 40 cycles were applied. The
119 primer sets used to detect the transcription of the *vajk* genes are listed in Table 1. The vajk-1 fw
120 and vajk-1 AB rev primers were used to amplify an 1031 bp fragment representing both, the
121 CG16886-RA and the CG16886-RB *vajk-1* transcripts; vajk-2 fw and vajk-2 rev were used for
122 detecting the *vajk-2* transcript (828 bp), vajk-3 fw and vajk-3 rev for the *vajk-3* transcript (849
123 bp) and vajk-4 fw and vajk-4 rev for the *vajk-4* transcript (1003 bp) respectively. To test the
124 quality of the cDNA isolated from blood cells we used the hemese fw and hemese rev primers,

which generate a 175 bp fragment. In case of genomic DNA contamination, a 354 bp PCR fragment was expected. To analyze the performance of the *vajk* specific primer sets, we also used vectors containing *vajk* cDNAs as templates in the PCR. The RH73259, RH04334, HL02234 and LD27203 cDNA vectors, were obtained from BDGP Gold cDNA Collection (https://dgrc.cgb.indiana.edu/vectors/Gold). Furthermore, in positive control reactions the housekeeping ribosomal gene specific rp49 fw and rp49 rev primer set was used, which amplifies a 316 bp fragment.

2.4. Generation of constructs encoding the recombinant Vajk proteins

To amplify the entire coding region of the *vajk* genes the RH73259 (*vajk-1*), the RH04334 (*vajk-2*), the HL02234 (*vajk-3*) and the LD27203 (*vajk-4*) cDNA clones were used as templates. In case of the *vajk-1* gene we used the vajk-1 fw and vajk-1 A rev primers (Table 1), which generate an 1137 bp fragment consisting the coding region of the CG16886-RA, without its stop codon. For the *vajk-2, vajk-3* and the *vajk-4* genes we used the same primer set as in case of the RT-PCR (Table1). Because the used primers carry adapter restriction sites the obtained fragments were digested with EcoRI and XbaI restriction enzymes and cloned in the EcoRI/XbaI site of the pB-FLAG vector (Zsámboki et al., 2013). The obtained pGC22, pGC16, pGC14 and pGC24 clones carry the *vajk-1,* the *vajk-2, the vajk-3 and the vajk-4* cDNA fragments respectively, followed in frame by the FLAG tag coding region, two stop codons and an additional NotI restriction site. DNA sequence analysis was done using a BigDye Terminator v3.1 Cycle Sequencing Kit (Invitrogen) and a 3500-Genetic Analyzer (Applied Biosystems). The obtained plasmids were digested with EcoRI and NotI restriction enzymes and after isolation of the fragments containing the whole fusion constructs, they were cloned in the EcoRI/NotI site of the pENTR3C vector (Invitrogen). pGC44 (*vajk-1*), pGC38 (*vajk-2*), the pGC42 (*vajk-3*) and the pGC46 (*vajk-4*) plasmids were obtained. With site-specific recombination using Gateway reaction the fusion constructs were further transferred into the pDEST^{TM}20 vector (Invitrogen). The pDEST^{TM}20 moreover provides a polyhedrin promoter and an N-terminal GST tag coding region to the constructs. The obtained pGC63 (*vajk-1*), pGC66 (*vajk-2*), the pGC69 (*vajk-3*) and the pGC72 (*vajk-4*) clones were used for transforming DH10Bac *E. coli* competent cells (Invitrogen). The DH10Bac cells contain a baculovirus shuttle vector (bacmid) that can

recombine with the pDEST20. The recombinant clones carrying the constructs (polyhedrin promoter, N-terminal GST tag, *vajk* coding region, C-terminal FLAG tag, stop codons), were selected according to the manufacturer's instructions.

To create the plasmid construct expressing the Vajk-FLAG recombinant proteins, the previously described EcoRI/NotI fragment carrying the *vajk* cDNA fragment and the FLAG-tag coding regions, were cloned into the EcoRI/NotI site of pFastBac^{TM}Dual vector (Invitrogen), downstream of the polyhedrin promoter. The obtained pGC51 (*vajk-1*), the pGC29 (*vajk-2*), the pGC32 (*vajk-3*) and the pGC52 (*vajk-4*) clones were used to generate bacmid DNS construct as described above.

2.5. Expression and isolation of the recombinant Vajk proteins

The recombinant Vajk proteins were expressed in a baculovirus expression system (Bac-to-Bac, Invitrogen). The obtained bacmid DNA constructs, described above, were isolated and Sf9 cells were transfected using Cellfectin II reagent (Invitrogen). Sf9 cells were grown in TNM-FH Insect Medium (Invitrogen) supplemented with 10% fetal calf serum (PAA Laboratories GmbH) and 5mM glutamine (PAN Biotech). The transfected Sf9 cells produced the recombinant Vajk proteins. For protein isolation the used lysis buffer contained 150 mM NaCl (Molar Chemicals Kft), 40 mM Tris (Reanal) pH 7.5, 10% glycerol (Merck), 0.1 % NP40 (Fluka), 5 mM EDTA (Sigma), 1mM PMSF (Sigma) and Complete Protease Inhibitor Cocktail (Roche) according to the manufacturer's instructions. Sf9 cells ($8x10^3$) were suspended in 1 ml lysis buffer and were three times sonicated for 1 min on ice, each sonication followed by freezing (in liquid nitrogen) and thawing. The protein lysates were centrifuged at 18000xg for 10 min at 4°C. Supernatants containing the GST-Vajk-FLAG recombinant proteins were incubated with Glutathione Sepharose 4B (GE Healthcare) 4 h at 4°C. Beads were washed 3 times with high salt buffer (1M NaCl, 40 mM Tris pH 7.5, 10% glycerol, 0.1 % NP40, 5 mM EDTA) and three times with low salt buffer (150 mM NaCl, 40 mM Tris pH 7.5, 10% glycerol, 0.1 % NP40, 5 mM EDTA). Elutions were done in low salt buffer containing 33mM gluthatione, reduced (Reanal). Eluted samples were boiled 5 min in sample buffer containing 0.6 M DTT (Thermo Scientific), loaded for SDS-PAGE and analyzed with Western blot using monoclonal mouse anti-FLAG M2 antibody (Sigma) at a dilution of 1:1000. Samples were loaded for SDS-PAGE and stained with

Coomassie Brilliant Blue. Gels were destained with 10% acetic acid followed by washing 4 h in sterile Millipore water. The respective protein bands were excised.

2.6. Immunization and analysis of antibody specificity

Gel sections containing the GST-Vajk-FLAG recombinant proteins were homogenized in sterile PBS buffer, and 10 µg recombinant protein was injected in Complete Freund Adjuvant (DIFCO) subcutaneously into BALB/c mice. Immunization was repeated monthly, to a total of four times in Incomplete Freund Adjuvant (DIFCO). One week after the fourth immunization antisera were prepared and tested using Western blot analysis on lysates of Sf9 cells, expressing the recombinant Vajk-FLAG proteins. Monoclonal mouse anti-FLAG antibodies (Sigma) were used as positive controls. In parallel the obtained antibodies were also tested on *Drosophila* samples. Although each antibody displayed the expected signals on Sf9 lysates, the anti Vajk2 and the anti Vajk3 antibodies were unreactive on *Drosophila* samples (data not shown).

2.7. Indirect immunofluorescence and immunohistochemistry

Localization of the Vajk proteins was analyzed with indirect immunofluorescence, as follows. The chorion of the embryos was removed with 2 min incubation in 21 g/l sodium hypochloride. Samples were washed with water extensively followed by vitelline membrane permeabilization and fixation for 40 min in a 1:1 mixture of n-heptane (Reanal) and 35% formaldehyde (Reanal). One volume of -20°C cold methanol was added to the samples and was shaken vigorously for one min to remove the vitelline membrane. Liquid components of the sample were removed and embryos were washed 3 times with cold methanol, followed by washing for 5 min in 2:1 metanol: PBT [PBS, 0.1% Triton X-100 (Sigma)]; 5 min in 1:1 methanol:PBT; two times 10 min in PBT and once 30 min in BBT [PBS, 0.1% Triton X-100, 0.1% BSA (Sigma)]. Primary antibodies were diluted 1:500 in PBT-N (PBT, 1% BSA, 5% fetal calf serum) and incubated with the samples overnight at 4 °C. Mouse anti β-tubulin E7 (Hybridoma bank) monoclonal antibody 1:100, and normal mouse serum, 1:500 were used as controls. Washing was performed 3 times for 5 min and once for 30 min in PBT. Alexa Fluor 488 goat anti mouse IgG (Invitrogen, 1:1000) was used as secondary antibody and DAPI (Sigma 1:400) for visualization of the nuclei, both

diluted in BBT. Incubation was done for 1 h at room temperature, followed by three 10 min washes in PBT. Samples were mounted in Fluoromount-G (SouthernBiotech). Cuticle preparations were fixed with 2% paraformaldehyde for 10 minutes, and blocked with 0.1% BSA in PBS supplemented with 0.01% Triton X-100 for 20 min. Following the incubation with the primary antibodies (as described above) samples were washed 3 times for 5 min in PBS, incubated with secondary antibodies for 45 min, washed 3 times for 5 min and mounted in Fluoromount-G. Samples were examined with an epifluorescence microscope (Zeiss Axioskope 2 MOT) or an Olympus FV1000 confocal LSM microscope. Images were edited and analyzed with the ImageJ program (http://imagej.nih.gov/ij/).

To analyze the tracheolae running along the indirect flight muscles of the vajk-1 silenced pupae, pharate adults were reared at 29°C and dissected from the pupal case at stage P15. After removal of their head and the abdomen, the remaining thoraxes were fixed for 2 hours in PBS with 4% formaldehyde. The thoraxes were then washed in PBS and soaked in PBS with 10% sucrose over night at 4°C, before being embedded in OCT and cryosectioned into 6-10 µm slices. The sections were labelled with anti-Nidogen (1:1000, kindly provided by Ruth Palmer, used previously by Wolfstetter et al., 2009 and a FITC-conjugated chitin-binding probe (1:500, New England Biolabs #P5211S) and viewed by confocal imaging.

For immunohistochemistry samples were fixed and incubated with the first antibody as described for the indirect immunofluorescence. The primary antibody was detected with Biotinylated Polyclonal Goat Anti-Mouse Immunoglobulin and Streptavidin/HRP (Dako Denmark A/S). The chromogen was 3-amino 9-ethylcharbasole (Sigma).

2.8. Western blot analysis

Whole organisms of different developmental stages were used to prepare lysates for Western blot analysis as described under section 2.5. The protein concentration of the samples was determined and 300 µg protein was loaded in each well of the 10% SDS-PAGE. The sample buffer contained DTT. Following the separation, proteins were transferred to a PVDF membrane (Millipore) using a wet electroblotting system (Bio-Rad). The membranes were blocked for 1 h with TBS buffer (10 mM Tris pH 7.5, 150 mM NaCl) containing 5% fat free milk powder, rinsed with TBS and incubated with the primary antibodies diluted in TBS containing 1% BSA, for 1 h.

The membranes were washed 3 times with TBS, 10 min each, followed by incubation with HRPO-anti-mouse IgG secondary antibodies (GE Healthcare) at a dilution of 1:5000. The reaction was visualized with ECL Plus (GE Healthcare) according to the manufacturer's instructions.

2.9. Chitin binding assay

Sf9 cells ($8x10^3$) producing Vajk1-FLAG, Vajk4-FLAG, and NimC1-FLAG (Zsámboki et al., 2013) recombinant proteins were lysed in 1 ml lysis buffer containing 150 mM NaCl (Molar Chemicals Kft), 40 mM Tris (Reanal) pH 7.5, 10% glycerol (Merck), 0.1 % NP40 (Fluka), 5 mM EDTA (Sigma), 1mM PMSF (Sigma) and Complete Protease Inhibitor Cocktail (Roche), according to the manufacturer's instructions. Each lysate was incubated for 2 h at 4°C with 40 µl Anti-Flag (DYKDDDDK) Affinity Gel (biotool.com). The affinity gels were washed 5 times in 1500 µl TBS and the bound material was eluted two times with 100mM glycine HCl pH 2.5, 250 µl each. The obtained samples were dialyzed on a cellulose membrane (Sigma-Aldrich) against chitin binding buffer (20mM Tris, 500mM NaCl, 0.05% Triton X-100 (Sigma), pH 7.5) overnight at 4°C. Each protein solution was incubated with 200 µl chitin beads (New England Biolabs), 6 h at 4°C, then the beads were washed 5 times with 1500 µl binding buffer. To elute chitin bound protein, the washed chitin beads were incubated at 4°C for 10 min with 300 µl of 8M urea (Sigma), twice. Fractions were pooled, precipitated with trichloracetic acid (CARLO ERBA) and washed two times with 1500 µl ice cold acetone. Precipitates were dissolved in DTT sample buffer and analyzed by SDS-PAGE and Western blot.

2.10. Electron microscopy

Tissue samples were fixed overnight at 4°C in a pH 7.4 solution of 3.2% paraformaldehyde, 0.5% glutaraldehyde, 1% sucrose, and 0.028% CaCl2 in 0.1 N sodium cacodylate. Postfixation was carried out in 0.5% osmium tetroxide for 1 h, followed by embedding in Durcupan (Fluka) according to the manufacturer's recommendations. 90 nm sections were made by Reichert Ultracut ultramicrotome and stained in Reynold's lead citrate. Sections were viewed in a JEM-1011; JEOL transmission electron microscope, and pictures taken with a Morada; Olympus camera and iTEM software (Olympus).

3. Results

3.1. Molecular characteristics of the Vajk proteins

The *D. melanogaster vajk-1*, *vajk-2*, *vajk-3* and *vajk-4* genes (*CG16886*, *CG16885*, *CG16884* and *CG30101*) encode proteins with relatively small size (between 270 and 373 amino acids) and similar sequence properties (Somogyi et al., 2010). The analysis of the Vajk1, Vajk2, Vajk3 and Vajk4 protein sequences reveals several repetitive segments (Fig. 1A). The presence of the low complexity regions and short repetitive sequences is particularly characteristic of cuticular proteins (Cornman and Willis 2009, Mun et al., 2015). Comparison of the Vajk sequences using the Kalign Multiple Sequence Alignment Tool (Lassmann and Sonnhammer, 2005), (http://www.ebi.ac.uk/Tools/msa/kalign/), revealed 34% to 50% identity. Moreover, we identified a consensus region in each Vajk protein, which also implies the presence of repetitive segments (Fig. 1B). According to flybase, *www.flybase.org*, Vajk1 has two isoforms, of which Vajk1-PA is 15 amino acids longer at the C-terminal region than Vajk1-PB.

By using NCBI and UniProt BLAST the Vajk proteins were found to be 49 to 80% identical to several cuticular proteins identified in other insect species, such as *Bombyx mori* (Ali et al., 2015; Futahashi et al., 2008; Suetsugu et al., 2013), *Danaus plexippus* (Zhan et al., 2011), *Operophtera brumata* (Derks et al., 2015), *Anopheles gambiae* (He et al., 2007), *Papilio polytes* (Futahashi et al., 2012) and *Papilio xuthus* (Futahashi and Fujiwara, 2008; Futahashi et al., 2012). The repetitive sequences found in the Vajk proteins (Fig. 1A) are present in these orthologs (Table 2.). Moreover, the consensus sequence found in the *Drosophila* Vajk proteins (Fig.1B) is also present in the orthologs (Fig. 1C).

3.2. Expression pattern of the vajk genes

As several Nimrod-cluster genes are expressed in hemocytes (Kurucz et al., 2003; Kurucz et al., 2007; Kurant et al., 2008; Zsámboki et al., 2013), we tested whether the *vajk* genes are also transcribed in these cells. RT-PCR analysis of RNA from hemocytes of the homozygous *l(3)mbn-1* mutant larvae, carrying all blood cell types, could not detect any of the *vajk* transcripts (Fig.

2A). Developmental analysis revealed that all four *vajk* transcripts are present in the embryo, in the first and second larval stages, in pupae and in one-day-old adults (Fig. 2B). In third instar larvae, low *vajk-2* and *vajk-3* mRNA levels were detected, while the *vajk-1* and *vajk-4* transcripts were undetectable. Only the *vajk-4* gene was still expressed in 10-day-old adults.

3.3. Vajk protein localization

To analyze the expression pattern of the Vajk proteins, the generated anti-Vajk sera were tested on Sf9 cells expressing the respective recombinant proteins, as well as on fly extracts obtained from different developmental stages. Although all four antibodies reacted with the corresponding Sf9 cells, antibodies to Vajk2 and Vajk3 did not react with fly extracts. For this reason we carried out subsequent analysis with the Vajk1 and Vajk4 specific antibodies.

In 16-17 stage embryos the Vajk1 protein is expressed in the esophagus, the ducts of the salivary glands, the trachea and the denticle belts of the ventral epidermis (Fig.3A). It is not expressed in any larval stages but is present in late, 72 h old pupae (P9-P11 pupal stages). Analysis of the late pupae (Fig. 4) for Vajk1 expression with indirect immunofluorescence revealed staining of a branching structure, the developing trachea, on the surface of the indirect flight muscles (Fig. 4A'). To confirm the tracheal expression we used the tracheoblast-specific *btl-GAL4* driver in combination with *UAS-GFPn*. As the *btl* promoter exhibited low activity at 72 h after puparium formation (at which stage the Vajk1 protein signal was the most intense), we applied lineage tracing (Ito et al., 1997; Honti et al., 2010) with the *btl-GAL4* driver. As shown in Fig. 4A the expression of the Vajk1 protein overlapped with the lineage traced cells. Surface plot analysis (Fig. 4C) of the images of the stained tracheoles revealed that the Vajk1 signal (far red) intensity is higher at the region located between the high intensity GFP signals, indicating that the Vajk1 protein is located inside the tracheolar tubes surrounded by tracheoblasts.

Indirect immunofluorescence analysis of the embryos for Vajk4 protein expression revealed that it is present in the denticle belts of the ventral epidermis and trachea of 16-17 stage embryos (Fig. 3B), similar to Vajk1. The protein was also detected in the epidermis of the 72 h old pupae, expressed homogenously in the cytoplasm of the hair cells and showed a granular distribution in the rest of the epidermal cells (Fig. 5A).

Western blot analysis of Vajk1 showed that both isoforms (41 and 43 kDa) were present in late embryos and in 72 h old pupae (P9-P11 pupal stages) (Fig. 6) in agreement with the indirect immunoflurescent data. The Vajk4 signal (38.1 kDa) was detected in the embryo, in the pupa and the adult (Fig. 6). Neither protein was detected in larvae (Fig. 6).

3.4. Silencing the vajk-1 and the vajk-4 genes with RNA interference

To silence the *vajk* genes, *UAS-vajk*-RNAi constructs and the ubiquitous *Da-GAL4* driver were used. For embryonic gene silencing the *alphaTub67C-GAL4* driver was applied, which possesses exclusive maternal expression.

The *vajk-1* gene was silenced by crossing the V102445/KK, V38024, V38026 and 16886-R3 lines to the ubiquitous driver at 29°C. The F1 generations exhibited a pupal lethal phenotype. When using the V102445/KK line lethality did not appear until the late pupal stage, hence we used this line to study Vajk1 protein levels in 72 h old pupae after RNA silencing. For this purpose we used *102445/KK/UAS-GFP;Da-GAL4/+* flies, where the GFP marker helped with accurate selection of the live pupae at 72 h of pupal development. We tested the presence of the Vajk1 protein using immunohistochemical staining and Western blot analysis in the *vajk-1* silenced flies. The protein was not detected in the trachea (Fig. 4D). The Western blot analysis confirmed this finding, as the Vajk1 protein was absent in lysates obtained from 72 h old pupae (Fig. 4E).

The *vajk-4* gene was silenced using the 15514 and the 3642 *UAS-RNAi* transgenic stocks. As revealed by indirect immunofluorescence (Fig. 5B) and Western-blot analysis (Fig. 5C), silencing with the *Da-GAL4* driver resulted in strong decrease of the Vajk4 signal in the epidermal cells of the 72 h old pupae, yet no visible phenotypic changes were observed in the pupae or adults.

Following gene silencing, both proteins showed decreased levels in the late embryos in immunofluorescence and Western blot analysis, however no phenotypic changes were observed (data not shown).

3.5. Vajk1 and Vajk4 proteins bind to chitin and absence of the Vajk1 causes remarkable changes in the structure of the cuticle

The insect cuticle is composed of two main layers, the epicuticle and the procuticle, providing a mechanical barrier and ensuring strength, as well as protecting the organism from dehydration and infection. The epicuticle, built up of lipids and fibrous proteins, is the thin upper layer, constituting the envelope of the cuticle. The procuticle is the thick inner layer, which is mainly composed of chitin and proteins. Chitin is a polymer of N-acetylglucosamine monomers and is embedded in a protein mesh. Several cuticular proteins have been shown to bind to chitin (Rebes and Willis, 2001; Dong et al., 2016). We analyzed whether the *D. melanogaster* Vajk1 and Vajk4 proteins also have chitin binding capacity. Vajk1-FLAG and Vajk4-FLAG recombinant proteins were produced in Sf9 cells and isolated by Anti-FLAG affinity gel, before being used in chitin binding assays. As a negative control we used NimC1-FLAG recombinant protein (Zsámboki et al., 2013). Samples were analyzed by SDS-PAGE and Western blot, using monoclonal mouse anti-FLAG antibodies. As shown in Fig. 7A the Vajk1 and Vajk4 proteins bind to chitin.

The cuticle also forms the internal lining of the tracheal tubes. The structure of this cuticle is different from the epidermal cuticle due to the presence of the taenidial forlds, forming a helical structure protruding into the lumen, providing stability and flexibility to the respiratory tubes. The taenidial folds include taenidia, which provide them reinforcement (Matusek et al., 2006; Öztürk-Colak et al., 2016). As the Vajk1 protein was shown to be present in the tracheolae running along the indirect flight muscle, we analyzed the structure of these tracheolae in *vajk-1* silenced pupae. A FITC-conjugated chitin-binding probe and an anti-Nidogen antibody were used on cross sections of the indirect flight muscles in P15 stage pupae to visualize the chitin and the basement membrane respectively in indirect immunofluorescence. A striking reduction was revealed in chitin-marked tracheoles innervating the muscles of the *vajk-1* silenced samples, compared to the parental strains (Fig. 7 B). Moreover electron microscopy pictures reveal remarkable differences in the appearance of the cuticle in Vajk1 silenced specimens (Fig. 7C). In the electron microscopy sections both parental control phenotypes are characterized by a strictly regular cuticular lining with an evenly spaced wave-like appearance of the epicuticular layer; the basal procuticular surface on the other hand is only very slightly wavy. As a result of this arrangement the thickness of the procuticle gets increased right under the crest of the wave, in addition, a taenidium can be observed in each of these taenidial folds (thicker regions). In the *vajk-1* silenced samples each cuticular layer follows the wavy course, the waves are less regular, the thickening of the procuticle and the taenidium is practically missing.

4. Discussion

We previously described that several Nimrod cluster genes encode proteins with roles in the innate immune responses of *D. melanogaster* (Kurucz et al., 2003; Kurucz et al., 2007; Zsámboki et al., 2013). As the cluster has been present in arthropod genomes for more than 300 million years (Somogyi et al., 2008; Somogyi et al., 2010) it is possible that the genes located in this cluster might be involved in a common biological process (Hurst et al., 2004; Lee and Sonnhammer, 2003), therefore we analyzed new candidates of the cluster, the *vajk* genes. We found that unlike the previously analyzed *nim* and *Hemese*, the *vajk* genes are not transcribed in hemocytes. Previously, it was shown that each Vajk protein contains an N-terminal signal peptide, suggesting that they are secreted and possess extracellular function. Their amino acid sequences contain about 20% valine, being the most abundant amino acid (Somogyi et al., 2010). Moreover here we found that they are homologous with other proteins described to be cuticle components. It is known that the insect cuticle, composed primarily of low complexity proteins (Cornman and Willis, 2009) and the polysaccharide chitin, represents the first barrier against invading microorganisms; meanwhile, it also protects the organism from environmental damage and dehydration (Tzou et al., 2000; Vincent and Wegst, 2004).

The cuticle of *D. melanogaster* is first produced during embryonic developmental stage 16 (Ostrowski et al., 2002), and it is present not only on the body surface, but also forms the internal lining of the tracheae, the salivary ducts and the gut (Moussian et al., 2006). We detected the presence of the Vajk1 protein in all these structures in late (16, 17) embryonic stages (Fig.3A). The Vajk4 protein showed a similar expression profile to Vajk1 (Fig.3B). It was observed in the denticle belt areas and in the trachea, but it could not be detected in the esophagus or the ducts of the salivary glands, where the Vajk1 was also expressed. The *vajk-4* (CG30101) RNA *in situ* hybridization data (available in the BDGP database) reveals RNA expression in the denticle belt and the trachea of embryos in 13-16 developmental stages (http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG30101), where in the later stages we detected the presence of the Vajk4 protein. Our Western blot data also supports the presence of the Vajk1 and Vajk4 proteins in the late embryonic developmental stages (Fig.6). Moreover, the BDGP and Fly-FISH databases also contain embryonic RNA *in situ* hybridization data on the *vajk-2* (CG16885)

(http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG16885) and the *vajk-3* (CG16884) (http://insitu.fruitfly.org/cgi-bin/ex/report.pl?ftype=1&ftext=CG16884). Signals can be observed in the denticle belts and the trachea, which suggests that the expression pattern of the *vajk* genes overlaps in these tissues.

Following embryogenesis, cuticle differentiation involves complete renewal during larval and pupal stages (Truman and Riddiford, 1999). The cuticular proteins are generally secreted by the epidermal epithelium (Ostrowski et al., 2002), but in the tracheae and tracheoles several cuticular proteins were shown to be transported through the epithelial cells from the hemolymph (Sass et al., 1994; Dong et al., 2014). We observed the presence of the Vajk1 in the 72 h old pupae in tracheoles connected to the developing indirect flight muscles. The tracheoles invading the indirect flight muscles of the pharate adult develop from the dorsal air sac already present in the third instar larvae (Cabernard et al., 2004). According to the literature, the indirect flight muscle invasion ends after the terminal branches enter the T-tubules, extending within the myocyte and encircling the mitochondria of the flight muscle. This process is completed at 70 h after puparium formation (Peterson and Krasnow, 2014), at which point the Vajk1 protein appears in the cuticular layer of the tracheoles. With indirect immunofluorescence the protein can be detected in 96 h old pupae, but the signal is weaker compared to that observed in 72 h old pupae, which may be the consequence of the conformational changes in the protein as a result of cuticle sclerotization, leading to the masking of the active epitopes for the antibody (Locke et al., 1994). We could not detect the presence of the Vajk1 protein with Western blot in 96 h old pupae, which may be due to the sclerotization that probably blocks the possibility for the Vajk1 to enter the protein lysates used for the Western blot. To increase the protein isolation efficiency, beside the homogenization of the pupae we also applied sonication of the samples but the Vajk1 protein could still not be detected in the protein lysates.

Vajk4 protein, similar to Vajk1, is not expressed in the third instar larva, and shows the highest expression level in 72 h old pupae, where indirect immunofluorescent analysis revealed its presence in the epithelial cells of the body surface. Although Western blot analysis still shows the presence of the protein in 96 h old pupae and in the adult, at these stages we could not detect the protein in the epithelial cells with indirect immunofluorescence.

Several Vajk protein orthologs are considered to be cuticular proteins of different insect species. Three Vajk orthologs of *Anopheles gambiae* (Fig. 1, Table 2) were isolated from cast

cuticles (He et al., 2007). Moreover it was formally shown that two Vajk orthologs, CPG12 and CPG13 from *Bombyx mori*, bound chitin (Dong et al., 2016). Using chitin binding assays, we show that the *D. melanogaster* Vajk1 and Vajk4 proteins also bind chitin (Fig. 7A). Furthermore immunofluorescence experiments revealed a striking reduction in chitin marked tracheoles innervating the indirect flight muscles of the *vajk-1* silenced flies (Fig. 7B). Significantly, these data were reinforced by electron microscopy, revealing a disordered structure of the cuticular lining of these tracheoles (Fig.7C). We hypothesize that in the absence of the Vajk1, the chitin cannot be properly embedded in the cuticle, hence resulting in a disorganized structure to the cuticular layer.

In our opinion the classification of the Vajk proteins in a cuticular protein family is not completely clear. Many Vajk protein orthologs (Fig. 1, Table 2) were assigned to the CPG family. Moreover the three Vajk orthologs from *Anopheles gambiae* (Fig. 1, Table 2) belong to the CPLCP family. Members of this family have a high density of PV and PY pairs (Cornman and Willis, 2009; Willis, 2010), which is also characteristic for each Vajk protein. Surprisingly the Vajk consensus sequence shown in Fig. 1C is present in the listed CPG and also CPLCP proteins.

In this work we show that all *vajk* genes are transcribed in the embryo, first and second stage larvae, pupae, as well as in one day old adults. Moreover, we show that the Vajk proteins are present in structures known to contain cuticle. It is conceivable that in the tracheoles running along the indirect flight muscles of the pharate adults, the presence of the Vajk1 protein is essential, because the observed cuticular structure is accompanied with pupal lethality. Moreover we suppose that the cuticular localization of the Vajk proteins might also serve as the first, mechanical barrier to protect the organism from the harmful environmental aspects, thus achieved a concerted action with the previously characterized proteins encoded by the Nimrod cluster.

Acknowledgements

We are grateful to Olga Kovalcsik, Anita Balázs, Anikó Képíró and Sarolta Pálfia for the technical help. We acknowledge to Kálmán Somogyi, Botond Sipos and József Mihály for their useful advices concerning the work. Our thanks are also due to Zoltán Györgypál for his help in bioinformatics analysis. This research was supported by grants from the Hungarian National

Science Foundation, GINOP-2.3.2-15-2016-00035 (ÉK), the Hungarian National Science Foundation, GINOP-2.3.2-15-2016-00001, NKFI NN118207 and NKFI K 120142 (IA), OTKA PD-115534 (VH), GINOP-2.3.2-15-2016-00032 (GJ), Carl Tryggers Stiftelse (MJW), the Stiftelsen Olle Engkvist Byggmästare (MJW) the Swedish Cancer foundation (AU), and the Knut and Alice Wallenberg Foundations (AU).

References

d Ali, M.S., Rahman, R.F., Swapon, A.H., 2015. Transcriptional regulation of cuticular protein glycine-rich13 gene expression in wing disc of *Bombyx mori*, Lepidoptera. J. Insect Sci. 15. doi:10.1093/jisesa/iev019

Cabernard, C., Neumann, M., Affolter, M., 2004. Cellular and molecular mechanisms involved in branching morphogenesis of the *Drosophila* tracheal system. J. Appl. Physiol. 97, 2347–2353. doi:10.1152/japplphysiol.00435.2004

Cornman, R.S., Willis, J.H., 2009. Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. Insect Mol. Biol. 18, 607–622. doi:10.1111/j.1365-2583.2009.00902.x

Derks, M.F.L., Smit, S., Salis, L., Schijlen, E., Bossers, A., Mateman, C., Pijl, A.S., de Ridder, D., Groenen, M.A.M., Visser, M.E., Megens, H.-J., 2015. The genome of winter moth (*Operophtera brumata*) provides a genomic perspective on sexual dimorphism and phenology. Genome Biol. Evol. 7, 2321–2332. doi:10.1093/gbe/evv145

Dong, B., Miao, G., Hayashi, S., 2014. A fat body-derived apical extracellular matrix enzyme is transported to the tracheal lumen and is required for tube morphogenesis in *Drosophila*. Development 141, 4104–4109. doi:10.1242/dev.109975

Dong Z., Zhang W., Zhang Y., Zhang X., Zhao P., Xia Q., 2016. Identification and characterization of novel chitin-binding proteins from the larval cuticle of silkworm, *Bombyx mori*. J. Proteome Res. 15, 1435−1445. doi: 10.1021/acs.jproteome.5b00943

Duressa, T.F., Boonen, K., Huybrechts, R., 2016. A quantitative peptidomics approach to unravel immunological functions of angiotensin converting enzyme in *Locusta migratoria*. Gen Comp Endocrinol. 235, 120-129. doi: 10.1016/j.ygcen.2016.06.024

525 Elliott, M.R., Ravichandran, K.S., 2008. Death in the CNS: six-microns-under. Cell 133, 393-
526 395. doi: 10.1016/j.cell.2008.04.014
527 Futahashi, R., Fujiwara, H., 2008. Identification of stage-specific larval camouflage associated
528 genes in the swallowtail butterfly, *Papilio xuthus*. Dev. Genes Evol. 218, 491–504.
529 doi:10.1007/s00427-008-0243-y
530 Futahashi, R., Okamoto, S., Kawasaki, H., Zhong, Y.-S., Iwanaga, M., Mita, K., Fujiwara, H.,
531 2008. Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*.
532 Insect Biochem. Mol. Biol. 38, 1138–1146.
533 Futahashi, R., Shirataki, H., Narita, T., Mita, K., Fujiwara, H., 2012. Comprehensive microarray-
534 based analysis for stage-specific larval camouflage pattern-associated genes in the
535 swallowtail butterfly, *Papilio xuthus*. BMC Biol. 10, 46. doi:10.1186/1741-7007-10-46
536 He, N., Botelho, J.M.C., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R., Willis,
537 J.H., 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass
538 spectrometry. Insect Biochem. Mol. Biol. 37, 135–146. doi:10.1016/j.ibmb.2006.10.011
539 Honti, V., Csordás, G., Márkus, R., Kurucz, É., Jankovics, F., Andó, I., 2010. Cell lineage tracing
540 reveals the plasticity of the hemocyte lineages and of the hematopoietic compartments in
541 *Drosophila melanogaster*. Mol. Immunol. 47, 1997–2004.
542 doi:10.1016/j.molimm.2010.04.017
543 Hurst, L.D., Pál, C., Lercher, M.J., 2004. The evolutionary dynamics of eukaryotic gene order.
544 Nat. Rev. Genet. 5, 299–310. doi:10.1038/nrg1319
545 Ibrahim, D.M., Biehs, B., Kornberg, T.B., Klebes, A., 2013. Microarray comparison of anterior
546 and posterior *Drosophila* wing imaginal disc cells identifies novel wing genes. G3 (Bethesda)
547 3, 1353–1362. doi:10.1534/g3.113.006569
548 Irving, P., Ubeda, J.-M., Doucet, D., Troxler, L., Lagueux, M., Zachary, D., Hoffmann, J.A.,
549 Hetru, C., Meister, M., 2005. New insights into *Drosophila* larval haemocyte functions
550 through genome-wide analysis. Cell. Microbiol. 7, 335–350. doi:10.1111/j.1462-
551 5822.2004.00462.x
552 Ito, K., Awano, W., Suzuki, K., Hiromi, Y., Yamamoto, D., 1997. The *Drosophila* mushroom
553 body is a quadruple structure of clonal units each of which contains a virtually identical set of
554 neurones and glial cells. Development 124, 761–771.

555 Konrad, L., Becker, G., Schmidt, A., Klöckner, T., Kaufer-Stillger, G., Dreschers, S., Edström,
556 J.E., Gateff, E., 1994. Cloning, structure, cellular localization, and possible function of the
557 tumor suppressor gene lethal(3)malignant blood neoplasm-1 of *Drosophila melanogaster*.
558 Dev. Biol. 163, 98-111. doi:10.1006/dbio.1994.1126
559 Kurant, E., Axelrod, S., Leaman, D., Gaul, U., 2008. Six-microns-under acts upstream of Draper
560 in the glial phagocytosis of apoptotic neurons. Cell 133, 498-509. doi:
561 10.1016/j.cell.2008.02.052
562 Kurucz, É., Márkus, R., Zsámboki, J., Folkl-Medzihradszky, K., Darula, Z., Vilmos, P., Udvardy,
563 A., Krausz, I., Lukacsovich, T., Gateff, E., Zettervall, C.-J., Hultmark, D., Andó, I., 2007a.
564 Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes.
565 Curr. Biol. 17, 649–654. doi:10.1016/j.cub.2007.02.041
566 Kurucz, É., Zettervall, C.-J., Sinka, R., Vilmos, P., Pivarcsi, A., Ekengren, S., Hegedüs, Z.,
567 Ando, I., Hultmark, D., 2003. Hemese, a hemocyte-specific transmembrane protein, affects
568 the cellular immune response in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. 100, 2622–2627.
569 doi:10.1073/pnas.0436940100
570 Lee, J.M., Sonnhammer, E.L.L., 2003. Genomic gene clustering analysis of pathways in
571 eukaryotes. Genome Res. 13, 875–882. doi:10.1101/gr.737703
572 Lionakis, M.S., Kontoyiannis D.P., 2012. *Drosophila melanogaster* as a model organism for
573 invasive aspergillosis. Methods Mol. Biol. 845, 455-468. doi: 10.1007/978-1-61779-539-
574 8_32
575 Locke, M., Kiss, A., Sass, M., 1994. The cuticular localization of integument peptides from
576 particular routing categories. Tissue Cell 26, 707–734. doi:10.1016/0040-8166(94)90055-8
577 Macours, N., Hens, K., Francis, C., De Loof, A., Huybrechts, R., 2003. Molecular evidence for
578 the expression of angiotensin converting enzyme in hemocytes of Locusta migratoria:
579 stimulation by bacterial lipopolysaccharide challenge. J. Insect. Physiol. 49, 739–746.
580 doi:10.1016/S0022-1910(03)00110-0
581 Matusek, T., Djiane, A., Jankovics, F., Brunner, D., Mlodzik, M., Mihály, J., 2006.
582 The *Drosophila* formin DAAM regulates the tracheal cuticle pattern through organizing the
583 actin cytoskeleton. Development. 133, 957-966. doi: 10.1242/dev.02266
584 Márkus, R., Lerner, Z., Honti, V., Csordás, G., Zsámboki, J., Cinege, G., Párducz, Á.,
585 Lukacsovich, T., Kurucz, É., Andó, I., 2015. Multinucleated giant hemocytes are effector

cells in cell-mediated immune responses of *Drosophila*. J. Innate Immun. 7, 340–353. doi:10.1159/000369618

Moussian, B., Seifarth, C., Müller, U., Berger, J., Schwarz, H., 2006. Cuticle differentiation during *Drosophila* embryogenesis. Arthropod Struct. Dev. 35, 137–152. doi:10.1016/j.asd.2006.05.003

Mun, S., Noh, M.Y., Dittmer, N.T., Muthukrishnan, S., Kramer, K.J., Kanost, M.R., Arakane, Y., 2015. Cuticular protein with a low complexity sequence becomes cross-linked during insect cuticle sclerotization and is required for the adult molt. Sci. Rep. 5, 10484. doi:10.1038/srep10484

Ostrowski, S., Dierick, H.A., Bejsovec, A., 2002. Genetic control of cuticle formation during embryonic development of *Drosophila melanogaster*. Genetics 161, 171–182.

Öztürk-Colak, A., Moussian, B., Araújo, S.J., Casanova, J., 2016. A feedback mechanism converts individual cell features into a supracellular ECM structure in *Drosophila* trachea. eLife. 5, e09373. doi:10.7554/eLife.09373.

Peterson, S.J., Krasnow, M.A., 2015. Subcellular trafficking of FGF controls tracheal invasion of *Drosophila* flight muscle. Cell 160, 313–323. doi:10.1016/j.cell.2014.11.043

Rebers, J.E., Willis J. H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. Insect Biochem. Mol. Biol. 31, 1083-1093. doi: http://doi.org/10.1016/S0965-1748(01)00056-X

Sass, M., Kiss, A., Locke, M., 1994. The localization of surface integument peptides in tracheae and tracheoles. J. Insect Physiol. 40, 561–575. doi:10.1016/0022-1910(94)90143-0

Sato, M., Kornberg, T.B., 2002. FGF is an essential mitogen and chemoattractant for the air sacs of the *Drosophila* tracheal system. Dev. Cell 3, 195–207. doi:10.1016/S1534-5807(02)00202-2

Shiga, Y., Tanaka-Matakatsu, M., Hayashi, S., 1996. A nuclear GFP/β-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*. Dev. Growth Differ. 38, 99–106.

Shklyar, B., Levy-Adam, F., Mishnaevski, K., Kurant, E., 2013. Caspase activity is required for engulfment of apoptotic cells. Mol. Cell Biol. 33, 3191-201. doi: 10.1128/MCB.00233-13

Somogyi, K., Sipos, B., Pénzes, Z., Kurucz, É, Zsámboki, J., Hultmark, D., Andó, I., 2008. Evolution of genes and repeats in the Nimrod superfamily. Mol. Biol. Evol. 25, 2337-2347. doi: 10.1093/molbev/msn180

Somogyi, K., Sipos, B., Pénzes, Z., Andó, I., 2010. A conserved gene cluster as a putative functional unit in insect innate immunity. FEBS Lett. 584, 4375–4378. doi:10.1016/j.febslet.2010.10.014

Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., Sezutsu, H., Osanai-Futahashi, M., Suzuki, M.G., Daimon, T., Shinoda, T., Taniai, K., Asaoka, K., Niwa, R., Kawaoka, S., Katsuma, S., Tamura, T., Noda, H., Kasahara, M., Sugano, S., Suzuki, Y., Fujiwara, H., Kataoka, H., Arunkumar, K.P., Tomar, A., Nagaraju, J., Goldsmith, M.R., Feng, Q., Xia, Q., Yamamoto, K., Shimada, T., Mita, K., 2013. Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect, *Bombyx mori*. G3 (Bethesda) 3, 1481–1492. doi:10.1534/g3.113.006239

Truman, J.W., Riddiford, L.M., 1999. The origins of insect metamorphosis. Nature 401, 447–452. doi:10.1038/46737

Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J.M., Lemaitre, B., Hoffmann, J.A., Imler, J.L., 2000. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. Immunity 13, 737–748. doi:10.1016/S1074-7613(00)00072-8

Vincent, J.F.V., Wegst, U.G.K., 2004. Design and mechanical properties of insect cuticle. Arthropod Struct. Dev. 33, 187–199. doi:10.1016/j.asd.2004.05.006

Wolfstetter, G., Shirinian, M., Stute, C., Grabbe, C., Hummel, T., Baumgartner, S., Palmer, R.H., Holz, A., 2009. Fusion of circular and longitudinal muscles in Drosophila is independent of the endoderm but further visceral muscle differentiation requires a close contact between mesoderm and endoderm. Mech. Dev. 126, 721-736. doi: 10.1016/j.mod.2009.05.001

Xi, Z., Gavotte, L., Xie, Y., Dobson, S.L., 2008. Genome-wide analysis of the interaction between the endosymbiotic bacterium *Wolbachia* and its *Drosophila* host. BMC Genomics 9, 1. doi: 10.1186/1471-2164-9-1

Zhan, S., Merlin, C., Boore, J.L., Reppert, S.M., 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147, 1171–1185. doi:10.1016/j.cell.2011.09.052

Zsámboki, J., Csordás, G., Honti, V., Pintér, L., Bajusz, I., Galgóczy, L., Andó, I., Kurucz, É., 2013. *Drosophila* Nimrod proteins bind bacteria. Cent. Eur. J. Biol. 8, 633–645. doi:10.2478/s11535-013-0183-4

Figure and table captions:

Fig. 1. Sequence analysis of the Vajk proteins and their orthologs. (A) The identical repetitive segments are marked with the same colors in bold and/or are underlined. (B) In the alignment, the identified consensus sequence is highlighted in bold and the sequences including repetitive segments are underlined. (C) The Vajk specific consensus sequence is also present in the Vajk orthologs of other species. The residues indicated with regular characters on gray background share similar properties with the consensus specific amino acids. Abbreviations: *Ag, Anopheles gambiae; Bm, Bombyx mori; Dp, Danaus plexippus*; *Ob*, *Operophtera brumata; Pp, Papilio polytes*; *Px, Papilio xuthus*. Sequence analysis was done using CLUSTALW Multiple Sequence Alignment (http://www.genome.jp/tools/clustalw/).

Fig. 2. Analysis of the *vajk* gene expression in the (A) hemocytes of the third instar *l(3)mbn-1* larvae and in (B) different developmental stages of the Oregon R *D. melanogaster*. Agarose gel electrophoresis of the RT-PCR products obtained with primer pairs specific for both RA and RB *vajk-1* transcripts, the *vajk-2, vajk-3 and vajk-4* transcripts, generating 1031, 828, 849, and 1003 bp fragments respectively. As positive controls for the cDNA, the *Hemese* specific (175 bp) (A) and rp49 specific (316 bp) (B) primer sets were used. To test the *vajk* specific primer sets, vectors containing *vajk* cDNAs were used as templates (positive controls). Negative control reactions include water as templates.

Fig 3. Localization of the Vajk1 (A) and Vajk4 (B) proteins in the 16-17 stage embryo. Both proteins were detected in the trachea (tr) and the denticle belt areas of the ventral epidermis (db). Moreover the Vajk1 protein was also detected in the esophagus (eg) and the ducts of the salivary glands (sd). As controls normal mouse serum (C) and anti-tubulin mouse monoclonal antibody (D) were used. The secondary antibody was Alexa Fluor 488 goat-anti-mouse IgG (green). Nuclear staining is blue. Images were obtained with Confocal LSM. The scalebar represents 100 µm.

Fig. 4. Expression of the Vajk1 protein in 72 h old pupae. (A, B) Vajk1 (far red) co-localizes with the *btl* promoter activity (green). Indirect flight muscles were analyzed with Confocal LSM.

The used line carried the *Btl-GAL4*, the *UAS-GFPn* and the *Flip-FRT* lineage tracing construct system (*Btl>AFG>GFP*). A, A' and A" images were stained with antibodies against the Vajk1 protein, and B, B' and B", with the normal mouse serum. As secondary antibodies Alexa Fluor 633 goat-anti-mouse IgG (far red) were used. The scalebar represents 50 µm. (C) Surface plot of *btl* activity and Vajk1 localization (ImageJ). The scalebars represents 8 µm. (D) Immunohistochemical analysis of the indirect flight muscle following RNA interference at 29 °C. After the treatment with the primary antibodies, streptavidin/HRP and biotinilated polyclonal goat anti-mouse immunoglobulins were used. The chromogen was 3-amino 9-ethylcharbasole. The scalebar represents 100 µm. (E) Western blot analysis of the lysates prepared after *vajk-1* RNA inhibition.

Fig. 5. Expression of the Vajk4 protein in 72 h old pupae. (A, B) Analysis of the protein with indirect immunofluorescence. (A) The protein is expressed in the epidermal tissue of the pharate adults. Hair cells are marked with arrows. Alexa Fluor 488 goat-anti mouse IgG (green) was used as secondary antibody. Nuclear staining is blue. (B) Analysis of the Vajk4 protein after gene silencing at 29 °C. Scalebars represent 100 µm. (C) Western blot analysis of the Vajk4 following gene silencing.

Fig. 6. Western blot analysis of the Vajk1 and Vajk4 proteins in different developmental stages of *D. melanogaster*.

Fig. 7. Functional analysis of the Vajk proteins. (A) Vajk1 and Vajk4 proteins bind to chitin. Chitin binding assays were carried out on recombinant, C terminally FLAG tagged, Vajk and NimC1 proteins as described in Materials and methods. Inp, input protein sample; Ft, flow through; W, the fifth wash of the chitin beads; Elu, fraction eluted with 8M urea. (B) Thoracic cross-sections at stage P15, showing dorsal longitudinal flight muscles labelled for Nidogen (Nido, red) to visualize the basement membrane and chitin (CBP, green) to mark the tracheal terminal branch tracheoles that innervate the muscles (arrows). The presence of CBP-labelled tracheoles is reduced upon expression of *vajk-1* RNAi compared to control flies. Scale bar represents 10 µm**.** (C) Electron micrographs of the longitudinally sectioned tracheloae localized in the indirect flight muscles. The epicuticle (ep), procuticle (pc), basal procuticular surface (ps)

and the taenidia (ta) are normally organized in the parental lines (*102445/KK* and *Da-GAL4*) as opposed to the remarkably disordered assembly in the *vajk-1* silenced samples (*102445/KK/+;Da-GAL4/+*). lm, indicates the tracheolar lumen. Scale bar represents 1 µm.

Table 1

Primers used in this work. Adapter restriction sites are marked with bold.

Table 2

Sequence analysis of the Vajk orthologs, concerning the repetitive blocks of the Vajk proteins. Abbreviations: +, present; -, not present; the species used are described in the legend of Fig. 1.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Primer	Sequence (5'-3')
vajk-1 fw	GAA**GAATTC**ATGAAGTCGATGCGCAGAG
vajk-1 A rev	GCC**TCTAGA**GTAACCGTAATATGAATGATAATC
vajk-1 AB rev	GTGGTGATGATCCTCCTCGATGTG
vajk-2 fw	ATC**GAATTC**ATGAAATCCATGCTCATTTT
vajk-2 rev	GTT**TCTAGA**CTTTTTGTGGAGGTAGCCAC
vajk-3 fw	GGC**GAATTC**ATGAAAGTATTCATCTGCTT
vajk-3 rev	TAG**TCTAGA**CTTGTGTAACCATGGTG
vajk-4 fw	GCA**GAATTC**ATGCGAATGTTCGTACTTCC
vajk-4 rev	TAG**TCTAGA**GTAGTGTCCATGGCCGTG
rp49 fw	GCATACAGGCCCAAGATCCGT
rp49 rev	CAATCTCCTTGCGCTTCTTG
hemese fw	TCAACTGACCGTCGTTTTCC
hemese rev	CCGTTTCACTTGGGGTTGAA

ACCEPTED MANUSCRIPT

	EKHVP	EKKIP	EKKVP	EVEKPY	EVKVPI	PVPY	PYEV	RPYPV	VIKK	VKVPV	VPAPYPVEK	VPQPYPVEK	VPVEV	VPVHV	VPVPV	YDRPVPV
Vajk1	+	-	-	-	-	+	+	-	+	+	+	-	-	+	+	+
Vajk2	-	-	-	-	-	+	-	-	-	+	-	+	+	+	+	-
Vajk3	-	-	-	-	-	-	+	+	-	+	-	-	+	+	+	-
Vajk4	-	+	+	+	+	+	+	-	+	+	-	-	-	+	-	-
Ag AGAP009758	+	-	+	-	-	+	+	-	+	+	+	-	-	+	+	+
Ag AGAP009759	+	-	+	-	-	+	+	+	-	+	-	+	-	+	+	-
Ag AGAP008817	+	+	+	-	+	+	+	-	-	+	-	-	-	-	-	-
Bm CPG12	+	+	+	-	-	+	-	+	-	+	+	+	+	+	-	-
Bm CPG13	+	-	+	-	-	+	+	+	-	+	+	+	-	+	+	-
Bm CPG24	+	+	+	+	+	+	+	-	+	+	-	-	-	-	+	-
Dp CPG12	+	-	+	-	-	+	-	+	-	+	+	+	-	+	+	-
Dp CPG24	+	-	+	+	+	+	+	-	+	+	-	-	-	-	-	-
Ob CPG11	-	-	+	-	-	+	-	+	-	+	-	-	+	+	+	-
Ob CPG12	+	+	-	-	-	+	-	+	-	+	+	+	-	+	+	-
Ob CPG24	+	-	+	+	+	+	+	-	+	+	-	-	-	-	-	-
Pp CPG11	+	-	+	-	-	+	-	-	-	+	+	-	+	+	+	-
Pp CPG12	+	-	+	-	-	+	-	+	-	+	+	+	-	+	+	-
Pp CPG24	+	-	+	+	+	+	+	-	+	+	-	-	-	-	+	-
Px CPG11	+	-	+	-	-	+	-	-	-	+	+	-	+	+	+	-
Px CPG12	+	-	+	-	-	+	-	+	-	+	+	+	-	+	+	-
Px CPG13	+	-	+	-	-	+	+	+	-	+	+	+	-	+	+	-

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

A

Vajk1
MKSMRTATSLALVLLLATSYVRAEDEKAAAPAEEKKAEPEAKAAEAAASEDTTKSKRGLHHYEDYHHHHVPHFPVHEEKTLTVIKKVPVPVPIEKIVHVPVEKHIHVPVK
VKVPKPYPVIKHIPYEVKEIVKVPYEVPAPYPVEKQVHVPVHVHYDRPVPVKVHVPAPYPVEKKVHVPVKVHVPAPYPVEKIVHYNVEKHVHVDKPYPVEKVVHYPVKVP
VDKPVPHYIDKPVPHYVDKPVPVPVIKKVPVPVHVPYDRPVPVHVEKPVPYEVKVHVPAPYPVIKEVPVKVEKHVPYPVKIPVEKPVHVHIEKHVPEYHEKHVTYKEPEF
HHKHIEEDHHHAPIHHHSHPIVEHEHEVEHEFASHDYHSYYGY

Vajk2
MKSMLIFGLVAMCVLVANASEEAPKKAVETAEPAEKKQEKRGIGHGLGYGYGPSAGGAILGSGIGVGVPVAPAVAELPTQVHTNTVVRTVQVPYQVERHVPYPVEKTVTY
PVKVPVPQPYPVEKIVHVPVKQIVKVPVEVPQPYPVEKVIRVPVKIPVDRPYTVHVDKPYPVPVEKPVPYTVEKRVIHKVPVHVERPVPYKVAVPVPVHVESHVKPAVAV
TIITVAAAPAIISIIGYSGIIGISGYGISGIIGISSYGSSSYGGSAIIGGYLIIKK

Vajk3
MKVFICLAALLVASACASKTEGEKVPLEKKLDKRGLLDLGYGYGHAGLDTGYLGHGSISGHGSYGHGYGLTGYSAPAAVAVGHSGPAIAVGHTAPAVAVHHAPAPYVISK
QADVHKTITITTKGIPVPVHVDRPYPVVHEKRVPVEVKVPVPQPYEVIRKVPVTVKEYVKVPVPVPQPYEVIRHEKVPVHVPVDRPVPVEVPRPYPVPVAKPYPVYVEKAV
NVQVPVHVDRPYPVYVKVPVVSHSVVKHAPTVAVSSYPVSAIGHDATVYSDHHGYHK

Vajk4
MRMFVLPCLAVCVALAHCGGAVEDKKAEGDGKTVEKRGLHLGDYHHYQPHHEHIKTVTIEKKIPVPYTVTKHVPYTVEKKIPYEVKVDVPQPYIVEKKVPVHVKEYVKVP
VHVPKPYEVIKKIPYEVKVPVDKPYEVKVPVPQPYEVIKKIPYEVKVPVPQPYEVIKKVPHEVKVEVPVPKPYEVIKKVPYEVKYEVEKPYDVEVPKPYDVEVEKPYTVV
VEKKVPYEVKVPVDKPYKVEVEKPYPVHVKVPVPQPYTVEKKVPYTVEKPVPYEVKVPIEKPIPVYTEVKVPIHKEIPVPEKYHVEVPIFKHHQEDHHDYHSHGHGHY

B

Vajk1 79-TLTVIKKVPVPVPIEKIVHVPVEKHIHVPVKVKVPKPYPVIKHIPYEVKEIVKVPYEVPAPYPVEKQVHVPVHVHYDRPVPVKV-163

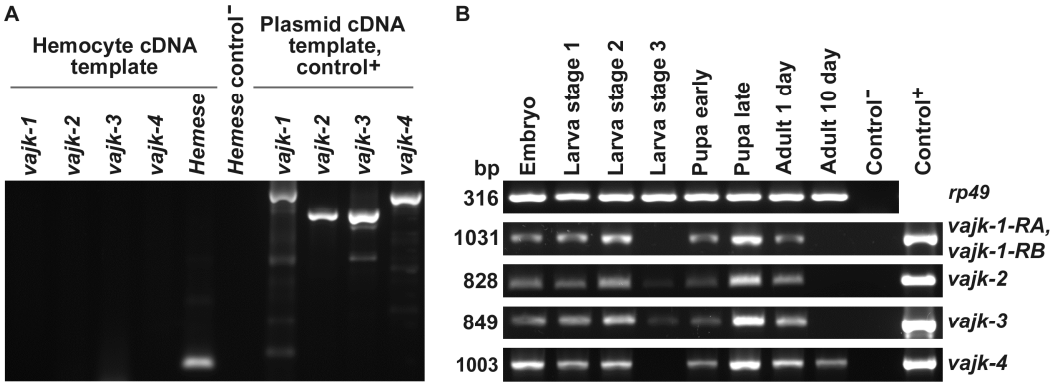
Vajk2 83-TNTVVRTVQVPYQVERHVPYPVEKTVTYPVKVPVPQPYPVEKIVHVPVKQIVKVPVEVPQPYPVEKVIRVPVKIPVDRPYTVHV-167

Vajk3 117-TITITKGIPVPVHVDRPYPVVHEKRVPVEVKVPVPQPYEVIRKVPVTVKEYVKVPVPVPQPYEVIRHEKVPVHVPVDRPVPVEV-201

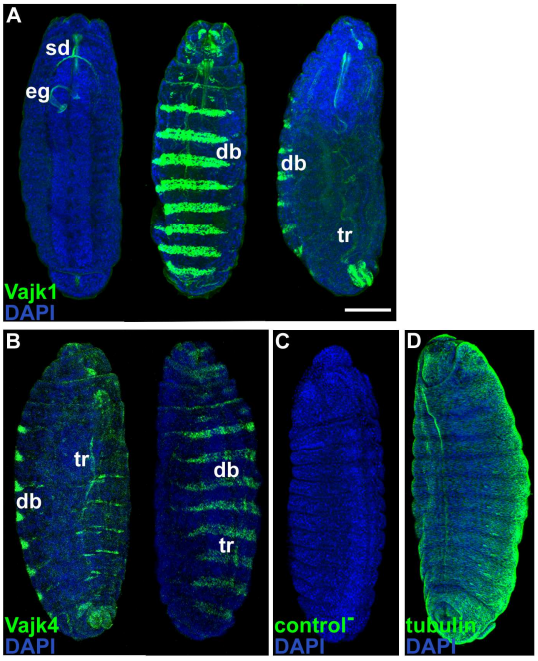
Vajk4 56-TVTIEKKIPVPYTVTKHVPYTVEKKIPYEVKVDVPQPYIVEKKVPVHVKEYVKVPVHVPKPYEVIKKIPYEVKVPVDKPYEVKV-140

C

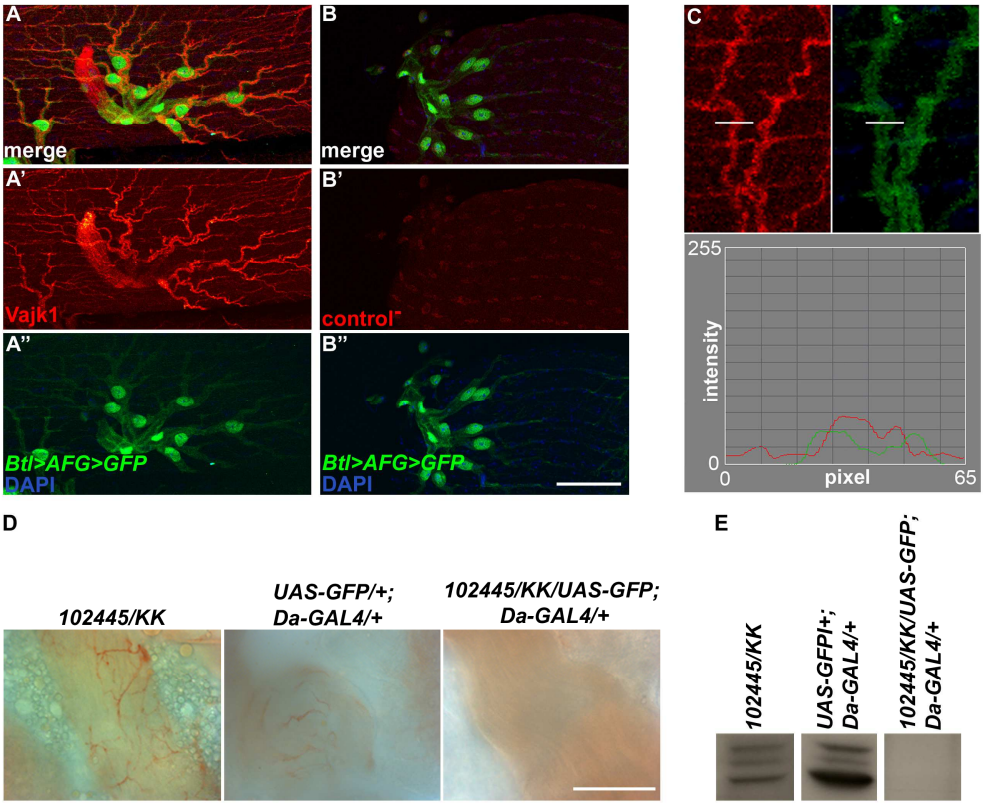
```
Consensus          T-T------VP-----------EK-----VKV-VP-PY-V-------VK--VKVP--VP-PY-V-------V----D-P--V-V
Ag AGAP009758   78-TLTVVKKVPVPYPVEKHIPVPVEKHVPVPVKVGVPKPYPVYKTVHYPVKEIVKVPVHVPAPYPVEKKVPYPVHVPYDRPVPVKV-161
Ag AGAP009759   92-TVTLVKGVPVPYTVEKHVPYTVEKHVPYPVKVPVPQPYPVVKHVPVHVKEYVKVPVHVPQPYPVEKKVPYPVHVPVDRPYPVKV-175
Ag AGAP008817   64-TVTIEKKIPVPYTVEKHVPYTVEKKVPYEVKVPIPQPYIVEKKVPVHYKEVVKYPVHVPAPYTVEKKVPVDKPYKVEVPKPYPV-147
Bm CPG12        86-TVTVVKKVPVPYPVEKHIPYPVEKKIPYPVKVHVPQPYPVVKHVPYPVKEIVKVPVHVPQPYPVEKKVPYPVHVPVDRPVPVKV-169
Bm CPG13        90-TITLVKGVPVPYAVDRPVPYPVEKHVPYPVKVAVPQPYEVVKHVPYHVKEYVKVPVHVPAPYPVEKKVPYPVHVPVDRPYPVKV-173
Bm CPG24       121-TVEVVKKVPVPYTVEKHVPYTVEKKVPYEVKVPVPQPYTVEKKVPFTVKEYVKYPVYVPEPYTVEKKVPYEVKVHVDKPYEVKV-204
Dp CPG12       125-TITVVKNVPVPYPVEKHIPYPVEKHIPVPVKVPVPQPYPVVKHVPYTVKELVKIPVHVPQPYPVEKKVPYPVHPVDRPVPVKV-208
Dp CPG24       124-TIEVIKKVPVPYTVEKHVPYTVEKKVPYEVKVPYPQPYTVEKRVPVNVKEYVKYPVHVPQPYVVEKKKVPYEVKVPVDKPYHVYV-208
Ob CPG11       105-TITTIKKVPVPYPVEKRIPYPVEKKVHYPVKVHVPHPYPVYKTVHFPVKETIKVPEYIKQPYPVTKHIPVEVKVHVDNPVPYKV-188
Ob CPG12        83-TVTIVKNVPVPYPVEKHIPYPVEKKIPYPVKVHVPQPYPVVKHVPYPVKEIVKVPVHVPQPYPVEKKIPYPVHPVDRPVPVKV-166
Ob CPG24       126-HVEVIKKVPVPYTVEKHVPYTVERKVPYEVKVPYPQPYTVEKKVPVTVKEYVKYPVYVPEPYTVEKKVPYEVKVHVDKPYEVKV-209
Pp CPG11        98-TITIVKKVPIPYPVEKHIPVPVEKHVPVPVKVHVPQPYPVVKTVHFPVKEYIKVPEYIKKPYPVTKHVPVPVEVHVDRPVPVKV-181
Pp CPG12        87-TITVVKNVPVPYPVEKHIPYPVEKQVPVPVKVPVPQPYPVVKHVPYPVKELVKVPVHVPQPYPVEKKVPYPVHVPVDRPVPVKV-170
Pp CPG24       112-TVEVIKKVPVPYTVEKHVPYTVEKKVPYEVKVPVPQPYTVEKKVPYTVKEYVKYPVYVPEPYTVEKKVPYEVKVPVDKPFEVKV-195
Px CPG11        98-TITIVKKVPLPYPVEKHIPVPVEKHVPVPVKVHVPQPYPVVKTVHFPVKEYIKVPEYIKKPYPVTKHVPVPVEVHVDRPVPVKV-181
Px CPG12        97-TITVVKNVPVPYPVEKHIPYPVEKQVPVPVKVPVPQPFPVVKHVPYPVKELVKVPVHVPQPYPVEKKVPYPVHPVDRPVPVKV-180
Px CPG13        92-TITLVKGVPVGVPVDRPVPYPVEKHVPYPVKVPVPQPYEVVKHVPVHVKEYVKVPVHVPAPYPVEKKVPYPVHPVDRPYPVKV-175
```



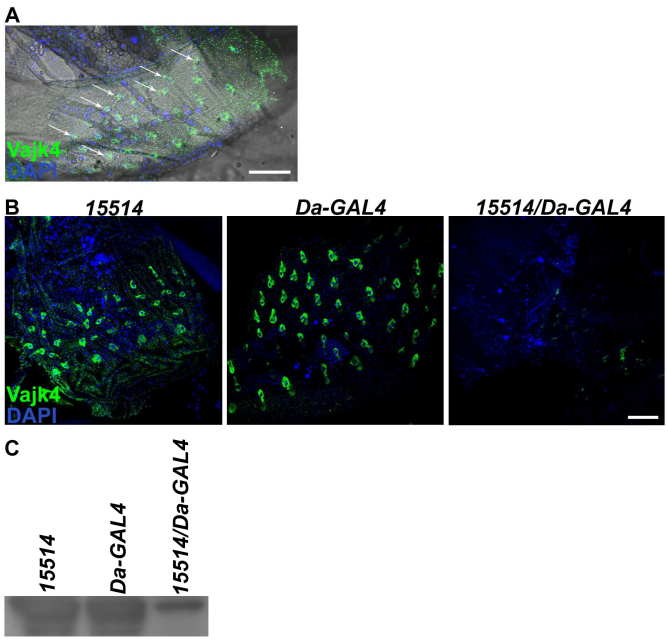
ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

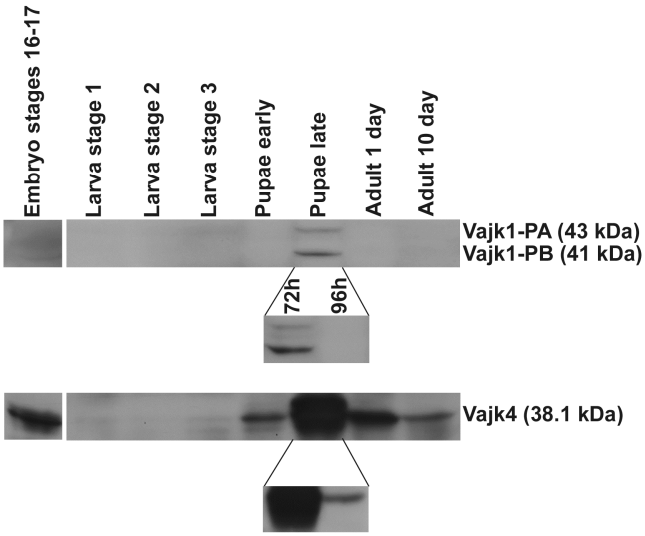


ACCEPTED MANUSCRIPT



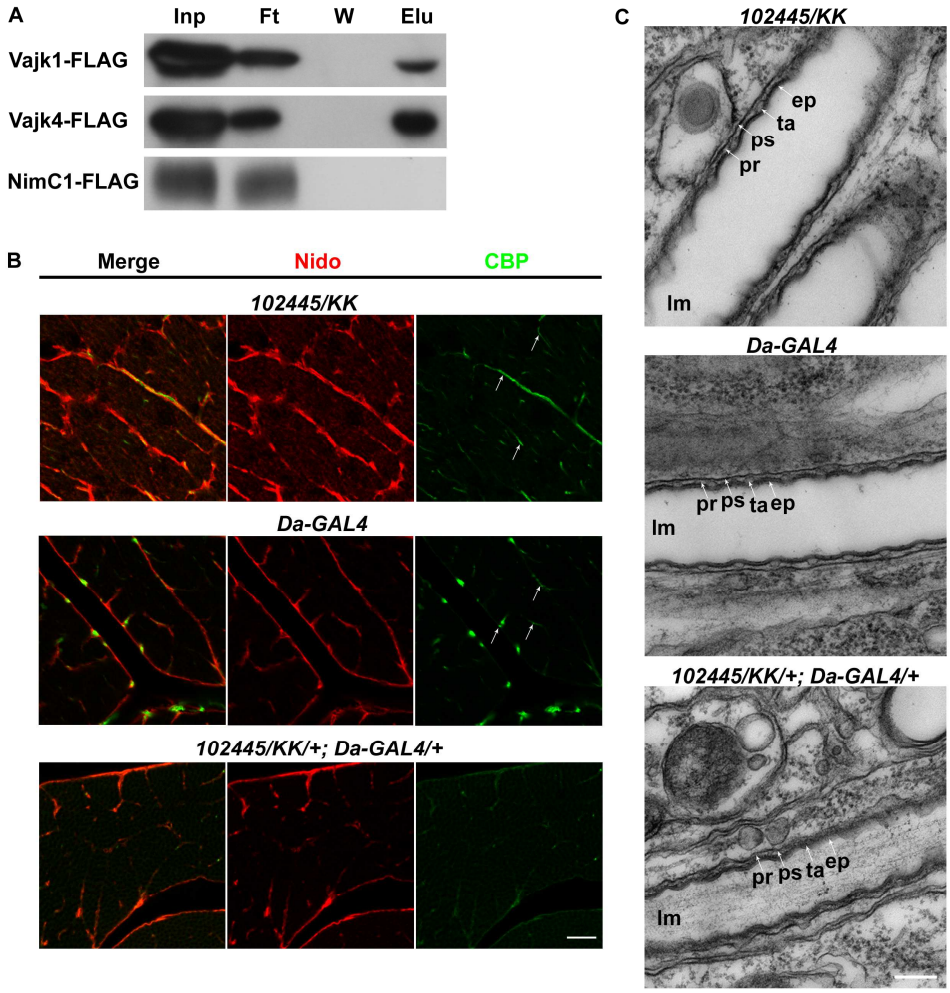
ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

- The Nimrod gene cluster is the largest synthenic unit of the *Drosophila* genome.
- The *Nim* and *Hemese* genes of the cluster are involved in the regulation of innate immunity and cell adhesion.
- Vajk proteins are encoded by genes in the Nimrod cluster and are expressed in cuticular structures and possibly contribute to barrier function.
- Genes of the Nimrod region may regulate different layers of defense in a concerted action.

ACCEPTED MANUSCRIPT