Repeated subfunctionalization of a modular antimicrobial peptide gene for neural function

M.A. Hanson* and B. Lemaitre

Global Health Institute, School of Life Science, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
* Corresponding author: M.A. Hanson (mark.hanson@epfl.ch)
ORCID IDs:
Hanson: https://orcid.org/0000-0002-6125-3672
Lemaitre: https://orcid.org/0000-0001-7970-1667

Abstract

Antimicrobial peptides (AMPs) are host-encoded antibiotics that combat invading pathogens. However recent studies have highlighted roles for AMPs in neurological contexts suggesting functions for these defence molecules beyond infection. Here we characterize the evolution of the *Drosophila Baramicin* (*Bara*) AMP gene family. During our immune study characterizing the *Baramicins*, we recovered multiple *Baramicin* paralogs in *Drosophila melanogaster* and other species, united by their N-terminal IM24 domain. Strikingly, some paralogs are no longer immune-induced. A careful dissection of the *Baramicin* family's evolutionary history indicates that these non-immune paralogs result from repeated events of duplication and subsequent truncation of the coding sequence from an immune-inducible ancestor. These truncations leave only the IM24 domain as the prominent gene product. Using mutation and targeted gene silencing, we demonstrate that two such genes are adapted for function in neural contexts in *D. melanogaster*, and show enrichment in the head for independent *Baramicin* genes in other species. The *Baramicin* evolutionary history reveals that the IM24 *Baramicin* domain is not strictly useful in an immune context. We thus provide a case study for how an AMP-encoding gene might play dual roles in both immune and non-immune processes via its multiple peptide products. We reflect on these findings to highlight a blind spot in the way researchers approach AMP research in in vivo contexts.

Introduction

Antimicrobial peptides (AMPs) are immune effectors best known for their role in defence against infection. These antimicrobials are commonly encoded as a polypeptide including both pro- and mature peptide domains (Zanetti 2005; Hanson and Lemaitre 2020). AMP genes frequently experience events of duplication and loss (Wang and Zhu 2011; Vilcinskas et al. 2013; Sackton et al. 2017; Hanson, Lemaitre, et al. 2019) and undergo rapid evolution at the sequence level (Tennessen 2005; Jiggins and Kim 2007; Hellgren et al. 2010; Halldórsdóttir and Árnason 2015;

Hanson et al. 2016; Chapman et al. 2019). The selective pressures that drive these evolutionary outcomes are likely the consequence of host-pathogen interactions (Unckless et al. 2016). However AMPs and AMP-like genes in various species have also recently been implicated in various non-immune roles in flies, nematodes, and emerging evidence in humans. These new contexts suggest that the evolutionary forces acting on AMP genes may not be driven strictly by trade-offs in host defence, but rather by conflicts between roles in immunity and other non-immune functions.

For instance, Diptericins are membrane-disrupting antimicrobial peptides of flies (Diptera) that are required for defence against infection by *Providencia* bacteria (Unckless et al. 2016; Hanson, Dostálová, et al. 2019). It was therefore surprising that the *D. melanogaster* gene *Diptericin B (DptB)* affects memory processes (Barajas-azpeleta et al. 2018). In this study, *DptB* derived from the fly fat body (analogous to the mammalian liver) regulated the ability of the fly to form long-term memory associations (Barajas-azpeleta et al. 2018). Another AMP-like gene, nemuri, regulates fly sleep and promotes survival upon infection (Toda et al. 2019). Studies in nematodes have also shown that an immune-induced polypeptide (NLP-29) binds to a G-protein coupled receptor (NPR-12) triggering neurodegeneration through activation of the NPR-12-dependent autophagy pathway (Lezi et al. 2018), and injury triggers epidermal AMPs including NLP-29 to promote sleep (Sinner et al. 2021). *Drosophila* AMPs have also recently been shown to regulate behaviours after seeing parasitoid wasps (Ebrahim et al. 2021), during feeding with different bacteria (Kobler et al. 2020), or following infection (Hanson et al. 2021). In humans, the *Cathelicidin* gene encodes the AMP LL-37, which is implicated in glia-mediated neuroinflammation and Alzheimer's disease (Lee et al. 2015; De Lorenzi et al. 2017), alongside evidence of Alzheimer's being an infectious syndrome (Dominy et al. 2019); though the importance of this process is debated (Abbott 2020). Notably, AMPs share a number of properties with classic neuropeptides (Brogden et al. 2005), further muddying the distinction between peptides of the immune and nervous systems.

We recently described a novel antifungal peptide gene of *Drosophila melanogaster* that we named *Baramicin A* (*BaraA*) (Hanson et al. 2021). A unique

aspect of *BaraA* is its precursor protein structure, which encodes a polypeptide cleaved into multiple mature products by interspersed furin cleavage sites. The use of furin cleavage sites to produce two mature peptides from a single polypeptide precursor is widespread in animal AMP genes (Gerdol et al. 2020; Hanson and Lemaitre 2020). However, *BaraA* represents an exceptional case as multiple tandem repeat peptides are produced from the translation of a single coding sequence, effectively resembling a "protein-based operon"; this tandem repeat structure has also been found in two other AMPs of bees and flies (Casteels-Josson et al. 1993; Hanson et al. 2016). The immature precursor protein of *D. melanogaster BaraA* encodes three types of domains: an IM24 domain, three tandem repeats of IM10-like domains, and an IM22 domain. *BaraA* mutants are susceptible to infection by fungi, and *in vitro* experiments suggest the *BaraA* IM10-like peptides have antifungal activity (Hanson et al. 2021). The other *Baramicin* domains encoding IM22 and IM24 remain uncharacterized. Curiously, *BaraA* deficient flies also display an erect wing behavioural phenotype upon immune stimulation even in the absence of infection, suggesting that *BaraA* products could have non-microbial targets (Hanson et al. 2021).

In this study, we describe the evolution of the Drosophilid *Baramicin* gene family. Three unique *Baramicin* genes (*BaraA, B, and C*) are present in the genome of *D. melanogaster*. Surprisingly, only *BaraA* is immune-induced, while *BaraB* and *BaraC* are enriched in the nervous system. Both *BaraB* and *BaraC* have truncations compared to the ancestral *Baramicin* gene, which focuses these genes towards producing the *Baramicin* IM24 domain. We found similar truncations in other species, and upon checking their patterns of expression, realized these overt gene structure changes correlate with loss of immune expression and enrichment in the nervous system. By resolving the genomic synteny of the various *Baramicin* genes in different species, we confirmed that these repeated truncations focusing on IM24 production stem from independent events (convergent evolution). The exaggerated 'protein operon' polypeptide nature of *Baramicin* draws attention to the unique roles that different mature peptides of AMP-encoding genes can play. Careful

attention paid to the multiple peptide products of AMP genes could explain how these immune effectors contribute to both immune and neurological processes.

Results

Baramicin is an ancestral immune effector

The *Baramicin A* gene was only recently described as encoding antifungal effectors by our group (Hanson et al. 2021), and another recent study also confirmed *Baramicin*'s important contribution to Toll immune defence (Huang et al. 2020). These initial characterizations were done only in *D. melanogaster*, and only focused on one *Baramicin* gene. We will therefore first provide a basic description of the immune *Baramicins* of other species and also the larger *Baramicin* gene family of *D. melanogaster* to establish that this is a classically immune gene family, and that deviations from immune function are derived.

In *D. melanogaster, BaraA* is regulated by the Toll immune signalling pathway (Huang et al. 2020; Hanson et al. 2021). Using BLAST, we recovered *BaraA*-like genes encoding each Baramicin peptide (IM24, IM10-like, and IM22) across the genus *Drosophila* and in the outgroup *Scaptodrosophila lebanonensis*. We performed infection experiments to confirm that these *BaraA*-like genes were immune-inducible by infecting the diverse species *D. melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, and *D. neotesteacea* (last common ancestor ~63mya (Tamura et al. 2004)) with *Micrococcus luteus* and *Candida albicans*, two microbes that stimulate the Toll pathway (**Fig.** 1A). In all five species, *BaraA*-like genes were immune-induced (**Fig.** 1B-F). We therefore confirm the ancestral *Baramicin* was an immune-induced gene.

The four D. melanogaster Baramicins: BaraA1, BaraA2, BaraB and BaraC

In *D. melanogaster*, we recovered four *Baramicin* genes. First, we realized that a duplication of *BaraA* is actively segregating in wild flies (**Fig. 2**A). The

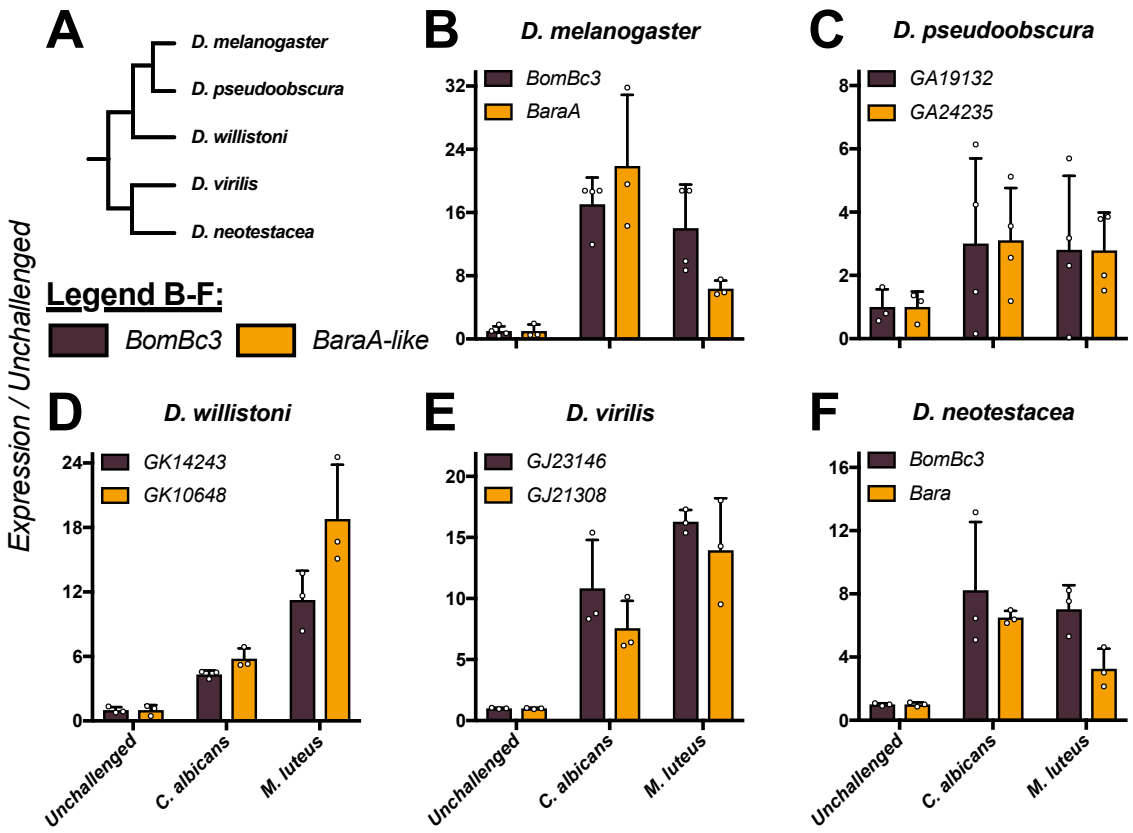

Figure 1: The ancestral *BaraA* gene was immune-induced. A) Cladogram of species used in B-F. B-F) Expression of *BomBc3* control genes (brown) or *BaraA-like* genes (orange) in diverse *Drosophila* species upon infection. In all cases, both *BomBc3* and *BaraA-like* genes are induced upon infection by either *C. albicans* yeast or *M. luteus* bacteria.

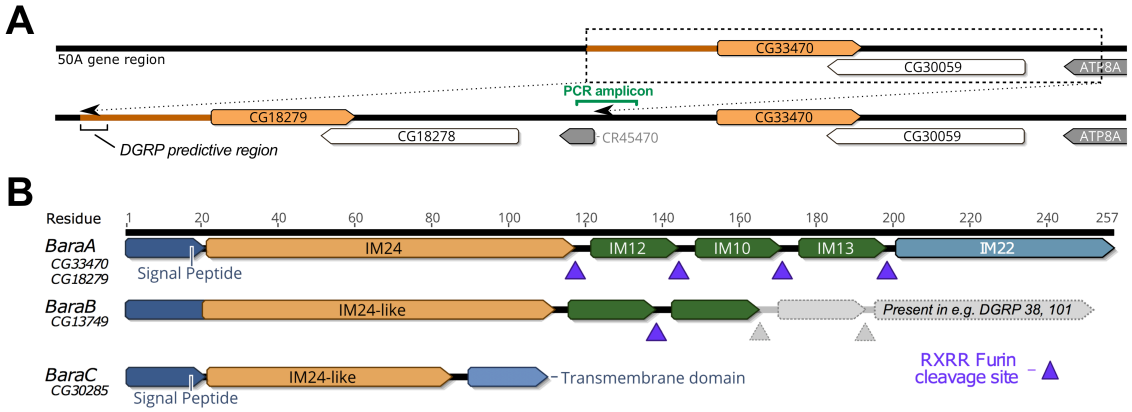

Figure 2: The *D. melanogaster Baramicin* genes. A) Schematic of the *BaraA* duplication. Using a PCR assay spanning the duplication-specific locus (PCR amplicon), we confirmed *BaraA* copy number is variable in various lab strains (Table S1) and wild-caught flies (Supplementary data file 1). B) *D. melanogaster* encodes two other *Baramicin* genes that we name *BaraB* and *BaraC*. These paralogs differ markedly in their precursor protein structure, which is truncated relative to *BaraA* including an actively segregating truncation event in the wild in *BaraB* (protein lengths implied by greyed out region).

D. melanogaster R6 genome assembly encodes two 100% identical *BaraA* genes (*CG33470* and *CG18279*, *BaraA1* and *BaraA2* respectively). We screened 132 DGRP lines for the *BaraA* duplication event, finding only ~14% (18/132) of strains were PCR-positive for two *BaraA* copies (supplementary data file 1). Perhaps as a consequence of the identical sequences of these two genes, this genome region is poorly resolved in RNA sequencing studies and the Drosophila Genetic Reference Panel (DGRP, see **Fig.** S1) (Mackay et al. 2012; Leader et al. 2018). Because this region is poorly resolved, it is unclear if our PCR assay might be sensitive to cryptic sequence variation. However our PCR screen nevertheless confirms that this region is variable in the wild, and we additionally note that common fly strains seem to differ in their *BaraA* copy number (Table S1), where extra gene copies correlated with increased expression after infection (see (Hanson et al. 2021) S10 Fig).

We also recovered two paralogous *Baramicin* genes in *D. melanogaster* through reciprocal BLAST searches: *CG13749* and *CG30285*, which we name *BaraB* and *BaraC* respectively (**Fig.** 2B). The three *Baramicin* gene loci are scattered on the right arm of chromosome II at cytological positions 44F9 (*BaraB*), 50A5 (*BaraA*), and 57F8 (*BaraC*). These paralogous *Baramicins* are united by the presence of the IM24 domain. In the case of *BaraB*, we additionally recovered a frameshift mutation (2R_4821599_INS) causing a premature stop segregating in the DGRP leading to the loss of IM13 and IM22 relative to the *BaraA* gene structure (**Fig. 2**B); this truncation is present in the Dmel_R6 genome assembly, but many DGRP strains encode a CDS with either a standard (e.g. DGRP38) or extended (e.g. DGRP101) IM22 domain (a DGRP *BaraB* alignment is provided in supplementary data file 2). Moreover, in contrast to BaraA, the initial IM10-like peptide of BaraB no longer follows a furin cleavage site, and encodes a serine (RSXR) in its IM10-like motif instead of the universal proline (RPXR) of *BaraA*-like IM10 peptides across the genus. Each of these mutations prevents the secretion of classical IM10-like and IM22 peptides by *BaraB*. Finally, *BaraC* encodes only IM24 tailed by a transmembrane domain at the C terminus (TMHMM v2.0 (Krogh et al. 2001)), and thus lacks both the IM10-like peptides and IM22 (**Fig. 2**B).

BaraB and BaraC are not immune-inducible

BaraA is strongly induced following microbial challenge (Fig. 1), being predominantly regulated by the Toll pathway with a minor input from the Immune Deficiency (Imd) pathway (Huang et al. 2020; Hanson et al. 2021). We therefore assayed the expression of *BaraB* and *BaraC* in wild-type flies, and also flies with defective Toll (spz^{rm7}) or Imd (Rel^{E20}) signalling to see if their basal expression relied on these pathways. Surprisingly, neither gene was induced upon infection regardless of microbial challenge (**Fig. 3**A and **Fig.** S2A-B). However *BaraC* levels were consistently reduced in spz^{rm7} mutants regardless of treatment (cumulative data in **Fig.** S2C, p = .005), suggesting *BaraC* basal expression is affected by Toll signalling. We next generated a novel time course of development from egg to adult to monitor the expression of the three *Baramicin* genes. We found that expression of all genes increased over development and reached their highest level in young adults (**Fig.** 3B). Of note, *BaraB* expression approached the lower limit of our assay's detection sensitivity at early life stages. However *BaraB* was robustly detected beginning at the pupal stage, indicating it is expressed during metamorphosis. *BaraC* expression also increased markedly between the L3 larval stage and pupal stage.

Here we reveal that *BaraA* is part of a larger gene family. While the *BaraA* gene was first described as an immune effector, the two *Baramicin* paralogs *BaraB* and *BaraC* are not induced by infection in *D. melanogaster*. Both *BaraB* and *BaraC* first see increased expression during pupation, and are ultimately expressed at their highest levels in adults.

Dmel\BaraB is required in the nervous system over the course of development

A simple interpretation of the truncated gene structure and low levels of *BaraB* expression is that this gene is undergoing pseudogenization. Indeed, AMP gene pseudogenization is common in insects including *Drosophila* (Quesada et al. 2005; Rolff and Schmid-Hempel 2016; Hanson, Lemaitre, et al. 2019). To explore

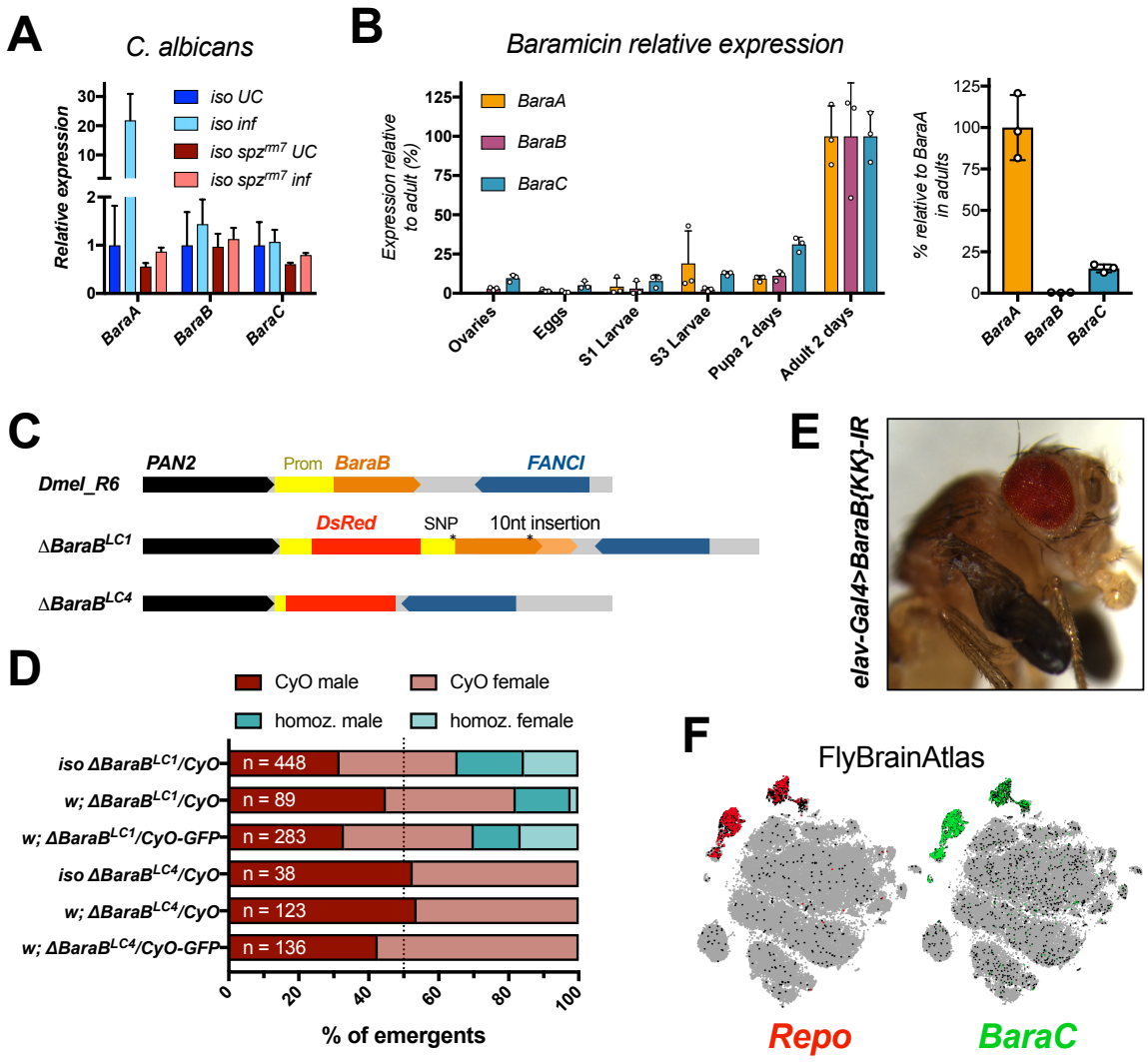

Figure 3: *D. melanogaster* non-immune *Baramicins* have neural functions. A) Only *BaraA* is immune-induced. *BaraB* and *BaraC* do not respond to infection, though basal *BaraC* expression relies on Toll signalling (Fig. S3C). B) Time series of whole animal *Baramicin* expression over the course of development. Expression values are normalized within each gene, with expression in the adult set to 100% (left panel). For context, normalizing each gene to *BaraA* in adults shows that *BaraC* and especially *BaraB* expression is much lower at a whole fly level (right panel). C) *ΔBaraB* mutations generated in this study. *ΔBaraB^{LC1}* is an incidental hypomorph with reduced *BaraB* expression (Fig. S3A), while *ΔBaraB^{LC4}* encodes the intended genetic knock in of a DsRed cassette. Both express DsRed in the eyes, ocelli, and abdomen. D) Partial lethality of *ΔBaraB^{LC1}* hypomorphs or complete lethality of *ΔBaraB^{LC4}* null flies in varied genetic backgrounds. E) Example of a nubbin-like wing, which is phenocopied by *BaraB* gene silencing using the neural driver *elav-Gal4*. F) The *BaraC* gene almost perfectly matches the expression pattern of glia specific *Repo*-expressing cells in single cell RNA sequencing of the adult fly brain (FlyBrainAtlas (Davie et al. 2018)).

BaraB function, we used two mutations for *BaraB* (*ΔBaraB^{LC1}* and *ΔBaraB^{LC4}*, generously gifted by S.A. Wasserman). These mutations were made using a CRISPR

double gRNA approach to replace the *BaraB* locus with sequence from the pHD-DsRed vector. The *ΔBaraB^{LC1}* and *ΔBaraB^{LC4}* mutations differ in their ultimate effect, as *ΔBaraB^{LC1}* is an incidental insertion of the DsRed cassette in the promoter of the gene. This disruption reduces gene expression, resulting in a hypomorph state (**Fig. S3**A). The *ΔBaraB^{LC4}* mutation however deletes the locus as intended, leading to *BaraB* null flies (**Fig. 3**C).

We further introgressed both *ΔBaraB* mutations into the DrosDel isogenic background (referred to as *iso*) for seven generations according to Ferreira et al. (Ferreira et al. 2014). At the same time, we combined the original *ΔBaraB* chromosomes with a CyO-GFP balancer chromosome in an arbitrary genetic background to distinguish homozygous/heterozygous larvae. In all cases, *ΔBaraB^{LC4}* homozygotes were entirely lethal during larval development, whereas the hypomorphic *ΔBaraB^{LC1}* flies allowed for homozygous adults to emerge (**Fig. 3**D). We further assessed *ΔBaraB^{LC1}* hypomorph viability using crosses between *ΔBaraB^{LC1}/CyO* heterozygous females and *ΔBaraB^{LC1}* homozygous males, which showed reduced viability and was exacerbated by rearing at 29°C (**Fig. S3**B). Using our *CyO-GFP* reporter to track genotypes in larvae revealed that the major lethal phase occurs primarily in the late larval and pupal stages (**Fig. S3**C-F), agreeing with a role for *BaraB* in larvae/pupae previously suggested by increased expression at this stage. Some emergent flies also exhibited locomotor defects, and/or a nubbin-like wing phenotype (FlyBase: FBrf0220532 and e.g. in **Fig. 3**E) where the wings were stuck in a shrivelled state for the remainder of the fly's lifespan. However, a plurality of *ΔBaraB^{LC1}* homozygotes successfully emerged, and unlike their siblings, had no immediate morphological or locomotory defects. The lifespan of morphologically normal *iso ΔBaraB^{LC1}* adults is nevertheless significantly shorter compared to wild-type flies and *iso ΔBaraB^{LC1}/CyO* siblings (**Fig. S4**G). We confirmed these developmental defects using ubiquitous gene silencing with *Actin5C-Gal4 (Act-Gal4)* to drive two *BaraB* RNAi constructs (*TRiP-IR* and *KK-IR*). Both constructs resulted in significant lethality and occurrence of nubbin-like wings (**Table** S2). Genomic deficiency crosses also confirmed significantly reduced numbers of eclosing *BaraB*-deficient flies at 25°C (n = 114, χ^2 p < .001) and 29°C (n = 63, χ^2 p <

.001) (**Fig. S3**H). Thus full gene deletion is lethal in the larval/pupal transition stage, and *BaraB* hypomorph flies suffer significant costs to fitness during development, and have reduced lifespan even following successful eclosion.

These data demonstrate a significant cost of *BaraB* disruption. While whole-fly *BaraB* expression is low, these results suggest that *BaraB* is not pseudogenized, and instead performs an integral developmental role. The fact that there is a bimodal outcome in hypomorph-like $\Delta BaraB^{LC1}$ adults (either severe defects or generally healthy) suggests *BaraB* is involved in passing some checkpoint during larval/pupal development. Flies deficient for *BaraB* may be more likely to fail at this developmental checkpoint, resulting in either lethality or developmental defects.

The Baramicin paralogs BaraB and BaraC are expressed in the nervous system

We next sought to determine in which tissue(s) *BaraB* is required. A previous screen using neural RNA interference highlighted *BaraB* for lethality effects (n = 15) (Neely et al. 2010). Given this preliminary result, and alongside our observed *BaraB* mutant locomotory defects, we started by silencing *BaraB* in the nervous system at 25°C or at 29°C for greater efficiency using the pan-neural *elav-Gal4* driver both the *TRiP-IR* and *KK-IR BaraB-IR* lines. We additionally combined this approach with *UAS-Dicer2* (*Dcr2*) to further strengthen gene silencing as used previously (Neely et al. 2010). In the event there was no lethality, it was expected that emerging *elav>TRiP-IR* flies would follow simple mendelian inheritance. However both *elav>TRiP-IR* and *elav>Dcr2, TRiP-IR* resulted in partial lethality and occasional nubbin-like wings (χ^2 $p < .02$, **Table** S2). Crosses using *KK-IR* used homozygous flies, and so we did not assess lethality using mendelian inheritance. However using this construct, no adults emerged when *elav>Dcr2, KK-IR* flies were reared at 29°C. Rare emergents (N = 11 after three experiments) occurred at 25°C, all of which bore nubbin-like wings. Using *elav-Gal4* at 29°C without *Dcr2*, we observed greater numbers of emerging adults, but 100% of flies had nubbin-like wings (**Fig. 3**E, **Table** S2). Finally, *elav>KK-IR* flies at 25°C suffered both partial lethality and nubbin-like wings, but normal-winged flies began emerging (χ^2 $p < .001$, **Table** S2).

This analysis indicates that *BaraB* is expressed in the nervous system, and this expression readily explains both the lethality and nubbin-like wing phenotypes. Moreover, we observed a consistent spectrum of developmental defects using *elav-Gal4>BaraB-IR* wherein strength of gene silencing correlates with severity of lethality and wing defect frequency. We additionally investigated the effect of *BaraB* RNAi using Gal4 drivers in non-neural tissues including the fat body (*c564-Gal4*), hemocytes (*hml-Gal4*), the gut (esg-Gal4), malpighian tubules (MyO-Gal4), the wing disc (*nubbin-Gal4*), and in myocytes (*mef2-gal4*) to no effect. We also screened neural drivers specific for glia (*Repo-Gal4*), motor neurons (*D42-, VGMN-,* and *OK6-Gal4*), and a recently-made *BaraA-Gal4* driver that is expressed in the larval ventral nervous system (Hanson et al. 2021). However all these *Gal4>BaraB-IR* flies were viable and never exhibited overt morphological defects.

We also screened for effects of *BaraC* disruption using ubiquitous *Act-Gal4* and neural *elav-Gal4>Dcr2* for developmental defects. However neither driver produced overt phenotypes in morphology or locomotor activity (not shown). Tissue-specific transcriptomic data indicate that *BaraC* is expressed in various neural tissues including the eye, brain, and the thoracic abdominal ganglion (**Fig. S4**A), but also the hindgut and rectal pads pointing to a complex expression pattern (Hammonds et al. 2013; Leader et al. 2018). We next searched FlyBrainAtlas (Davie et al. 2018) to narrow down which neural subtypes *BaraB and BaraC* were expressed in. *BaraB* expressing cells were few and showed only low expression in this dataset. However *BaraC* was robustly expressed in all glial cell types, fully overlapping the glia marker *Repo* (**Fig. 3**F). To confirm the observation that *BaraC* was expressed in glia, we compared the effects of *BaraC* RNA silencing (*BaraC-IR*) using *Act-Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia) drivers on *BaraC* expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that of control flies (**Fig. S4**B). By comparison *elav-Gal4* reduced *BaraC* expression to ~63% that of controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls (overall controls vs. neural/glia-IR, p = .002).

Collectively, our results support the notion that *BaraC* is expressed in the nervous system, and are consistent with *BaraC* expression being most localized to glial cells.

Extensive genomic turnover of the Baramicin gene family

Our results thus far show that *BaraA*-like genes are consistently immune-induced in all *Drosophila* species (**Fig.** 1), however the two paralogs *Dmel\BaraB* and *Dmel\BaraC* are not immune-induced, and are truncated in a fashion that deletes some or all of the antifungal IM10-like peptides (**Fig. 2**B). These two *Baramicins* are now enriched in the nervous system (**Fig. 3**E-F). In the case of *BaraB*, a role in the nervous system is evidenced by severe defects recapitulated using pan-neural RNA silencing. In the case of *BaraC*, nervous system expression is evidenced by a clear overlap with *Repo*-expressing cells.

While *BaraA*-like genes are conserved throughout the genus *Drosophila*, *BaraB* is conserved only in Melanogaster group flies, and *BaraC* is found only in Melanogaster and Obscura group flies, indicating that both paralogs stem from duplication events of a *BaraA*-like ancestor (**Fig. 4**). To determine the ancestry of each *D. melanogaster Baramicin* gene, we traced their evolutionary history by analyzing genomic synteny through hierarchical orthologous groups (Train et al. 2019). Ancestry tracing revealed that these three loci ultimately stem from a single-locus ancestor encoding only one *Baramicin* gene that resembled *Dmel\BaraA* (**Fig. 4**A). This is evidenced by the presence of only a single *BaraA*-like gene in the outgroup *S. lebanonensis*, and also in multiple lineages of the subgenus Drosophila (**Fig. 4**B). Indeed, the general *BaraA* gene structure encoding IM24, tandem repeats of IM10-like peptides, and IM22 is conserved in *S. lebanonensis* and all *Drosophila* species (**Fig. 4**C). On the other hand, the *Dmel\BaraC* gene comes from an ancient duplication restricted to the subgenus Sophophora, and *Dmel\BaraB* resulted from a more recent duplication found only in the Melanogaster group (**Fig. 4**B).

We originally recovered outgroup *Baramicins* assayed for immune induction (**Fig.** 1) through reciprocal BLAST searches. However following genomic synteny analysis, we realized that the *D. willistoni BaraA*-like gene *Dwil\GK10648* is syntenic

with the *Dmel\BaraC* locus (**Fig. 4**A), yet this gene is immune-induced (**Fig.** 1D) and retains a *BaraA*-like gene structure (**Fig. 4**C). On the other hand, *Dwil\GK10645* is found at the locus syntenic with *BaraA*, but has undergone an independent truncation to encode just an IM24 peptide (similar to *Dmel\BaraC*). Thus these two *D. willistoni* genes have evolved similar to *D. melanogaster BaraA/BaraC*, but in a vice versa fashion. This suggests a pattern of convergent evolution with two key points: **i)** the duplication event producing *Dmel\BaraA* and *Dmel\BaraC* originally copied a full-length *BaraA*-like gene to both loci, and **ii)** the derivation of an IM24-specific gene structure has occurred more than once (*Dmel\BaraC* and *Dwil\GK10645*). Indeed, another independent IM24-specific *Baramicin* gene is present in *D. virilis* (*Dvir\GJ25897*), which is a direct sister of the *BaraA*-like gene *Dvir\GJ21309* (the signal peptides of these genes is identical at the nucleotide level, and see **Fig. 4**C). Thus *Baramicins* in both *D. willistoni* and *D. virilis* have convergently evolved towards an IM24-focused protein structure resembling *Dmel\BaraC*. We checked the expression of these truncated *Baramicins* in each species upon infection. As was the case for *Dmel\BaraC*, neither gene is immune-induced (**Fig. S5**A-C). Given the glial expression of *Dmel\BaraC,* we reasoned that the heads of adult flies (rich in nerve tissue) should be enriched in *BaraC* compared to whole animals. Indeed we saw a significant enrichment of *BaraC* in the heads of *D. melanogaster* males compared to whole flies, which was not the case for *BaraA* (**Fig. S5**D). When we checked the heads of *D. willistoni* and *D. virilis*, we indeed saw a consistent and significant enrichment in the head for the IM24-specific genes *Dwil\GK10645* and *Dvir\GJ25897*, while *BaraA*-like genes were more variable in expression (**Fig. S5**E-F).

Thus, multiple independent IM24-specific *Baramicins* are not immune induced and are more specifically enriched in the head. In the case of *Dmel\BaraC*, this is likely due to expression in glia. Strikingly, we observe a parallel evolution of expression pattern and gene structure in *Baramicins* of *D. willistoni* and *D. virilis*. These expression data are summarized in **Fig. 4**C. Genomic synteny shows the gene structure and immune expression of *BaraA* are the ancestral state, and *Dmel\BaraB* and *Dmel\BaraC* are paralogs derived from independent duplication events.

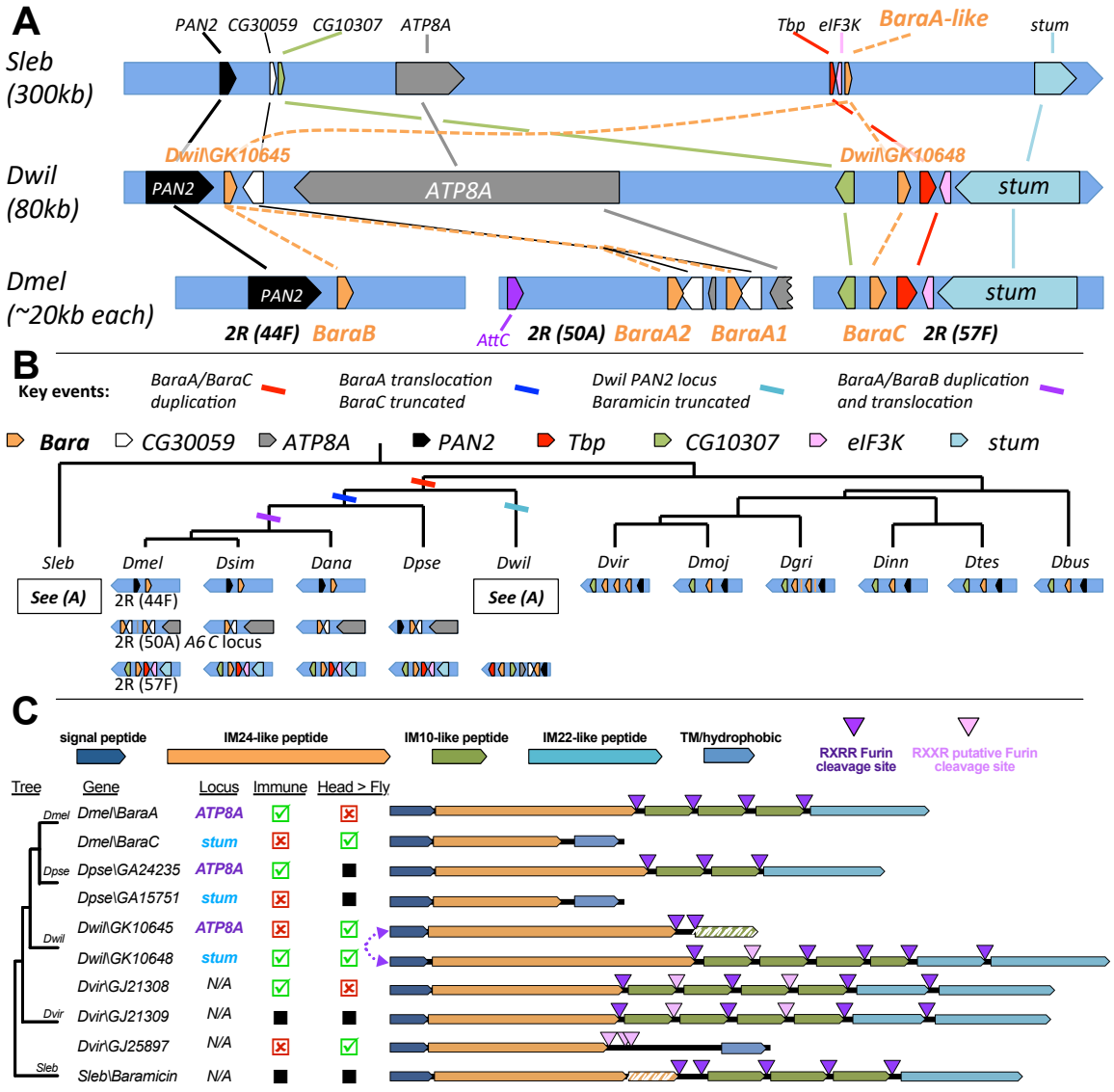

Figure 4: *Baramicin* evolutionary history. A) Detailed map of genomic neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D. melanogaster*, detailing inferred duplication, inversion, and translocation events. Gene names are given as found in *D. melanogaster*. B) Cladogram and genomic loci further detailing the series of events leading to the extant *Baramicin* loci of *Drosophila* species. Loci in *S. lebanonensis* and flies in the subgenus *Drosophila* encode only one *Baramicin* gene, indicating the ancestral drosophilid likely encoded only one *Baramicin*. C) IM24-specific *Baramicins* arose from convergent evolution both in gene structure and expression profile. Genomic loci are described here as *ATP8A* or *stum* reflecting prominent genes neighbouring the *Baramicins* (see Fig. 4A). Expression dynamics relating to immune-induction or enrichment in the head (checked boxes) are shown in Fig. 2 and Fig. S5. The *Baramicin* loci in *D. willistoni* are syntenic with *D. melanogaster*, but evolved in a vice versa fashion (purple arrow). The *D. virilis Baramicins GJ21309* and *GJ25897* are direct sister genes (100% identity at N-terminus).

Residue 29 in the IM24 domain evolves in lineage-specific fashions

Multiple independent *Baramicin* genes have lost both IM10-like and IM22 peptides, converge on loss of immune induction, and are enriched in the head. Taken together, these truncations and expression patterns suggest that the IM10-like peptides and IM22 are strictly useful during the immune response, consistent with a recently described antifungal role for IM10-like peptides (Hanson et al. 2021). Inversely, non-immune *Baramicin* genes have repeatedly and independently truncated to encode primarily IM24. We could not generate a reasonable model of the IM24 peptide conformation using Phyre2 (Kelley et al. 2015), QUARK, or TASSER protein modelling methodologies (Zhang et al. 2016). AlphaFold (Jumper et al. 2021) also has only low confidence estimates for the mature structure of *D. melanogaster Baramicins*. The IM24 domain unites the *Baramicin* gene family, making its apparent non-immune functional roles in *BaraB* and *BaraC* intriguing. Failing to model the protein, we next asked if we could highlight any residues in this traditionally immune peptide that might correlate with immune or non-immune gene lineages to gain insight into what governs the IM24-specific gene preference for neural expression.

To do this, we screened for positive selection (elevated non-synonymous mutation rate) in the IM24 domain using the HyPhy package implemented in Datamonkey.org (Delport et al. 2010) using separate codon alignments of *Baramicin* IM24 domains beginning at their conserved Q^1 starting residue. As is recommended with the HyPhy package (Delport et al. 2010), we employed multiple statistical approaches including Likelihood (FEL), Bayesian (FUBAR), and Count-based (SLAC) analyses to ensure patterns in selection analyses were robust to different methods of investigation. Specifically, we used locus-specific alignments (e.g. genes at the *stum* locus in **Fig. 4**B were all analyzed together) independent of overall gene structure to ensure IM24 evolution reflected locus-specific evolution. FEL, FUBAR, and SLAC site-specific analyses each suggest strong purifying selection in many residues of the IM24 domain (data in **supplementary data file 3**), agreeing with the general protein structure of IM24 being broadly conserved (**Fig.** 5A). However one residue (site 29) was consistently highlighted as evolving under positive selection

using each type of statistical approach for genes located at the Sophophora *ATP8A* locus (*BaraA* genes and *Dwil\GK10645*: p-adj < .05; **Fig. 5**A). This site is universally Proline in *Baramicin* genes located at the stum locus (*BaraC*-like) and in the outgroup *S. lebanonensis*, but is variable in both the *BaraA* (commonly Threonine) and *BaraB* (commonly Valine) lineages. Both the *S. lebanonensis* and the two *D. willistoni Baramicins* encode Proline at site 29 independent of gene structure, suggesting Proline is the ancestral state. We also note that two sites on either side of site 29 (site 27 and site 31) similarly diverge by lineage in an otherwise highly conserved region of the IM24 domain. FUBAR analysis (but not FEL or SLAC) similarly found evidence of positive selection at site 31 in the *BaraA* locus genes (p-adj = .026). Thus this neighbouring site could also be evolving in a non-random fashion. Similar analyses of the *BaraB* and *stum* loci *Baramicins* did not find evidence of site-specific positive selection.

While the structure of IM24 is unknown, HyPhy analysis highlights site 29 as a key residue in IM24 that diverged in *Baramicin* lineage-specific fashions. This ancestrally Proline residue has settled on a Threonine in most *BaraA*-like genes of Obscura and Melanogaster group flies, and a Valine in most *BaraB* genes, which are unique to the Melanogaster group.

Another IM24 domain in Baramicin lineages varies through relaxed selection

Visual inspection of aligned IM24 proteins makes it evident that the overall IM24 domain is broadly conserved, except in sites 40-48 (**Fig. 5**A). This motif uniquely encodes the residues ^{40}HHASSPAD^{48} in *Dmel\BaraB*. Given the severe cost of *BaraB* mutation, intriguingly the H^{40} and D^{48} residues are not found in any other *Baramicin* genes. The three C-terminal residues of this motif are also diagnostic for each gene lineage (*BaraA, BaraB,* and *BaraC* have RGE, PXE, or (S/N)GQ respectively; **Fig. 5**A). However even with additional branch-site selection analyses (aBSREL and BUSTED (Murrell et al. 2015)), we found no evidence of positive selection at the ^{40}HHASSPAD^{48} homologous domain (supplementary data file 3).

Thus while IM24 residues 40-48 are variable across lineages, this motif is not evolving with elevated non-synonymous change. We suspect instead that this motif

is diversifying due to relaxed selection as six of nine sites in the *BaraA* locus analysis failed to reach significance ($p < .05$) for purifying selection in e.g. SLAC analysis (supplementary data file 3). It is nevertheless striking that this region is so variable given the conservation of residues upstream and downstream of sites 40-48. This pattern should have implications on the IM24 protein functional motifs, which future protein folding investigations may decipher.

Overt IM24 structural change best explains loss of immune induction

Site 29 varies in lineage-specific fashions, encoding a derived Valine residue in most species' BaraB IM24 domains. If the Valine at site 29 explains the *BaraB* functional divergence relative to its sister *BaraA* lineage, this could suggest that *BaraB* has long functioned in a neural role common to most Melanogaster group flies. To this end, we performed infection experiments in diverse species across the Melanogaster group to see if their *BaraB* genes had similarly lost immune induction (see **Fig.** S6 for qPCR data). Surprisingly, we instead found that the non-immune expression of *Dmel\BaraB* is extremely recent, as Melanogaster sister species like *D. sechellia* and *D. mauritiana* nevertheless encode immune inducible *BaraB* loci (summary in **Fig.** 5B). However, we also found that *D. simulans BaraB* lacked immune induction, despite being most closely related to *D. sechellia*. This drew our attention instead to the overall protein structure of the various *BaraB* genes. A striking feature of the Dmel\BaraB protein is the absence of a signal peptide structure (**Fig.** 2B). Signal peptide sequence is conserved in all *Baramicin* lineages, except in *BaraB* of *D. melanogaster* and also *D. simulans* (last common ancestor ~3mya (Chakraborty et al. 2021)). Indeed despite *D. simulans* being more closely related to *D. sechellia* and *D. mauritiana*, both *Dmel\BaraB* and *Dsim\BaraB* encode a homologous N-terminus of parallel length (**Fig.** 5A). We also found that *D. yakuba BaraB* is not immune-responsive, but note that *D. yakuba* has an insertion upstream of residue 40 that elongates the IM24 domain (**Fig.** 5B black X boxes), and its sister species *D. erecta* encodes multiple indels and premature stops suggesting *BaraB* is pseudogenized in this lineage. While an insertion is also present in *D. sechellia BaraB* in the signal peptide, this is still predicted to allow secretion (SignalP 5.0), suggesting *Dsec\BaraB* is a functional immune protein.

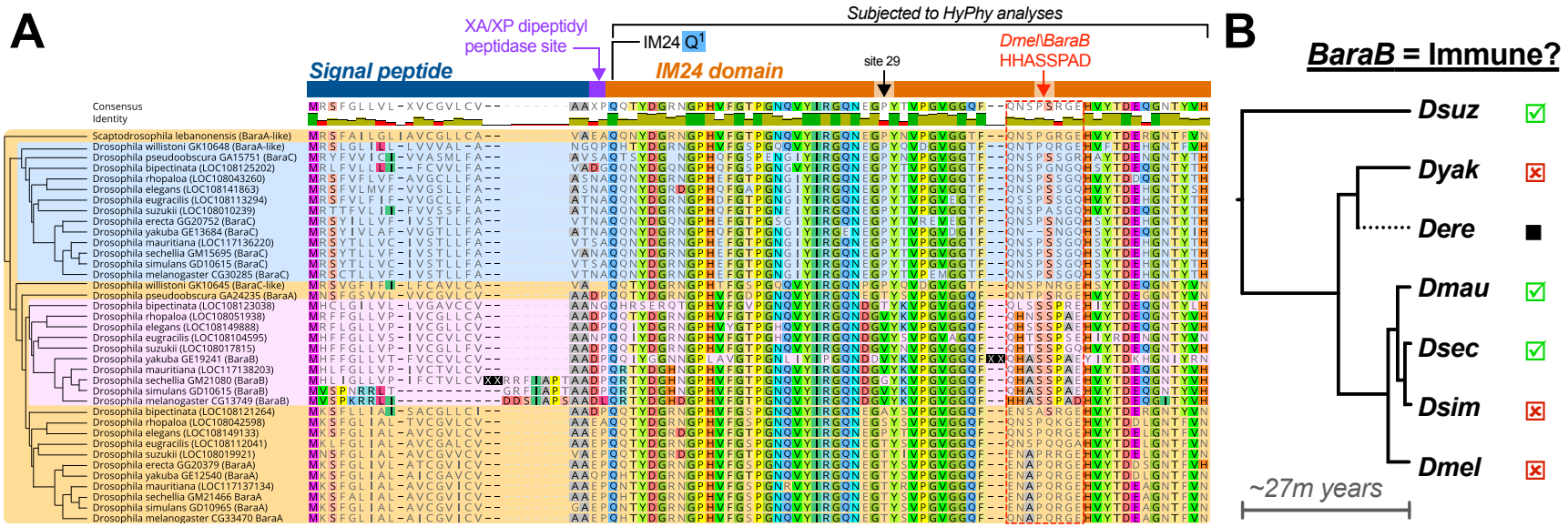

Figure 5: *BaraB* and IM24 rapid evolution. A) Evolution of the signal peptide and core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*. Arrows indicate site 29 and the *D. melanogaster* [40]HHASSPAD[48] domain. Insertion events in *D. yakuba* and *D. sechellia BaraB* are denoted as XX to save space. The *D. sechellia* signal peptide is predicted to remain functional (SignalP 5.0). The cladogram on the left shows genomic relatedness (by speciation and locus) independent of sequence similarity. Background colouring is included to show birth of novel *Baramicin* loci/lineages. B) *BaraB* immune or non-immune expression by phylogeny. The *D. erecta BaraB* gene is pseudogenized by multiple premature stop codons, and the *D. yakuba* gene is not immune-induced and encodes a 9-residue insertion in the IM24 peptide bordering the HHASSPAD domain (see XX site in A). However the *BaraB* genes of *D. suzukii*, and both *D. mauritiana* and *D. sechellia* remain inducible by infection, while the *BaraB* genes of *D. simulans* and *D. melanogaster* are not and are expressed at very low levels (Fig. S6). This pattern suggests that *BaraB* of *D. melanogaster* (and *D. simulans*) acquired its non-immune role only recently, and is correlated with the loss of the *BaraB* signal peptide.

Loss of the BaraB signal peptide is therefore more specifically associated with loss of immune expression in the Melanogaster species complex (*D. simulans, D. sechellia, D. mauritiana,* and *D. melanogaster*). The last common ancestor of *D. simulans, D. sechellia,* and *D. mauritiana* is estimated to be just ~250,000 years ago, and these species diverged from *D. melanogaster* ~3 million years ago (Chakraborty et al. 2021). The fact that *D. simulans* uniquely encodes this *Dmel\BaraB*-like sequence suggests it was either introgressed from one species to the other prior to the complete development of hybrid inviability, or reflects incomplete lineage sorting of this locus in the Melanogaster species complex. In either case, this points to an extremely young age for the novel function of *Dmel\BaraB* in the nervous system. This loss of the signal peptide also occurs alongside a segregating allele that truncates the mature *Dmel\BaraB* sequence (**Fig.** 2B), a pattern commonly found in *Baramicins* derived for neural expression (**Fig.** 4C). This reinforces the fact that the *Dmel\BaraB* gene had an immune function so recently that some wild flies still produce the immune-relevant IM10-like and IM22 Baramicin peptides despite neural expression of *BaraB*.

BaraB evolution therefore reinforces that the core driver of *Baramicin* functional divergence is not based on minor sequence changes, but rather correlates with overt protein structural change. *BaraB* has a mutation affecting secretion, while *BaraC* now encodes a transmembrane domain, which should cause it to insert itself into either its endogenous glial cell membrane or a neighbouring cell (e.g. a neuron). In both cases IM24 is preferentially expressed and localized to the nervous system.

Discussion

We recently showed that *BaraA* deletion causes infected flies to display an erect wing behavioural phenotype (Hanson et al. 2021). Notably, flies displayed erect wing even when heat-killed bacteria were injected, indicating this behaviour depends only on the triggering of the immune response in the absence of *BaraA*, and not on active infection. Thus *BaraA* likely interacts with some host target(s) to prevent this behaviour during the immune response.

Here we find the *Baramicin* IM24 domain has a predilection for interactions with the nervous system. We speculate that the immune-mediated production of IM24 by *BaraA* could protect the nervous system from autoimmune activity, which occurs in the absence of *BaraA* to cause erect wing display. A notable aspect of this hypothesis is it proposes that some peptides of AMP genes are responsible for microbe killing, while others are co-secreted with the intent of preventing autoimmune toxicity. For now this remains speculation, however it will be interesting to clarify the mature structure of IM24 and determine what partner(s) IM24 binds to. In this regard, we highlight site 29 as an important residue for IM24 function, and suggest that while residues 40-48 are variable, the sequence at this motif does not experience the same sort of evolutionary selection. One possibility to explain these evolutionary patterns is that site 29 is exposed in some way that an IM24 binding partner uses. Meanwhile residues 40-48 could act as a sort of linker between the two ends of the IM24 domain where the length of these residues is important, but the exact sequence is malleable. Indeed, we found that *D. yakuba BaraB* independently lost immune induction alongside an insertion at site 40.

Antimicrobial peptide genes have recently been implicated by a number of studies in neural functions, regulating processes like memory, sleep, taste aversion, behaviour, and neurodegeneration. The properties of these immune peptides share many features with classic neuropeptides, including cationic charge and amphipathicity (Brogden et al. 2005). Nevertheless it is unclear why AMPs can play dual roles in either immunity or neural function. By characterizing the evolution of the *Baramicin* gene family, we provide insight on how an ancestrally immune AMP gene has adapted itself for neural function on a repeated basis. The mechanism through which *Baramicin* achieves its neural effect is specific to the IM24 domain, as the antifungal IM10-like peptides and IM10-related IM22 peptide are consistently lost in *Baramicin* lineages now specific to neural expression. This realization is made possible by the exaggerated polypeptide structure of *Baramicins,* which focuses the interpretation on how different peptides of this modular protein structure may play roles in either neurology or immunity. Other AMP genes similarly encode polypeptides, but do not have so many tandem repeats of identical peptides and so

the polypeptide nature of their precursor protein is easily glossed over. The polypeptide nature of these genes is lacking from the current conversation on AMP involvements in the nervous system, despite AMP genes of fruit flies and other animals encoding furin-cleaved polypeptides (Hanson and Lemaitre 2020).

One human AMP recently implicated in chronic neuroinflammatory disease is the Cathelicidin LL-37 (Lee et al. 2015; De Lorenzi et al. 2017; Moir et al. 2018). Like *Baramicin*, the *Cathelicidin* gene family is unified by its N-terminal "Cathelin" domain. However to date no one has described antimicrobial activity of the Cathelin domain in vitro (Zanetti 2005). Instead, *Cathelicidin* research has focused almost exclusively on the mature peptide LL-37 at the C-terminus of mammalian *Cathelicidin* genes. Reflecting on *Baramicin* evolution and the implication of *Cathelicidin* in neurodegenerative diseases, what does the Cathelin domain do? While this study was conducted in fruit flies, we hope we have emphasized the importance of considering each peptide of AMP genes for in vivo function. This is relevant to neural processes even if the gene is typically thought of for its role in innate immunity. Indeed, recent studies of *Drosophila* AMPs have emphasized that in vitro activity does not always predict the interactions that these genes can have in vivo (Clemmons et al. 2015; Hanson, Dostálová, et al. 2019). Care should be taken not to conflate in vitro activity with realized in vivo function. Most studies focus on AMPs specifically in an immune role, but this is akin to 'looking for your keys under the streetlight.' To understand AMP functions in vivo, genetic approaches will be necessary that allow a more global view of gene function.

In summary, we characterize how an ancestral AMP-encoding gene has repeatedly evolved for neural expression by truncating its protein sequence to express just one peptide. It will be interesting to consider the functions of AMP genes in neural processes not simply at the level of the gene, but at the level of the mature peptides produced by that gene. Given the polypeptide of many AMP gene structures and commonalities between AMPs and neuropeptides, these canonical immune effectors may be adapted for neural function more often than appreciated.

Acknowledgements

We would like to thank Maria Litovchenko for advice, Ana Marija Jakšić for generously providing DGRP flies, Rob Unckless for stimulating discussion, Huang et al. (Huang et al. 2020) for collaborative cooperation, Brian McCabe for consultation, and Florent Masson, Hannah Westlake, and the anonymous reviewers and the editors at MBE for commentary on our initial manuscript. This research was supported by Sinergia grant CRSII5_186397 awarded to Bruno Lemaitre. The *BaraB*^{LC1} and *BaraB*^{LC4} mutations were graciously provided by Steven Wasserman and generated by Lianne Cohen, who we also thank for their critical involvement in characterizing *Baramicin A*.

Materials and Methods

DGRP population screening and bioinformatics analyses

Genomic sequence data were downloaded from GenBank default reference assemblies and Kim et al. (Kim et al. 2021), and DGRP sequence data from http://dgrp2.gnets.ncsu.edu/ (Mackay et al. 2012). Sequence comparisons and alignment figures were prepared using Geneious R10 (Kearse et al. 2012), Prism 7, and Inkscape. Alignments were performed using MUSCLE or MAFFT followed by manual curation, and phylogenetic analyses were performed to validate sequence patterns using the Neighbour Joining, PhyML, RaxML, and MrBayes plugins in Geneious. *BaraA* copy number screening was performed using primers specific to the duplication and *CG30059* control primers for DNA extraction (primers in supplementary data file 5). We found a significant correlation between *BaraA* PCR status and variant sites starting at 2R_9293471_SNP and extending to 2R_9293576_SNP (Pearson's correlation matrix: $0.0001 < \text{p-value} < 0.005$ at all nine sites), however the status of genetic variants at this site is poorly resolved and so we cannot be confident that our ~14% estimate for the *BaraA* duplication in the DGRP would hold true if long-read sequencing was employed. DGRP annotation of the *BaraA* locus in Fig. S1 was generated using the UCSC *D. melanogaster* DGRP2 genome browser. Selection analyses were performed using the HyPhy package implemented in datamonkey.org (Delport et al. 2010). Codon alignments of the IM24 domain used in **Fig. 5**A are included as a .fasta file in supplementary data file 3 alongside outputs from FEL, FUBAR, SLAC, and aBSREL selection analyses.

Fly genetics

The *BaraB*^{LC1} and *BaraB*^{LC4} mutations were generated using CRISPR with two gRNAs and an HDR vector by cloning 5' and 3' region-homologous arms into the pHD-DsRed vector, and consequently *ΔBaraB* flies express DsRed in their eyes, ocelli, and abdomen. The following PAM sites were used for CRISPR bordering the *BaraB* region. Slashes indicate the cut site: 5': GCGGGCAACAGATGTGTTCA/GGG 3': GTCCATTGCTTATTCAAAAA/TGG. These mutants were generated in the laboratory of Steve Wasserman by Lianne Cohen, who graciously allowed their use in this study. All fly stocks including Gal4 and RNAi lines are listed in **supplementary data file 4**. Experiments were performed at 25°C unless otherwise

indicated. When possible, genetic crosses of 6-8 males and 6-8 females were performed in both directions to test for an effect of the X or Y chromosomes on *BaraB*-mediated lethality; crosses in both directions yielded similar results in all cases and reported data are pooled results. Fly diet consisted of a nutrient-rich lab standard food: 3.72g agar, 35.28g cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL propionic acid, 15.9mL moldex, and H_2O to 600mL.

Infection experiments

Bacteria and yeast were grown to mid-log phase shaking at 200rpm in their respective growth media (LB, BHI, or YPG) and temperature conditions, and then pelleted by centrifugation to concentrate microbes. Resulting cultures were diluted to OD = 200 at 600nm before infections to measure gene expression. The following microbes were grown at 37°C: *Escherichia coli strain 1106* (LB) and *Candida albicans* (YPG). *Micrococcus luteus* was grown at 29°C in LB. For **Fig.** 1 and S2, pooled fly samples were collected either 6 hours post-infection (*E. coli*) or 24 hours post-infection (*C. albicans, M. luteus*) prior to RNA extraction on pools of 5 adult males. These timepoints correspond to the maximal expression inputs of the Imd (6hpi) or Toll (24hpi) NF-κB signalling pathways, which are most specifically induced by Gram-negative bacteria (Imd) or Gram-positive bacteria or fungi (Toll) (Lemaitre et al. 1997). Flies were pricked in the thorax as described in (Hanson, Dostálová, et al. 2019).

RNA extractions were performed using TRIzol™, Ambion DNAse treatment, and PrimeScript RT according to manufacturer's protocols. RT-qPCR was performed using PowerUP SYBR Green master mix with primers listed in supplementary data file 5. Gene expression differences were analyzed using the PFAFFL method (Pfaffl 2001). For gene expression experiments requiring dissection of heads, pools of 20 males were used for either whole flies or heads dissected in ice-cold PBS and transferred immediately to a tube kept on dry ice.

Selection analysis using HyPhy package

Codon aligned nexus tree files were generated using either the Neighbour-joining (1000 bootstraps) or PhyML (100 bootstraps) methods including proteins beyond those shown in **Fig. 5**. These tree files were analyzed using the HyPhy package with only 174nt pertaining to just the IM24 domain codons included. The cladogram in **Fig. 5**A is manually drawn from known species divergences (Kim et al. 2021). Use of either tree building method was chosen for convenience to best reflect known lineage sorting, as use of just 174nt was too information poor to resolve exact phylogenetic relatedness reliably. Tree files were qualitatively screened to ensure topologies broadly matched known species sortings, and thus ensure only relevant comparisons were made given the genomic synteny analysis in **Fig. 4** is principally informative of true gene lineages. HyPhy analyses were run separately for each *Baramicin* lineage within their clade, defined by genomic synteny; i.e. based on locus (e.g. ATP8A locus), and not considering convergent gene structures. We used three site-specific analyses (FEL, FUBAR, and SLAC) that use three independent

677 statistical approaches (Likelihood, Bayesian, and Count-based methods
678 respectively). We also employed both BUSTED and aBSREL branch-site analyses,
679 which are likelihood methods that differ in their approach of testing whole-
680 phylogeny selection or branch-specific comparisons respectively; an anology might
681 be performing analysis of variance (ANOVA) at the level of the entire ANOVA, or
682 comparing multiple groups against each other and subsequently using multiple test
683 correction. Each tree was rooted using the *Scaptodrosophila lebanonensis Baramicin*
684 as an outgroup with ancestral characteristics; we did not include *Baramicins* of the
685 subgenus Drosophila as including these resulted in long-branch attraction of the
686 Willistoni group *Baramicins* to subgenus Drosophila lineages, which would
687 confound relevant phylogenetic comparisons. When applicable, all internal branches
688 were assessed for potential selection. For *Baramicins* of the *ATP8A* locus, one site
689 (site 29) was highlighted as experiencing positive selection using FEL, FUBAR, and
690 SLAC analyses (p-adj = .011, .013, and .039 respectively). Additionally, site 31 was
691 also highlighted by FUBAR (p-adj = .026), but not FEL or SLAC analyses (p-adj > .05).
692 BUSTED analysis also supported diversifying selection in the *BaraA* lineage (ATP8A
693 locus, LRT p-adj = .008), indicating at least one site on at least one test branch has
694 experienced diversifying selection within the ATP8A lineage. The aBSREL branch-
695 site analysis specifically highlights the branch distinguishing the Willistoni group
696 *Baramicins* from the other Sophophora species (p-adj = .0045), suggesting variation
697 between these branches drives the signals of diversifying selection in the BUSTED
698 analysis. This result is intuitive, as we find a parallel but opposite evolution of
699 Baramicin protein structure in *Baramicins* of the *ATP8A* locus in *D. willistoni*
700 compared with *Baramicins* of other Sophophora species. Furthermore, in whole-
701 gene phylogenies, both *D. willistoni Baramicins* cluster together, supporting the
702 notion that these two daughter genes have evolved independent from the selection
703 that shaped the orthologues of *Dmel\BaraA* and *Dmel\BaraC*, also seen in qPCR data
704 that showed both genes were significantly enriched in the head (**Fig.** S6). This
705 phylogenetic clustering of the two *D. willistoni Baramicins* holds true when
706 additional *Baramicins* from recently sequenced genomes of the Willistoni group are
707 included (from (Kim et al. 2021) in supplementary data file 3), indicating this is
708 characteristic of the Willistoni group lineage and not specific to *D. willistoni*.

Supplementary figures and tables

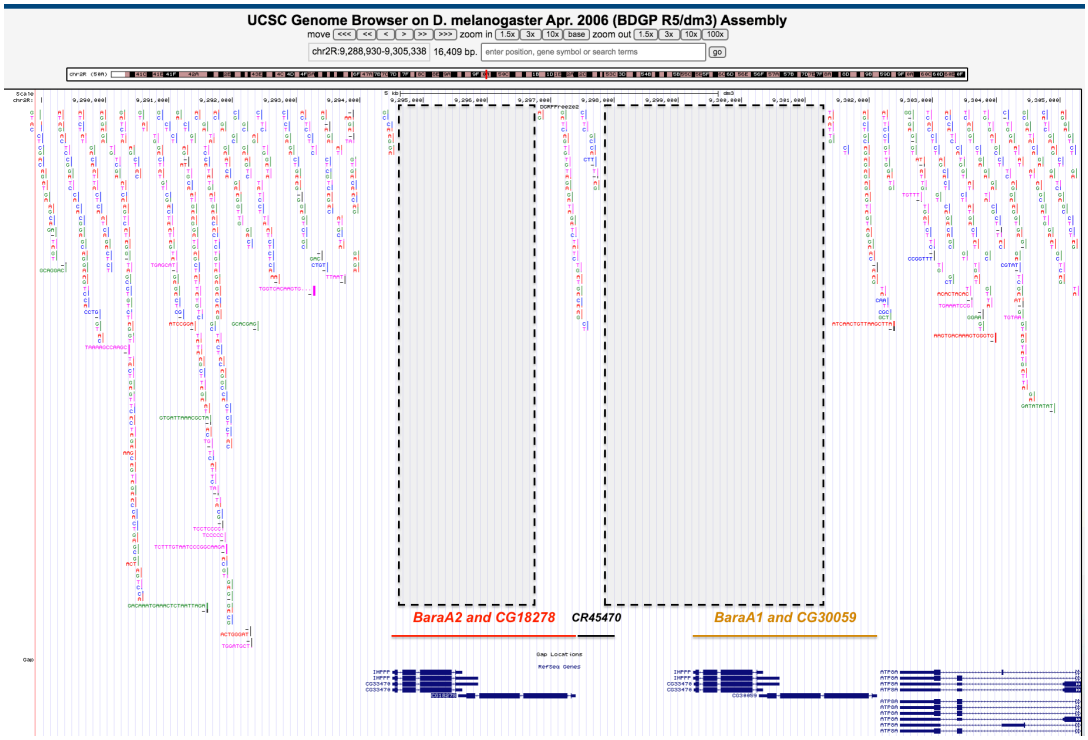

Figure S1: The *BaraA* locus is poorly resolved in DGRP genome assemblies. The *BaraA1* and *BaraA2* gene regions are totally devoid of mapped variants (dashed boxes). We speculate this is due to an artefact during genomic assembly, where reads mapping equally to the two identical *BaraA* genes were discarded as non-specific. This would explain why *BaraA* is typically discarded in RNAseq datasets using such measures in their pipeline, but not in microarray data from De Gregorio et al. (De Gregorio et al. 2002) where it is called “IM10.”

Genotype	Inferred copy #	Gtype	Junction PCR band	CG30059 band
Wild-type	**2**	*OR-R*	**YES**	**YES**
	2	*iso w1118*	**YES**	**YES**
	2	*Exelexis*	**YES**	**YES**
	1	*w VDRC*	NO	**YES**
	1	*yw*	NO	**YES**
	1	*Canton S*	NO	**YES**

Table S1: the *BaraA* duplication is variable in common lab stocks. Inferred *BaraA* copy numbers are given for six different lab stocks including *Oregon R* (*OR-R*), *iso DrosDel* (*iso* w^{1118}), *Exelexis*, a w^{1118} line from the VDRC, a *yellow white* (*yw*) line used previously by our group, and *Canton S*. All samples were additionally screened for *CG30059* as a DNA extraction positive control.

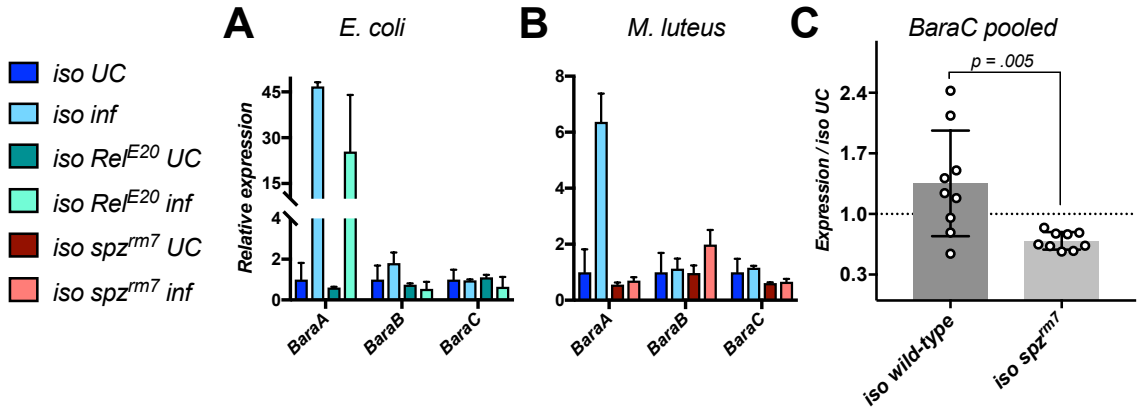

Figure S2: Additional assays of Baramicin expression upon infection. A) Neither *BaraB* nor *BaraC* are regulated by the Imd pathway, which is specifically stimulated by *E. coli* infection. **B)** Neither *BaraB* nor *BaraC* are induced after infection by *M. luteus*. **C)** *BaraC* levels were consistently depressed in spz^{rm7} flies in the unchallenged condition (UC) or upon infection with *C. albicans* (**Fig.** 3B) or *M. luteus* (**Fig.** S4B). Data here are pooled for *iso* wild type or *iso* spz^{rm7} flies without regard for infection treatment (student's t, p = .005).

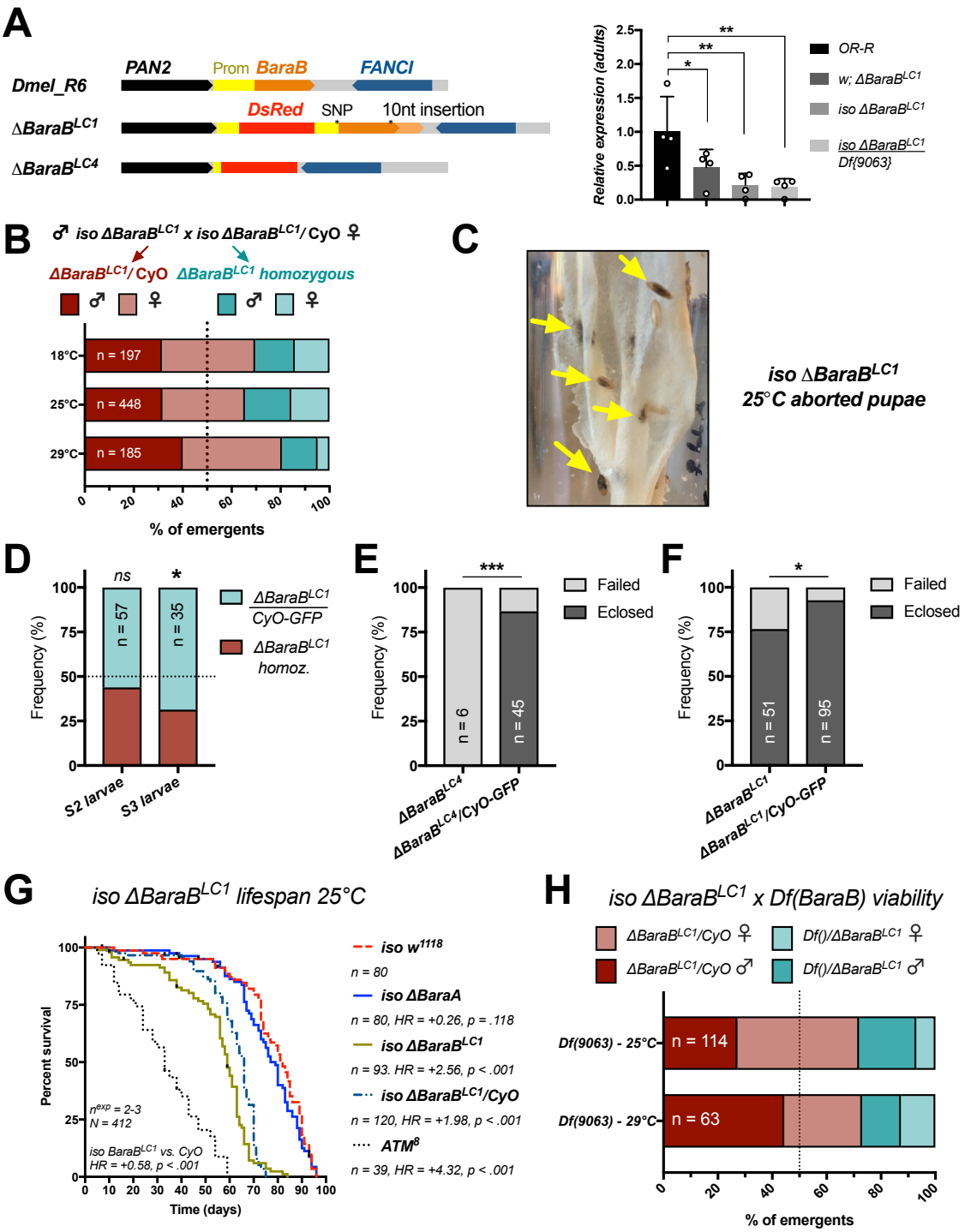

Figure S3: *BaraB* mutation is highly deleterious, even in *ΔBaraB^{LC1}* hypomorphs. A) Diagram of *BaraB* mutant loci and qPCR showing the *ΔBaraB^{LC1}* is a hypomorph mutation. Under our normal qPCR assay conditions, *BaraB* expression is not detected in *ΔBaraB^{LC1}* homozygotes. However using highly concentrated cDNA beyond our assay's valid range (100ng/10µL reaction), we could detect *BaraB* transcript in *ΔBaraB^{LC1}* flies. Quantification shown here is intended only to show

745 that *BaraB* transcript can be recovered from *ΔBaraB^{LC1}* homozygotes, and to give a
746 sense of relative whole-fly expression levels. **B)** Emergent frequencies of *ΔBaraB^{LC1}*
747 flies at different temperatures. **C)** Aborted pupae (yellow arrows) are a common
748 occurrence in *ΔBaraB* vials, and sometimes contain fully-developed adults that
749 simply never eclosed. In **D-F:** ns = not significant, * = p < .05, *** p = < .001. **D)** The
750 ratio of *ΔBaraB^{LC1}*/CyO-GFP to *ΔBaraB^{LC1}* homozygous larvae drops between the S2
751 and S3 larval stages (χ^2, p = .515 and p = .012 respectively). **E)** Frequency of
752 successfully eclosing adults using *BaraB^{LC4}*/CyO-GFP flies. **F)** Frequency of
753 successfully eclosing adults using *BaraB^{LC1}*/CyO-GFP flies. **G)** *BaraB* mutation
754 negatively affects lifespan. *iso ΔBaraB^{LC1}* homozygotes suffer reduced lifespan even
755 relative to their *iso ΔBaraB^{LC1}/CyO* siblings. By comparison, *iso ΔBaraA* flies that
756 used the same vector for mutant generation live as wild-type. ATM^{8} flies suffer
757 precocious neurodegeneration and are included as short-lived controls (Petersen et
758 al. 2013). **H)** *ΔBaraB^{LC1}* crossed to the genomic deficiency line (*Df(9063)*) supports a
759 partial-lethal effect of *BaraB* mutation.
760

Parents and temperature	Offspring	Sex	# eclosed	X^2	p-value
TRiP					
y1 v1; P{TRiP.HMJ23624}attP40/CyO	Act-Gal4 / CyO	m	26	25.687	**p < .001**
Act-Gal4/CyO-GFP ; +		f	30		
25°C	TRiP / CyO	m	26		
		f	29		
	Act-Gal4 > BaraB-IR{TRiP}	m	2		
		f	18		
y1 v1; P{TRiP.HMJ23624}attP40/CyO male	elav>Dcr2 ; +/CyO	m	67	10.037	**p < .02**
elav>Dcr2 ;; female		f	62		
25°C	elav>Dcr2 ; +/BaraB-IR	m	38		
		f	47		
y1 v1; P{TRiP.HMJ23624}attP40/CyO male	elav-Gal4; +/CyO	m	29	13.593	**p < .01**
elav-Gal4> ;; female		f	46		
25°C	elav-Gal4 ; +/BaraB-IR	m	21		
		f	22		
KK					
P{KK112854}VIE-260B	Act-Gal4 ; +/CyO	m	27	45.187	**p < .001**
Act-Gal4/CyO-GFP ; +		f	55		
25°C	Act-Gal4 > BaraB-IR{KK}	m	14		
		f	11		
P{KK112854}VIE-260B male	elav>Dcr2 ; +/BaraB-IR	m	0	11.000	**p < .001**
elav>Dcr2 ;; female		f	11		
25°C					
P{KK112854}VIE-260B male	normal wing	m	0	87.766	**p < .001**
elav-Gal4> ;; female		f	0		
29°C	nubbin-like wing	m	28		
		f	57		
P{KK112854}VIE-260B male	normal wing	m	37	32.133	**p < .001**
elav-Gal4> ;; female		f	47		
25°C	nubbin-like wing	m	31		
		f	5		

Table S2: *BaraB* RNAi summary statistics. Crosses used either the TRiP or KK *BaraB-IR* lines, driven by either *Actin5C-Gal4* or *elav-Gal4,* sometimes including *UAS-Dcr2*. Rearing at 29°C and inclusion of *UAS-Dcr2* increases the strength of RNA silencing.

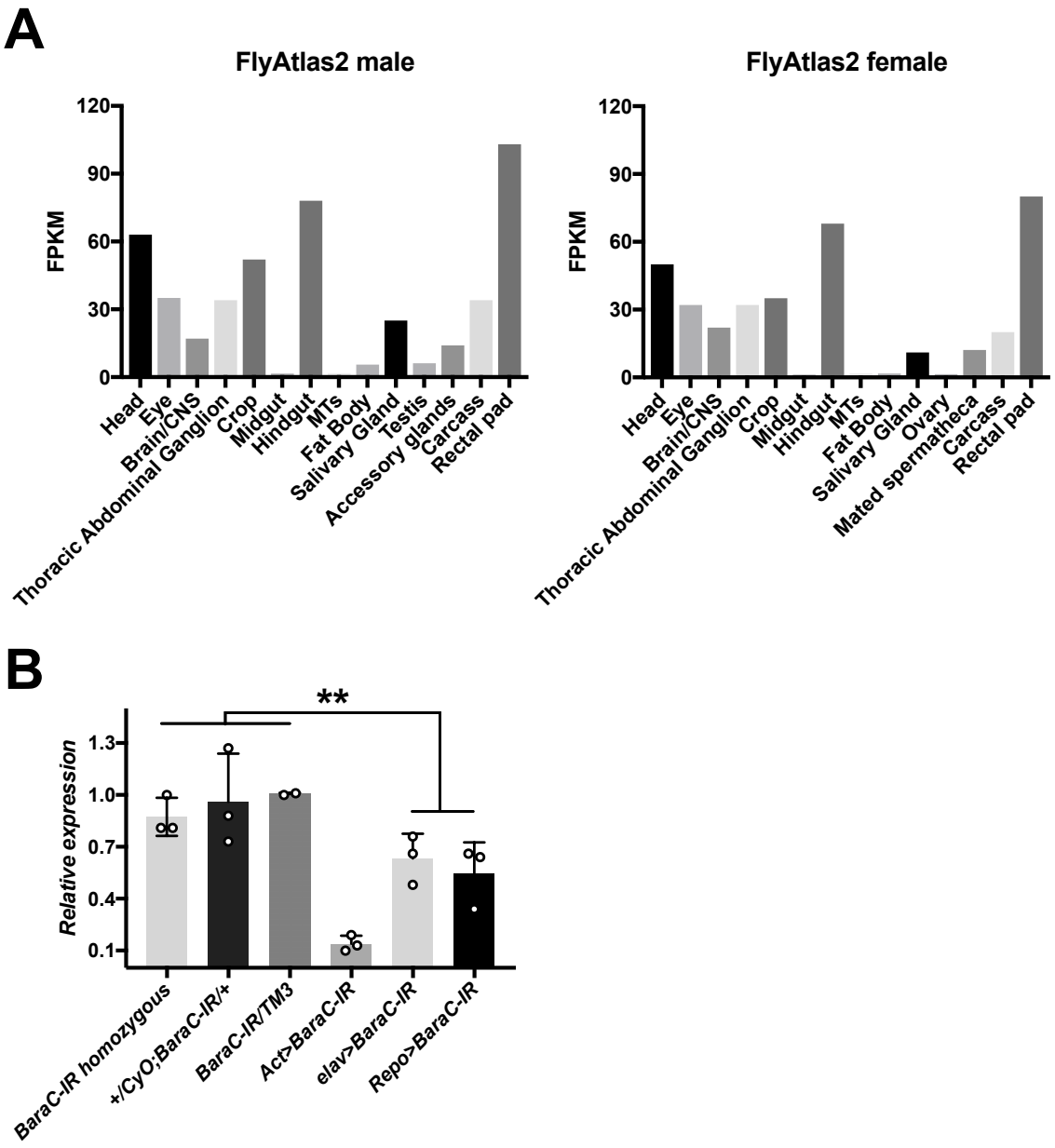

Figure S4: *BaraC* is expressed in the nervous system, but also the hindgut and rectal pads. A) FlyAtlas2 expression data for *BaraC*. **B)** RT-qPCR of *BaraC* in whole flies using different Gal4 drivers to express *BaraC* RNAi. *BaraC* is knocked down by both the *elav-Gal4* and *Repo-Gal4* nervous system drivers. Cumulatively, nervous system drivers significantly depress *BaraC* expression compared to *BaraC-IR* controls (student's t, $p < .01$). Ubiquitous knockdown using *Act>BaraC-IR* provides a comparative knockdown to better understand the strength of nervous system-specific knockdowns at the whole fly level.

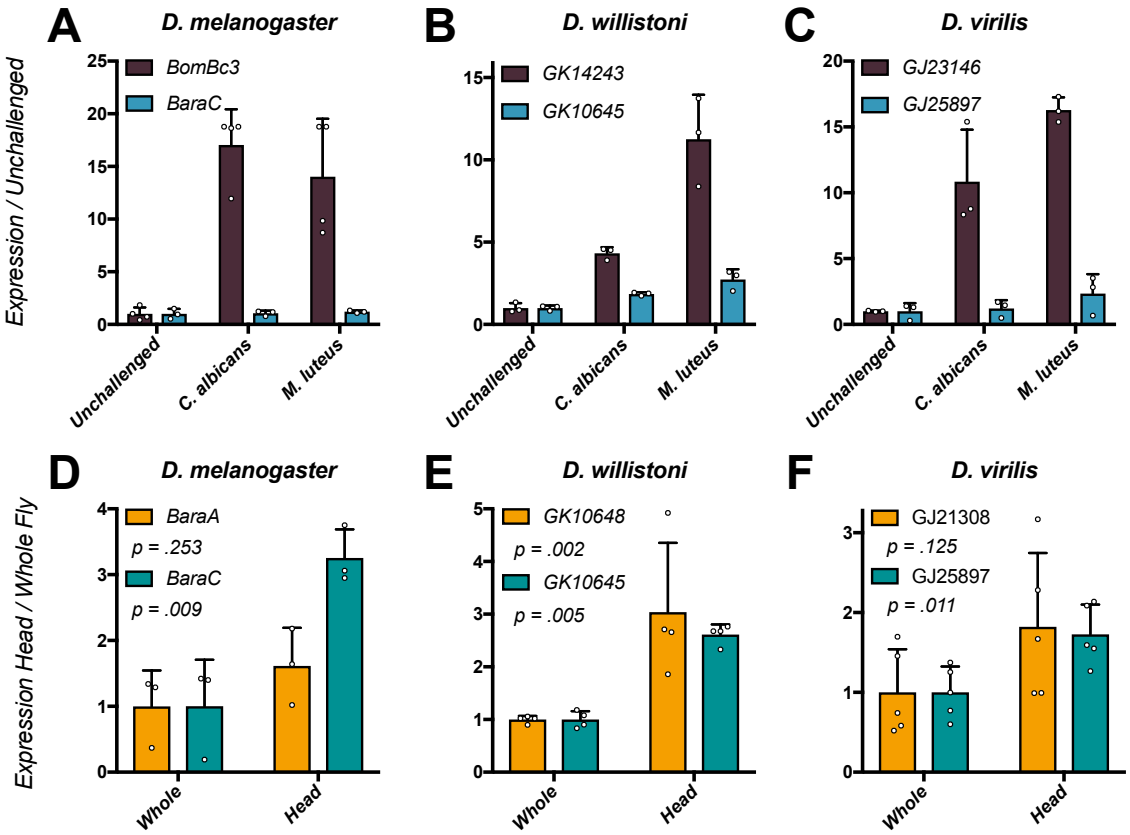

Figure S5: RT-qPCR of *Baramicin* genes in diverse species. A-C) Independent IM24-specific genes in *D. melanogaster* (A), *D. willistoni* (B), and *D. virilis* (C) are not induced by infection. *BomBc3* is included as an immune-induced control. **D-F)** The independent IM24-specific genes (blue) of *D. melanogaster* (D), *D. willistoni* (E), and *D. virilis* (F) are each enriched in the head relative to whole flies. *BaraA-like* genes (orange) were expressed more stochastically in the head, but also generally showed an enrichment pattern relative to whole flies (not always significant). Each data point represents an independent pooled sample from 20 male flies. Data were analyzed using one-way ANOVA with Holm's-Sidak multiple test correction.

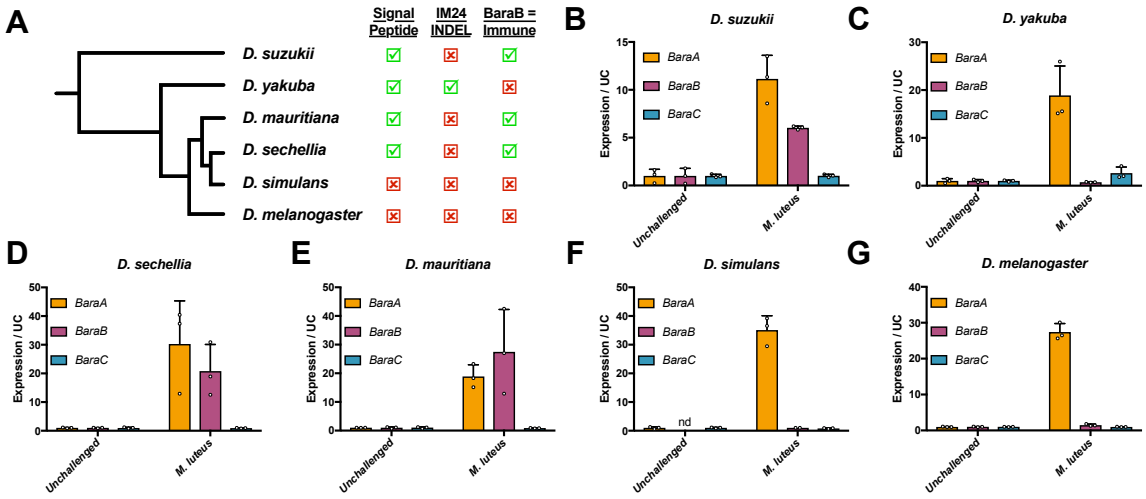

Figure S6: The *D. melanogaster BaraB* gene acquired its non-immune role only recently. A) Cladogram of the Melanogaster species group. The presence of a functional signal peptide (Fig. 5A), and the disruption of the *D. yakuba* IM24 peptide by an in-frame insertion is noted. A summary of whether *BaraB* is an immune-induced orthologue (B-G) is annotated. **B-G)** *Baramicin* expression data from Melanogaster group flies either unchallenged or infected with *M. luteus*. *BaraB* is immune-induced in *D. suzukii, D. sechellia,* and *D. mauritiana*, but not in *D. simulans* and *D. melanogaster,* which both lack signal peptide structures. *Drosophila yakuba BaraB* is not immune-induced (C), has an insertion event in its IM24 peptide (Fig. 5A), and its sister species *D. erecta* has pseudogenized its *BaraB* orthologue (Fig. 5B), suggesting pseudogenization may explain the lack of immune induction in *D. yakuba BaraB*.

References:

Abbott A. 2020. Are infections seeding some cases of Alzheimer's disease? *Nature* 587:22–25.

Barajas-azpeleta R, Wu J, Gill J, Welte R. 2018. Antimicrobial peptides modulate long-term memory. *PLoS Genetics*:1–26.

Brogden KA, Guthmiller JM, Salzet M, Zasloff M. 2005. The nervous system and innate immunity: The neuropeptide connection.

Casteels-Josson K, Capaci T, Casteels P, Tempst P. 1993. Apidaecin multipeptide precursor structure: a putative mechanism for amplification of the insect antibacterial response. *The EMBO journal* 12:1569–1578.

Chakraborty M, Chang C-H, Khost DE, Vedanayagam J, Adrion JR, Liao Y, Montooth KL, Meiklejohn CD, Larracuente AM, Emerson JJ. 2021. Evolution of genome structure in the Drosophila simulans species complex. *Genome Res* 31:380–396.

Chapman JR, Hill T, Unckless RL. 2019. Balancing selection drives maintenance of genetic variation in Drosophila antimicrobial peptides. *Genome Biology and Evolution* 11:2691–2701.

Clemmons AW, Lindsay SA, Wasserman SA. 2015. An Effector Peptide Family Required for Drosophila Toll-Mediated Immunity. *PLoS Pathogens* 11.

Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, Aibar S, Makhzami S, Christiaens V, Bravo González-Blas C, et al. 2018. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* 174:982-998.e20.

De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B. 2002. The Toll and Imd pathways are the major regulators of the immune response in Drosophila. *EMBO Journal* 21:2568–2579.

De Lorenzi E, Chiari M, Colombo R, Cretich M, Sola L, Vanna R, Gagni P, Bisceglia F, Morasso C, Lin JS, et al. 2017. Evidence that the human innate immune peptide LL-37 may be a binding partner of amyloid-β and inhibitor of fibril assembly. *Journal of Alzheimer's Disease* 59:1213–1226.

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26:2455–2457.

Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, Nguyen M, Haditsch U, Raha D, Griffin C, et al. 2019. Porphyromonas gingivalis in

836 Alzheimer's disease brains: Evidence for disease causation and treatment
837 with small-molecule inhibitors. *Science Advances* 5.

838 Ebrahim SAM, Talross GJS, Carlson JR. 2021. Sight of parasitoid wasps accelerates
839 sexual behavior and upregulates a micropeptide gene in Drosophila. *Nat*
840 *Commun* 12:2453.

841 Ferreira ÁG, Naylor H, Esteves SS, Pais IS, Martins NE, Teixeira L. 2014. The Toll-
842 dorsal pathway is required for resistance to viral oral infection in Drosophila.
843 *PLoS Pathog.* 10:e1004507.

844 Gerdol M, Schmitt P, Venier P, Rocha G, Rosa RD, Destoumieux-Garzón D. 2020.
845 Functional Insights From the Evolutionary Diversification of Big Defensins.
846 *Front Immunol* 11:758.

847 Halldórsdóttir K, Árnason E. 2015. Trans-species polymorphism at antimicrobial
848 innate immunity cathelicidin genes of Atlantic cod and related species. *PeerJ*
849 3:e976.

850 Hammonds AS, Bristow CA, Fisher WW, Weiszmann R, Wu S, Hartenstein V, Kellis M,
851 Yu B, Frise E, Celniker SE. 2013. Spatial expression of transcription factors in
852 Drosophila embryonic organ development. *Genome Biol.* 14:R140.

853 Hanson MA, Cohen LB, Marra A, Iatsenko I, Wasserman SA, Lemaitre B. 2021. The
854 Drosophila Baramicin polypeptide gene protects against fungal infection.
855 *PLoS Pathog* 17:e1009846.

856 Hanson MA, Dostálová A, Ceroni C, Poidevin M, Kondo S, Lemaitre B. 2019. Synergy
857 and remarkable specificity of antimicrobial peptides in vivo using a
858 systematic knockout approach. *eLife* 8.

859 Hanson MA, Hamilton PT, Perlman SJ. 2016. Immune genes and divergent
860 antimicrobial peptides in flies of the subgenus Drosophila. *BMC evolutionary*
861 *biology* 16:228.

862 Hanson MA, Lemaitre B. 2020. New insights on Drosophila antimicrobial peptide
863 function in host defense and beyond. *Curr. Opin. Immunol.* 62:22–30.

864 Hanson MA, Lemaitre B, Unckless RL. 2019. Dynamic Evolution of Antimicrobial
865 Peptides Underscores Trade-Offs Between Immunity and Ecological Fitness.
866 *Frontiers in Immunology* 10:2620.

867 Hellgren O, Sheldon BC, Buckling A. 2010. In vitro tests of natural allelic variation of
868 innate immune genes (avian ??-defensins) reveal functional differences in
869 microbial inhibition. *Journal of Evolutionary Biology* 23:2726–2730.

870 Huang J, Lou Y, Liu J, Bulet P, Jiao R, Hoffmann JA, Liegeois S, Li Z, Ferrandon D. 2020.
871 The BaramicinA gene is required at several steps of the host defense against
872 Enterococcus faecalis and Metarhizium robertsii in a septic wound infection
873 model in Drosophila melanogaster. bioRxiv Available from:
874 http://biorxiv.org/lookup/doi/10.1101/2020.11.23.394809

875 Jiggins FM, Kim KW. 2007. A screen for immunity genes evolving under positive
876 selection in Drosophila. *Journal of Evolutionary Biology* 20:965–970.

877 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool
878 K, Bates R, Žídek A, Potapenko A, et al. 2021. Highly accurate protein
879 structure prediction with AlphaFold. *Nature*.

880 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper
881 A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and
882 extendable desktop software platform for the organization and analysis of
883 sequence data. *Bioinformatics*.

884 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web
885 portal for protein modeling, prediction and analysis. *Nature protocols*
886 10:845–858.

887 Kim BY, Wang J, Miller DE, Barmina O, Delaney EK, Thompson A, Comeault AA,
888 Peede D, D'Agostino ERR, Pelaez J, et al. 2021. Highly contiguous assemblies
889 of 101 drosophilid genomes. *eLife* 10:e66405.

890 Kobler JM, Rodriguez Jimenez FJ, Petcu I, Grunwald Kadow IC. 2020. Immune
891 Receptor Signaling and the Mushroom Body Mediate Post-ingestion Pathogen
892 Avoidance. *Curr Biol* 30:4693-4709.e3.

893 Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane
894 protein topology with a hidden Markov model: application to complete
895 genomes. *J Mol Biol* 305:567–580.

896 Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. 2018. FlyAtlas 2: A new version
897 of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq
898 and sex-specific data. *Nucleic Acids Research* 46:D809–D815.

899 Lee M, Shi X, Barron AE, McGeer E, McGeer PL. 2015. Human antimicrobial peptide
900 LL-37 induces glial-mediated neuroinflammation. *Biochemical Pharmacology*
901 94:130–141.

902 Lemaitre B, Reichhart JM, Hoffmann JA. 1997. Drosophila host defense: differential
903 induction of antimicrobial peptide genes after infection by various classes of
904 microorganisms. *Proceedings of the National Academy of Sciences of the*
905 *United States of America* 94:14614–14619.

906 Lezi E, Zhou T, Koh S, Chuang M, Sharma R, Pujol N, Chisholm AD, Eroglu C,
907 Matsunami H, Yan D. 2018. An Antimicrobial Peptide and Its Neuronal
908 Receptor Regulate Dendrite Degeneration in Aging and Infection. *Neuron*
909 97:125-138.e5.

910 Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y,
911 Magwire MM, Cridland JM, et al. 2012. The Drosophila melanogaster Genetic
912 Reference Panel. *Nature* 482:173–178.

913 Moir RD, Lathe R, Tanzi RE. 2018. The antimicrobial protection hypothesis of
914 Alzheimer's disease. *Alzheimer's & Dementia* 14:1602–1614.

915 Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner
916 T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic
917 selection. *Mol Biol Evol* 32:1365–1371.

918 Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, Griffin RS, Belfer I,
919 Dai F, Smith SB, et al. 2010. A Genome-wide Drosophila screen for heat
920 nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell*
921 143:628–638.

922 Petersen AJ, Katzenberger RJ, Wassarman DA. 2013. The innate immune response
923 transcription factor relish is necessary for neurodegeneration in a Drosophila
924 model of ataxia-telangiectasia. *Genetics* 194:133–142.

925 Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time
926 RT-PCR. *Nucleic Acids Res.* 29:e45.

927 Quesada H, Ramos-Onsins SE, Aguad?? M. 2005. Birth-and-death evolution of the
928 Cecropin multigene family in Drosophila. *Journal of Molecular Evolution* 60:1–
929 11.

930 Rolff J, Schmid-Hempel P. 2016. Perspectives on the evolutionary ecology of
931 arthropod antimicrobial peptides. *Philosophical Transactions of the Royal*
932 *Society B: Biological Sciences* 371.

933 Sackton TB, Lazzaro BP, Clark AG, Wittkopp P. 2017. Rapid expansion of immune-
934 related gene families in the house fly, musca domestica. *Molecular Biology*
935 *and Evolution*.

936 Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. 2021. Innate Immunity
937 Promotes Sleep through Epidermal Antimicrobial Peptides. *Current Biology*
938 31:564-577.e12.

939 Tamura K, Subramanian S, Kumar S. 2004. Temporal Patterns of Fruit Fly
940 (Drosophila) Evolution Revealed by Mutation Clocks. *Molecular Biology and*
941 *Evolution* 21:36–44.

942 Tennessen JA. 2005. Molecular evolution of animal antimicrobial peptides:
943 Widespread moderate positive selection.

944 Toda H, Williams JA, Gulledge M. 2019. A sleep-inducing gene, nemuri , links sleep
945 and immune function in Drosophila. 515:509–515.

946 Train C-M, Pignatelli M, Altenhoff A, Dessimoz C. 2019. iHam and pyHam: visualizing
947 and processing hierarchical orthologous groups. *Bioinformatics (Oxford,*
948 *England)* 35:2504–2506.

949 Unckless RL, Howick VM, Lazzaro BP. 2016. Convergent Balancing Selection on an
950 Antimicrobial Peptide in Drosophila. *Current Biology* 26:257–262.

951 Vilcinskas A, Mukherjee K, Vogel H. 2013. Expansion of the antimicrobial peptide
952 repertoire in the invasive ladybird Harmonia axyridis. *Proceedings of the*
953 *Royal Society B: Biological Sciences*.

954 Wang Y, Zhu S. 2011. The defensin gene family expansion in the tick Ixodes
955 scapularis. *Developmental and Comparative Immunology*.

956 Zanetti M. 2005. The role of cathelicidins in the innate host defenses of mammals.
957 *Curr Issues Mol Biol* 7:179–196.

958 Zhang W, Yang J, He B, Walker SE, Zhang H, Govindarajoo B, Virtanen J, Xue Z, Shen
959 H-B, Zhang Y. 2016. Integration of QUARK and I-TASSER for Ab Initio Protein
960 Structure Prediction in CASP11. *Proteins* 84 Suppl 1:76–86.

961