Repeated truncation of a modular antimicrobial peptide gene for neural context

M.A. Hanson* and B. Lemaitre

Global Health Institute, School of Life Science, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

*Author for Correspondence: Mark Hanson, Global Health Institute, School of Life Science, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (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. These genes commonly encode multiple products as post-translationally cleaved polypeptides. Recent studies have highlighted roles for AMPs in neurological contexts suggesting functions for these defence molecules beyond infection. During our immune study characterizing the antimicrobial peptide gene *Baramicin*, we recovered multiple *Baramicin* paralogs in *Drosophila melanogaster* and other species, united by their N-terminal IM24 domain. Not all paralogs were immune-induced. Here, through careful dissection of the *Baramicin* family's evolutionary history, we find 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. Surprisingly, using mutation and targeted gene silencing we demonstrate that two such genes are adapted for function in neural contexts in *D. melanogaster*. We also 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.

Key words: AMPs, Drosophila, Immunity, Baramicin, Neurology

Significance statement:

Antimicrobial peptides are immune proteins recently implicated in neurological roles. To date little attention has been paid to the contributions of different gene products in this function. Here we show that an antimicrobial peptide gene encodes multiple products with either immune-specific or neurological roles.

46 # Introduction

47 Antimicrobial peptides (AMPs) are immune effectors best known for their
48 role in defence against infection. These antimicrobials are commonly encoded as a
49 polypeptide including both pro- and mature peptide domains (Zanetti 2005; Hanson
50 and Lemaitre 2020). AMP genes frequently experience events of duplication and
51 loss (Wang and Zhu 2011; Vilcinskas et al. 2013; Sackton et al. 2017; Hanson,
52 Lemaitre, et al. 2019) and undergo rapid evolution at the sequence level (Tennessen
53 2005; Jiggins and Kim 2007; Hellgren et al. 2010; Halldórsdóttir and Árnason 2015;
54 Hanson et al. 2016; Chapman et al. 2019). The selective pressures that drive these
55 evolutionary outcomes are likely the consequence of host-pathogen interactions
56 (Unckless et al. 2016). However AMPs and AMP-like genes in many species have also
57 recently been implicated in non-immune roles in flies, nematodes, and humans.
58 These new contexts suggest that the evolutionary forces acting on AMP genes may
59 not be driven strictly by trade-offs in host defence, but rather by conflicts between
60 roles in immunity and other non-immune functions.

61 For instance, Diptericins are membrane-disrupting antimicrobial peptides of
62 flies (Diptera) that are required for defence against infection by *Providencia* bacteria
63 (Unckless et al. 2016; Hanson, Dostálová, et al. 2019). It was therefore surprising
64 that the *D. melanogaster* gene *Diptericin B (DptB)* affects memory processes
65 (Barajas-azpeleta et al. 2018). In this study, *DptB* derived from the fly fat body
66 (analogous to the mammalian liver) regulated the ability of the fly to form long-term
67 memory associations (Barajas-azpeleta et al. 2018). Another AMP-like gene, nemuri,
68 regulates fly sleep and promotes survival upon infection (Toda et al. 2019). Studies
69 in nematodes have also shown that an immune-induced polypeptide (NLP-29) binds
70 to a G-protein coupled receptor (NPR-12) triggering neurodegeneration through
71 activation of the NPR-12-dependent autophagy pathway (Lezi et al. 2018), and
72 injury triggers epidermal AMPs including NLP-29 to promote sleep (Sinner et al.
73 2021). *Drosophila* AMPs have also recently been shown to regulate behaviours after
74 seeing parasitoid wasps (Ebrahim et al. 2021), during feeding with different bacteria
75 (Kobler et al. 2020), or following infection (Hanson et al. 2021). In humans, the
76 *Cathelicidin* gene encodes the AMP LL-37, which is implicated in glia-mediated

77 neuroinflammation and Alzheimer's disease (Lee et al. 2015; De Lorenzi et al. 2017).
78 Indeed recent evidence suggests Alzheimer's disease is an infectious syndrome
79 (Dominy et al. 2019), though the importance of this process is debated (Abbott
80 2020). Notably, AMPs share a number of properties with classic neuropeptides
81 (Brogden et al. 2005), further muddying the distinction between peptides of the
82 immune and nervous systems.

83 We recently described a novel antifungal peptide gene of *Drosophila*
84 *melanogaster* that we named *Baramicin A* (*BaraA*) (Hanson et al. 2021). A unique
85 aspect of *BaraA* is its precursor protein structure, which encodes a polypeptide
86 cleaved into multiple mature products by interspersed furin cleavage sites. The use
87 of furin cleavage sites to produce more than one mature peptide from a single
88 polypeptide precursor is widespread in animal AMP genes (Gerdol et al. 2020;
89 Hanson and Lemaitre 2020), including multiple peptide repeats in bees and other
90 flies (Casteels-Josson et al. 1993; Hanson et al. 2016). However, *BaraA* represents an
91 exceptional case as many tandem repeat peptides are cleaved by furin from a single
92 precursor protein, effectively resembling a "protein-based operon". The immature
93 precursor protein of *D. melanogaster BaraA* encodes three types of domains: an
94 IM24 domain, three tandem repeats of IM10-like domains, and an IM22 domain.
95 *BaraA* mutants are susceptible to infection by fungi, and *in vitro* experiments
96 suggest the *BaraA* IM10-like peptides have antifungal activity (Hanson et al. 2021).
97 The other *Baramicin* domains encoding IM22 and IM24 remain uncharacterized.
98 Curiously, *BaraA* deficient flies also display an erect wing behavioural phenotype
99 upon immune stimulation even in the absence of infection, suggesting that *BaraA*
100 products could have non-microbial targets (Hanson et al. 2021).

101 In this study, we describe the evolution of the Drosophilid *Baramicin* gene
102 family. Three unique *Baramicin* genes (*BaraA, B, and C*) are present in the genome of
103 *D. melanogaster*. Surprisingly, only *BaraA* is immune-induced, while *BaraB* and
104 *BaraC* are enriched in the nervous system. Both *BaraB* and *BaraC* have truncations
105 compared to the ancestral *Baramicin* gene, and these two genes effectively encode
106 just the *Baramicin* IM24 domain. We found similar truncations in other species, and
107 confirmed loss of immune expression for IM24-specific *Baramicins* of other species.

108 We also confirmed enrichment in the head or nervous system for IM24-specific
109 genes in *D. melanogaster* and other species. We resolved the genomic ancestry of the
110 *Baramicins*, which confirmed that these repeated truncations creating IM24-specific
111 genes came from independent events (convergent evolution). The complex 'protein
112 operon' polypeptide nature of *Baramicin* draws attention to how different sub-
113 peptides can be adapted to context-specific roles, like in immunity or neurology.
114 Attention to the multiple peptide products of AMP genes could explain how these
115 immune effectors affect both immune and neurological processes.
116

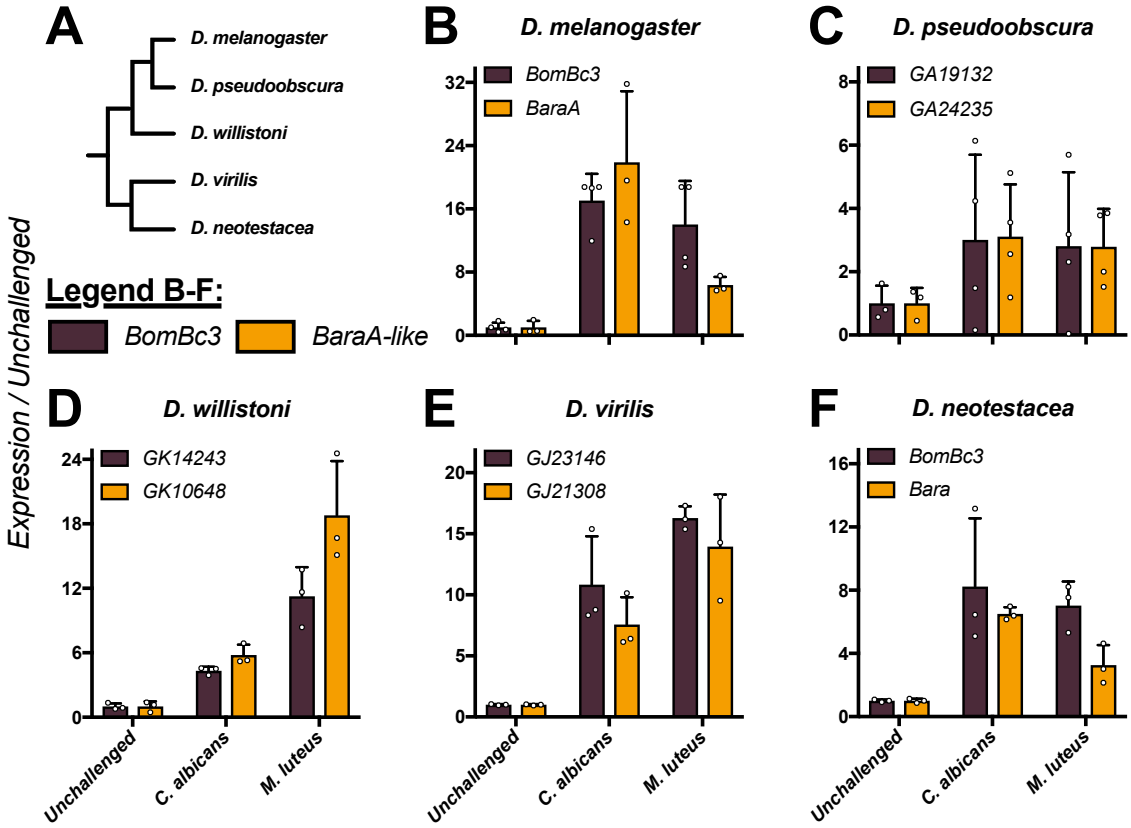
117 # Results

118 ***Baramicin is an ancestral immune effector***

119 The *Baramicin A* gene was only recently described as encoding antifungal
120 effectors by our group (Hanson et al. 2021), and another recent study also
121 confirmed *Baramicin*'s important contribution to Toll immune defence (Huang et al.
122 2020). These initial characterizations were done only in *D. melanogaster*, and
123 focused on one *Baramicin* gene (*BaraA*). We will therefore first provide a basic
124 description of the immune *Baramicins* of other species and also the larger *Baramicin*
125 gene family of *D. melanogaster* to establish that this is a classically immune gene
126 family. This is relevant to non-immune duplicate genes discussed later.

127 In *D. melanogaster, BaraA* is regulated by the Toll immune signalling pathway
128 (Huang et al. 2020; Hanson et al. 2021). Using BLAST, we recovered *BaraA*-like
129 genes encoding each Baramicin peptide (IM24, IM10-like, and IM22) across the
130 genus *Drosophila* and in the outgroup *Scaptodrosophila lebanonensis*. In many
131 species, this was the only *Baramicin* gene present, suggesting *Dmel\BaraA*
132 resembles the ancestral *Baramicin* structure. We performed infection experiments
133 to confirm that *BaraA*-like genes were immune-inducible in the diverse species *D.*
134 *melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, and *D. neotesteacea* (last
135 common ancestor ~53mya (Suvorov et al. 2021)) with *Micrococcus luteus* and
136 *Candida albicans*, two microbes that stimulate the Toll pathway (**Fig.** 1A). In all five
137 species, *BaraA*-like genes were immune-induced (**Fig.** 1B-F). We therefore confirm

138 the ancestral *Baramicin* was an immune-induced gene. Deviations from immune
139 function derived.
140



141

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

146

147 ***The D. melanogaster genome encodes up to four Baramicins: BaraA1, BaraA2,***
148 ***BaraB and BaraC***

149 In *D. melanogaster*, we recovered four *Baramicin* genes. First, we realized
150 that a duplication of *BaraA* is actively segregating in wild flies (**Fig. 2**A). The *D.*
151 *melanogaster* R6 genome assembly encodes two 100% identical *BaraA* genes
152 (*CG33470* and *CG18279*, *BaraA1* and *BaraA2* respectively). We screened 132 DGRP
153 lines for the *BaraA* duplication event, finding only ~14% (18/132) of strains were
154 PCR-positive for two *BaraA* copies (supplementary data file 1). Perhaps as a

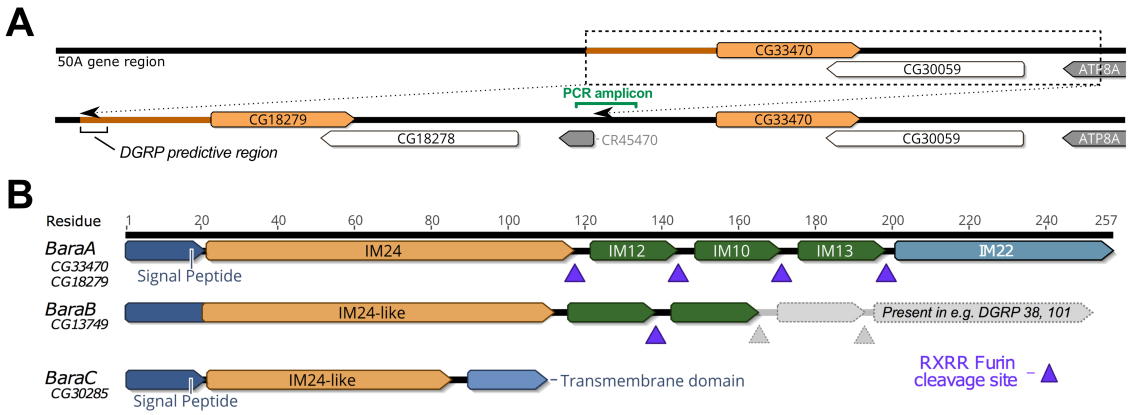


Figure 2: The *D. melanogaster Baramicin* genes. A) Schematic of the *BaraA* duplication. Using a PCR assay spanning the duplication-specific locus (PCR amplicon), we confirmed *BaraA* copy number is variable in various lab strains (Hanson et al. 2021) and wild-caught flies (Supplementary data file 1). B) *D. melanogaster* encodes two other *Baramicin* genes that we name *BaraB* and *BaraC*. These paralogs differ markedly in their precursor protein structure through truncation of the C-terminus relative to *BaraA.* The *BaraB* truncation is segregating in the DGRP (greyed out region, and see supplementary data file 2).

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, where extra gene copies correlates 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

181 DGRP *BaraB* alignment is provided in supplementary data file 2). Moreover, in
182 contrast to BaraA, the initial IM10-like peptide of BaraB no longer follows a furin
183 cleavage site, and encodes a serine (RSXR) in its IM10-like motif instead of the
184 universal proline (RPXR) of *BaraA*-like IM10 peptides across the genus. Each of
185 these mutations prevents the secretion of classical IM10-like and IM22 peptides by
186 *BaraB*. Finally, *BaraC* encodes only IM24 tailed by a transmembrane domain at the
187 C terminus (TMHMM v2.0 (Krogh et al. 2001)), and thus lacks both the IM10-like
188 peptides and IM22 (**Fig. 2**B).
189
190 ***BaraB and BaraC are not immune-inducible***
191 *BaraA* is strongly induced following microbial challenge (Fig. 1), being
192 predominantly regulated by the Toll pathway with a minor input from the Immune
193 Deficiency (Imd) pathway (Huang et al. 2020; Hanson et al. 2021). We therefore
194 assayed the expression of *BaraB* and *BaraC* in wild-type flies, and also flies with
195 defective Toll (spz^{rm7}) or Imd (Rel^{E20}) signalling to see if their basal expression relied
196 on these pathways. Surprisingly, neither gene was induced upon infection
197 regardless of microbial challenge (**Fig. 3**A and **Fig.** S2A-B). Of note: *BaraC* levels
198 were consistently reduced in spz^{rm7} mutants regardless of treatment (cumulative
199 data in **Fig.** S2C, p = .005), suggesting *BaraC* basal expression is affected by Toll
200 signalling. We next generated a novel time course of development from egg to adult
201 to monitor the expression of the three *Baramicin* genes. We found that expression of
202 all genes increased over development and reached their highest level in young
203 adults (**Fig.** 3B). Of note, *BaraB* expression approached the lower limit of our assay's
204 detection sensitivity at early life stages. However *BaraB* was robustly detected
205 beginning at the pupal stage, indicating it is expressed during metamorphosis. *BaraC*
206 expression also increased markedly between the L3 larval stage and pupal stage.
207 Collectively we reveal that *BaraA* is part of a larger gene family. While the
208 *BaraA* gene was first described as an immune effector, the two *Baramicin* paralogs
209 *BaraB* and *BaraC* are not induced by infection in *D. melanogaster*. Both *BaraB* and
210 *BaraC* first see increased expression during pupation, and are ultimately expressed
211 at their highest levels in adults.

212

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

214 A simple interpretation of the truncated gene structure and low levels of
215 *BaraB* expression is that this gene is undergoing pseudogenization. Indeed, AMP
216 gene pseudogenization is common in insects including *Drosophila* (Quesada et al.
217 2005; Rolff and Schmid-Hempel 2016; Hanson, Lemaitre, et al. 2019). To explore
218 *BaraB* function, we used two mutations for *BaraB* ($\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$,
219 generously gifted by S.A. Wasserman). These mutations were made using a CRISPR
220 double gRNA approach to replace the *BaraB* locus with sequence from the pHD-
221 DsRed vector. The $\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$ mutations differ in their ultimate effect,
222 as $\Delta BaraB^{LC1}$ is an incidental insertion of the DsRed cassette in the promoter of the
223 gene. This disruption reduces gene expression, resulting in a hypomorph state (**Fig.**
224 **S3**A). The $\Delta BaraB^{LC4}$ mutation however deletes the locus as intended, leading to
225 *BaraB* null flies (**Fig. 3**C).

226 We further introgressed both *ΔBaraB* mutations into the DrosDel isogenic
227 background (referred to as *iso*) for seven generations according to Ferreira et al.
228 (Ferreira et al. 2014). At the same time, we combined the original *ΔBaraB*
229 chromosomes with a CyO-GFP balancer chromosome in a mixed genetic background
230 to distinguish homozygous/heterozygous larvae. In all cases, $\Delta BaraB^{LC4}$ homozygous
231 individuals failed to develop to the adult stage, whereas homozygous $\Delta BaraB^{LC1}$
232 adults emerged (**Fig. 3**D). We next assessed *BaraB* hypomorph viability by crossing
233 $\Delta BaraB^{LC1}$ homozygous males to $\Delta BaraB^{LC1}/CyO$ heterozygous females. The resulting
234 offspring ratio departed from mendelian inheritance, and was exacerbated by
235 rearing at 29°C (**Fig. S3**B). Using our *CyO-GFP* reporter to track hetero- vs.
236 homozygous larvae revealed that the major lethal phase occurs primarily in the late
237 larval and pupal stages (**Fig. S3**C-F), consistent with a role for *BaraB* in the larva/

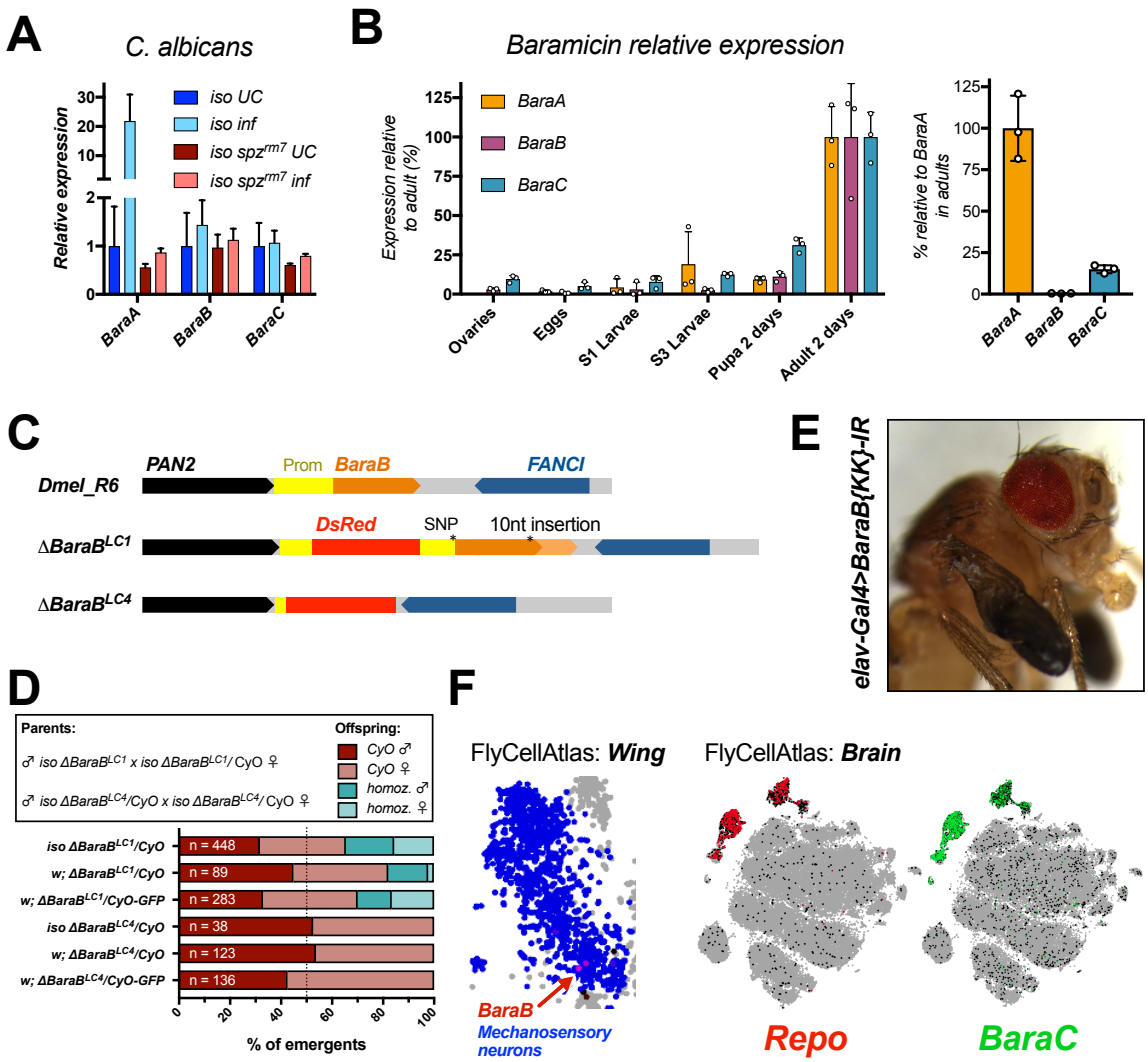


Figure 3: *D. melanogaster BaraB and BaraC* have neural functions. A) Only *BaraA* is immune-induced. *BaraB* and *BaraC* do not respond to infection, though basal *BaraC* expression relies on Toll signalling (Fig. S3C). B) Time series of whole animal *Baramicin* expression over the course of development. 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} has the intended genetic replacement of *BaraB* with a DsRed cassette. Both mutations cause expression of 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) The nubbin-like wing phenotype is copied by *BaraB* gene silencing using the neural driver *elav-Gal4*. F) Specific wing mechanosensory neurons seem to express *BaraB* in adults. Meanwhile 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 (Li et al. 2021).

254 pupa stage as suggested by an increase in expression at this stage (**Fig.** 3B). Some
255 *ΔBaraB^{LC1}* homozygous flies also exhibited locomotor defects, and/or a nubbin-like
256 wing phenotype (FlyBase: FBrf0220532 and e.g. in **Fig. 3**E) where the wings were
257 stuck in a shrivelled state for the remainder of the fly's lifespan. However, a
258 proportion of *ΔBaraB^{LC1}* homozygotes successfully emerged, and unlike their
259 siblings, had no immediate morphological or locomotory defects. The lifespan of
260 morphologically normal *iso ΔBaraB^{LC1}* adults is nevertheless significantly shorter
261 compared to wild-type flies and *iso ΔBaraB^{LC1}/CyO* siblings (**Fig. S4**G). We confirmed
262 these developmental defects using ubiquitous gene silencing with *Actin5C-Gal4 (Act-*
263 *Gal4)* to drive two *BaraB* RNAi/interfering RNA (IR) constructs (*TRiP-IR* and *KK-IR*).
264 Both constructs resulted in significant lethality and occurrence of nubbin-like wings
265 (**Table** S1). Genomic deficiency crosses also confirmed significantly reduced
266 numbers of eclosing *BaraB*-deficient flies at 25°C ($n = 114$, χ^2 $p < .001$) and 29°C ($n =$
267 63, χ^2 $p < .001$) (**Fig. S3**H). We therefore conclude that full gene deletion causes
268 lethality at the larva-pupa transition stage, and *BaraB* hypomorph flies suffer
269 significant costs to fitness during development, and have reduced lifespan even
270 following successful eclosion.
271 These data indicate *BaraB* is unlikely to be pseudogenized. While whole-fly
272 *BaraB* expression is low, *BaraB* appears to perform an integral developmental role.
273 The fact that there is a bimodal outcome in hypomorph-like *ΔBaraB^{LC1}* adults (either
274 severe defects or generally healthy) suggests *BaraB* is involved in passing some
275 checkpoint during larval/pupal development. Flies deficient for *BaraB* may be more
276 likely to fail at this developmental checkpoint, resulting in either lethality or
277 developmental defects.
278
279 ***Baramicin B plays an important role in the nervous system***
280 We next sought to determine in which tissue(s) *BaraB* is required. A previous
281 screen using neural *elav-Gal4* driven RNA interference highlighted *BaraB* silencing
282 for lethality effects ($n = 15$) (Neely et al. 2010). Given *BaraB* mutant locomotory
283 defects, we started by silencing *BaraB* in the nervous system using the pan-neural

284 *elav-Gal4* driver with both the *TRiP-IR* and *KK-IR BaraB-IR* lines *(IR = interfering*
285 *RNA)*. We also used a combination of *UAS-Dicer2 (Dcr2)* and/or 29°C for greater
286 silencing efficiency. In all cases, *BaraB-IR* driven by *elav-Gal4* caused a significant (p
287 $< .02$) departure from mendelian inheritance in lethality and nubbin-like wing
288 presentation (Table S1). Moreover the frequency of both lethality and the nubbin-
289 like wing phenotype was increased with increasing strength of gene silencing.
290 This analysis suggests that *BaraB* plays an important role in the nervous system,
291 explaining both the lethality and nubbin-like wing phenotypes. Interestingly, *BaraB*
292 is expressed in specific subset of mechanosensory neuron cells in the wing in
293 FlyCellAtlas (Li et al. 2021) (**Fig.** 3F), despite very low levels of *BaraB* expression in
294 other FlyCellAtlas tissue datasets. We additionally investigated the effect of *BaraB*
295 silencing in non-neural tissues including the fat body (*c564-Gal4*), hemocytes (*hml-*
296 *Gal4*), the gut (esg-Gal4, MyO-Gal4), the wing disc (*nubbin-Gal4*), and in myocytes
297 (*mef2-gal4*), which did not present with increased lethality or nubbin-like wings. We
298 also screened neural drivers specific for glia (*Repo-Gal4*), motor neurons (*D42-,*
299 *VGMN-,* and *OK6-Gal4*), and a *BaraA-Gal4* driver (Hanson et al. 2021) that could
300 overlap *BaraB*-expressing cells. However all these *Gal4>BaraB-IR* flies were viable
301 and never exhibited overt morphological defects.
302
303 ***Baramicin C is expressed in glia***
304 Tissue-specific transcriptomic data indicate that *BaraC* is expressed in
305 various neural tissues including the eye, brain, and the thoracic abdominal ganglion
306 (**Fig. S4**A), but also the hindgut and rectal pads pointing to a complex expression
307 pattern (Hammonds et al. 2013; Leader et al. 2018). We next searched FlyCellAtlas
308 (Li et al. 2021) to narrow down which neural subtypes *BaraB and BaraC* were
309 expressed in. *BaraB* expressing cells were few and showed only low expression in
310 this dataset. However *BaraC* was robustly expressed in all glial cell types, fully
311 overlapping the glia marker *Repo* (**Fig. 3**F). To confirm the observation that *BaraC*
312 was expressed in glia, we compared the effects of *BaraC* RNA silencing (*BaraC-IR*)
313 using *Act-Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia) drivers on

314 *BaraC* expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that of control
315 flies (**Fig. S4**B). By comparison *elav-Gal4* reduced *BaraC* expression to ~63% that of
316 controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls (overall
317 controls vs. neural/glia-IR, p = .002). We also screened for overt lethality, and
318 locomotor or developmental defects upon *BaraC* silencing using ubiquitous *Act-Gal4*
319 and neural *elav-Gal4>Dcr2* or *Repo-Gal4*. However *BaraC* silencing never produced
320 overt phenotypes in morphology or locomotor activity (not shown).
321 Collectively, our results support the notion that *BaraC* is expressed in the
322 nervous system, and are consistent with *BaraC* expression being most localized to
323 glial cells.
324
325 ***Repeated genomic turnover of the Baramicin gene family***
326 Our results thus far show that *BaraA*-like genes are consistently immune-
327 induced in all *Drosophila* species (**Fig.** 1), however the two paralogs *Dmel\BaraB*
328 and *Dmel\BaraC* are not immune-induced, and are truncated in a fashion that
329 deletes some or all of the antifungal IM10-like peptides (**Fig. 2**B). These two
330 *Baramicins* are now enriched in the nervous system (**Fig. 3**E-F). In the case of *BaraB*,
331 a role in the nervous system is evidenced by severe defects recapitulated using pan-
332 neural RNA silencing. In the case of *BaraC*, nervous system expression is evidenced
333 by a clear overlap with *Repo*-expressing cells.
334 While *BaraA*-like genes are conserved throughout the genus *Drosophila*,
335 *BaraB* is conserved only in Melanogaster group flies, and *BaraC* is found only in
336 Melanogaster and Obscura group flies, indicating that both paralogs stem from
337 duplication events of a *BaraA*-like ancestor (**Fig. 4**). To determine the ancestry of
338 each *D. melanogaster Baramicin* gene, we traced their evolutionary history by
339 analyzing genomic synteny through hierarchical orthologous groups (Train et al.
340 2019). Ancestry tracing revealed that these three loci ultimately stem from a single-
341 locus ancestor encoding only one *Baramicin* gene that resembled *Dmel\BaraA* (**Fig.**
342 **4**A). This is evidenced by the presence of only a single *BaraA*-like gene in the
343 outgroup *S. lebanonensis*, and also in multiple lineages of the subgenus Drosophila
344 (**Fig. 4**B). Indeed, the general *BaraA* gene structure encoding IM24, tandem repeats

345 of IM10-like peptides, and IM22 is conserved in *S. lebanonensis* and all *Drosophila*
346 species (**Fig. 4**C). On the other hand, the *Dmel\BaraC* gene comes from an ancient
347 duplication restricted to the subgenus Sophophora, and *Dmel\BaraB* resulted from a
348 more recent duplication found only in the Melanogaster group (**Fig. 4**B).
349 We originally found outgroup *Baramicins* by reciprocal BLAST searches, and
350 screened *BaraA-like* genes encoding the full suite of Baramicin peptides for immune
351 induction (i.e. encoding IM24, IM10-likes, and IM22: expression in **Fig.** 1). However
352 following genomic synteny analysis, we realized that the *D. willistoni BaraA*-like
353 gene *Dwil\GK10648* is syntenic with the *Dmel\BaraC* locus (**Fig. 4**A), yet this gene is
354 immune-induced (**Fig.** 1D) and retains a *BaraA*-like gene structure (**Fig. 4**C). On the
355 other hand, *Dwil\GK10645* is found at the locus syntenic with *BaraA*, but has
356 undergone an independent truncation to encode just an IM24 peptide (similar to
357 *Dmel\BaraC*). Thus these two *D. willistoni* genes have evolved similar to *D.*
358 *melanogaster BaraA/BaraC*, but in a vice versa fashion. This suggests a pattern of
359 convergent evolution with two key points: **i)** the duplication event producing
360 *Dmel\BaraA* and *Dmel\BaraC* originally copied a full-length *BaraA*-like gene to the
361 *BaraC* locus, and **ii)** the derivation of an IM24-specific gene structure has occurred
362 more than once (*Dmel\BaraC* and *Dwil\GK10645*). Indeed, another independent
363 IM24-specific *Baramicin* gene is present in *D. virilis* (*Dvir\GJ25897*), which is a direct
364 sister of the *BaraA*-like gene *Dvir\GJ21309* (the signal peptides of these genes is
365 identical at the nucleotide level, and see **Fig. 4**C). Thus *Baramicins* in both *D.*
366 *willistoni* and *D. virilis* have convergently evolved towards an IM24-specific protein
367 structure resembling *Dmel\BaraC*. We checked the expression of these truncated
368 *Baramicins* in each species upon infection. As was the case for *Dmel\BaraC*, neither
369 gene is immune-induced (**Fig. S5**A-C). Given the glial expression of *Dmel\BaraC,* we
370 reasoned that the heads of adult flies (rich in nerve tissue) should be enriched in
371 *BaraC* compared to whole animals. Indeed we saw a significant enrichment of *BaraC*
372 in the heads of *D. melanogaster* males compared to whole flies, which was not the
373 case for *BaraA* (**Fig. S5**D). When we checked the heads of *D. willistoni* and *D. virilis*,
374 we indeed saw a consistent and significant enrichment in the head for the IM24-

375 specific genes *Dwil\GK10645* and *Dvir\GJ25897*, while *BaraA*-like genes were more
376 variable in expression (**Fig. S5**E-F).
377 Genomic synteny shows the gene structure and immune expression of *BaraA*
378 are the ancestral state, and *Dmel\BaraB* and *Dmel\BaraC* are paralogs derived from
379 independent duplication events. Strikingly, we observe a parallel evolution of
380 expression pattern and gene structure in *Baramicins* of *D. willistoni* and *D. virilis*.
381 Moreover these independent IM24-specific *Baramicins* across species are not
382 immune induced, and are enriched in the head. Expression data across genes and
383 species are shown in **Fig.** S5 and summarized in **Fig. 4**C.
384
385 ***Residue 29 in the IM24 domain evolves in lineage-specific fashions***
386 Thus far we have shown that IM24-specific genes are expressed in the
387 nervous system, yet IM24 is the only peptide domain conserved across all *Baramicin*
388 genes. We were therefore interested to better understand the properties of IM24 to
389 know if any evolutionary patterns might distinguish the IM24 domains of nervous
390 system-expressed genes from IM24 domains of immune-induced genes. We were
391 unable to model the protein satisfactorily with various protein prediction
392 techniques, preventing a 3D comprehension of the IM24 peptide. Therefore we
393 asked if we could highlight any residues in this traditionally immune peptide that
394 might correlate with nervous system or immune-induced gene lineages to better
395 understand what aspect of IM24 contributes to it being retained in neural contexts.
396 To do this, we screened for positive selection (elevated non-synonymous
397 mutation rate) in the IM24 domain using the HyPhy package implemented in
398 Datamonkey.org (Delport et al. 2010) using separate codon alignments of *Baramicin*
399 IM24 domains beginning at their conserved Q^1 starting residue. As is recommended
400 with the HyPhy package (Delport et al. 2010), we employed multiple statistical
401 approaches including Likelihood (FEL), Bayesian (FUBAR), and Count-based (SLAC)
402 analyses to ensure patterns in selection analyses were robust to different methods
403 of investigation. Specifically, we used locus-specific alignments (e.g. genes at the
404 *stum* locus in **Fig. 4**B were all analyzed together) to ensure IM24 evolution reflected
405 locus-specific evolution. FEL, FUBAR, and SLAC site-specific analyses each suggest

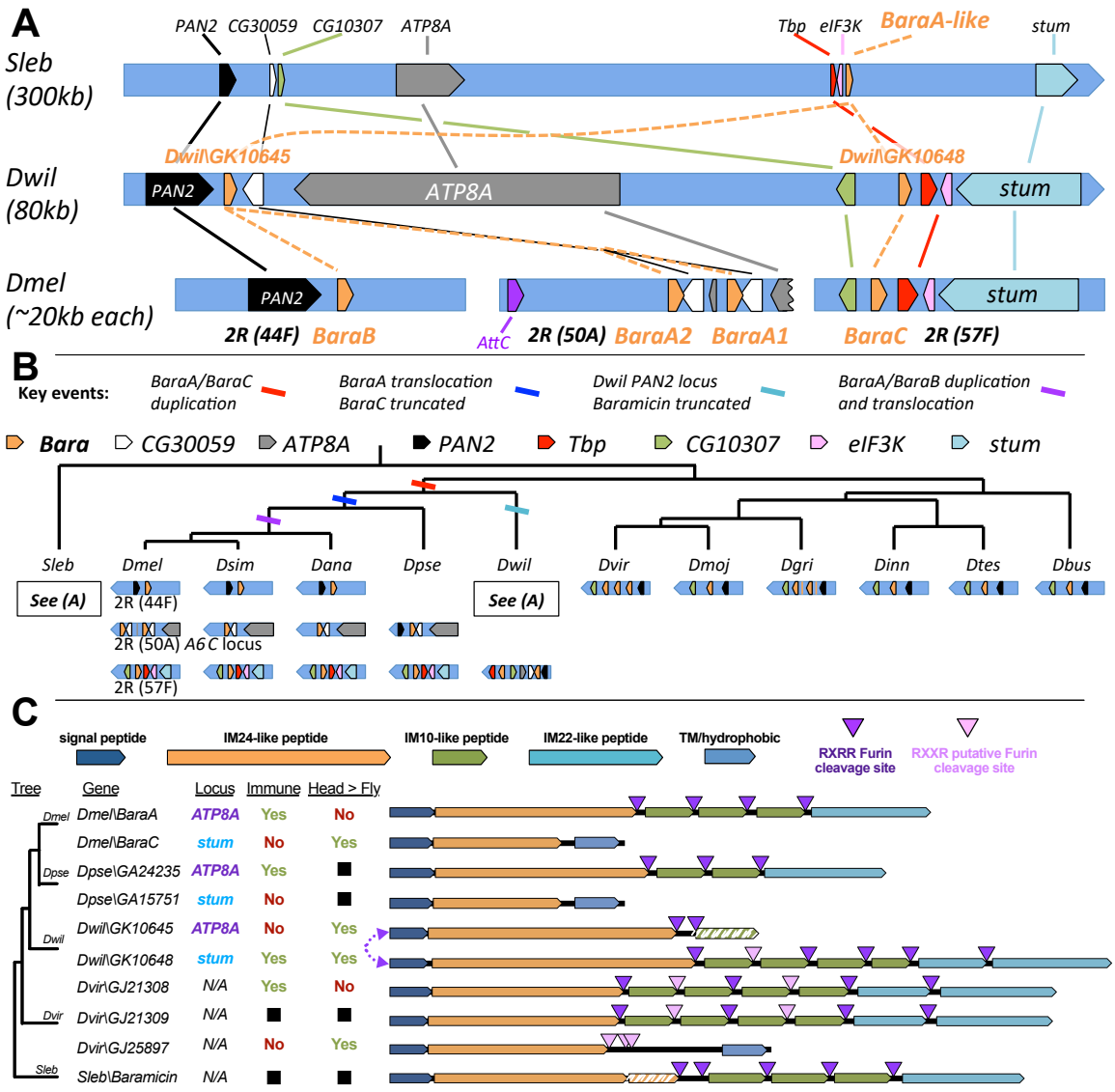


Figure 4: *Baramicin* evolutionary history. A) Detailed map of genomic neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D. melanogaster*, detailing inferred duplication, inversion, and translocation events. Gene names are given as found in *D. melanogaster*. B) Cladogram 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).

422 strong purifying selection in many residues of the IM24 domain (p-adj < .05, data in
423 **supplementary data file 3**), agreeing with the general protein structure of IM24
424 being broadly conserved (**Fig.** 5A). However one residue (site 29) was consistently
425 highlighted as evolving under positive selection using each type of statistical
426 approach for genes located at the Sophophora *ATP8A* locus (*BaraA* genes and
427 *Dwil\GK10645*: p-adj < .05; **Fig. 5**A). This site is universally Proline in *Baramicin*
428 genes located at the stum locus (*BaraC*-like), in both *D. willistoni Baramicins*, and in
429 the outgroup *S. lebanonensis*, suggesting Proline is the ancestral state. However this
430 residue diverges in both the *BaraA* (commonly Threonine) and *BaraB* (commonly
431 Valine) lineages. We also note that two sites on either side of site 29 (site 27 and site
432 31) similarly diverge by lineage in an otherwise highly conserved region of the IM24
433 domain. FUBAR analysis (but not FEL or SLAC) similarly found evidence of positive
434 selection at site 31 in the *BaraA* locus genes (p-adj = .026). Thus this neighbouring
435 site could also be evolving in a non-random fashion. Similar analyses of the *BaraB*
436 and *stum* loci *Baramicins* did not find evidence of site-specific positive selection.
437 We highlight site 29 as a key residue in IM24 that diverged in *Baramicin* in
438 lineage-specific fashions. This ancestrally Proline residue has settled on a Threonine
439 in most *BaraA*-like genes of Obscura and Melanogaster group flies, and a Valine in
440 most *BaraB* genes, which are unique to the Melanogaster group. The ancestral
441 Proline residue is found in both *D. willistoni Baramicins*, alongside significant
442 enrichment of both genes in the head, despite only one gene being immune-induced
443 (**Fig.** 4C, **Fig.** S5). Thus it is unclear how this site contributes to tissue-specific
444 *Baramicin* functions, but Threonine and Valine residues evolved in the *BaraA* and
445 *BaraB* lineages.
446
447 ***Another IM24 domain in Baramicin lineages varies through relaxed selection***
448 Visual inspection of aligned IM24 proteins makes it evident that the overall
449 IM24 domain is broadly conserved, except in sites 40-48 (**Fig. 5**A). However a motif
450 aligned to residues ^{40}HHASSPAD^{48} of *Dmel\BaraB* departs in lineage-specific
451 fashions; the three C-terminal residues of this motif are also diagnostic of each gene
452 lineage (*BaraA, BaraB,* and *BaraC* have RGE, PXE, or (S/N)GQ respectively; **Fig. 5**A).

453 However even with additional branch-site selection analyses (aBSREL and BUSTED
454 (Murrell et al. 2015)), we found no evidence of positive selection at this motif, and in
455 fact many residues also failed to show evidence of purifying selection. For instance,
456 six of nine sites in the *BaraA* locus analysis failed to reach significance ($p < .05$) for
457 purifying selection in SLAC analysis (supplementary data file 3).
458 Given an absence of positive selection, and many residues failing to reach
459 significance for purifying selection at the residue 40-48 motif, we this motif is
460 diversifying due to drift effects through relaxed selection. The high conservation of
461 IM24 residues up- and downstream of this motif is nevertheless striking. One
462 possible explanation may be that these residues act as a linker between the two
463 functional parts of IM24 up- and downstream of residues 40-48. Perhaps supporting
464 this interpretation, we found that *D. yakuba BaraB* independently lost immune
465 induction alongside an insertion at site 40 (**Fig.** 5). Such speculation awaits
466 validation by robust protein modelling efforts.
467
468 ***Overt structural change best explains Baramicin loss of immune induction***
469 We found that site 29 evolves rapidly in *Baramicin* lineages, but this site is
470 common to both immune and non-immune Baramicin genes (e.g. in *D. willistoni*).
471 Thus IM24 sequence variation does not explain why IM24-specific Baramicins lose
472 immune inducibility. However within the *BaraA/BaraB* lineage, we observed that
473 *BaraB* genes commonly encode Valine at IM24 site 29, compared to Threonine in
474 *BaraA*. As the *BaraB* locus is derived from an ancestral immune-induced (**Fig.** 4A), it
475 is unclear if other *BaraB* genes are immune-inducible, and thus what IM24
476 evolutionary patterns (like Valine at site 29) might predict *BaraB* functional
477 divergence.
478 We therefore performed infection experiments in diverse species across the
479 Melanogaster group to see if their *BaraB* genes had similarly lost immune induction
480 (see **Fig.** S6 for qPCR data). Surprisingly, we found that the non-immune expression
481 of *Dmel\BaraB* is extremely recent, as Melanogaster sister species like *D. sechellia*
482 and *D. mauritiana* encode immune inducible *BaraB* loci (summary in **Fig.** 5B).
483 However, we also found that *D. simulans BaraB* lacked immune induction, despite *D.*

484 *simulans* being most closely related to *D. sechellia* (Chakraborty et al. 2021). Thus
485 IM24 sequence evolution does not predict immune induction.
486 This drew our attention to the overall protein structure of the various extant *BaraB*
487 genes. A striking feature of the Dmel\BaraB protein is the absence of a functional
488 signal peptide (**Fig.** 2B). This signal peptide sequence is conserved in all *Baramicin*
489 lineages, except in *Dmel/BaraB* and also *Dsim/BaraB*. Indeed despite *D. simulans*
490 being more closely related to *D. sechellia* and *D. mauritiana*, both *Dmel\BaraB* and
491 *Dsim\BaraB* encode a homologous N-terminus of parallel length (**Fig.** 5A). We also
492 found that *Dyak\BaraB* is not immune-responsive, but note that *D. yakuba* has an
493 insertion upstream of residue 40 that elongates the IM24 domain (**Fig.** 5B black X
494 boxes), and its sister species *D. erecta* encodes multiple indels and premature stops
495 suggesting *BaraB* is pseudogenized in this lineage. Loss of the BaraB signal peptide
496 is therefore more specifically associated with loss of immune expression in the
497 Melanogaster species complex (*D. simulans, D. sechellia, D. mauritiana,* and *D.*
498 *melanogaster*). The last common ancestor of *D. simulans, D. sechellia,* and *D.*
499 *mauritiana* is estimated to be just ~250,000 years ago, and these species diverged
500 from *D. melanogaster* ~3 million years ago (Chakraborty et al. 2021). The fact that *D.*
501 *simulans* uniquely encodes this *Dmel\BaraB*-like sequence suggests it was either
502 introgressed from one species to the other prior to the complete development of
503 hybrid inviability, or reflects incomplete lineage sorting of this locus in the
504 Melanogaster species complex.
505 Overall, we find no evidence to suggest IM24 domain sequence has evolved
506 drastically to allow for function in the nervous system. The Proline residue at site 29
507 of neural *Baramicin* genes appears to be the ancestral state. Rather than small
508 sequence changes, overt structural changes like truncation to focus on the IM24
509 domain and loss of a signal peptide in *Dmel\BaraB* are associated with *Baramicins*
510 expressed in the nervous system. After duplication, *Baramicin* daughter lineages
511 have repeatedly derived neural-specific expression through subfunctionalization of
512 the IM24 domain from the overall precursor protein. Importantly, this finding
513 suggests that the ancestral *Baramicin* encoded peptides with distinct roles in either
514 the immune response or the nervous system.

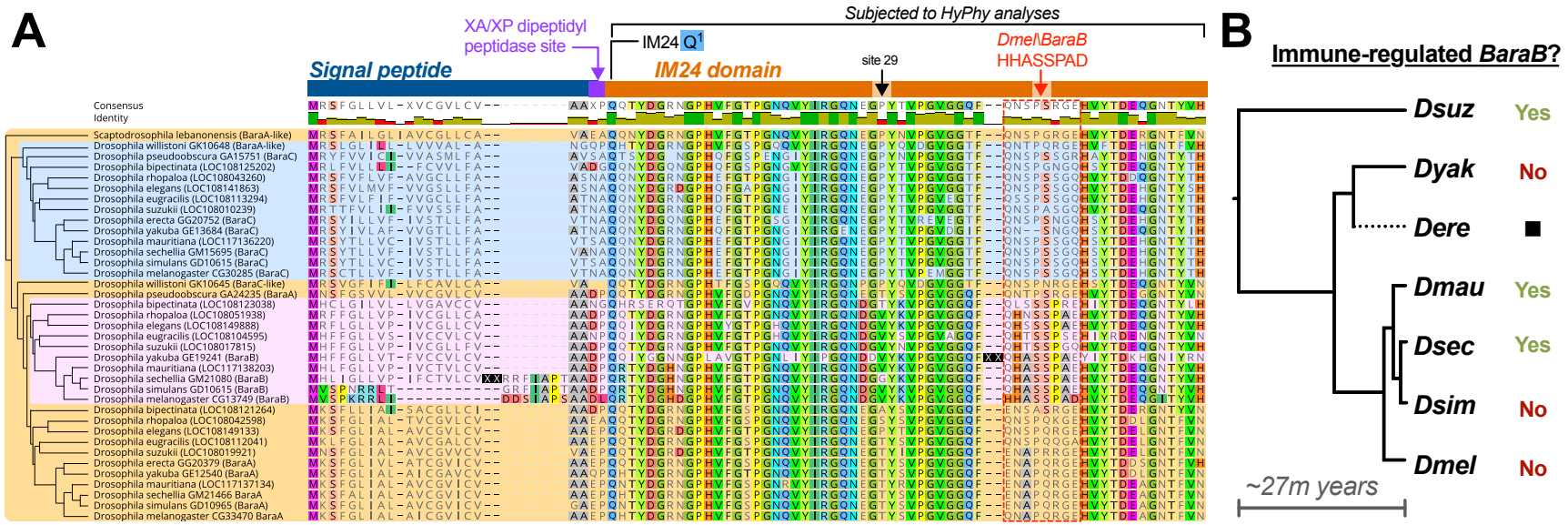


Figure 5: *BaraB* and IM24 rapid evolution. A) Evolution of the signal peptide and core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*. Arrows indicate site 29 and the *D. melanogaster* ^{40}HHASSPAD^{48} domain. Insertion events in *D. yakuba* and *D. sechellia BaraB* are denoted as XX to save space. While longer than other Baramicin proteins, 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.

526 # Discussion

527 Recent studies have suggested AMP genes drive behaviour or disease
528 through interactions with the nervous system. Many of these genes encode
529 polypeptides with multiple mature products. To date, little attention has been paid
530 to AMP genes in these neural contexts at the level of the sub-peptides they encode.
531 Here we demonstrate that the *Baramicin* antimicrobial peptide gene of *Drosophila*
532 ancestrally encodes distinct peptides involved with either the nervous system or the
533 immune response. Given the ancestral *Baramicin* encoded all Baramicin peptide
534 domains, this antimicrobial peptide gene has always produced sub-peptides that
535 interact with either the nervous system (IM24) or invading pathogens (IM10-like,
536 IM22). These peptides are maturated from a longer precursor protein, accomplished
537 via furin cleavage. Importantly, this suggests that AMP genes can mediate these
538 distinct neural and immune roles via specialized sub-peptides, and not necessarily
539 due to dual action of a single peptide.

540 The goal of generating different gene products is commonly accomplished
541 through alternate transcript isoforms. For instance, the insect *doublesex* gene
542 determines sexual differentiation, with male- or female-specific isoforms mediated
543 by alternative splicing. Select hymenopteran lineages have even lost coding
544 sequence for the male-specific exon (Baral et al. 2019). These context-specific
545 isoforms are similar in theme to loss of immune-relevant peptides in non-immune
546 *Baramicin* genes, except that in *Baramicins* this is achieved through gene
547 duplication. Moreover the 'protein operon' structure of immune-induced *Baramicins*
548 can act as a mechanism to allow peptides with distinct roles to be produced
549 simultaneously from a single mRNA transcript.

550 There is building evidence that immune-induced AMPs and AMP-like genes
551 affect the nervous system. Loss of Metchnikowin protects flies from
552 neurodegeneration after traumatic brain injury (Swanson et al. 2020), *Induced by*
553 *infection (IBIN)* regulates behavioural changes in flies after seeing parasitoid wasps
554 (Ebrahim et al. 2021), epidermal nematode AMPs trigger motor neuron autophagy
555 (Lezi et al. 2018) and sleep (Sinner et al. 2021) after infection, and loss of *Diptericin*

556 *B* produced by the fat body leads to memory deficits in *Drosophila* (Barajas-azpeleta
557 et al. 2018). This last example is intriguing, as *Diptericin B* also encodes a
558 polypeptide maturated by furin cleavage, and its effect on memory was derived from
559 peptide secreted into the hemolymph by the fat body and not from neural
560 expression. Similarly, we recently found that *BaraA* deletion causes infected flies to
561 display an erect wing behavioural phenotype, which was independent of active
562 infection, and could be rescued by priming the hemolymph with *BaraA* expressed by
563 the fat body (Hanson et al. 2021). Thus some component of *BaraA* likely interacts
564 with some host target(s) to prevent this behaviour during the immune response.
565 The present study suggests this could be due to action of BaraA IM24, given
566 evidence for IM10-like peptides as antifungal peptides (Hanson et al. 2021), and loss
567 of these peptides in nervous system-specific *Baramicins*. However AMPs and
568 neuropeptides have many similar features, including cationic charge and
569 amphipathicity (Brogden et al. 2005). Thus while we present evidence that IM24 of
570 different *Baramicin* genes underlies *Baramicin* interactions with the nervous
571 system, we cannot exclude the possibility that IM24 is also antimicrobial.
572 One human AMP recently implicated in chronic neuroinflammatory disease is
573 the Cathelicidin LL-37 (Lee et al. 2015; De Lorenzi et al. 2017; Moir et al. 2018). Like
574 *Baramicin*, the *Cathelicidin* gene family is unified by its N-terminal "Cathelin"
575 domain. However to date no one has described antimicrobial activity of the Cathelin
576 domain in vitro (Zanetti 2005). Instead, *Cathelicidin* research has focused almost
577 exclusively on the mature peptide LL-37 at the C-terminus of mammalian
578 *Cathelicidin* genes. Reflecting on *Baramicin* evolution and the implication of
579 *Cathelicidin* in neurodegenerative diseases, what does the Cathelin domain do?
580 While this study was conducted in fruit flies, we hope we have emphasized the
581 importance of considering each peptide of AMP genes for in vivo function. This is
582 relevant to neural processes even if the gene is typically thought of for its role in
583 innate immunity. Indeed, recent studies of *Drosophila* AMPs have emphasized that in
584 vitro activity does not always predict the interactions that these genes can have in
585 vivo (Clemmons et al. 2015; Hanson, Dostálová, et al. 2019). Care should be taken
586 not to conflate in vitro activity with realized in vivo function. Most studies focus on

587 AMPs specifically in an immune role, but this is akin to 'looking for your keys under
588 the streetlight.' To understand AMP functions in vivo, genetic approaches will be
589 necessary that allow a more global view of gene function.
590 In summary, we reflect on the structural characteristic of AMP genes through
591 the lens of *Baramicins*. We found that one sub-peptide of the immune-induced
592 *Baramicin* ancestor is readily adapted for functions relating to the nervous system.
593 Meanwhile, other sub-peptides known to suppress fungi are repeatedly lost in
594 daughter genes that lack immune inducibility, suggesting they are irrelevant to
595 neural functions. As AMP genes commonly encode polypeptides maturated by furin
596 cleavage (including *Baramicin*), it will be interesting to consider the functions of
597 AMP genes in neural processes not simply at the level of the gene, but at the level of
598 the mature peptides produced by that gene. This consideration may explain how
599 some polypeptide immune effectors play dual roles in disparate contexts.
600
601 **Data availability statement**
602 All data are included within the manuscript and supplementary files.

Acknowledgements

We would like to thank Maria Litovchenko for advice, Ana Marija Jakšić for generously providing DGRP flies, 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: $0.0001 < p\text{-value} < 0.005$ 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

648 performed in both directions to test for an effect of the X or Y chromosomes on
649 *BaraB*-mediated lethality; crosses in both directions yielded similar results in all
650 cases and reported data are pooled results. Fly diet consisted of a nutrient-rich lab
651 standard food: 3.72g agar, 35.28g cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL
652 propionic acid, 15.9mL moldex, and H_2O to 600mL.
653
654 ***Infection experiments***
655 Bacteria and yeast were grown to mid-log phase shaking at 200rpm in their
656 respective growth media (LB, BHI, or YPG) and temperature conditions, and then
657 pelleted by centrifugation to concentrate microbes. Resulting cultures were diluted
658 to OD = 200 at 600nm before infections to measure gene expression. The following
659 microbes were grown at 37°C: *Escherichia coli strain 1106* (LB) and *Candida albicans*
660 (YPG). *Micrococcus luteus* was grown at 29°C in LB. For **Fig.** 1 and S2, pooled fly
661 samples were collected either 6 hours post-infection (*E. coli*) or 24 hours post-
662 infection (*C. albicans, M. luteus*) prior to RNA extraction on pools of 5 adult males.
663 These timepoints correspond to the maximal expression inputs of the Imd (6hpi) or
664 Toll (24hpi) NF-κB signalling pathways, which are most specifically induced by
665 Gram-negative bacteria (Imd) or Gram-positive bacteria or fungi (Toll) (Lemaitre et
666 al. 1997). Flies were pricked in the thorax as described in (Hanson, Dostálová, et al.
667 2019).
668 RNA extractions were performed using TRIzol™, Ambion DNAse treatment,
669 and PrimeScript RT according to manufacturer's protocols. RT-qPCR was performed
670 using PowerUP SYBR Green master mix with primers listed in supplementary data
671 file 5. Gene expression differences were analyzed using the PFAFFL method (Pfaffl
672 2001). For gene expression experiments requiring dissection of heads, pools of 20
673 males were used for either whole flies or heads dissected in ice-cold PBS and
674 transferred immediately to a tube kept on dry ice.
675
676 ***Selection analysis using HyPhy package***
677
678 Codon aligned nexus tree files were generated using either the Neighbour-
679 joining (1000 bootstraps) or PhyML (100 bootstraps) methods including proteins
680 beyond those shown in **Fig. 5**. These tree files were analyzed using the HyPhy
681 package with only 174nt pertaining to just the IM24 domain codons included. The
682 cladogram in **Fig. 5**A is manually drawn from known species divergences (Kim et al.
683 2021). Use of either tree building method was chosen for convenience to best reflect
684 known lineage sorting, as use of just 174nt was too information poor to resolve
685 exact phylogenetic relatedness reliably. Tree files were qualitatively screened to
686 ensure topologies broadly matched known species sortings, and thus ensure only
687 relevant comparisons were made given the genomic synteny analysis in **Fig. 4** is
688 principally informative of true gene lineages. HyPhy analyses were run separately
689 for each *Baramicin* lineage within their clade, defined by genomic synteny; i.e. based
690 on locus (e.g. ATP8A locus), and not considering convergent gene structures. We
691 used three site-specific analyses (FEL, FUBAR, and SLAC) that use three independent
692 statistical approaches (Likelihood, Bayesian, and Count-based methods

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

724 **References:**
725

726 Abbott A. 2020. Are infections seeding some cases of Alzheimer's disease? *Nature*
727 587:22–25.

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

730 Baral S, Arumugam G, Deshmukh R, Kunte K. 2019. Genetic architecture and sex-
731 specific selection govern modular, male-biased evolution of *doublesex*. *Sci.*
732 *Adv.* 5:eaau3753.

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

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

738 Chakraborty M, Chang C-H, Khost DE, Vedanayagam J, Adrion JR, Liao Y, Montooth
739 KL, Meiklejohn CD, Larracuente AM, Emerson JJ. 2021. Evolution of genome
740 structure in the Drosophila simulans species complex. *Genome Res* 31:380–
741 396.

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

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

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

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

754 Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: A
755 suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*
756 26:2455–2457.

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

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

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

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

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

770 Halldórsdóttir K, Árnason E. 2015. Trans-species polymorphism at antimicrobial
771 innate immunity cathelicidin genes of Atlantic cod and related species. *PeerJ*
772 3:e976.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

816 Lee M, Shi X, Barron AE, McGeer E, McGeer PL. 2015. Human antimicrobial peptide
817 LL-37 induces glial-mediated neuroinflammation. *Biochemical Pharmacology*
818 94:130–141.

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

823 Lezi E, Zhou T, Koh S, Chuang M, Sharma R, Pujol N, Chisholm AD, Eroglu C,
824 Matsunami H, Yan D. 2018. An Antimicrobial Peptide and Its Neuronal
825 Receptor Regulate Dendrite Degeneration in Aging and Infection. *Neuron*
826 97:125-138.e5.

827 Li H, Janssens J, De Waegeneer M, Kolluru SS, Davie K, Gardeux V, Saelens W, David
828 F, Brbić M, Leskovec J, et al. 2021. Fly Cell Atlas: a single-cell transcriptomic

829 atlas of the adult fruit fly. Genomics Available from:
830 http://biorxiv.org/lookup/doi/10.1101/2021.07.04.451050

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

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

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

839 Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, Griffin RS, Belfer I,
840 Dai F, Smith SB, et al. 2010. A Genome-wide Drosophila screen for heat
841 nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell*
842 143:628–638.

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

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

848 Quesada H, Ramos-Onsins SE, Aguad?? M. 2005. Birth-and-death evolution of the
849 Cecropin multigene family in Drosophila. *Journal of Molecular Evolution* 60:1–
850 11.

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

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

857 Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. 2021. Innate Immunity
858 Promotes Sleep through Epidermal Antimicrobial Peptides. *Current Biology*
859 31:564-577.e12.

860 Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK,
861 Waddell P, Lang M, Courtier-Orgogozo V, et al. 2021. Widespread
862 introgression across a phylogeny of 155 Drosophila genomes. *Current*
863 *Biology*:S0960982221014962.

Swanson LC, Rimkus SA, Ganetzky B, Wassarman DA. 2020. Loss of the Antimicrobial Peptide Metchnikowin Protects Against Traumatic Brain Injury Outcomes in Drosophila melanogaster. *G3 (Bethesda)* 10:3109–3119.

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

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

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

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

Van Wetering S, Mannesse-Lazeroms SPG, Van Sterkenburg MAJA, Hiemstra PS. 2002. Neutrophil defensins stimulate the release of cytokines by airway epithelial cells: Modulation by dexamethasone. In: Inflammation Research. Vol. 51. p. 8–15.

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

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

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