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;

42 Hanson et al. 2016; Chapman et al. 2019). The selective pressures that drive these
43 evolutionary outcomes are likely the consequence of host-pathogen interactions
44 (Unckless et al. 2016). However AMPs and AMP-like genes in various species have
45 also recently been implicated in various non-immune roles in flies, nematodes, and
46 emerging evidence in humans. These new contexts suggest that the evolutionary
47 forces acting on AMP genes may not be driven strictly by trade-offs in host defence,
48 but rather by conflicts between roles in immunity and other non-immune functions.
49 For instance, Diptericins are membrane-disrupting antimicrobial peptides of
50 flies (Diptera) that are required for defence against infection by *Providencia* bacteria
51 (Unckless et al. 2016; Hanson, Dostálová, et al. 2019). It was therefore surprising
52 that the *D. melanogaster* gene *Diptericin B (DptB)* affects memory processes
53 (Barajas-azpeleta et al. 2018). In this study, *DptB* derived from the fly fat body
54 (analogous to the mammalian liver) regulated the ability of the fly to form long-term
55 memory associations (Barajas-azpeleta et al. 2018). Another AMP-like gene, nemuri,
56 regulates fly sleep and promotes survival upon infection (Toda et al. 2019). Studies
57 in nematodes have also shown that an immune-induced polypeptide (NLP-29) binds
58 to a G-protein coupled receptor (NPR-12) triggering neurodegeneration through
59 activation of the NPR-12-dependent autophagy pathway (Lezi et al. 2018), and
60 injury triggers epidermal AMPs including NLP-29 to promote sleep (Sinner et al.
61 2021). *Drosophila* AMPs have also recently been shown to regulate behaviours after
62 seeing parasitoid wasps (Ebrahim et al. 2021), during feeding with different bacteria
63 (Kobler et al. 2020), or following infection (Hanson et al. 2021). In humans, the
64 *Cathelicidin* gene encodes the AMP LL-37, which is implicated in glia-mediated
65 neuroinflammation and Alzheimer's disease (Lee et al. 2015; De Lorenzi et al. 2017),
66 alongside evidence of Alzheimer's being an infectious syndrome (Dominy et al.
67 2019); though the importance of this process is debated (Abbott 2020). Notably,
68 AMPs share a number of properties with classic neuropeptides (Brogden et al.
69 2005), further muddying the distinction between peptides of the immune and
70 nervous systems.
71 We recently described a novel antifungal peptide gene of *Drosophila*
72 *melanogaster* that we named *Baramicin A* (*BaraA*) (Hanson et al. 2021). A unique

73 aspect of *BaraA* is its precursor protein structure, which encodes a polypeptide
74 cleaved into multiple mature products by interspersed furin cleavage sites. The use
75 of furin cleavage sites to produce two mature peptides from a single polypeptide
76 precursor is widespread in animal AMP genes (Gerdol et al. 2020; Hanson and
77 Lemaitre 2020). However, *BaraA* represents an exceptional case as multiple tandem
78 repeat peptides are produced from the translation of a single coding sequence,
79 effectively resembling a "protein-based operon"; this tandem repeat structure has
80 also been found in two other AMPs of bees and flies (Casteels-Josson et al. 1993;
81 Hanson et al. 2016). The immature precursor protein of *D. melanogaster BaraA*
82 encodes three types of domains: an IM24 domain, three tandem repeats of IM10-like
83 domains, and an IM22 domain. *BaraA* mutants are susceptible to infection by fungi,
84 and *in vitro* experiments suggest the *BaraA* IM10-like peptides have antifungal
85 activity (Hanson et al. 2021). The other *Baramicin* domains encoding IM22 and
86 IM24 remain uncharacterized. Curiously, *BaraA* deficient flies also display an erect
87 wing behavioural phenotype upon immune stimulation even in the absence of
88 infection, suggesting that *BaraA* products could have non-microbial targets (Hanson
89 et al. 2021).
90 In this study, we describe the evolution of the Drosophilid *Baramicin* gene
91 family. Three unique *Baramicin* genes (*BaraA, B, and C*) are present in the genome of
92 *D. melanogaster*. Surprisingly, only *BaraA* is immune-induced, while *BaraB* and
93 *BaraC* are enriched in the nervous system. Both *BaraB* and *BaraC* have truncations
94 compared to the ancestral *Baramicin* gene, which focuses these genes towards
95 producing the *Baramicin* IM24 domain. We found similar truncations in other
96 species, and upon checking their patterns of expression, realized these overt gene
97 structure changes correlate with loss of immune expression and enrichment in the
98 nervous system. By resolving the genomic synteny of the various *Baramicin* genes in
99 different species, we confirmed that these repeated truncations focusing on IM24
100 production stem from independent events (convergent evolution). The exaggerated
101 'protein operon' polypeptide nature of *Baramicin* draws attention to the unique
102 roles that different mature peptides of AMP-encoding genes can play. Careful

103 attention paid to the multiple peptide products of AMP genes could explain how
104 these immune effectors contribute to both immune and neurological processes.
105

106 Results

107 ***Baramicin is an ancestral immune effector***

108 The *Baramicin A* gene was only recently described as encoding antifungal
109 effectors by our group (Hanson et al. 2021), and another recent study also
110 confirmed *Baramicin*'s important contribution to Toll immune defence (Huang et al.
111 2020). These initial characterizations were done only in *D. melanogaster*, and only
112 focused on one *Baramicin* gene. We will therefore first provide a basic description of
113 the immune *Baramicins* of other species and also the larger *Baramicin* gene family of
114 *D. melanogaster* to establish that this is a classically immune gene family, and that
115 deviations from immune function are derived.

116 In *D. melanogaster, BaraA* is regulated by the Toll immune signalling pathway
117 (Huang et al. 2020; Hanson et al. 2021). Using BLAST, we recovered *BaraA*-like
118 genes encoding each Baramicin peptide (IM24, IM10-like, and IM22) across the
119 genus *Drosophila* and in the outgroup *Scaptodrosophila lebanonensis*. We performed
120 infection experiments to confirm that these *BaraA*-like genes were immune-
121 inducible by infecting the diverse species *D. melanogaster*, *D. pseudoobscura*, *D.*
122 *willistoni*, *D. virilis*, and *D. neotesteacea* (last common ancestor ~63mya (Tamura et
123 al. 2004)) with *Micrococcus luteus* and *Candida albicans*, two microbes that stimulate
124 the Toll pathway (**Fig.** 1A). In all five species, *BaraA*-like genes were immune-
125 induced (**Fig.** 1B-F). We therefore confirm the ancestral *Baramicin* was an immune-
126 induced gene.
127

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

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

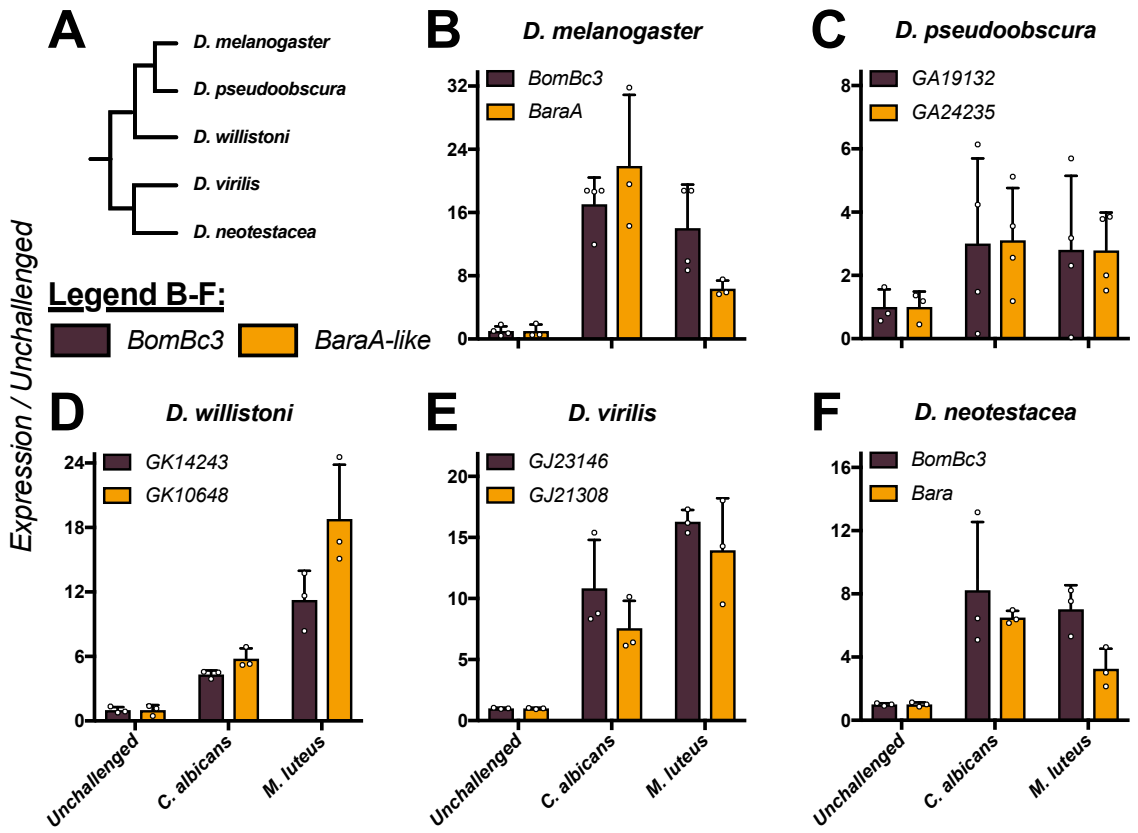

Figure 1: The ancestral *BaraA* gene was immune-induced. A) Cladogram of species used in B-F. B-F) Expression of *BomBc3* control genes (brown) or *BaraA-like* genes (orange) in diverse *Drosophila* species upon infection. In all cases, both *BomBc3* and *BaraA-like* genes are induced upon infection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

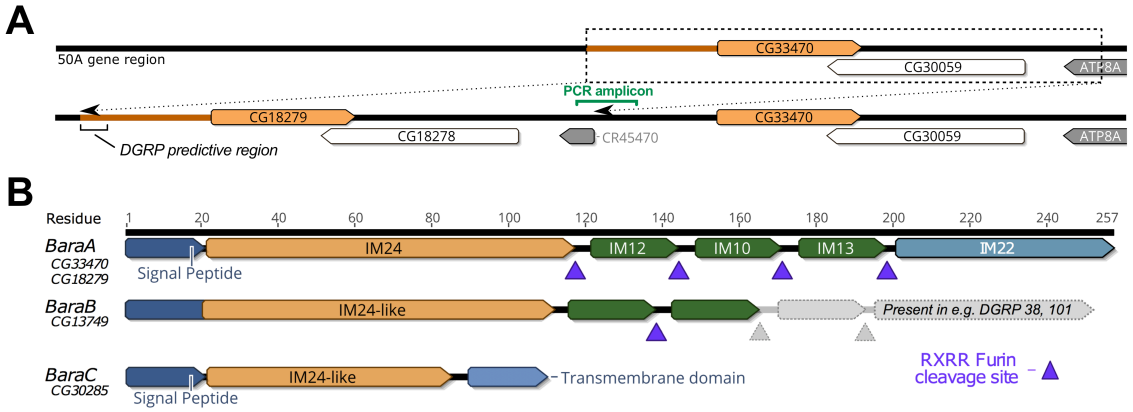

Figure 2: The *D. melanogaster Baramicin* genes. A) Schematic of the *BaraA* duplication. Using a PCR assay spanning the duplication-specific locus (PCR amplicon), we confirmed *BaraA* copy number is variable in various lab strains (Table S1) and wild-caught flies (Supplementary data file 1). B) *D. 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).

146 *D. melanogaster* R6 genome assembly encodes two 100% identical *BaraA* genes
147 (*CG33470* and *CG18279*, *BaraA1* and *BaraA2* respectively). We screened 132 DGRP
148 lines for the *BaraA* duplication event, finding only ~14% (18/132) of strains were
149 PCR-positive for two *BaraA* copies (supplementary data file 1). Perhaps as a
150 consequence of the identical sequences of these two genes, this genome region is
151 poorly resolved in RNA sequencing studies and the Drosophila Genetic Reference
152 Panel (DGRP, see **Fig.** S1) (Mackay et al. 2012; Leader et al. 2018). Because this
153 region is poorly resolved, it is unclear if our PCR assay might be sensitive to cryptic
154 sequence variation. However our PCR screen nevertheless confirms that this region
155 is variable in the wild, and we additionally note that common fly strains seem to
156 differ in their *BaraA* copy number (Table S1), where extra gene copies correlated
157 with increased expression after infection (see (Hanson et al. 2021) S10 Fig).
158 We also recovered two paralogous *Baramicin* genes in *D. melanogaster*
159 through reciprocal BLAST searches: *CG13749* and *CG30285*, which we name *BaraB*
160 and *BaraC* respectively (**Fig.** 2B). The three *Baramicin* gene loci are scattered on the
161 right arm of chromosome II at cytological positions 44F9 (*BaraB*), 50A5 (*BaraA*),
162 and 57F8 (*BaraC*). These paralogous *Baramicins* are united by the presence of the
163 IM24 domain. In the case of *BaraB*, we additionally recovered a frameshift mutation
164 (2R_4821599_INS) causing a premature stop segregating in the DGRP leading to the
165 loss of IM13 and IM22 relative to the *BaraA* gene structure (**Fig. 2**B); this truncation
166 is present in the Dmel_R6 genome assembly, but many DGRP strains encode a CDS
167 with either a standard (e.g. DGRP38) or extended (e.g. DGRP101) IM22 domain (a
168 DGRP *BaraB* alignment is provided in supplementary data file 2). Moreover, in
169 contrast to BaraA, the initial IM10-like peptide of BaraB no longer follows a furin
170 cleavage site, and encodes a serine (R<u>S</u>XR) in its IM10-like motif instead of the
171 universal proline (R<u>P</u>XR) of *BaraA*-like IM10 peptides across the genus. Each of
172 these mutations prevents the secretion of classical IM10-like and IM22 peptides by
173 *BaraB*. Finally, *BaraC* encodes only IM24 tailed by a transmembrane domain at the
174 C terminus (TMHMM v2.0 (Krogh et al. 2001)), and thus lacks both the IM10-like
175 peptides and IM22 (**Fig. 2**B).
176

177 ***BaraB and BaraC are not immune-inducible***

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

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

199

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

201 A simple interpretation of the truncated gene structure and low levels of
202 *BaraB* expression is that this gene is undergoing pseudogenization. Indeed, AMP
203 gene pseudogenization is common in insects including *Drosophila* (Quesada et al.
204 2005; Rolff and Schmid-Hempel 2016; Hanson, Lemaitre, et al. 2019). To explore

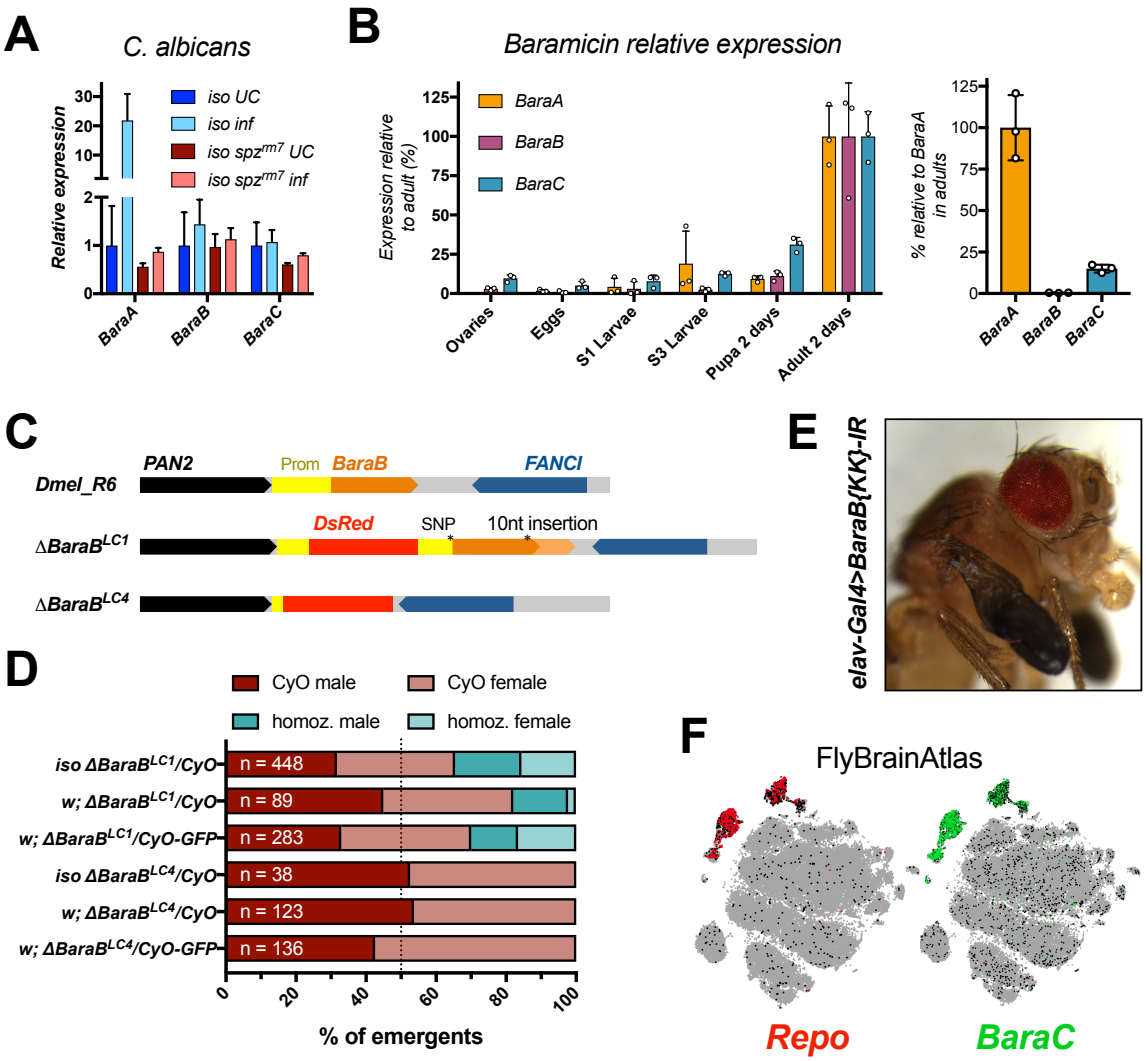

Figure 3: *D. melanogaster* non-immune *Baramicins* have neural functions. A) Only *BaraA* is immune-induced. *BaraB* and *BaraC* do not respond to infection, though basal *BaraC* expression relies on Toll signalling (Fig. S3C). B) Time series of whole animal *Baramicin* expression over the course of 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

222 double gRNA approach to replace the *BaraB* locus with sequence from the pHD-
223 DsRed vector. The *ΔBaraB*^{LC1} and *ΔBaraB*^{LC4} mutations differ in their ultimate effect,
224 as *ΔBaraB*^{LC1} is an incidental insertion of the DsRed cassette in the promoter of the
225 gene. This disruption reduces gene expression, resulting in a hypomorph state (**Fig.**
226 **S3**A). The *ΔBaraB*^{LC4} mutation however deletes the locus as intended, leading to
227 *BaraB* null flies (**Fig. 3**C).
228 We further introgressed both *ΔBaraB* mutations into the DrosDel isogenic
229 background (referred to as *iso*) for seven generations according to Ferreira et al.
230 (Ferreira et al. 2014). At the same time, we combined the original *ΔBaraB*
231 chromosomes with a CyO-GFP balancer chromosome in an arbitrary genetic
232 background to distinguish homozygous/heterozygous larvae. In all cases, *ΔBaraB*^{LC4}
233 homozygotes were entirely lethal during larval development, whereas the
234 hypomorphic *ΔBaraB*^{LC1} flies allowed for homozygous adults to emerge (**Fig. 3**D).
235 We further assessed *ΔBaraB*^{LC1} hypomorph viability using crosses between
236 *ΔBaraB*^{LC1}*/CyO* heterozygous females and *ΔBaraB*^{LC1} homozygous males, which
237 showed reduced viability and was exacerbated by rearing at 29°C (**Fig. S3**B). Using
238 our *CyO-GFP* reporter to track genotypes in larvae revealed that the major lethal
239 phase occurs primarily in the late larval and pupal stages (**Fig. S3**C-F), agreeing with
240 a role for *BaraB* in larvae/pupae previously suggested by increased expression at
241 this stage. Some emergent flies also exhibited locomotor defects, and/or a nubbin-
242 like wing phenotype (FlyBase: FBrf0220532 and e.g. in **Fig. 3**E) where the wings
243 were stuck in a shrivelled state for the remainder of the fly's lifespan. However, a
244 plurality of *ΔBaraB*^{LC1} homozygotes successfully emerged, and unlike their siblings,
245 had no immediate morphological or locomotory defects. The lifespan of
246 morphologically normal *iso ΔBaraB*^{LC1} adults is nevertheless significantly shorter
247 compared to wild-type flies and *iso ΔBaraB*^{LC1}*/CyO* siblings (**Fig. S4**G). We confirmed
248 these developmental defects using ubiquitous gene silencing with *Actin5C-Gal4 (Act-*
249 *Gal4)* to drive two *BaraB* RNAi constructs (*TRiP-IR* and *KK-IR*). Both constructs
250 resulted in significant lethality and occurrence of nubbin-like wings (**Table** S2).
251 Genomic deficiency crosses also confirmed significantly reduced numbers of
252 eclosing *BaraB*-deficient flies at 25°C (n = 114, χ^2 p < .001) and 29°C (n = 63, χ^2 p <

253 .001) (**Fig. S3**H). Thus full gene deletion is lethal in the larval/pupal transition stage,
254 and *BaraB* hypomorph flies suffer significant costs to fitness during development,
255 and have reduced lifespan even following successful eclosion.
256 These data demonstrate a significant cost of *BaraB* disruption. While whole-
257 fly *BaraB* expression is low, these results suggest that *BaraB* is not pseudogenized,
258 and instead performs an integral developmental role. The fact that there is a
259 bimodal outcome in hypomorph-like $\Delta BaraB^{LC1}$ adults (either severe defects or
260 generally healthy) suggests *BaraB* is involved in passing some checkpoint during
261 larval/pupal development. Flies deficient for *BaraB* may be more likely to fail at this
262 developmental checkpoint, resulting in either lethality or developmental defects.
263
264 ***The Baramicin paralogs BaraB and BaraC are expressed in the nervous system***
265 We next sought to determine in which tissue(s) *BaraB* is required. A previous
266 screen using neural RNA interference highlighted *BaraB* for lethality effects (n = 15)
267 (Neely et al. 2010). Given this preliminary result, and alongside our observed *BaraB*
268 mutant locomotory defects, we started by silencing *BaraB* in the nervous system at
269 25°C or at 29°C for greater efficiency using the pan-neural *elav-Gal4* driver both the
270 *TRiP-IR* and *KK-IR BaraB-IR* lines. We additionally combined this approach with
271 *UAS-Dicer2* (*Dcr2*) to further strengthen gene silencing as used previously (Neely et
272 al. 2010). In the event there was no lethality, it was expected that emerging
273 *elav>TRiP-IR* flies would follow simple mendelian inheritance. However both
274 *elav>TRiP-IR* and *elav>Dcr2, TRiP-IR* resulted in partial lethality and occasional
275 nubbin-like wings (χ^2 $p < .02$, **Table** S2). Crosses using *KK-IR* used homozygous
276 flies, and so we did not assess lethality using mendelian inheritance. However using
277 this construct, no adults emerged when *elav>Dcr2, KK-IR* flies were reared at 29°C.
278 Rare emergents (N = 11 after three experiments) occurred at 25°C, all of which bore
279 nubbin-like wings. Using *elav-Gal4* at 29°C without *Dcr2*, we observed greater
280 numbers of emerging adults, but 100% of flies had nubbin-like wings (**Fig. 3**E,
281 **Table** S2). Finally, *elav>KK-IR* flies at 25°C suffered both partial lethality and
282 nubbin-like wings, but normal-winged flies began emerging (χ^2 $p < .001$, **Table** S2).

283 This analysis indicates that *BaraB* is expressed in the nervous system, and this
284 expression readily explains both the lethality and nubbin-like wing phenotypes.
285 Moreover, we observed a consistent spectrum of developmental defects using *elav-*
286 *Gal4>BaraB-IR* wherein strength of gene silencing correlates with severity of
287 lethality and wing defect frequency. We additionally investigated the effect of *BaraB*
288 RNAi using Gal4 drivers in non-neural tissues including the fat body (*c564-Gal4*),
289 hemocytes (*hml-Gal4*), the gut (esg-Gal4), malpighian tubules (MyO-Gal4), the wing
290 disc (*nubbin-Gal4*), and in myocytes (*mef2-gal4*) to no effect. We also screened
291 neural drivers specific for glia (*Repo-Gal4*), motor neurons (*D42-, VGMN-,* and *OK6-*
292 *Gal4*), and a recently-made *BaraA-Gal4* driver that is expressed in the larval ventral
293 nervous system (Hanson et al. 2021). However all these *Gal4>BaraB-IR* flies were
294 viable and never exhibited overt morphological defects.
295 We also screened for effects of *BaraC* disruption using ubiquitous *Act-Gal4*
296 and neural *elav-Gal4>Dcr2* for developmental defects. However neither driver
297 produced overt phenotypes in morphology or locomotor activity (not shown).
298 Tissue-specific transcriptomic data indicate that *BaraC* is expressed in various
299 neural tissues including the eye, brain, and the thoracic abdominal ganglion (**Fig.**
300 **S4**A), but also the hindgut and rectal pads pointing to a complex expression pattern
301 (Hammonds et al. 2013; Leader et al. 2018). We next searched FlyBrainAtlas (Davie
302 et al. 2018) to narrow down which neural subtypes *BaraB and BaraC* were
303 expressed in. *BaraB* expressing cells were few and showed only low expression in
304 this dataset. However *BaraC* was robustly expressed in all glial cell types, fully
305 overlapping the glia marker *Repo* (**Fig. 3**F). To confirm the observation that *BaraC*
306 was expressed in glia, we compared the effects of *BaraC* RNA silencing (*BaraC-IR*)
307 using *Act-Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia) drivers on
308 *BaraC* expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that of control
309 flies (**Fig. S4**B). By comparison *elav-Gal4* reduced *BaraC* expression to ~63% that of
310 controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls (overall
311 controls vs. neural/glia-IR, p = .002).

312 Collectively, our results support the notion that *BaraC* is expressed in the
313 nervous system, and are consistent with *BaraC* expression being most localized to
314 glial cells.
315
316 ***Extensive genomic turnover of the Baramicin gene family***
317 Our results thus far show that *BaraA*-like genes are consistently immune-
318 induced in all *Drosophila* species (**Fig.** 1), however the two paralogs *Dmel\BaraB*
319 and *Dmel\BaraC* are not immune-induced, and are truncated in a fashion that
320 deletes some or all of the antifungal IM10-like peptides (**Fig. 2**B). These two
321 *Baramicins* are now enriched in the nervous system (**Fig. 3**E-F). In the case of *BaraB*,
322 a role in the nervous system is evidenced by severe defects recapitulated using pan-
323 neural RNA silencing. In the case of *BaraC*, nervous system expression is evidenced
324 by a clear overlap with *Repo*-expressing cells.
325 While *BaraA*-like genes are conserved throughout the genus *Drosophila*,
326 *BaraB* is conserved only in Melanogaster group flies, and *BaraC* is found only in
327 Melanogaster and Obscura group flies, indicating that both paralogs stem from
328 duplication events of a *BaraA*-like ancestor (**Fig. 4**). To determine the ancestry of
329 each *D. melanogaster Baramicin* gene, we traced their evolutionary history by
330 analyzing genomic synteny through hierarchical orthologous groups (Train et al.
331 2019). Ancestry tracing revealed that these three loci ultimately stem from a single-
332 locus ancestor encoding only one *Baramicin* gene that resembled *Dmel\BaraA* (**Fig.**
333 **4**A). This is evidenced by the presence of only a single *BaraA*-like gene in the
334 outgroup *S. lebanonensis*, and also in multiple lineages of the subgenus Drosophila
335 (**Fig. 4**B). Indeed, the general *BaraA* gene structure encoding IM24, tandem repeats
336 of IM10-like peptides, and IM22 is conserved in *S. lebanonensis* and all *Drosophila*
337 species (**Fig. 4**C). On the other hand, the *Dmel\BaraC* gene comes from an ancient
338 duplication restricted to the subgenus Sophophora, and *Dmel\BaraB* resulted from a
339 more recent duplication found only in the Melanogaster group (**Fig. 4**B).
340 We originally recovered outgroup *Baramicins* assayed for immune induction
341 (**Fig.** 1) through reciprocal BLAST searches. However following genomic synteny
342 analysis, we realized that the *D. willistoni BaraA*-like gene *Dwil\GK10648* is syntenic

343 with the *Dmel\BaraC* locus (**Fig. 4**A), yet this gene is immune-induced (**Fig.** 1D) and
344 retains a *BaraA*-like gene structure (**Fig. 4**C). On the other hand, *Dwil\GK10645* is
345 found at the locus syntenic with *BaraA*, but has undergone an independent
346 truncation to encode just an IM24 peptide (similar to *Dmel\BaraC*). Thus these two
347 *D. willistoni* genes have evolved similar to *D. melanogaster BaraA/BaraC*, but in a
348 vice versa fashion. This suggests a pattern of convergent evolution with two key
349 points: **i)** the duplication event producing *Dmel\BaraA* and *Dmel\BaraC* originally
350 copied a full-length *BaraA*-like gene to both loci, and **ii)** the derivation of an IM24-
351 specific gene structure has occurred more than once (*Dmel\BaraC* and
352 *Dwil\GK10645*). Indeed, another independent IM24-specific *Baramicin* gene is
353 present in *D. virilis* (*Dvir\GJ25897*), which is a direct sister of the *BaraA*-like gene
354 *Dvir\GJ21309* (the signal peptides of these genes is identical at the nucleotide level,
355 and see **Fig. 4**C). Thus *Baramicins* in both *D. willistoni* and *D. virilis* have
356 convergently evolved towards an IM24-focused protein structure resembling
357 *Dmel\BaraC*. We checked the expression of these truncated *Baramicins* in each
358 species upon infection. As was the case for *Dmel\BaraC*, neither gene is immune-
359 induced (**Fig. S5**A-C). Given the glial expression of *Dmel\BaraC,* we reasoned that
360 the heads of adult flies (rich in nerve tissue) should be enriched in *BaraC* compared
361 to whole animals. Indeed we saw a significant enrichment of *BaraC* in the heads of *D.*
362 *melanogaster* males compared to whole flies, which was not the case for *BaraA* (**Fig.**
363 **S5**D). When we checked the heads of *D. willistoni* and *D. virilis*, we indeed saw a
364 consistent and significant enrichment in the head for the IM24-specific genes
365 *Dwil\GK10645* and *Dvir\GJ25897*, while *BaraA*-like genes were more variable in
366 expression (**Fig. S5**E-F).
367 Thus, multiple independent IM24-specific *Baramicins* are not immune
368 induced and are more specifically enriched in the head. In the case of *Dmel\BaraC*,
369 this is likely due to expression in glia. Strikingly, we observe a parallel evolution of
370 expression pattern and gene structure in *Baramicins* of *D. willistoni* and *D. virilis*.
371 These expression data are summarized in **Fig. 4**C. Genomic synteny shows the gene
372 structure and immune expression of *BaraA* are the ancestral state, and *Dmel\BaraB*
373 and *Dmel\BaraC* are paralogs derived from independent duplication events.

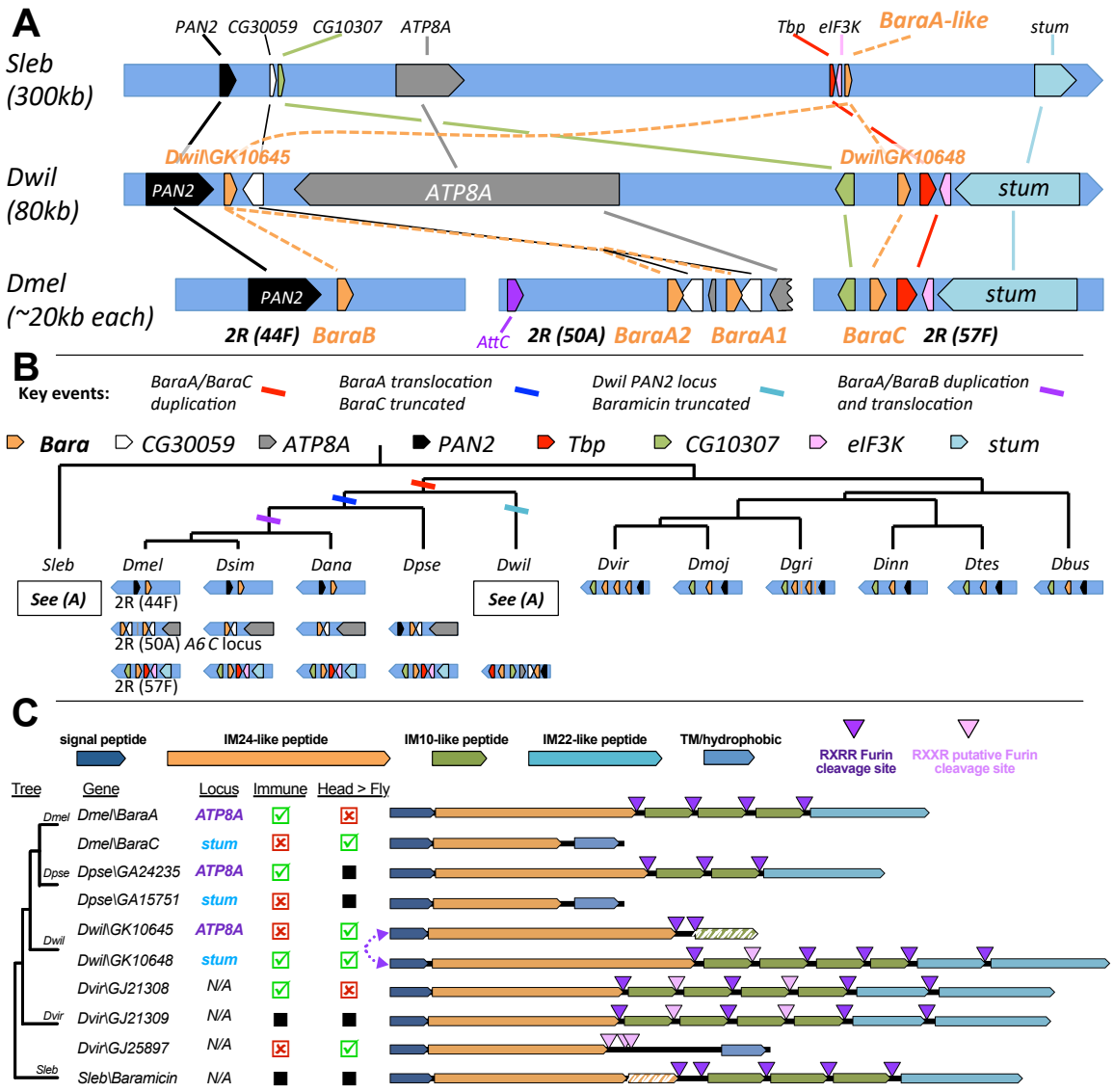

Figure 4: *Baramicin* evolutionary history. A) Detailed map of genomic neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D. melanogaster*, detailing inferred duplication, inversion, and translocation events. Gene names are given as found in *D. melanogaster*. B) Cladogram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

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

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

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

421 using each type of statistical approach for genes located at the Sophophora *ATP8A*
422 locus (*BaraA* genes and *Dwil\GK10645*: p-adj < .05; **Fig. 5**A). This site is universally
423 Proline in *Baramicin* genes located at the stum locus (*BaraC*-like) and in the
424 outgroup *S. lebanonensis*, but is variable in both the *BaraA* (commonly Threonine)
425 and *BaraB* (commonly Valine) lineages. Both the *S. lebanonensis* and the two *D.*
426 *willistoni Baramicins* encode Proline at site 29 independent of gene structure,
427 suggesting Proline is the ancestral state. We also note that two sites on either side of
428 site 29 (site 27 and site 31) similarly diverge by lineage in an otherwise highly
429 conserved region of the IM24 domain. FUBAR analysis (but not FEL or SLAC)
430 similarly found evidence of positive selection at site 31 in the *BaraA* locus genes (p-
431 adj = .026). Thus this neighbouring site could also be evolving in a non-random
432 fashion. Similar analyses of the *BaraB* and *stum* loci *Baramicins* did not find evidence
433 of site-specific positive selection.
434 While the structure of IM24 is unknown, HyPhy analysis highlights site 29 as
435 a key residue in IM24 that diverged in *Baramicin* lineage-specific fashions. This
436 ancestrally Proline residue has settled on a Threonine in most *BaraA*-like genes of
437 Obscura and Melanogaster group flies, and a Valine in most *BaraB* genes, which are
438 unique to the Melanogaster group.
439
440 ***Another IM24 domain in Baramicin lineages varies through relaxed selection***
441 Visual inspection of aligned IM24 proteins makes it evident that the overall
442 IM24 domain is broadly conserved, except in sites 40-48 (**Fig. 5**A). This motif
443 uniquely encodes the residues ^{40}HHASSPAD^{48} in *Dmel\BaraB*. Given the severe cost
444 of *BaraB* mutation, intriguingly the H^{40} and D^{48} residues are not found in any other
445 *Baramicin* genes. The three C-terminal residues of this motif are also diagnostic for
446 each gene lineage (*BaraA, BaraB,* and *BaraC* have RGE, PXE, or (S/N)GQ
447 respectively; **Fig. 5**A). However even with additional branch-site selection analyses
448 (aBSREL and BUSTED (Murrell et al. 2015)), we found no evidence of positive
449 selection at the ^{40}HHASSPAD^{48} homologous domain (supplementary data file 3).
450 Thus while IM24 residues 40-48 are variable across lineages, this motif is not
451 evolving with elevated non-synonymous change. We suspect instead that this motif

452 is diversifying due to relaxed selection as six of nine sites in the *BaraA* locus analysis
453 failed to reach significance ($p < .05$) for purifying selection in e.g. SLAC analysis
454 (supplementary data file 3). It is nevertheless striking that this region is so variable
455 given the conservation of residues upstream and downstream of sites 40-48. This
456 pattern should have implications on the IM24 protein functional motifs, which
457 future protein folding investigations may decipher.
458
459 ***Overt IM24 structural change best explains loss of immune induction***
460 Site 29 varies in lineage-specific fashions, encoding a derived Valine residue
461 in most species' BaraB IM24 domains. If the Valine at site 29 explains the *BaraB*
462 functional divergence relative to its sister *BaraA* lineage, this could suggest that
463 *BaraB* has long functioned in a neural role common to most Melanogaster group
464 flies. To this end, we performed infection experiments in diverse species across the
465 Melanogaster group to see if their *BaraB* genes had similarly lost immune induction
466 (see **Fig.** S6 for qPCR data). Surprisingly, we instead found that the non-immune
467 expression of *Dmel\BaraB* is extremely recent, as Melanogaster sister species like *D.*
468 *sechellia* and *D. mauritiana* nevertheless encode immune inducible *BaraB* loci
469 (summary in **Fig.** 5B). However, we also found that *D. simulans BaraB* lacked
470 immune induction, despite being most closely related to *D. sechellia*. This drew our
471 attention instead to the overall protein structure of the various *BaraB* genes. A
472 striking feature of the Dmel\BaraB protein is the absence of a signal peptide
473 structure (**Fig.** 2B). Signal peptide sequence is conserved in all *Baramicin* lineages,
474 except in *BaraB* of *D. melanogaster* and also *D. simulans* (last common ancestor
475 ~3mya (Chakraborty et al. 2021)). Indeed despite *D. simulans* being more closely
476 related to *D. sechellia* and *D. mauritiana*, both *Dmel\BaraB* and *Dsim\BaraB* encode a
477 homologous N-terminus of parallel length (**Fig.** 5A). We also found that *D. yakuba*
478 *BaraB* is not immune-responsive, but note that *D. yakuba* has an insertion upstream
479 of residue 40 that elongates the IM24 domain (**Fig.** 5B black X boxes), and its sister
480 species *D. erecta* encodes multiple indels and premature stops suggesting *BaraB* is
481 pseudogenized in this lineage. While an insertion is also present in *D. sechellia*
482 *BaraB* in the signal peptide, this is still predicted to allow secretion (SignalP 5.0),
483 suggesting *Dsec\BaraB* is a functional immune protein.

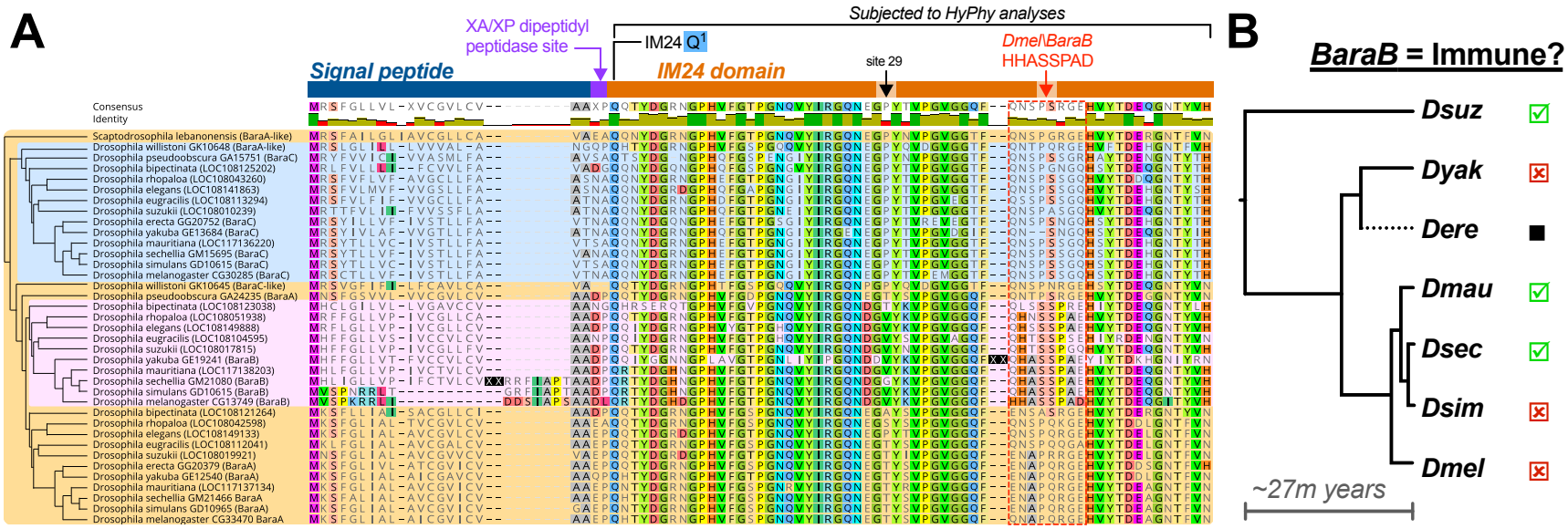

Figure 5: *BaraB* and IM24 rapid evolution. A) Evolution of the signal peptide and core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*. Arrows indicate site 29 and the *D. melanogaster* ^{40}HHASSPAD^{48} domain. Insertion events in *D. yakuba* and *D. sechellia BaraB* are denoted 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.

496 Loss of the BaraB signal peptide is therefore more specifically associated
497 with loss of immune expression in the Melanogaster species complex (*D. simulans, D.*
498 *sechellia, D. mauritiana,* and *D. melanogaster*). The last common ancestor of *D.*
499 *simulans, D. sechellia,* and *D. mauritiana* is estimated to be just ~250,000 years ago,
500 and these species diverged from *D. melanogaster* ~3 million years ago (Chakraborty
501 et al. 2021). The fact that *D. simulans* uniquely encodes this *Dmel\BaraB*-like
502 sequence suggests it was either introgressed from one species to the other prior to
503 the complete development of hybrid inviability, or reflects incomplete lineage
504 sorting of this locus in the Melanogaster species complex. In either case, this points
505 to an extremely young age for the novel function of *Dmel\BaraB* in the nervous
506 system. This loss of the signal peptide also occurs alongside a segregating allele that
507 truncates the mature *Dmel\BaraB* sequence (**Fig.** 2B), a pattern commonly found in
508 *Baramicins* derived for neural expression (**Fig.** 4C). This reinforces the fact that the
509 *Dmel\BaraB* gene had an immune function so recently that some wild flies still
510 produce the immune-relevant IM10-like and IM22 Baramicin peptides despite
511 neural expression of *BaraB*.
512 *BaraB* evolution therefore reinforces that the core driver of *Baramicin*
513 functional divergence is not based on minor sequence changes, but rather correlates
514 with overt protein structural change. *BaraB* has a mutation affecting secretion, while
515 *BaraC* now encodes a transmembrane domain, which should cause it to insert itself
516 into either its endogenous glial cell membrane or a neighbouring cell (e.g. a neuron).
517 In both cases IM24 is preferentially expressed and localized to the nervous system.
518

519 # Discussion

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

526 Here we find the *Baramicin* IM24 domain has a predilection for interactions
527 with the nervous system. We speculate that the immune-mediated production of
528 IM24 by *BaraA* could protect the nervous system from autoimmune activity, which
529 occurs in the absence of *BaraA* to cause erect wing display. A notable aspect of this
530 hypothesis is it proposes that some peptides of AMP genes are responsible for
531 microbe killing, while others are co-secreted with the intent of preventing
532 autoimmune toxicity. For now this remains speculation, however it will be
533 interesting to clarify the mature structure of IM24 and determine what partner(s)
534 IM24 binds to. In this regard, we highlight site 29 as an important residue for IM24
535 function, and suggest that while residues 40-48 are variable, the sequence at this
536 motif does not experience the same sort of evolutionary selection. One possibility to
537 explain these evolutionary patterns is that site 29 is exposed in some way that an
538 IM24 binding partner uses. Meanwhile residues 40-48 could act as a sort of linker
539 between the two ends of the IM24 domain where the length of these residues is
540 important, but the exact sequence is malleable. Indeed, we found that *D. yakuba*
541 *BaraB* independently lost immune induction alongside an insertion at site 40.
542 Antimicrobial peptide genes have recently been implicated by a number of
543 studies in neural functions, regulating processes like memory, sleep, taste aversion,
544 behaviour, and neurodegeneration. The properties of these immune peptides share
545 many features with classic neuropeptides, including cationic charge and
546 amphipathicity (Brogden et al. 2005). Nevertheless it is unclear why AMPs can play
547 dual roles in either immunity or neural function. By characterizing the evolution of
548 the *Baramicin* gene family, we provide insight on how an ancestrally immune AMP
549 gene has adapted itself for neural function on a repeated basis. The mechanism
550 through which *Baramicin* achieves its neural effect is specific to the IM24 domain, as
551 the antifungal IM10-like peptides and IM10-related IM22 peptide are consistently
552 lost in *Baramicin* lineages now specific to neural expression. This realization is made
553 possible by the exaggerated polypeptide structure of *Baramicins,* which focuses the
554 interpretation on how different peptides of this modular protein structure may play
555 roles in either neurology or immunity. Other AMP genes similarly encode
556 polypeptides, but do not have so many tandem repeats of identical peptides and so

557 the polypeptide nature of their precursor protein is easily glossed over. The
558 polypeptide nature of these genes is lacking from the current conversation on AMP
559 involvements in the nervous system, despite AMP genes of fruit flies and other
560 animals encoding furin-cleaved polypeptides (Hanson and Lemaitre 2020).
561 One human AMP recently implicated in chronic neuroinflammatory disease is
562 the Cathelicidin LL-37 (Lee et al. 2015; De Lorenzi et al. 2017; Moir et al. 2018). Like
563 *Baramicin*, the *Cathelicidin* gene family is unified by its N-terminal "Cathelin"
564 domain. However to date no one has described antimicrobial activity of the Cathelin
565 domain in vitro (Zanetti 2005). Instead, *Cathelicidin* research has focused almost
566 exclusively on the mature peptide LL-37 at the C-terminus of mammalian
567 *Cathelicidin* genes. Reflecting on *Baramicin* evolution and the implication of
568 *Cathelicidin* in neurodegenerative diseases, what does the Cathelin domain do?
569 While this study was conducted in fruit flies, we hope we have emphasized the
570 importance of considering each peptide of AMP genes for in vivo function. This is
571 relevant to neural processes even if the gene is typically thought of for its role in
572 innate immunity. Indeed, recent studies of *Drosophila* AMPs have emphasized that in
573 vitro activity does not always predict the interactions that these genes can have in
574 vivo (Clemmons et al. 2015; Hanson, Dostálová, et al. 2019). Care should be taken
575 not to conflate in vitro activity with realized in vivo function. Most studies focus on
576 AMPs specifically in an immune role, but this is akin to 'looking for your keys under
577 the streetlight.' To understand AMP functions in vivo, genetic approaches will be
578 necessary that allow a more global view of gene function.
579 In summary, we characterize how an ancestral AMP-encoding gene has
580 repeatedly evolved for neural expression by truncating its protein sequence to
581 express just one peptide. It will be interesting to consider the functions of AMP
582 genes in neural processes not simply at the level of the gene, but at the level of the
583 mature peptides produced by that gene. Given the polypeptide of many AMP gene
584 structures and commonalities between AMPs and neuropeptides, these canonical
585 immune effectors may be adapted for neural function more often than appreciated.
586

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: $0.0001 < p\text{-value} < 0.005$ 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

632 indicated. When possible, genetic crosses of 6-8 males and 6-8 females were
633 performed in both directions to test for an effect of the X or Y chromosomes on
634 *BaraB*-mediated lethality; crosses in both directions yielded similar results in all
635 cases and reported data are pooled results. Fly diet consisted of a nutrient-rich lab
636 standard food: 3.72g agar, 35.28g cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL
637 propionic acid, 15.9mL moldex, and H_2O to 600mL.
638
639 ***Infection experiments***
640 Bacteria and yeast were grown to mid-log phase shaking at 200rpm in their
641 respective growth media (LB, BHI, or YPG) and temperature conditions, and then
642 pelleted by centrifugation to concentrate microbes. Resulting cultures were diluted
643 to OD = 200 at 600nm before infections to measure gene expression. The following
644 microbes were grown at 37°C: *Escherichia coli strain 1106* (LB) and *Candida albicans*
645 (YPG). *Micrococcus luteus* was grown at 29°C in LB. For **Fig.** 1 and S2, pooled fly
646 samples were collected either 6 hours post-infection (*E. coli*) or 24 hours post-
647 infection (*C. albicans, M. luteus*) prior to RNA extraction on pools of 5 adult males.
648 These timepoints correspond to the maximal expression inputs of the Imd (6hpi) or
649 Toll (24hpi) NF-κB signalling pathways, which are most specifically induced by
650 Gram-negative bacteria (Imd) or Gram-positive bacteria or fungi (Toll) (Lemaitre et
651 al. 1997). Flies were pricked in the thorax as described in (Hanson, Dostálová, et al.
652 2019).
653 RNA extractions were performed using TRIzol™, Ambion DNAse treatment,
654 and PrimeScript RT according to manufacturer's protocols. RT-qPCR was performed
655 using PowerUP SYBR Green master mix with primers listed in supplementary data
656 file 5. Gene expression differences were analyzed using the PFAFFL method (Pfaffl
657 2001). For gene expression experiments requiring dissection of heads, pools of 20
658 males were used for either whole flies or heads dissected in ice-cold PBS and
659 transferred immediately to a tube kept on dry ice.
660
661 ***Selection analysis using HyPhy package***
662
663 Codon aligned nexus tree files were generated using either the Neighbour-
664 joining (1000 bootstraps) or PhyML (100 bootstraps) methods including proteins
665 beyond those shown in **Fig. 5**. These tree files were analyzed using the HyPhy
666 package with only 174nt pertaining to just the IM24 domain codons included. The
667 cladogram in **Fig. 5**A is manually drawn from known species divergences (Kim et al.
668 2021). Use of either tree building method was chosen for convenience to best reflect
669 known lineage sorting, as use of just 174nt was too information poor to resolve
670 exact phylogenetic relatedness reliably. Tree files were qualitatively screened to
671 ensure topologies broadly matched known species sortings, and thus ensure only
672 relevant comparisons were made given the genomic synteny analysis in **Fig. 4** is
673 principally informative of true gene lineages. HyPhy analyses were run separately
674 for each *Baramicin* lineage within their clade, defined by genomic synteny; i.e. based
675 on locus (e.g. ATP8A locus), and not considering convergent gene structures. We
676 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

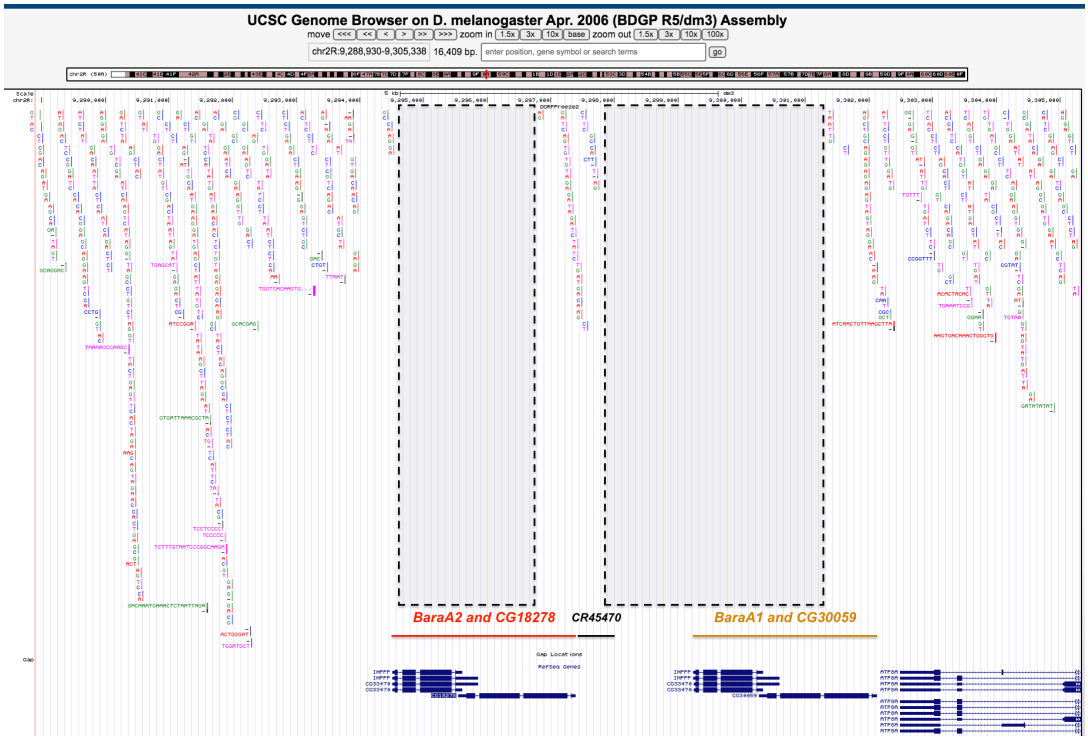

Figure S1: The *BaraA* locus is poorly resolved in DGRP genome assemblies. The *BaraA1* and *BaraA2* gene regions are totally devoid of mapped variants (dashed boxes). We speculate this is due to an artefact during genomic assembly, where reads mapping equally to the two identical *BaraA* genes were 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.

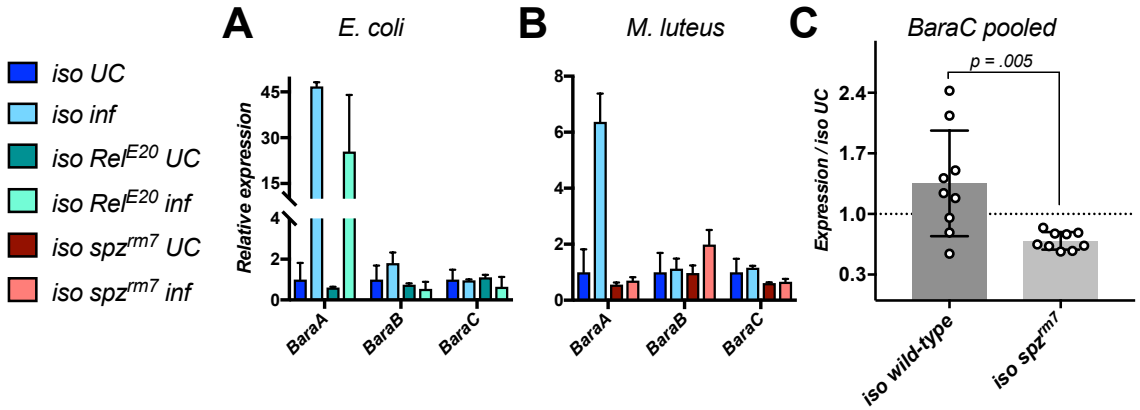

Figure S2: Additional assays of Baramicin expression upon infection. A) Neither *BaraB* nor *BaraC* are regulated by the Imd pathway, which is specifically stimulated by *E. coli* infection. **B)** Neither *BaraB* nor *BaraC* are induced after infection by *M. luteus*. **C)** *BaraC* levels were consistently 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).

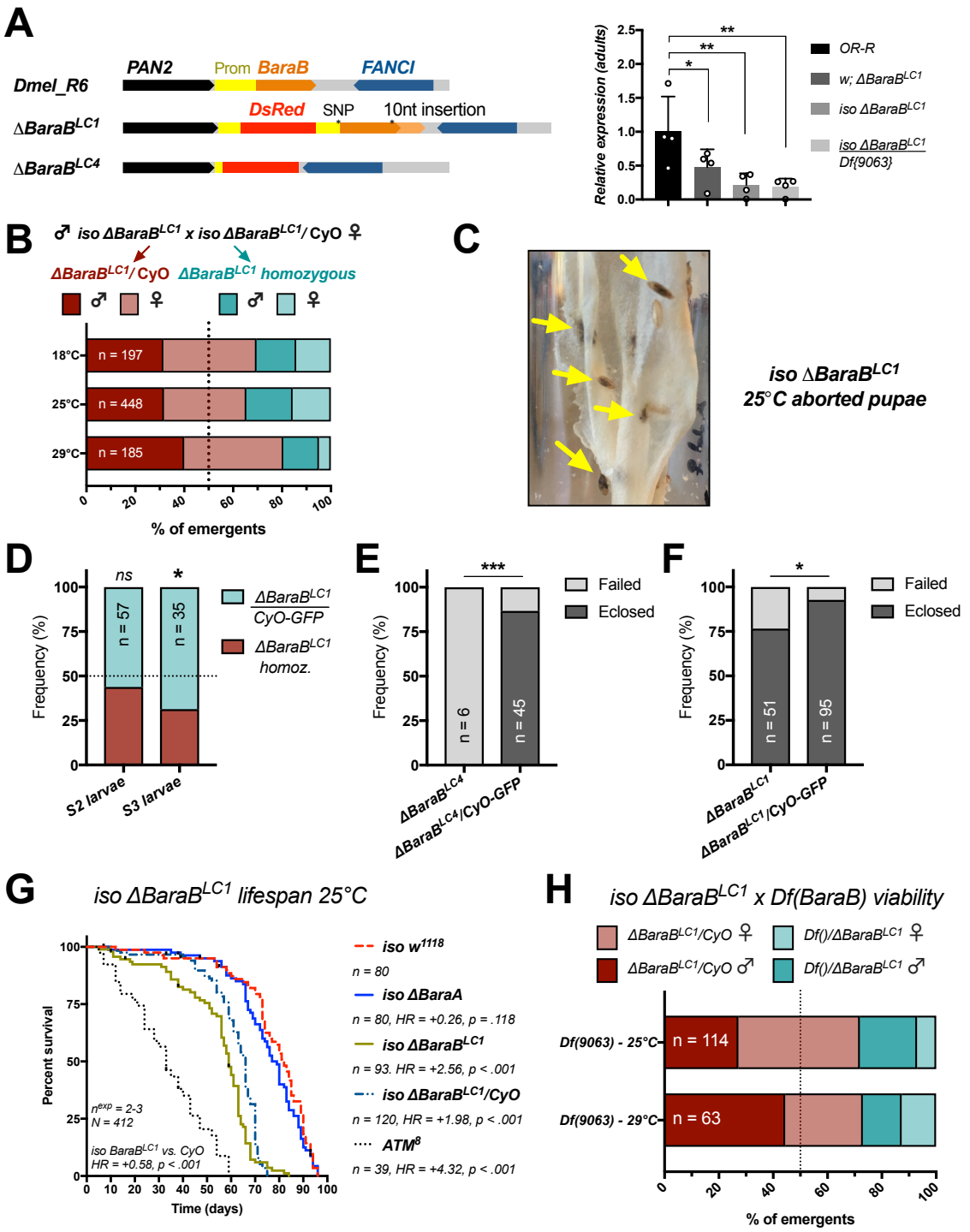

Figure S3: *BaraB* mutation is highly deleterious, even in *ΔBaraB*^{LC1} hypomorphs. A) Diagram of *BaraB* mutant loci and qPCR showing the *ΔBaraB*^{LC1} is a hypomorph mutation. Under our normal qPCR assay conditions, *BaraB* expression is not detected in *ΔBaraB*^{LC1} homozygotes. However using 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.

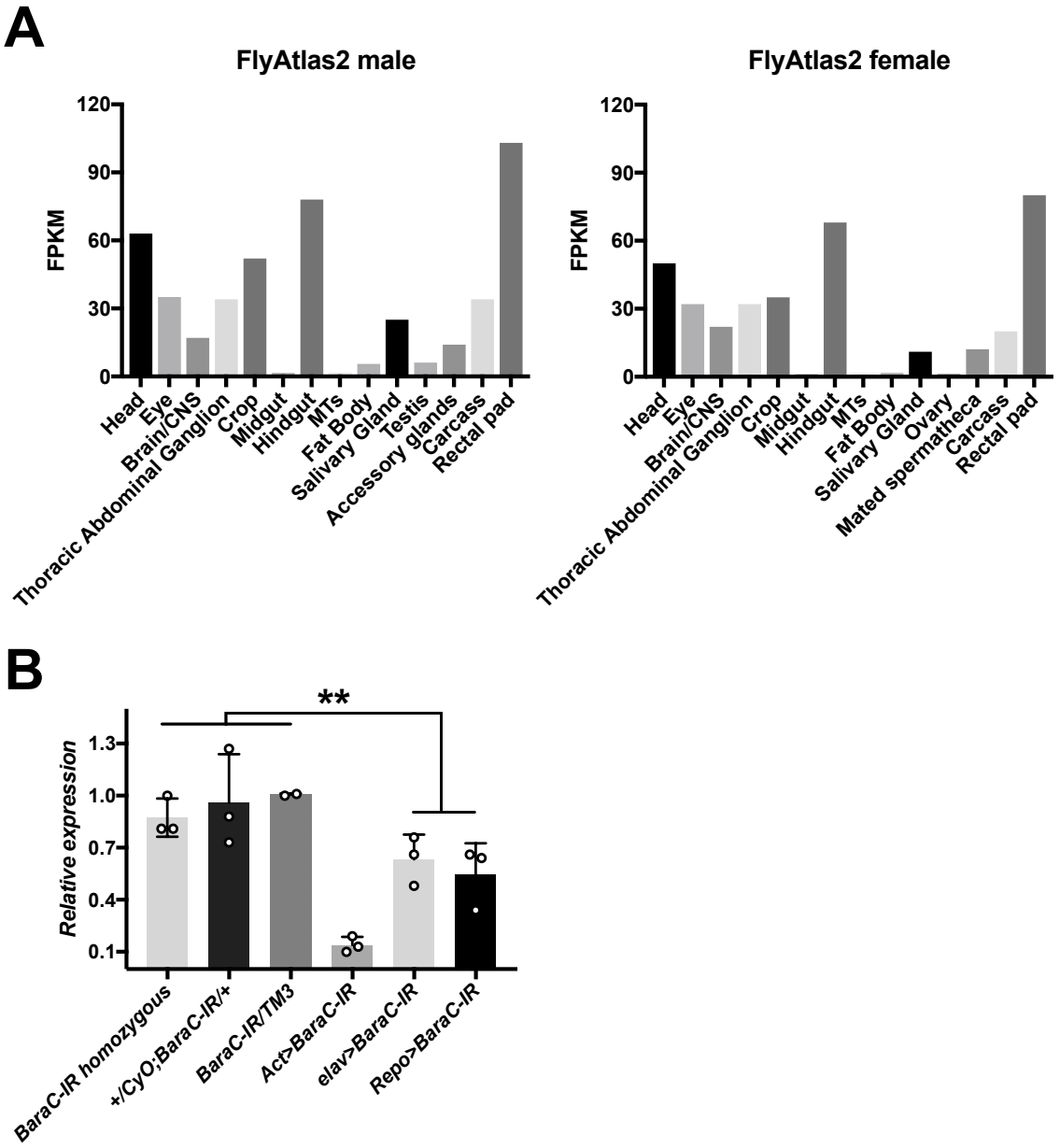

Figure S4: *BaraC* is expressed in the nervous system, but also the hindgut and rectal pads. A) FlyAtlas2 expression data for *BaraC*. **B)** RT-qPCR of *BaraC* in whole flies using different Gal4 drivers to express *BaraC* RNAi. *BaraC* is knocked down by both the *elav-Gal4* and *Repo-Gal4* nervous 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.

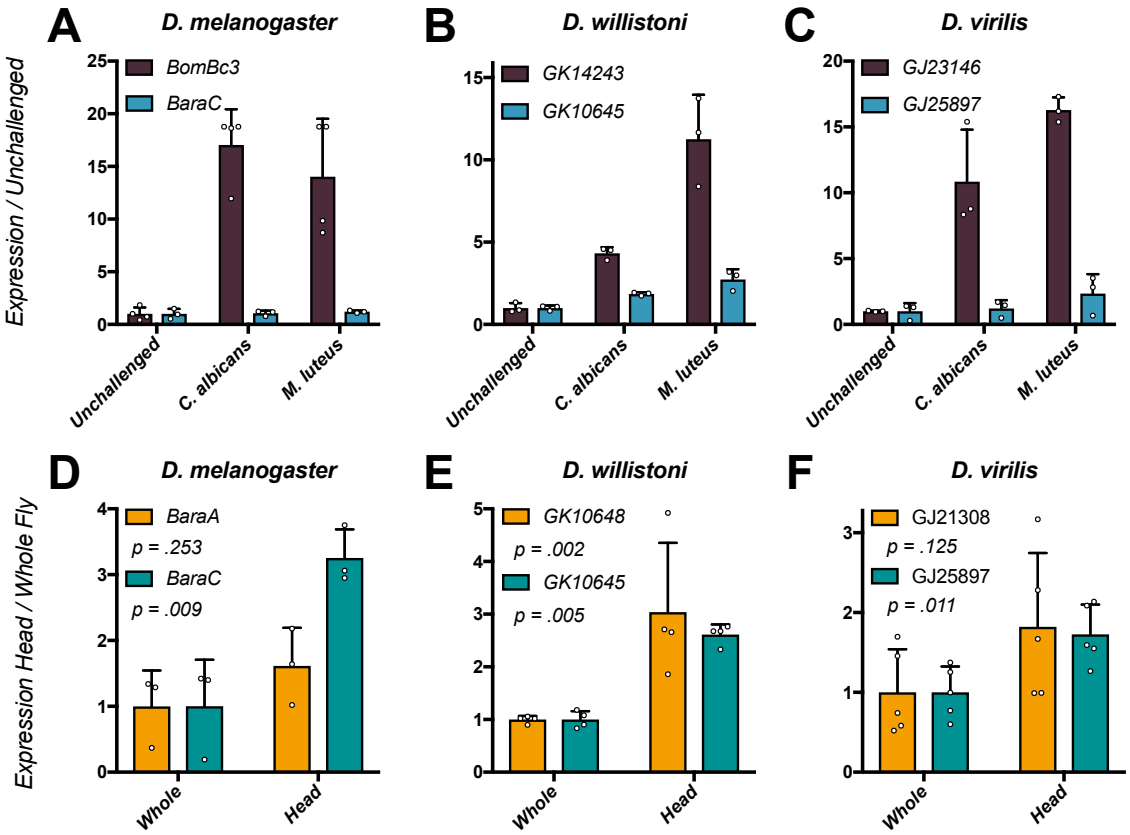

Figure S5: RT-qPCR of *Baramicin* genes in diverse species. A-C) Independent IM24-specific genes in *D. melanogaster* (A), *D. willistoni* (B), and *D. virilis* (C) are not induced by infection. *BomBc3* is included as an immune-induced control. **D-F)** The independent IM24-specific genes (blue) of *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.

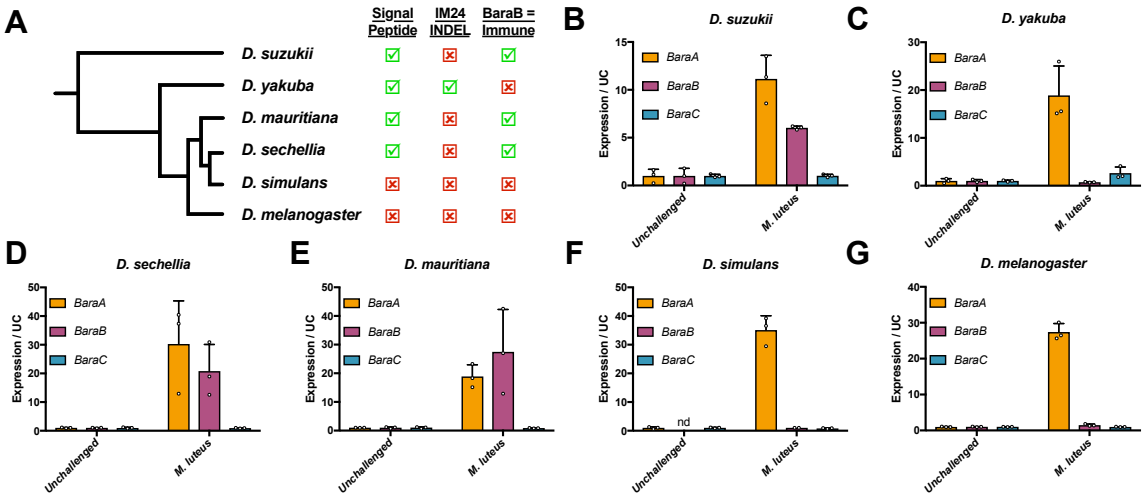

Figure S6: The *D. melanogaster BaraB* gene acquired its non-immune role only recently. A) Cladogram of the Melanogaster species group. The presence of a functional signal peptide (Fig. 5A), and the disruption of the *D. yakuba* IM24 peptide by an in-frame insertion is noted. A summary of whether *BaraB* 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*.

801 **References:**
802

803 Abbott A. 2020. Are infections seeding some cases of Alzheimer's disease? *Nature*
804 587:22–25.

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

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

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

812 Chakraborty M, Chang C-H, Khost DE, Vedanayagam J, Adrion JR, Liao Y, Montooth
813 KL, Meiklejohn CD, Larracuente AM, Emerson JJ. 2021. Evolution of genome
814 structure in the Drosophila simulans species complex. *Genome Res* 31:380–
815 396.

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

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

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

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

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

831 Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: A
832 suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*
833 26:2455–2457.

834 Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, Nguyen M,
835 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