Extensive duplication and convergent evolution of an antimicrobial polypeptide gene

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. AMPs are encoded by short genes that tend to evolve rapidly, likely responding to host-pathogen arms races. 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. We describe a recent duplication of the immune-induced *BaraA* locus, and convergent polymorphisms in *BaraA* peptides that suggest dynamic selective pressures. We further recover 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, leaving 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*. Using the *Baramicin* evolutionary history, we reveal unique properties of the different *Baramicin* domains. In doing so, we provide a case study for how an AMP-encoding gene might play dual roles in both immune and non-immune processes.

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 (Semple and Dorin 2012; Gerdol et al. 2020; 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 (Sackton et al. 2007; Lazzaro et al. 2020). This rapid evolution is driven by positive selection, and acts to either maintain genetic

42 variation (balancing selection), or differentiate AMP sequence (diversifying
43 selection) (Tennessen 2005; Jiggins and Kim 2007; Hellgren et al. 2010;
44 Halldórsdóttir and Árnason 2015; Hanson et al. 2016; Chapman et al. 2019). The
45 selective pressures that drive these evolutionary outcomes are likely the
46 consequence of host-pathogen interactions. One striking example is the
47 maintenance of two alleles of *Diptericin A* in *Drosophila* that predict defence against
48 infection by *Providencia rettgeri* bacteria (Unckless et al. 2016). However AMPs and
49 AMP-like genes in various species have also recently been implicated in non-
50 immune roles as regulators of sleep (Toda et al. 2019; Sinner et al. 2021), memory
51 (Barajas-azpeleta et al. 2018), and neurodegeneration (De Lorenzi et al. 2017;
52 Pineda and Burré 2017; Gosztyla et al. 2018; Lezi et al. 2018; Dominy et al. 2019) in
53 both flies and nematodes. These new contexts suggest that the evolutionary forces
54 acting on AMP genes may not be driven strictly by trade-offs in host defence, but
55 rather by conflicts between roles in immunity and other non-immune functions.
56 For instance, Diptericins are membrane-disrupting antimicrobial peptides of
57 flies (Diptera) that are required for defence against infection (Unckless et al. 2016;
58 Hanson, Dostálová, et al. 2019). It was therefore surprising that the *D. melanogaster*
59 gene *Diptericin B (DptB)* affects memory processes (Barajas-azpeleta et al. 2018). In
60 this study, *DptB* derived from the fly fat body (analogous to the mammalian liver)
61 regulated the ability of the fly to form long-term memory associations (Barajas-
62 azpeleta et al. 2018). Another study in nematodes showed that an immune-induced
63 polypeptide (NLP-29) binds to a G-protein coupled receptor (NPR-12) triggering
64 neurodegeneration through activation of the NPR-12-dependent autophagy
65 pathway (Lezi et al. 2018). The two *Drosophila* AMP-like effectors *GNBP-like3* and
66 *IBIN* have also been shown to affect neuronal processes derived from expression
67 within the nervous system (Barajas-azpeleta et al. 2018; Ebrahim et al. 2021). In
68 humans, the *Cathelicidin* gene encodes the AMP LL-37, which is implicated in glia-
69 mediated neuroinflammation and Alzheimer's disease (Lee et al. 2015; De Lorenzi et
70 al. 2017), alongside evidence of Alzheimer's as being stimulated by pathogen-
71 derived inflammation (Dominy et al. 2019). Thus, immune effectors best known for
72 their roles in the systemic immune response can have surprising interactions with

73 host neurology, derived either from expression within neuronal tissue or through
74 production via immune-responsive tissue.
75 We recently described a novel antifungal peptide gene of *Drosophila*
76 *melanogaster* that we named *Baramicin A* (*BaraA*) (Hanson et al. 2020). A unique
77 aspect of *BaraA* is its precursor protein structure, which encodes a polypeptide
78 cleaved into multiple mature products by interspersed furin cleavage sites. The use
79 of furin cleavage sites to produce two mature peptides from a single polypeptide
80 precursor is common in animal AMP genes (Gerdol et al. 2020; Hanson and Lemaitre
81 2020). However, *BaraA* represents an exceptional case as multiple tandem repeat
82 peptides are produced from the translation of a single coding sequence, effectively
83 resembling a "protein-based operon"; this tandem repeat structure has also been
84 found in two other AMPs of flies and bees (Casteels-Josson et al. 1993; Hanson et al.
85 2016). The immature precursor protein of *D. melanogaster BaraA* encodes three
86 types of domains: an IM24 domain, three tandem repeats of IM10-like domains, and
87 an IM22 domain (**Fig.** 1A). *BaraA* mutants are susceptible to infection by fungi, and
88 preliminary *in vitro* experiments suggest the *BaraA* IM10-like peptides have
89 antifungal activity (Hanson et al. 2020; Huang et al. 2020). The remaining *Baramicin*
90 domains encoding IM22 and IM24 remain uncharacterized. Curiously, *BaraA*
91 deficient flies also display an erect wing behavioural phenotype upon infection,
92 suggesting that *BaraA* could have additional roles beyond immunity (Hanson et al.
93 2020).
94 In this study, we describe the evolution of the immune-induced *BaraA* locus.
95 We then characterize two other *Baramicin* genes in the genome of *D. melanogaster*
96 (*BaraB* and *BaraC)*, providing insight on the function of each *Baramicin* domain.
97 Surprisingly, only *BaraA* is immune-induced, while *BaraB* and *BaraC* are enriched in
98 the nervous system. Moreover independent *Baramicin*s in other *Drosophila* have
99 convergently evolved to resemble *BaraC*, both in expression and gene structure. We
100 resolved the history of *Baramicin* genomic translocations, revealing properties of
101 each protein domain. We suggest that distinct *Baramicin* domains perform different
102 functions with either host protein targets or antimicrobial activity. By characterizing
103 the distinct roles of sub-peptides in an AMP gene, we provide a case study for how

other polypeptide AMP-encoding genes (such as fly *DptB* or human *Cathelicidin*) could play dual roles in both immune and non-immune processes.

Results and Discussion

A duplication of the BaraA gene is present in some laboratory and wild strains

The *D. melanogaster BaraA* gene encodes three key types of mature peptide (**Fig.** 1A): one IM24 peptide, three IM10-like peptides, and one IM22 peptide that is structurally related to the IM10-like peptides (Hanson et al. 2020). We recently found that *BaraA* contributes to the *Drosophila* defence against filamentous fungi (Hanson et al. 2020). During this investigation, we found that while the *BaraA* locus in the *D. melanogaster* R6 reference genome encodes two copies of *BaraA* (*CG33470* and *CG18279*, *BaraA1* and *BaraA2* respectively), other fly strains carry only one *BaraA* gene (**Fig.** 1B). Several lines of evidence indicate that this duplication is a very recent evolutionary event. First: in the Dmel_R6 genome assembly, 3,380bp of contiguous sequence in the *BaraA1* and *BaraA2* loci are 100% identical, including the regions upstream (1,140bp) and downstream (892bp) of the 5' and 3' UTRs. Second: this duplication is not found in close relatives such as *Drosophila simulans* and *Drosophila sechellia*. And third: this duplication is not fixed in laboratory (**Table** S1) or wild flies (below).

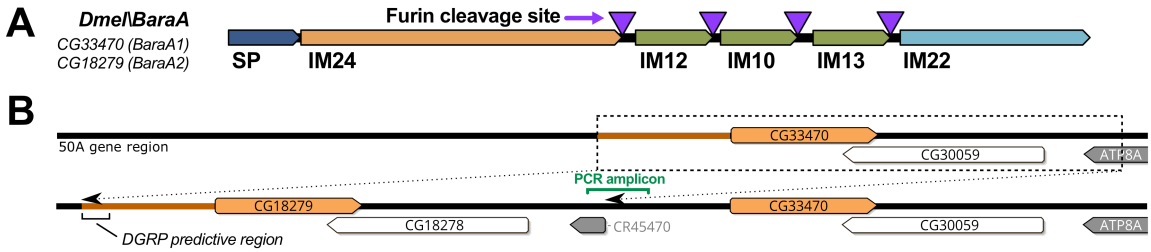


Figure 1: The *BaraA* duplication is not fixed in wild flies. A) Gene structure of *BaraA*, including its three key domain types: IM24 (orange), IM10-like (green), and IM22 (teal). SP = signal peptide. Purple triangles indicate furin cleavage sites. B) Depiction of the *BaraA* duplication. The PCR amplicon used to detect presence of the *BaraA* duplication event, and the region in the DGRP of unknown variant calls correlated with the duplication are annotated (detailed in Fig. S1).

The *Drosophila* Genetic Reference Panel (DGRP) is a genome-sequenced collection of wild-caught flies from the USA (Mackay et al. 2012). As the *D. melanogaster* R6 genome assembly encodes two copies of *BaraA*, we reasoned that mapping the genomes of DGRP strains with variable copy number to this reference genome would result in a failure to call genetic variants ("call = unknown") in the *BaraA* gene region. Indeed, a sharp spike in unknown variant calls occurs just upstream of *BaraA2* in the region that is identical to *BaraA1* (**Fig.** S1A and predictive region in **Fig.** 1B). We performed a PCR-based screen of 132 DGRP strains using primers specific to the duplication (PCR amplicon in **Fig.** 1B), finding 18/132 flies were PCR-positive for the duplication (~14%). Next, we used Pearson's correlation analysis to ask if genetic variant status (annotated or unknown) correlated with PCR-positive genotypes from our screen. We recovered a significant correlation between *BaraA* PCR status and a string of nine genetic variants from 2R_9293471_SNP to 2R_9293576_SNP (Pearson's correlation: $p < 5e^{-3}$ at all nine sites). Specifically, flies with annotated alleles were always PCR-negative (53/53), while 17/18 PCR-positive lines had unknown annotations throughout these nine sites. However 61/114 DGRP strains had unknown calls despite also being PCR-negative for the duplication. Thus, unknown call status is an imperfect but informative predictor of the duplication event. Finally we asked if *BaraA* copy number might inform on infection outcome in data from Wang et al. (Wang et al. 2017). Agreeing with an antifungal role for *BaraA*, we recovered striking correlations between these same nine variant sites and susceptibility to infection by *Metarhizium anisopliae* fungi (.09 < Pearson's r < .16, and .03 < p-value < .10; supplementary information and supplementary data file 1). Thus, considering variation at the *BaraA* locus can improve interpretation of DGRP responses to infection.

Here we annotate *BaraA* copy number in both lab strains and wild-caught flies. Given this region's poor resolution, cryptic SNPs or indels could have interfered with our PCR assay. Nevertheless, we reveal an unusual pattern in the DGRP that provides insight into how *BaraA* copy number seemingly affects antifungal defence. Importantly, our results highlight how lack of information (i.e.

162 unknown variant status) can have predictive value. It will be important to consider
163 *BaraA* copy number in infection studies, or perhaps even behavioural responses
164 where AMP genes have previously been implicated (Barajas-azpeleta et al. 2018;
165 Hanson et al. 2020; Kobler et al. 2020; Ebrahim et al. 2021).
166
167 ***Baramicin A is an ancestral immune effector with convergent polymorphisms***
168 In *D. melanogaster, BaraA* is regulated by the Toll immune pathway. Using
169 BLAST, we recovered *BaraA*-like genes encoding each Baramicin peptide (IM24,
170 IM10-like, and IM22) across the genus *Drosophila* and in the outgroup
171 *Scaptodrosophila lebanonensis*. We next performed infection experiments to confirm
172 that these *BaraA*-like genes were immune-inducible. We infected the diverse species
173 *D. melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, and *D. neotesteacea* (last
174 common ancestor ~50mya (Kumar et al. 2017)) with *Micrococcus luteus* and
175 *Candida albicans*, two microbes that stimulate the Toll pathway (**Fig.** 2A). In all five
176 species, *BaraA*-like genes were immune-induced, confirming that the ancestral
177 *Baramicin* was an immune-induced gene (**Fig.** 2B-F). The predominant peptide
178 products of *D. melanogaster BaraA* are the three IM10-like peptides IM10, IM12, and
179 IM13, which can inhibit fungal growth in vitro (Hanson et al. 2020).
180 Antimicrobial peptides can display polymorphisms found in different species,
181 suggesting selective pressure (e.g. balancing selection) that favours one allele or the
182 other in a context-dependent manner. Such polymorphisms are commonly found in
183 Dipteran AMPs (Unckless and Lazzaro 2016; Chapman et al. 2019; Hanson,
184 Lemaitre, et al. 2019). We found that the 23-residue IM10-like peptides of *Baramicin*
185 genes display signatures of convergent evolution, similar to other fly AMPs (**Fig.** S2
186 and supplementary information). Of particular note, an ^{6}RPXR^{9} motif universally
187 encodes aspartate (D) or glycine (G) at residue 8, which has convergently evolved
188 multiple times (**Fig.** S2C). Future studies of these peptides should incorporate
189 variation of IM10-like peptide residues into activity assays and interpretations of
190 host-pathogen dynamics.

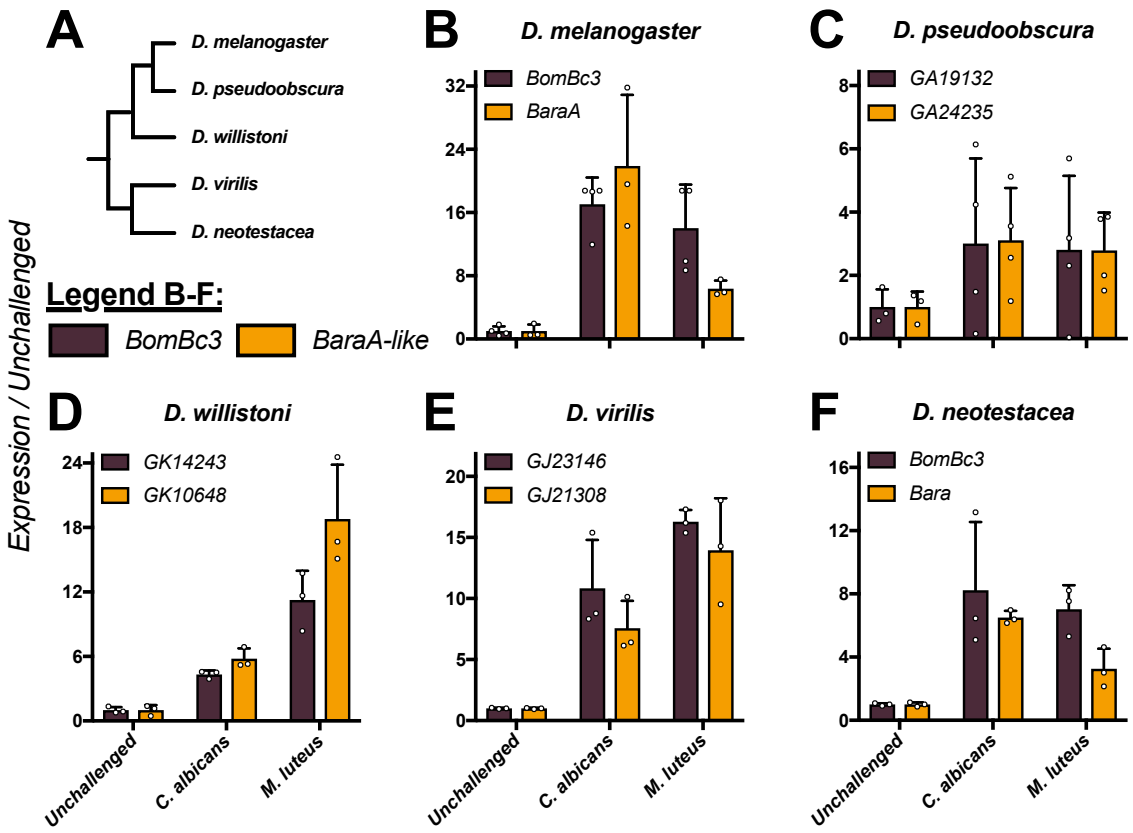


Figure 2: The ancestral *BaraA* gene was immune-induced. A) Cladogram of species used in B-F. B-F) Expression of *BomBc3* control genes (brown) or *BaraA-like* genes (orange) in diverse *Drosophila* species upon infection. In all cases, both *BomBc3* and *BaraA-like* genes are induced upon infection by either *C. albicans* yeast or *M. luteus* bacteria.

Identification of BaraB and BaraC, two non-immune Baramicin paralogs

While investigating *Baramicin* evolution across species, we also recovered two paralogous *Baramicin* genes in *D. melanogaster* through reciprocal BLAST searches: *CG13749* and *CG30285*, which we name *BaraB* and *BaraC* respectively. The three *Baramicin* genes are scattered on the right arm of chromosome II at cytological positions 44F9 (*BaraB*), 50A5 (*BaraA*), and 57F8 (*BaraC*). While *BaraA*-like genes are conserved throughout the genus *Drosophila*, *BaraB* is conserved only in Melanogaster group flies, and *BaraC* is found only in Melanogaster and Obscura group flies, indicating that both paralogs stem from duplication events of a *BaraA*-like ancestor (a summary of the *Baramicin* loci is shown later). In the *D. melanogaster* reference genome, these paralogous *Baramicin* genes are united by

208 the presence of the IM24 domain (**Fig. 3**A). In the case of *BaraB*, we additionally
209 recovered a frameshift mutation (2R_4821599_INS) causing a premature stop
210 segregating in the DGRP leading to the loss of IM13 and IM22 relative to the *BaraA*
211 gene structure (**Fig. 3**A); this truncation is present in the Dmel_R6 genome
212 assembly, but many DGRP strains encode a CDS with either a standard (e.g.
213 DGRP38) or extended (e.g. DGRP101) IM22 domain (a DGRP *BaraB* alignment is
214 provided in supplementary data file 2). Moreover, in contrast to BaraA, the initial
215 IM10-like peptide of BaraB no longer follows a furin cleavage site (**Fig. 3**A), and
216 encodes a serine (RSXR) in its IM10-like motif instead of the proline (RPXR) of
217 *BaraA*-like genes. Each of these mutations prevents the secretion of classical IM10-
218 like and IM22 peptides. Finally, *BaraC* encodes only IM24 tailed by a
219 transmembrane domain at the C terminus (TMHMM v2.0 (Krogh et al. 2001)), and
220 thus lacks both the IM10-like peptides and IM22 (**Fig. 3**A).
221 *BaraA* is strongly induced following microbial challenge, being
222 predominantly regulated by the Toll pathway with a minor input from the Immune
223 Deficiency (Imd) pathway (Hanson et al. 2020; Huang et al. 2020). We assayed the
224 expression of *BaraB* and *BaraC* in wild-type flies, and also flies with defective Toll
225 (spz^{rm7}) or Imd (Rel^{E20}) signalling to see if their basal expression relied on these
226 pathways. Surprisingly, neither gene was induced upon infection regardless of
227 microbial challenge (**Fig. 3**B and **Fig.** S3A-B). However *BaraC* levels were reduced in
228 spz^{rm7} mutants regardless of treatment (cumulative data in **Fig.** S3C, p = .005),
229 suggesting that *BaraC* basal expression is affected by Toll signalling. We next
230 generated a novel time course of development from egg to adult to monitor the
231 expression of the three *Baramicin* genes. We found that expression of all genes
232 increased over development and reached their highest level in young adults (**Fig.**
233 3C). Of note, *BaraB* expression approached the lower limit of our assay's detection
234 sensitivity at early life stages. However *BaraB* was robustly detected beginning at
235 the pupal stage, indicating it is expressed during metamorphosis. *BaraC* expression
236 also increased markedly between the L3 larval stage and pupal stage.

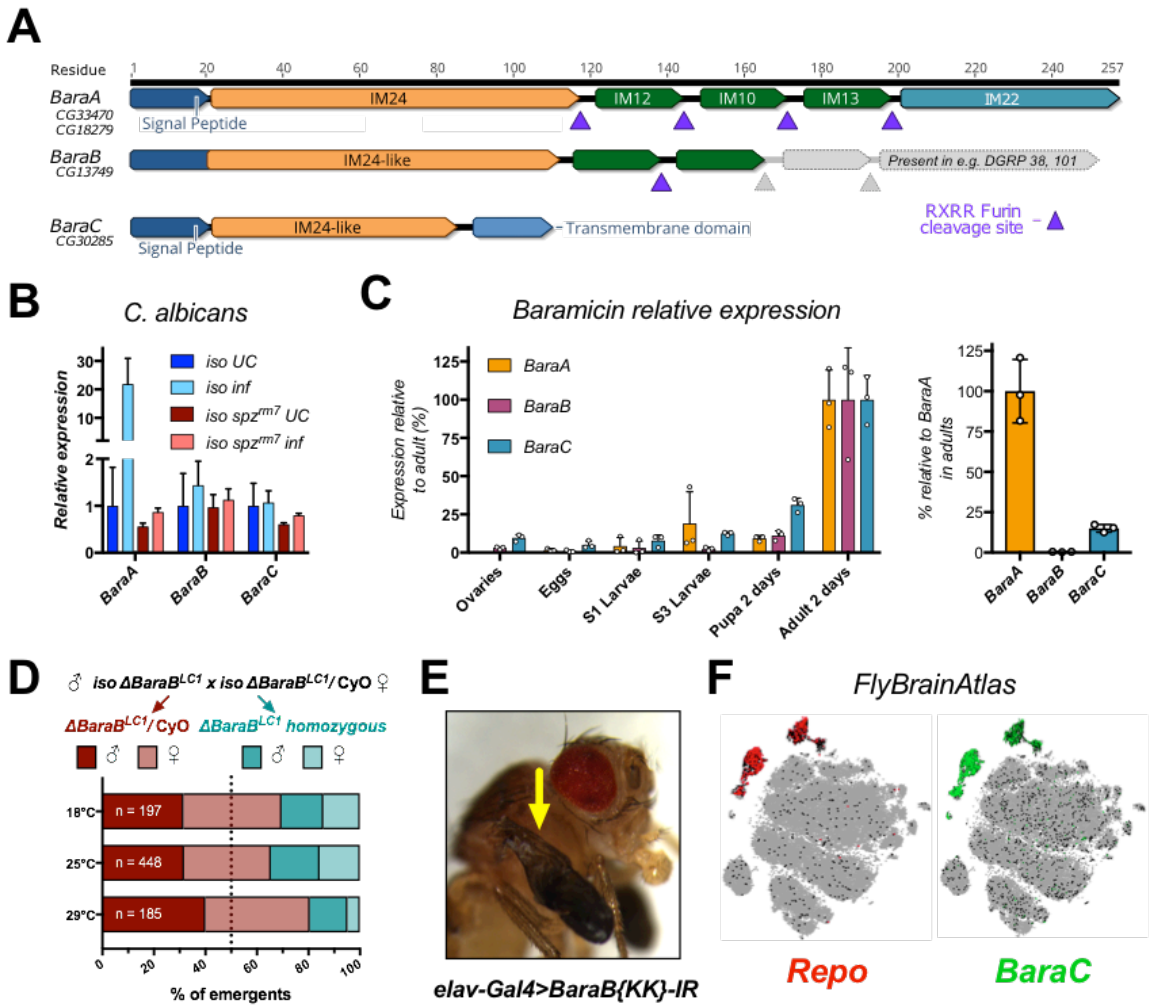


Figure 3: The *D. melanogaster Baramicin* gene family. A) Two *Baramicin* paralogs in addition to *BaraA* are present in the Dmel_R6 reference genome, united by the presence of an IM24 peptide. Relative to *BaraA*, two frameshift mutations in *BaraB* are segregating in the DGRP. One results in the same mature protein annotated in the Dmel_R6 reference genome (coloured domains), while the other elongates the IM22 peptide (e.g. *DGRP101*). Most strains carry a *BaraB* gene of similar length to *BaraA* (e.g. DGRP38). Additionally, the *BaraB* gene does not encode a functional signal peptide. B) Only *BaraA* is immune-induced. *BaraB* and *BaraC* do not respond to infection, though basal *BaraC* expression relies on Toll signalling (Fig. S3C). C) 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). D) The $\Delta BaraB^{LC1}$ mutation is partial lethal, an effect that is exaggerated by rearing flies at 29°C. 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 (Davie et al. 2018).

255 Here we reveal that *BaraA* is part of a larger gene family. While the *BaraA*
256 gene was first described as an immune effector, the two more phylogenetically-
257 restricted *Baramicin* paralogs *BaraB* and *BaraC* are not induced by infection in *D.*
258 *melanogaster*. Both *BaraB* and *BaraC* first see increased expression during pupation,
259 and are ultimately expressed at their highest levels in adults.
260
261 ***Dmel\BaraB is required in the nervous system over the course of development***
262 A simple interpretation of the truncated gene structure and low levels of
263 *BaraB* expression is that this gene is undergoing pseudogenization. To explore
264 *BaraB* function, we used two mutations for *BaraB* ($\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$,
265 generously gifted by S.A. Wasserman). These mutations were made using a CRISPR
266 double gRNA approach to replace the *BaraB* locus with sequence from the pHD-
267 DsRed vector. The $\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$ mutations differ in their ultimate effect,
268 as $\Delta BaraB^{LC1}$ is an incidental insertion of the DsRed cassette in the promoter of the
269 gene. This disruption reduces gene expression, resulting in a hypomorph state (**Fig.**
270 **S4**A-B). The $\Delta BaraB^{LC4}$ mutation however deletes the entire locus, leading to *BaraB*
271 null flies (**Fig. S4**A). We further introgressed both mutations into the DrosDel
272 isogenic background (referred to as *iso*) for seven generations. $\Delta BaraB^{LC4}$
273 homozygotes are lethal during larval development, whereas the hypomorphic
274 $\Delta BaraB^{LC1}$ flies allow for homozygous adults to emerge (**Fig. S4**C). Thus we used
275 $\Delta BaraB^{LC1}$ for experiments involving adults. Isogenization was performed as
276 described in Ferreira et al. (Ferreira et al. 2014). We assessed $\Delta BaraB^{LC1}$ mutant
277 viability using crosses between $\Delta BaraB^{LC1}/CyO$ heterozygous females and $\Delta BaraB^{LC1}$
278 homozygous males. Under mendelian ratios of inheritance, this cross is expected to
279 result in a 1:1 ratio of $\Delta BaraB^{LC1}$/CyO and $\Delta BaraB^{LC1}$ homozygous progeny.
280 However, *BaraB* deficient flies showed reduced viability using both mutations
281 independent of genetic background (**Fig. S4**C). This reduced viability was
282 exacerbated by rearing at 29°C, reaching a lethality rate over 50% (**Fig. 3**D). Using a
283 *CyO-GFP* reporter to track genotype in larvae revealed that the lethal phase occurs
284 primarily in the late larval and pupal stages (**Fig. S4**D-G), agreeing with a role for
285 *BaraB* in larvae/pupae previously suggested by increased expression at this stage.

286 Some emergent flies also exhibited a nubbin-like wing phenotype (FlyBase:
287 FBrf0220532 and e.g. in **Fig. 3**E), where the wings were stuck in a shrivelled state
288 for the remainder of the fly's lifespan. However, a plurality of *ΔBaraB*^{LC1}
289 homozygotes successfully emerged, and unlike their siblings, had no immediate
290 morphological or locomotory defects. The lifespan of morphologically normal *iso*
291 *ΔBaraB*^{LC1} adults is nevertheless significantly shorter compared to wild-type flies
292 and *iso ΔBaraB*^{LC1}*/CyO* siblings (**Fig. S4**H). We confirmed these developmental
293 effects using ubiquitous gene silencing with *Actin5C-Gal4 (Act-Gal4)* to drive two
294 *BaraB* RNAi constructs (*TRiP-IR* and *KK-IR*). Both constructs resulted in significant
295 lethality and occurrence of nubbin-like wings (**Table** S2). Additionally, we crossed
296 *iso ΔBaraB*^{LC1} to a genomic deficiency line (BDSC9063) at both 25°C and 29°C. These
297 crosses resulted in significantly reduced numbers of eclosing *ΔBaraB*^{LC1}*/Df(9063)*
298 flies at 25°C ($n = 114$, χ^2 $p < .001$) and 29°C ($n = 63$, χ^2 $p < .001$) (**Fig. S4**I).
299 These data demonstrate a significant consequence of *BaraB* disruption on
300 fitness. While whole-fly *BaraB* expression is low, these results suggest that *BaraB* is
301 not pseudogenized, and instead performs an integral function during development.
302 The fact that there is a bimodal outcome in hypomorph-like *ΔBaraB*^{LC1} adults (either
303 severe defects or generally healthy) suggests *BaraB* is involved in passing some
304 checkpoint during larval/pupal development. Flies deficient for *BaraB* are more
305 likely to fail at this developmental checkpoint, resulting in either lethality or
306 developmental defects.
307
308 ***The Baramicin paralogs BaraB and BaraC are expressed in the nervous system***
309 Given that whole-fly *BaraB* depletion produced lethality and nubbin-like
310 wings, we next sought to determine in which tissue(s) *BaraB* is required. *BaraB*
311 mutant locomotory defects led us to suspect an involvement in the nervous system.
312 We silenced *BaraB* in the nervous system at 25°C or at 29°C for greater efficiency,
313 using a pan-neural *elav-Gal4* driver, and for *BaraB-IR* used both the *TRiP-IR* and *KK-*
314 *IR* lines. We additionally combined this approach with *UAS-Dicer2* (*Dcr2*) to further
315 strengthen gene silencing (Neely et al. 2010). We used the *TRiP-IR* line with a *CyO*

316 balancer chromosome (*elav-Gal4 * TRiP-IR/CyO*), allowing the same simple
317 mendelian approach to assess lethality as used previously. Both *elav>TRiP-IR* and
318 *elav>Dcr2, TRiP-IR* caused partial lethality and occasional nubbin-like wings (χ^2 p <
319 .02, **Table** S2). For *KK-IR*, we instead crossed homozygous flies and so we did not
320 assess lethality using mendelian inheritance. However, no adults emerged when
321 *elav>Dcr2, KK-IR* flies were reared at 29°C. Rare emergents (n^{total} = 11 after three
322 experiments) occurred at 25°C, all of which bore nubbin-like wings. Using *elav-Gal4*
323 at 29°C without *Dcr2*, we observed greater numbers of emerging adults, but 100%
324 of flies had nubbin-like wings (**Fig. 3**E, **Table** S2). Finally, *elav>KK-IR* flies at 25°C
325 suffered both partial lethality and nubbin-like wings, but normal-winged flies began
326 emerging (χ^2 p < .001, **Table** S2). This analysis indicates that *BaraB* is expressed in
327 the nervous system, and this expression readily explains both the lethality and
328 nubbin-like wing phenotypes. Moreover, we observed a consistent spectrum of
329 developmental defects using *elav-Gal4>BaraB-IR* wherein strength of gene silencing
330 correlates with severity of defects.
331 *BaraB* is not the only *Baramicin* gene now specialized for expression in the
332 nervous system. Tissue-specific transcriptomic data indicate that *BaraC* is expressed
333 in various neural tissues including the eye, brain, and the thoracic abdominal
334 ganglion (**Fig. S5**A) (Leader et al. 2018). However *BaraC* is also highly enriched in
335 the hindgut and rectal pads, pointing to a complex expression pattern (Hammonds
336 et al. 2013; Leader et al. 2018). We next searched FlyBrainAtlas (Davie et al. 2018)
337 to narrow down which neural subtypes *BaraC* was expressed in. Strikingly, *BaraC*
338 was expressed in all glial cell types, fully overlapping the glia marker *Repo* (**Fig. 3**F).
339 To confirm these observations, we compared the effects of *BaraC* RNA silencing
340 (*BaraC-IR*) using *Act-Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia)
341 drivers on *BaraC* expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that
342 of control flies (**Fig. S5**B). By comparison *elav-Gal4* reduced *BaraC* expression to
343 ~63% that of controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls
344 (overall controls vs. neural/glia-IR, p = .002). Collectively, our results support the

345 notion that *BaraC* is expressed in the nervous system, and are consistent with *BaraC*
346 expression being most localized to glial cells.
347
348 ***Extensive genomic turnover of the Baramicin gene family***
349 The results thus far show that *BaraA*-like genes are consistently immune-
350 induced in all *Drosophila* species (**Fig.** 2), and the *BaraA* antifungal IM10-like
351 peptides evolve rapidly (**Fig.** S2). However, the two phylogenetically-restricted
352 paralogs *Dmel\BaraB* and *Dmel\BaraC* are not immune-induced, and are truncated
353 in a fashion that deletes some or all of the IM10-like peptides (**Fig. 3**A). These two
354 *Baramicins* are now enriched in the nervous system (**Fig. 3**E-F). In the case of *BaraB*,
355 a role in the nervous system is evidenced by severe defects recapitulated using pan-
356 neural RNA silencing. In the case of *BaraC*, nervous system expression is evidenced
357 by a clear overlap with *Repo*-expressing cells.
358 The three *D. melanogaster Baramicin* genes are united by their IM24 domain,
359 and are scattered across different loci along chromosome 2R, suggesting a series of
360 complex genetic events. To determine the ancestry of each *D. melanogaster*
361 *Baramicin* gene, we traced their evolutionary history by analyzing genomic synteny.
362 This allowed us to determine ancestry beyond sequence similarity-based
363 approaches. Using this strategy, we traced a series of genomic duplications,
364 inversions, and locus translocation events that ultimately resulted in the three
365 *Baramicin* loci of *D. melanogaster*. Our analysis reveals that these three loci
366 ultimately stem from a single-locus ancestor encoding only one *Baramicin* gene that
367 resembled *Dmel\BaraA* (**Fig. 4**A). This is evidenced by the presence of only a single
368 *BaraA*-like gene in the outgroup *S. lebanonensis*, and also in multiple lineages of the
369 subgenus Drosophila (**Fig. 4**B). Indeed, the general *BaraA* gene structure encoding
370 IM24, tandem repeats of IM10-like peptides, and IM22 is conserved in *S.*
371 *lebanonensis* and all *Drosophila* species (**Fig. 4**C). On the other hand, the *Dmel\BaraC*
372 gene comes from an ancient duplication restricted to the subgenus Sophophora, and
373 *Dmel\BaraB* resulted from a more recent duplication found only in the
374 Melanogaster group (**Fig. 4**B).

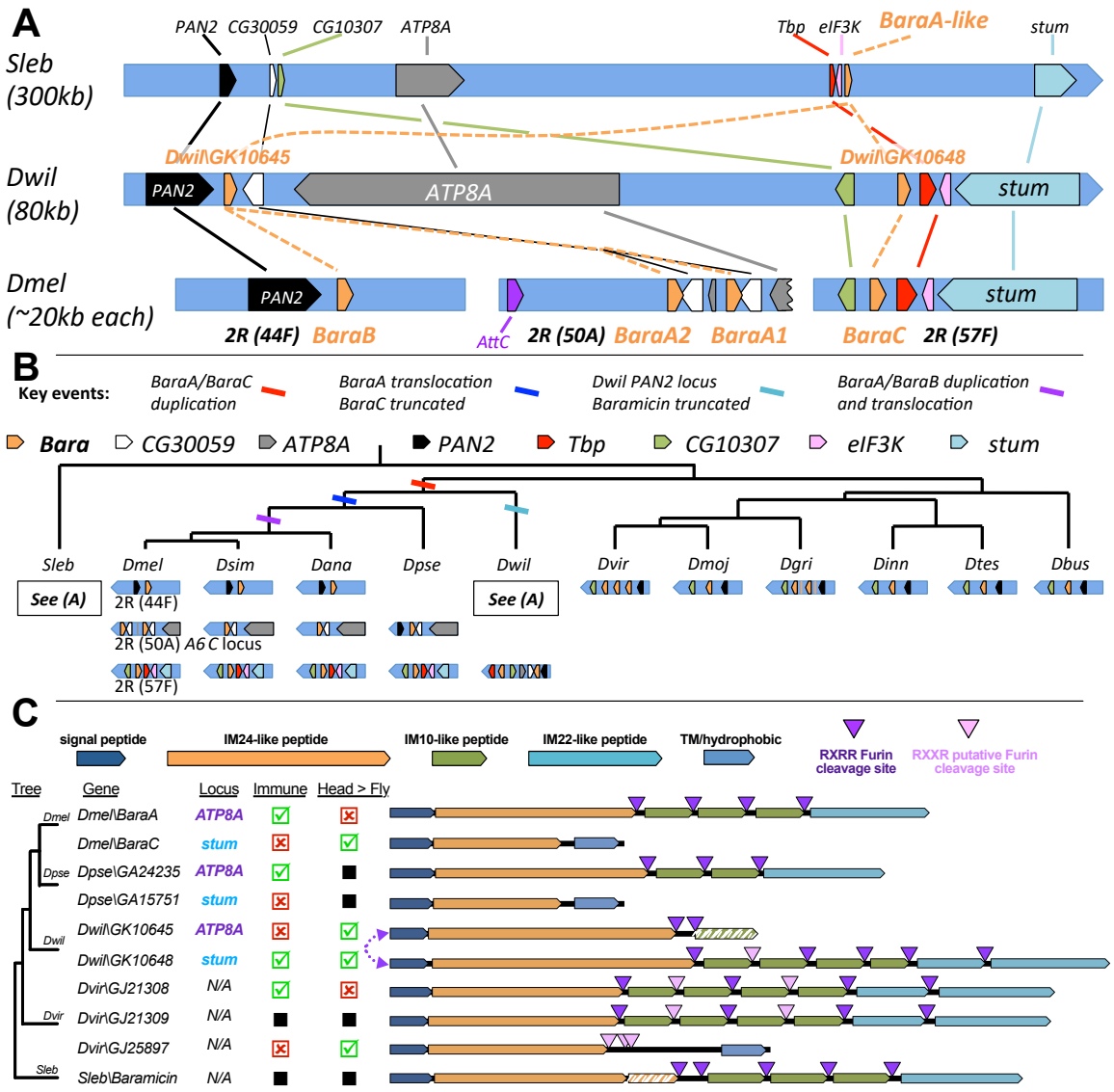


Figure 4: *Baramicin* evolutionary history. A) Detailed map of genomic neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D. melanogaster*, detailing inferred duplication, inversion, and translocation events. Gene names are given as found in *D. melanogaster*. B) Cladogram 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. 5A). Expression dynamics relating to immune-induction or enrichment in the head (checked boxes) are shown in Fig. 2 and Fig. S6. 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 This genomic synteny analysis revealed a striking observation: the *D.*
391 *willistoni* gene *Dwil\GK10648* is syntenic with the *Dmel\BaraC* locus (**Fig. 4**A), yet it
392 is immune-induced (**Fig.** 2D) and retains a *BaraA*-like gene structure (**Fig. 4**C).
393 Meanwhile, the other *D. willistoni Baramicin Dwil\GK10645* is found at the locus
394 syntenic with *BaraA*, yet this gene has undergone an independent truncation to
395 encode just an IM24 peptide (similar to *Dmel\BaraC*). Thus these two *D. willistoni*
396 genes have evolved similar to *D. melanogaster BaraA/BaraC*, but in a vice versa
397 fashion. This suggests a pattern of convergent evolution with two key points: **i)** the
398 duplication event producing *Dmel\BaraA* and *Dmel\BaraC* originally copied a full-
399 length *BaraA*-like gene to both loci, and **ii)** the derivation of an IM24-specific gene
400 structure has occurred more than once (*Dmel\BaraC* and *Dwil\GK10645*). Indeed,
401 another independent IM24-specific *Baramicin* gene is present in *D. virilis*
402 (*Dvir\GJ25897*), which is a direct sister of the *BaraA*-like gene *Dvir\GJ21309* (**Fig.**
403 **4**C). We evaluated the expression of these truncated *Baramicins* encoding only IM24
404 in both *D. willistoni* and *D. virilis* upon infection. As was the case for *Dmel\BaraC*,
405 neither gene is immune-induced (**Fig. S6**A-C). Given the glial expression of
406 *Dmel\BaraC,* we reasoned that the heads of adult flies (rich in nerve tissue) should
407 be enriched in *BaraC* compared to whole animals. Indeed we saw a significant
408 enrichment of *BaraC* in the heads of *D. melanogaster* males compared to whole flies,
409 which was not the case for *BaraA* (**Fig. S6**D). We further saw a consistent and
410 significant enrichment in the head for the IM24-specific genes *Dwil\GK10645* and
411 *Dvir\GJ25897*, while *BaraA*-like genes were more variable in expression (**Fig. S6**E-
412 F).
413 Thus, multiple independent IM24-specific *Baramicins* apparently do not play
414 an immune role, and are enriched in the head. In the case of *Dmel\BaraC*, this is
415 likely due to expression in glia. Strikingly, we observe an independent but parallel
416 evolution of expression pattern and gene structure in *Baramicins* of *D. willistoni* and
417 *D. virilis*. A summary of these expression data is included in **Fig. 4**C. Phylogenetic
418 and genomic synteny analysis reveals that the gene structure and immune
419 expression pattern of *BaraA* reflect the ancestral state, while *Dmel\BaraB* and
420 *Dmel\BaraC* are paralogs derived from independent duplication events. These two

421 genes, which are truncated and delete both IM10-like and IM22 domains, have also
422 lost immune function.
423
424 ***Evolution of IM24 in lineage-specific fashions***
425 Multiple independent *Baramicin* genes have lost both IM10-like and IM22
426 peptides, and also converge on loss of immune induction. Taken together, these
427 truncations and expression patterns suggest that the IM10-like peptides and IM22
428 are strictly useful during the immune response, consistent with a recently described
429 antifungal role for IM10-like peptides (Hanson et al. 2020). Inversely, non-immune
430 *Baramicin* genes have repeatedly and independently truncated to encode primarily
431 IM24. This suggests that IM24 is adaptable to different contexts – a protein domain
432 with utility extending beyond the immune response.
433 Thus far, we have discussed IM24 primarily as a unifying feature of the
434 *Baramicin* gene family. To understand how this universal domain differs across gene
435 lineages, we next investigated the evolution of the IM24 domain at the sequence
436 level. We analyzed IM24 domain evolution via multiple statistical approaches
437 (including Likelihood (FEL), Bayesian (FUBAR), and Count-based (SLAC) analyses)
438 using the HyPhy package implemented in Datamonkey.org, (Delport et al. 2010)
439 with separate codon alignments of *BaraA, BaraB,* or *BaraC* IM24 domains beginning
440 at their conserved Q^1 starting residue. We collected additional genomic sequence
441 data from recently-sequenced fly species (Kim et al. 2021) to better inform on the
442 sequence evolution of these genes and loci. We employed analyses of each lineage
443 using locus-specific alignments (e.g. genes at the *stum* locus in **Fig. 4**B were all
444 analyzed together). Globally, FEL, FUBAR, and SLAC site-specific analyses suggest
445 strong purifying selection across the IM24 domain (data in **supplementary data**
446 **file 3**), agreeing with the general protein structure of IM24 being broadly conserved.
447 However one site (residue 29) was consistently highlighted as evolving under
448 positive selection using each type of statistical approach for those genes located at
449 the Sophophora *ATP8A* locus (*BaraA* genes and *Dwil\GK10645*: p-adj < .05; **Fig. 5**A).
450 This site is universally Proline in *Baramicin* genes located at the stum locus (*BaraC*-
451 like), but is variable in both the *BaraA* and *BaraB* lineages (**Fig. 5**A). We also note

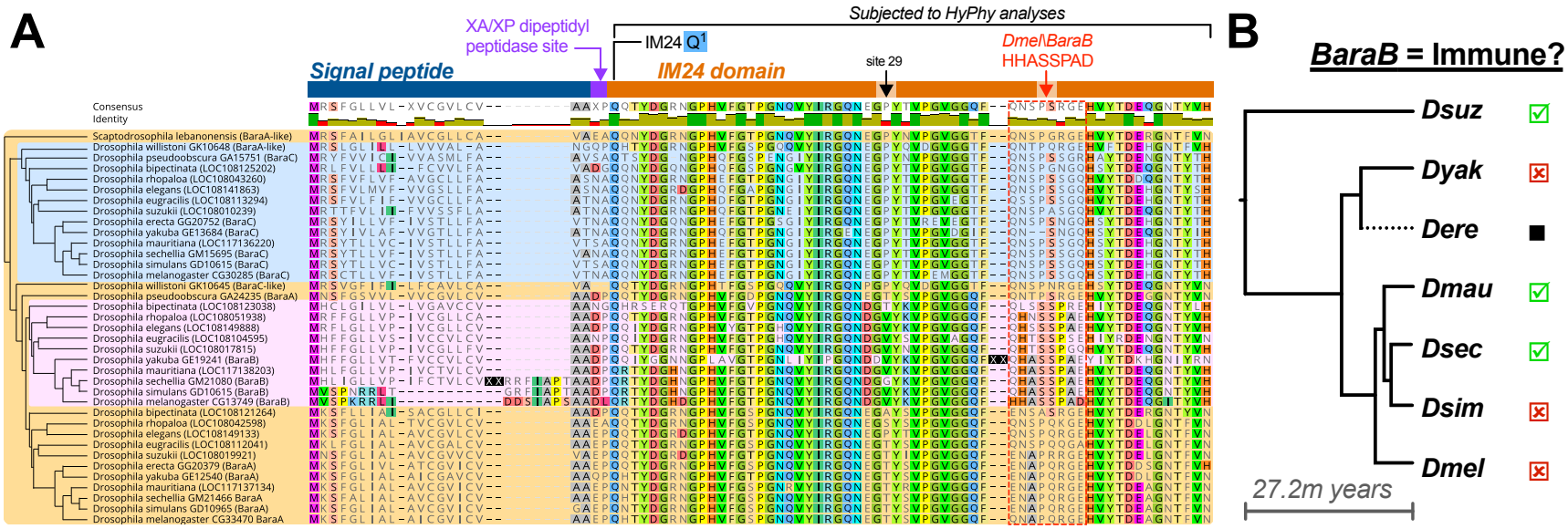


Figure 5: *BaraB* and IM24 rapid evolution. A) Evolution of the signal peptide and core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*. Arrows indicate site 29 and the *D. melanogaster* [40]HHASSPAD[48] domain. Insertion events in *D. yakuba* and *D. sechellia BaraB* are denoted 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) Proposed timeframe for gain of novel *BaraB* function. The *D. erecta BaraB* gene is pseudogenized, and the *D. yakuba* gene is not immune-induced and encodes a 9-residue insertion in the IM24 peptide bordering the HHASSPAD domain (see 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. S8). 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. Speciation timeframes adapted from Timetree.org (Kumar et al. 2017).

464 that two sites on either side of site 29 (site 27 and site 31) similarly diverge by
465 lineage in an otherwise highly conserved region of the IM24 domain. FUBAR
466 analysis (but not FEL or SLAC) also found evidence of positive selection at site 31 (p-
467 adj = .026), suggesting this neighbouring site could also be evolving in a non-random
468 fashion. Similar analyses of the *BaraB* and *stum* loci *Baramicins* did not find evidence
469 of site-specific positive selection.

470 We next assessed the potential for diversifying selection across the
471 phylogeny using two branch-site methods (aBSREL and BUSTED (Murrell et al.
472 2015)). Both branch-site analyses found evidence of episodic diversifying selection
473 in different lineages. For the *ATP8A* locus *Baramicins (BaraA* genes and
474 *Dwil\GK10645)*, BUSTED whole-phylogeny analysis found evidence of diversifying
475 selection in at least one branch across the phylogeny (p-adj = .008). aBSREL analysis
476 also finds evidence of diversifying selection specifically at the node delineating the
477 Willistoni group from other Sophophora (**Fig.** S7). This node corresponds to the
478 parallel and opposite evolution of BaraA-like and BaraC-like protein structures at
479 this locus in *D. melanogaster* compared to *D. willistoni* (**Fig. 4**C), indicating both gene
480 structure and sequence evolved rapidly in these branches. Finally one other string of
481 residues is also striking: a specific motif with lineage-restricted patterns occurs at
482 sites 40-48, numbered according to the Q^1 starting residue. In *Dmel\BaraB*, this
483 motif encodes the residues ^{40}HHASSPAD^{48}. The three C-terminal residues of this
484 motif are diagnostic for each gene lineage (*BaraA, BaraB,* and *BaraC* have RGE, PXE,
485 or (S/N)GQ respectively; **Fig. 5**A). We also recovered a signature of diversifying
486 selection in the stum locus *Baramicin* gene of *Drosophila paulistorum* using recent
487 genomic sequence for strains L06 and L12 (see (Kim et al. 2021) and supplementary
488 information), suggesting population-level data may reveal additional sites of
489 interest.

490 Globally, *Baramicin* IM24 peptides have experienced positive selection. Site-
491 specific analyses highlight residue 29 in the *ATP8A* locus lineage. We further found
492 evidence that, alongside major structural changes specific to gene lineages, the IM24
493 sequence itself evolved under diversifying selection. Thus, not only are IM10-like
494 and IM22 peptides independently lost in *Dmel\BaraC* and *Dwil\GK10645*, alongside

495 parallel loss of immune induction, but their IM24 domains also evolved rapidly
496 following the speciation event distinguishing these lineages. Another 8-residue
497 motif encoding "HHASSPAD" in Dmel\BaraB is diagnostic of *Baramicin* lineages. We
498 could not generate a reasonable model of the IM24 peptide conformation using
499 Phyre2 (Kelley et al. 2015), QUARK, or TASSER protein modelling methodologies
500 (Zhang et al. 2016). However the pattern of residues at site 29 and neighbouring
501 residues at site 27 and site 31 may suggest this region has some importance in
502 interactions with targets of IM24. The Dmel\BaraB HHASSPAD motif also shows
503 striking patterns in different gene lineages, though we did not find evidence of
504 positive selection at this motif.
505
506 ***A timeframe for BaraB neofunctionalization***
507 Selection analyses support the notion that IM24 evolves rapidly in different
508 lineages. One other striking feature of the Dmel\BaraB protein is the absence of a
509 signal peptide structure (SignalP 5.0 (Almagro Armenteros et al. 2019)). Signal
510 peptide sequence is conserved in all *Baramicin* lineages, except in *BaraB* of *D.*
511 *melanogaster* and also *D. simulans* (last common ancestor 5.9mya (Kumar et al.
512 2017)); these species encode a conserved region of parallel length (**Fig.** 5A). This is
513 surprising, as *D. simulans* is phylogenetically more closely related to *D. sechellia* (last
514 common ancestor 3.3mya) and *D. mauritiana* (last common ancestor 4.5mya), yet *D.*
515 *sechellia* and *D. mauritiana* retain signal peptide structures in their *BaraB* genes; we
516 note that the *D. sechellia* BaraB signal peptide is divergent due to an insertion event
517 (**Fig.** 5A), but is still predicted to enable protein secretion (SignalP 5.0). We also
518 previously noted that a genetic variant truncating the *BaraB* CDS is segregating in
519 wild *D. melanogaster* (Fig. 3A and supplementary data file 2). However we
520 recovered only full-length *Baramicin* gene structures encoding IM24, IM10-like
521 peptides, and IM22 in *BaraB* orthologues of other species (except *D. erecta* where
522 the *BaraB* orthologue appears pseudogenized by the loss of the *BaraB* start codon
523 and multiple indel events).
524 We reason that the *D. simulans BaraB* N-terminus – which mirrors that of *D.*
525 *melanogaster* – was either introgressed from one species to the other prior to the

526 complete development of hybrid inviability, or reflects incomplete lineage sorting of
527 this locus in the Melanogaster species complex. In either scenario, two major events
528 affect the *D. melanogaster* BaraB protein structure: i) loss of signal peptide
529 sequence, and ii) premature stop variants truncating the precursor protein. Patterns
530 of residue conservation at IM24 site 29 or in the BaraB ^{40}HHASSPAD^{48} domain may
531 also determine different species' *BaraB* gene function. We were curious if any of
532 these events correlated with the loss of an immune role, presumably alongside
533 neofunctionalization of *BaraB* for its specific role in the nervous system. To this end,
534 we designed qPCR assays for *Baramicin* genes in phylogenetically diverse members
535 of the Melanogaster species group including *D. suzukii, D. yakuba, D. mauritiana, D.*
536 *sechellia,* and *D. simulans* (last common ancestor 27.2mya (Kumar et al. 2017)). We
537 then infected these flies with *M. luteus* bacteria, known to activate the Toll pathway.
538 We found that the *BaraB* orthologues of *D. suzukii*, *D. mauritiana* and *D. sechellia* all
539 remain immune-induced, indicating that the ancestral *BaraB* gene was still immune-
540 regulated following its genomic translocation (**Fig.** 5B). However like *Dmel\BaraB*,
541 importantly the *BaraB* gene of *D. simulans* was not immune-induced and is
542 expressed at a very low level despite direct sister species having immune-induced
543 *BaraB* genes.
544 These observations provide a precise phylogenetic timeframe confirming
545 that the last point where the *Dmel\BaraB* lineage encoded a functional signal
546 peptide was in the last common ancestor of the Melanogaster species complex
547 *5*.9mya. The *D. melanogaster* and *D. simulans BaraB* genes appear to have lost
548 immune regulation only recently, as *BaraB* is robustly expressed and immune-
549 induced in the Melanogaster sister species *D. mauritiana* and *D. sechellia*. Taken
550 together, we demonstrate not only that this AMP-encoding gene has repeatedly
551 evolved non-immune roles in multiple species, but that one such event unfolded
552 very recently in the Melanogaster species complex with significant implications on
553 fitness during development. We therefore characterize the precise timeframe
554 wherein an AMP-encoding gene transitioned to perform a novel function in the
555 nervous system, and highlight key sequence and structural changes correlated with
556 this neofunctionalization event.

557 ## Conclusion and Future Directions

558 The *Baramicin* gene family has undergone a varied history of genomic
559 duplication, translocation, and complex sequence evolution. The *BaraA* duplication
560 may be especially relevant to immunity studies given it is variable in common lab
561 stocks. Our initial exploration using data from a previous DGRP infection screen
562 suggests that *BaraA* copy number can explain a great deal of DGRP strain response
563 to infection (also see supplementary information). As *BaraA* underlies a sexually
564 dimorphic "erect wing" response to infection (Hanson et al. 2020), and other short
565 immune peptides have recently been implicated in behavioural regulation (Barajas-
566 azpeleta et al. 2018; Ebrahim et al. 2021; Sinner et al. 2021), *BaraA* copy number
567 variation may be an important consideration beyond immune function.

568 The ancestral *BaraA*-like gene structure is comprised of multiple distinct
569 peptide domains (i.e. IM24, IM10-like peptides, IM22). Two recent studies suggest
570 the *BaraA* gene (Huang et al. 2020) and specifically the IM10-like peptides have
571 antifungal properties (Hanson et al. 2020). We found a striking convergent
572 polymorphism in the IM10-like peptides (**Fig.** S2) similar to polymorphisms found
573 in other AMPs that can have drastic effects on host-pathogen interactions (Unckless
574 et al. 2016). We also discovered multiple non-immune *Baramicin* paralogs in various
575 *Drosophila* species that have convergently lost IM10-like peptides, reinforcing a role
576 for the IM10-like peptides specifically in the immune response. Development of
577 these peptides or other AMPs as therapeutics should consider the consequences of
578 such polymorphisms on antimicrobial activity.

579 The IM24 domain unites the *Baramicin* gene family. It was therefore
580 surprising that multiple *Baramicin* genes retain this domain, but have lost immune
581 induction. In the case of *D. melanogaster BaraB* and *BaraC* we found enrichment in
582 the nervous system, where *BaraB* knockdown by the pan-neural driver *elav-Gal4*
583 phenocopies both the lethality and nubbin-like wing phenotype. We additionally
584 investigated the effect of *BaraB* RNAi using Gal4 drivers in non-neural tissues
585 including the fat body (*c564-Gal4*), hemocytes (*hml-Gal4*), the gut (esg-Gal4),
586 malpighian tubules (MyO-Gal4), the wing disc (*nubbin-Gal4*), and in myocytes (*mef2-*

587 *gal4*) to no effect. We also screened neural drivers specific for glia (*Repo-Gal4*),
588 motor neurons (*D42-*, *VGMN-*, and *OK6-Gal4*), and a recently-made *BaraA-Gal4*
589 driver that is expressed in the larval ventral nervous system (Hanson et al. 2020).
590 However all these *Gal4>BaraB-IR* flies were viable and never exhibited overt
591 morphological defects. While we did not identify the specific neural cell type that
592 expresses *BaraB*, our Gal4 approach rules out various options including tissue
593 overlapping in expression with *BaraA*. Future study will benefit from systematically
594 knocking down *BaraB* in neuronal subtypes, and perhaps exploring *BaraB* evolution
595 in the Melanogaster species complex.
596 Adding to the complex patterns in IM24 evolution and gene function, the
597 cellular location of each Baramicin peptide is unclear. The Dmel\BaraB protein has
598 lost the otherwise conserved signal peptide structure of the *Baramicin* gene family, a
599 mutation that prevents secretion. Certainly *BaraA* IM24 is secreted, as this peptide
600 was first described for its robust extracellular signature as a component in immune-
601 induced hemolymph (Levy et al. 2004). Likewise, *Dmel\BaraC* bears a couple key
602 signatures indicating it is likely to reach the cell membrane: i) it encodes a signal
603 peptide that should enable entry into the secretory pathway, and ii) its IM24 domain
604 is preceded by an "SA" dipeptidyl peptidase site. XA or XP motifs are cleaved by
605 dipeptidyl peptidases at the cell membrane to maturate secreted peptides (Barrett
606 et al. 2004), and indeed XA or XP motifs are seen preceding the Q^1 residue in all
607 IM24 domains of BaraA and BaraC proteins (**Fig. 5**A), and an "SP" motif precedes
608 each of the *D. melanogaster* IM10-like peptides (Hanson et al. 2020). Given IM24
609 conservation in genes with totally different expression profiles, the general IM24
610 protein is not strictly tied to immunity. In this regard, AMPs and neuropeptides
611 share many features including: small size, cationic net charge, and amphipathicity
612 (Brogden 2005). Indeed Dmel\BaraA IM24 is both relatively short (96 AA) and
613 cationic (+2.4 at pH=7). However, while IM24 sequence is broadly conserved (**Fig.**
614 5A), changes present in Dmel\BaraC make it strikingly anionic by comparison: the
615 net charge of the putative extracellular part is -4.6 at pH=7 assuming the same N-
616 terminal pGlu modification found in BaraA (Krogh et al. 2001). The glia-expressed
617 *Dmel\BaraC* encodes a transmembrane tail that is found throughout the

618 Melanogaster and Obscura groups. This domain should enable insertion into either
619 the endogenous cell membrane, or the membrane of a neighbouring cell. The latter
620 possibility is intriguing, as glia envelope and regulate neurons. Discovering the
621 binding partner(s) of any of *Baramicin* IM24 proteins should be exciting, as this will
622 highlight a protein scaffold with either immune and/or neural functional domains.
623 While *Baramicins* are restricted to drosophilids, they are not the only AMP
624 gene family united by a domain not necessarily functioning in the immune response.
625 One human AMP recently implicated in chronic neuroinflammatory disease is the
626 Cathelicidin LL-37 (Lee et al. 2015; De Lorenzi et al. 2017; Moir et al. 2018). The
627 *Cathelicidin* gene family is unified by its N-terminal "Cathelin" domain, however this
628 region has no antimicrobial activity in vitro (Zanetti 2005). Instead, the
629 Cathelicidins encode their mature AMPs (e.g. LL-37) at the protein C-terminus. The
630 role of innate immunity and specifically antimicrobial peptides in neurological
631 disorders is a budding and exciting field of exploration (Barajas-azpeleta et al. 2018;
632 Gosztyla et al. 2018; Lezi et al. 2018; Masuzzo et al. 2019; Toda et al. 2019; Ebrahim
633 et al. 2021; Sinner et al. 2021). The *Baramicins* add to this growing list of immune-
634 neural interactions by providing an evolutionary case study demonstrating not all
635 peptides of AMP-encoding genes are immune-specific. Indeed, most AMP studies
636 have a peculiar blind spot in this regard, as *in vivo* experiments typically disrupt
637 entire genes, while ex vivo or in vitro approaches typically manipulate only the
638 canonical mature AMP. Future studies on AMPs relations to disease processes
639 should benefit greatly from dissecting the functions of AMPs at the gene level from
640 the specific functions of their encoded peptides, which could act in totally separate
641 tissues and contexts.
642 In summary, we characterize how an ancestral AMP-encoding gene has
643 evolved under a number of unique evolutionary scenarios. In light of the repeated
644 derivation of IM24-specific *Baramicin* genes – including at least two specifically
645 involved in neural processes – it will be interesting to consider the functions of AMP
646 genes in neural processes not simply at the level of the gene, but at the level of the
647 mature peptides produced by that gene. The distinct functional roles of the IM24
648 domain in neural processes and IM10-like peptides in immunity suggest that

649 selection on AMP genes may go beyond host-pathogen interactions. In the case of
650 *Baramicin* and other recently-characterized invertebrate AMPs, natural selection
651 may impose trade-offs that contrast pressures imposed by essential roles in host
652 processes (e.g. in neural tissues) with maintenance of a competent host defence.
653 Given the many commonalities between AMPs and neuropeptides, these canonical
654 immune effectors might be adapted for non-immune functions more commonly than
655 appreciated.
656
657 **Acknowledgements**
658 We would like to thank Maria Litovchenko for advice, Ana Marija Jakšić for
659 generously providing DGRP flies, Rob Unckless for stimulating discussion, Huang et
660 al. (Huang et al. 2020) for collaborative cooperation, and Florent Masson, Hannah
661 Westlake, and the anonymous reviewers for a thorough reading of the manuscript
662 and appreciated feedback. This research was supported by Sinergia grant
663 CRSII5_186397 awarded to Bruno Lemaitre. The *BaraB*^{LC1} and *BaraB*^{LC4} mutations
664 were graciously provided by Steven Wasserman and generated by Lianne Cohen,
665 who we also thank for their critical involvement in characterizing *Baramicin A*.
666
667 # Materials and Methods
668 ***DGRP population screening and bioinformatics analyses***
669 Genomic sequence data were downloaded from GenBank default reference
670 assemblies and Kim et al. (Kim et al. 2021), and DGRP sequence data from
671 http://dgrp2.gnets.ncsu.edu/ (Mackay et al. 2012). Sequence comparisons and
672 alignment figures were prepared using Geneious R10 (Kearse et al. 2012), Prism 7,
673 and Inkscape. Alignments were performed using MUSCLE or MAFFT followed by
674 manual curation, and phylogenetic analyses were performed to validate sequence
675 patterns using the Neighbour Joining, PhyML, RaxML, and MrBayes plugins in
676 Geneious. Figure S1B was generated using the UCSC *D. melanogaster* DGRP2 genome
677 browser. Selection analyses were performed using the HyPhy package implemented
678 in datamonkey.org (Delport et al. 2010). Codon alignments of the IM24 domain used
679 in **Fig. 5**A is included as a .fasta file in **supplementary data file 3** alongside outputs
680 from FEL, FUBAR, SLAC, and aBSREL selection analyses.
681
682 ***Fly genetics***
683 The *BaraB*^{LC1} and *BaraB*^{LC4} mutations were generated using CRISPR with two
684 gRNAs and an HDR vector by cloning 5' and 3' region-homologous arms into the
685 pHD-DsRed vector, and consequently *ΔBaraB* flies express DsRed in their eyes,
686 ocelli, and abdomen. The following PAM sites were used for CRISPR bordering the
687 *BaraB* region. Slashes indicate the cut site: 5': GCGGGCAACAGATGTGTTCA/GGG
688 3': GTCCATTGCTTATTCAAAAA/TGG. These mutants were generated in the
689 laboratory of Steve Wasserman by Lianne Cohen, who graciously allowed their use
690 in this study. All fly stocks including Gal4 and RNAi lines are listed in

691 **supplementary data file 4**. Experiments were performed at 25°C unless otherwise
692 indicated. When possible, genetic crosses of 6-8 males and 6-8 females were
693 performed in both directions to test for an effect of the X or Y chromosomes on
694 *BaraB*-mediated lethality; crosses in both directions yielded similar results in all
695 cases and reported data are pooled results. Fly diet consisted of a nutrient-rich lab
696 standard food: 3.72g agar, 35.28g cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL
697 propionic acid, 15.9mL moldex, and H_2O to 600mL.
698
699 ***Infection experiments***
700 Infection experiments and qPCR analyses were performed as previously
701 described (Hanson, Dostálová, et al. 2019). Additional details are available in the
702 supplementary information.
703

704 ## Figure captions
705
706 **Figure 1: The *BaraA* duplication is not fixed in wild flies. A)** Gene structure of
707 *BaraA*, including its three key domain types: IM24 (orange), IM10-like (green), and
708 IM22 (teal). SP = signal peptide. Purple triangles indicate furin cleavage sites. **B)**
709 Depiction of the *BaraA* duplication. The PCR amplicon used to detect presence of the
710 *BaraA* duplication event, and the region in the DGRP of unknown variant calls
711 correlated with the duplication are annotated (detailed in Fig. S1).
712
713 **Figure 2: The ancestral *BaraA* gene was immune-induced. A)** Cladogram of
714 species used in B-F. **B-F)** Expression of *BomBc3* control genes (brown) or *BaraA-like*
715 genes (orange) in diverse *Drosophila* species upon infection. In all cases, both
716 *BomBc3* and *BaraA-like* genes are induced upon infection by either *C. albicans* yeast
717 or *M. luteus* bacteria.
718
719 **Figure 3: The *D. melanogaster Baramicin* gene family. A)** Two *Baramicin*
720 paralogs in addition to *BaraA* are present in the Dmel_R6 reference genome, united
721 by the presence of an IM24 peptide. Relative to *BaraA*, two frameshift mutations in
722 *BaraB* are segregating in the DGRP. One results in the same mature protein
723 annotated in the Dmel_R6 reference genome (coloured domains), while the other
724 elongates the IM22 peptide (e.g. *DGRP101*). Most strains carry a *BaraB* gene of
725 similar length to *BaraA* (e.g. DGRP38). Additionally, the *BaraB* gene does not encode
726 a functional signal peptide. **B)** Only *BaraA* is immune-induced. *BaraB* and *BaraC* do
727 not respond to infection, though basal *BaraC* expression relies on Toll signalling
728 (**Fig.** S3C). **C)** Time series of whole animal *Baramicin* expression over the course of
729 development. Expression values are normalized within each gene, with expression
730 in the adult set to 100% (left panel). For context, normalizing each gene to *BaraA* in
731 adults shows that *BaraC* and especially *BaraB* expression is much lower at a whole
732 fly level (right panel). **D)** The $\Delta BaraB^{LC1}$ mutation is partial lethal, an effect that is
733 exaggerated by rearing flies at 29°C. **E)** Example of a nubbin-like wing, which is
734 phenocopied by *BaraB* gene silencing using the neural driver *elav-Gal4*. **F)** The
735 *BaraC* gene almost perfectly matches the expression pattern of glia specific *Repo*-
736 expressing cells in single cell RNA sequencing of the adult fly brain (Davie et al.
737 2018).
738
739 **Figure 4: *Baramicin* evolutionary history. A)** Detailed map of genomic
740 neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D.*
741 *melanogaster*, detailing inferred duplication, inversion, and translocation events.
742 Gene names are given as found in *D. melanogaster*. **B)** Cladogram and genomic loci
743 further detailing the series of events leading to the extant *Baramicin* loci of
744 *Drosophila* species. Loci in *S. lebanonensis* and flies in the subgenus *Drosophila*
745 encode only one *Baramicin* gene, indicating the ancestral drosophilid likely encoded
746 only one *Baramicin*. **C)** IM24-specific *Baramicins* arose from convergent evolution
747 both in gene structure and expression profile. Genomic loci are described here as
748 *ATP8A* or *stum* reflecting prominent genes neighbouring the *Baramicins* (see **Fig.**

749 5A). Expression dynamics relating to immune-induction or enrichment in the head
750 (checked boxes) are shown in **Fig.** 2 and **Fig.** S6. The *Baramicin* loci in *D. willistoni*
751 are syntenic with *D. melanogaster*, but evolved in a vice versa fashion (purple
752 arrow). The *D. virilis Baramicins GJ21309* and *GJ25897* are direct sister genes (100%
753 identity at N-terminus).
754
755 **Figure 5: *BaraB* and IM24 rapid evolution. A)** Evolution of the signal peptide and
756 core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*.
757 Arrows indicate site 29 and the *D. melanogaster* ^{40}HHASSPAD^{48} domain. Insertion
758 events in *D. yakuba* and *D. sechellia BaraB* are denoted as XX to save space. The *D.*
759 *sechellia* signal peptide is predicted to remain functional (SignalP 5.0). The
760 cladogram on the left shows genomic relatedness (by speciation and locus)
761 independent of sequence similarity. Background colouring is included to show birth
762 of novel *Baramicin* loci/lineages. **B)** Proposed timeframe for gain of novel *BaraB*
763 function. The *D. erecta BaraB* gene is pseudogenized, and the *D. yakuba* gene is not
764 immune-induced and encodes a 9-residue insertion in the IM24 peptide bordering
765 the HHASSPAD domain (see **A**). However the *BaraB* genes of *D. suzukii*, and both *D.*
766 *mauritiana* and *D. sechellia* remain inducible by infection, while the *BaraB* genes of
767 *D. simulans* and *D. melanogaster* are not, and are expressed at very low levels (Fig.
768 S8). This pattern suggests that *BaraB* of *D. melanogaster* (and *D. simulans*) acquired
769 its non-immune role only recently, and is correlated with the loss of the *BaraB* signal
770 peptide. Speciation timeframes adapted from Timetree.org (Kumar et al. 2017).
771
772

Supplementary figures and tables

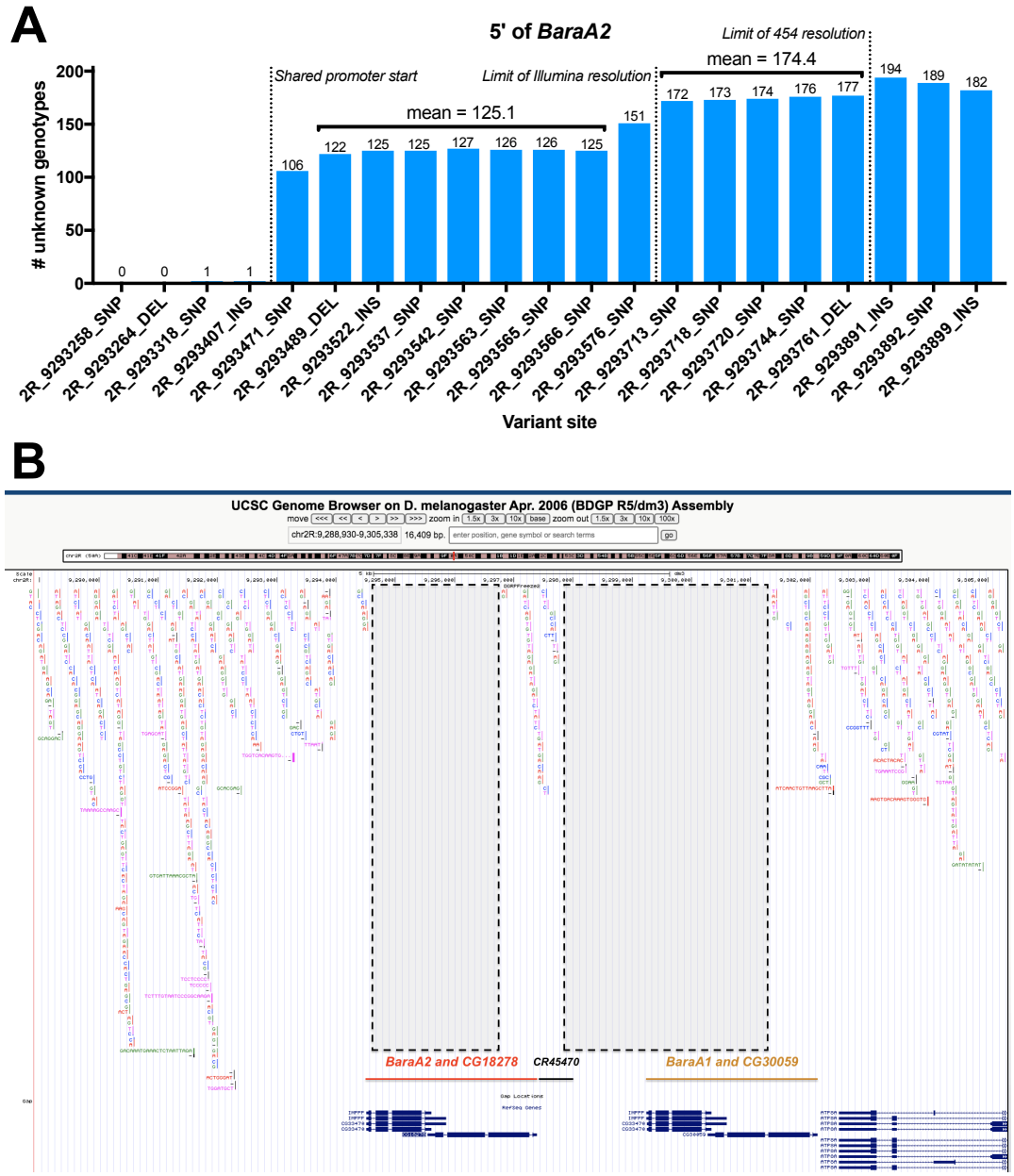
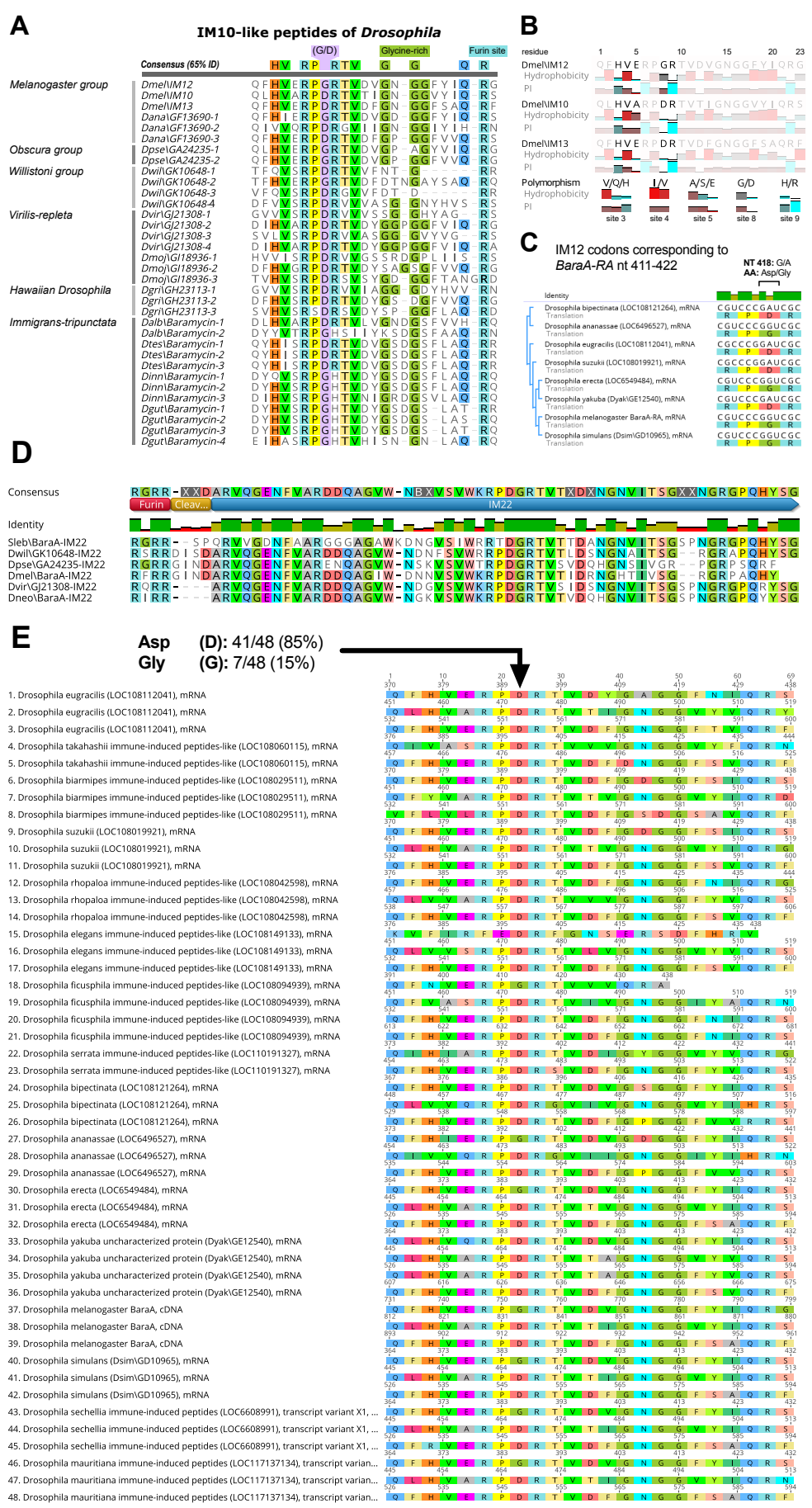


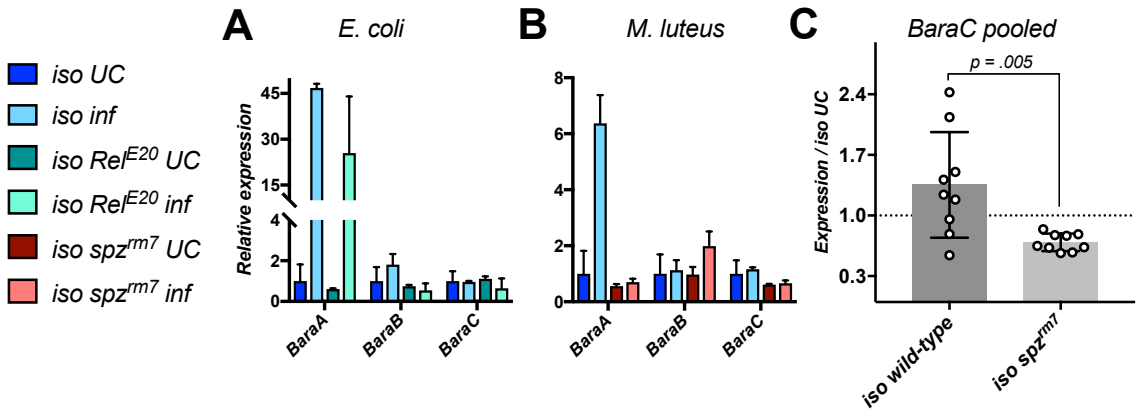
Figure S1: The *BaraA* locus is poorly resolved in DGRP genome assemblies. A) The number of unknown variant calls rises immediately after the start of the identical promoter region shared by *BaraA1* and *BaraA2,* stabilizing at ~125/205 DGRP lines. Approaching the limits of Illumina resolution (~300bp), a second spike in unknown calls stabilizes at ~174/205 lines corresponding to the limits of 454 sequencing (~450bp), after which unknown call frequency seems more randomly distributed. **B)** 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.

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**

786
787 **Table S1: the *BaraA* duplication is variable in common lab stocks.** Inferred
788 *BaraA* copy numbers are given for six different lab stocks including *Oregon R* (*OR-R*),
789 *iso DrosDel* (*iso* w^{1118}), *Exelexis*, a w^{1118} line from the VDRC, a *yellow white* (*yw*) line
790 used previously by our group, and *Canton S*. All samples were additionally screened
791 for *CG30059* as a DNA extraction positive control.
792


A
IM10-like peptides of Drosophila
(G/D)
Glycine-rich
Furin site
Consensus (65% ID)
Melanogaster group
Obscura group
Willistoni group
Virilis-repleta
Hawaiian Drosophila
Immigrans-tripunctata
B
residue
Dmel\IM12
Dmel\IM10
Dmel\IM13
Hydrophobicity
PI
Polymorphism
V/Q/H
I/V
A/S/E
G/D
H/R
site 3
site 4
site 5
site 8
site 9
C
IM12 codons corresponding to BaraA-RA nt 411-422
NT 418: G/A
AA: Asp/Gly
Identity
Drosophila bipectinata (LOC108121264), mRNA
Drosophila ananassae (LOC6496527), mRNA
Drosophila eugracilis (LOC108112041), mRNA
Drosophila suzukii (LOC108019921), mRNA
Drosophila erecta (LOC6549484), mRNA
Drosophila yakuba (Dyak\GE12540), mRNA
Drosophila melanogaster BaraA-RA, mRNA
Drosophila simulans (Dsim\GD10965), mRNA
Translation
D
Consensus
Furin
Cleav...
IM22
Identity
Sleb\BaraA-IM22
Dwil\GK10648-IM22
Dpse\GA24235-IM22
Dmel\BaraA-IM22
Dvir\GJ21308-IM22
Dneo\BaraA-IM22
E
Asp (D): 41/48 (85%)
Gly (G): 7/48 (15%)
1. Drosophila eugracilis (LOC108112041), mRNA
2. Drosophila eugracilis (LOC108112041), mRNA
3. Drosophila eugracilis (LOC108112041), mRNA
4. Drosophila takahashii immune-induced peptides-like (LOC108060115), mRNA
5. Drosophila takahashii immune-induced peptides-like (LOC108060115), mRNA
6. Drosophila biarmipes immune-induced peptides-like (LOC108029511), mRNA
7. Drosophila biarmipes immune-induced peptides-like (LOC108029511), mRNA
8. Drosophila biarmipes immune-induced peptides-like (LOC108029511), mRNA
9. Drosophila suzukii (LOC108019921), mRNA
10. Drosophila suzukii (LOC108019921), mRNA
11. Drosophila suzukii (LOC108019921), mRNA
12. Drosophila rhopaloa immune-induced peptides-like (LOC108042598), mRNA
13. Drosophila rhopaloa immune-induced peptides-like (LOC108042598), mRNA
14. Drosophila rhopaloa immune-induced peptides-like (LOC108042598), mRNA
15. Drosophila elegans immune-induced peptides-like (LOC108149133), mRNA
16. Drosophila elegans immune-induced peptides-like (LOC108149133), mRNA
17. Drosophila elegans immune-induced peptides-like (LOC108149133), mRNA
18. Drosophila ficusphila immune-induced peptides-like (LOC108094939), mRNA
19. Drosophila ficusphila immune-induced peptides-like (LOC108094939), mRNA
20. Drosophila ficusphila immune-induced peptides-like (LOC108094939), mRNA
21. Drosophila ficusphila immune-induced peptides-like (LOC108094939), mRNA
22. Drosophila serrata immune-induced peptides-like (LOC110191327), mRNA
23. Drosophila serrata immune-induced peptides-like (LOC110191327), mRNA
24. Drosophila bipectinata (LOC108121264), mRNA
25. Drosophila bipectinata (LOC108121264), mRNA
26. Drosophila bipectinata (LOC108121264), mRNA
27. Drosophila ananassae (LOC6496527), mRNA
28. Drosophila ananassae (LOC6496527), mRNA
29. Drosophila ananassae (LOC6496527), mRNA
30. Drosophila erecta (LOC6549484), mRNA
31. Drosophila erecta (LOC6549484), mRNA
32. Drosophila erecta (LOC6549484), mRNA
33. Drosophila yakuba uncharacterized protein (Dyak\GE12540), mRNA
34. Drosophila yakuba uncharacterized protein (Dyak\GE12540), mRNA
35. Drosophila yakuba uncharacterized protein (Dyak\GE12540), mRNA
36. Drosophila yakuba uncharacterized protein (Dyak\GE12540), mRNA
37. Drosophila melanogaster BaraA, cDNA
38. Drosophila melanogaster BaraA, cDNA
39. Drosophila melanogaster BaraA, cDNA
40. Drosophila simulans (Dsim\GD10965), mRNA
41. Drosophila simulans (Dsim\GD10965), mRNA
42. Drosophila simulans (Dsim\GD10965), mRNA
43. Drosophila sechellia immune-induced peptides (LOC6608991), transcript variant X1, ...
44. Drosophila sechellia immune-induced peptides (LOC6608991), transcript variant X1, ...
45. Drosophila sechellia immune-induced peptides (LOC6608991), transcript variant X1, ...
46. Drosophila mauritiana immune-induced peptides (LOC117137134), transcript varian...
47. Drosophila mauritiana immune-induced peptides (LOC117137134), transcript varian...
48. Drosophila mauritiana immune-induced peptides (LOC117137134), transcript varian...

794 **Figure S2: Convergent evolution maintains a polymorphism in Baramicin**
795 **IM10-like peptides. A)** Alignment of IM10-like peptides from diverse *Drosophila*
796 species. The G/D polymorphic site is highlighted in lavender (residue 8). Broadly
797 conserved polymorphisms are also seen at sites 3 (H/V/Q), 4 (V/I), 5 (A/E/S), and 9
798 (R/H). **B)** The IM10-like polymorphic residues confer different hydrophobicity and
799 isoelectric points (PI). The three *D. melanogaster* IM10-like peptides are given as
800 examples. Bar height indicates greater hydrophobicity/PI. **C)** The syntenic codon of
801 IM12 in Melanogaster group flies encodes either a D or G residue owing to
802 convergent G-A nucleotide transitions. The D residue is far more common in
803 additional outgroups (**Fig.** S2E), suggesting the ancestral state encoded D, and G was
804 derived at least twice. **D)** IM22 encodes a unique peptide with a domain similar to
805 the IM10-like ^{6}RP(G/D)R^{9} motif, except both D and G are encoded in tandem. There
806 is a furin cleavage site upstream of IM22 in all *BaraA-like* genes. In *S. lebanonensis*
807 there is an SP dipeptidyl peptidase site that is also predicted to be cleaved off. In *D.*
808 *melanogaster BaraA* the GIND motif is also cleaved off by an unknown process
809 (Hanson et al. 2020). In subgenus Drosophila flies, the mature IM22 peptide begins
810 immediately following the furin cleavage site. **E)** Presentation of the residue 8 G/D
811 polymorphism in all IM10-like peptides (i.e. IM10, IM12, IM13) of Melanogaster
812 group flies. 41/48 encode an Aspartic acid, while 7/48 encode a Glycine. However,
813 in the Quinaria species group, only Glycine residues are observed (**Fig.** S2A),
814 suggesting G/D frequency is lineage-specific.
815
816
817
818



819
820 **Figure S3: Additional assays of Baramicin expression upon infection. A)**
821 Neither *BaraB* nor *BaraC* are regulated by the Imd pathway, which is specifically
822 stimulated by *E. coli* infection. **B)** Neither *BaraB* nor *BaraC* are induced after
823 infection by *M. luteus*. **C)** *BaraC* levels were consistently depressed in *spz^{rm7}* flies in
824 the unchallenged condition (UC) or upon infection with *C. albicans* (**Fig.** 3B) or *M.*
825 *luteus* (**Fig.** S4B). Data here are pooled for *iso* wild type or *iso spz^{rm7}* flies without
826 regard for infection treatment (student's t, p = .005).

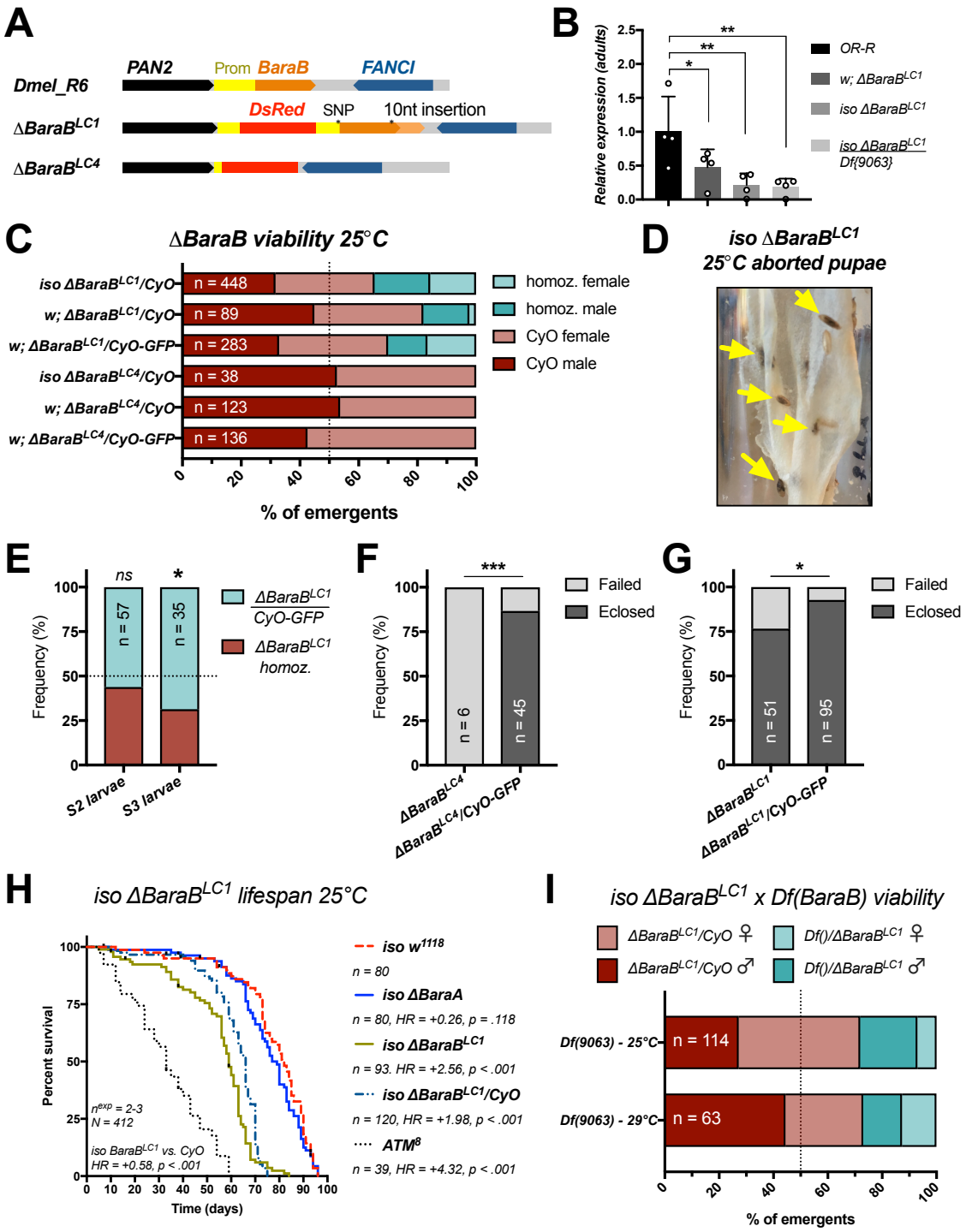


Figure S4: *BaraB* mutation is highly deleterious and penetrance of effect depends on genetic background. A) Diagram of *BaraB* mutant loci. The *ΔBaraB*^{LC1} mutation encodes an unintended insertion in the *BaraB* promoter, and both a SNP in the promoter and frameshift in protein product (asterisks). The *ΔBaraB*^{LC4} mutation encodes the intended knock-in of a DsRed cassette in place of the *BaraB* gene producing a full gene knockout. **B)** *ΔBaraB*^{LC1} is a *BaraB* hypomorph. Under normal

834 qPCR assay conditions, *BaraB* expression is not detected. However using highly
835 concentrated cDNA beyond our assay's valid range (100ng/10μL reaction), we could
836 detect *BaraB* transcript in *ΔBaraB*^{LC1} flies. Quantification shown here is intended
837 only to show that *BaraB* transcript can be recovered from *ΔBaraB*^{LC1} homozygotes,
838 and to give a sense of relative whole-fly expression levels. **C)** Emergent frequencies
839 of *ΔBaraB* flies in a variety of genetic backgrounds. *iso ΔBaraB*^{LC1} flies were
840 backcrossed for seven generations. *w; ΔBaraB/CyO* flies represent the original
841 background the mutation was generated in, while *ΔBaraB/CyO-GFP* flies have had
842 additional genetic variation introduced in their 1st and 3rd chromosomes. Rare
843 homozygous L3 larvae were observed in *ΔBaraB*^{LC4}/CyO-GFP vials (<5% of L3
844 larvae), but these larvae never successfully eclosed. **D)** Aborted pupae (yellow
845 arrows) are a common occurrence in *ΔBaraB* vials, and sometimes contain fully-
846 developed adults that simply never eclosed. In **E-G:** ns = not significant, * = $p < .05$,
847 *** $p = < .001$. **E)** The ratio of *ΔBaraB*^{LC1}/CyO-GFP to *ΔBaraB*^{LC1} homozygous larvae
848 drops between the S2 and S3 larval stages (χ^2, $p = .515$ and $p = .012$ respectively). **F)**
849 Frequency of successfully eclosing adults using *BaraB*^{LC4}/CyO-GFP flies. **G)**
850 Frequency of successfully eclosing adults using *BaraB*^{LC1}/CyO-GFP flies. **H)** *BaraB*
851 mutation negatively affects lifespan. *iso ΔBaraB*^{LC1} homozygotes suffer reduced
852 lifespan even relative to their *iso ΔBaraB*^{LC1}*/CyO* siblings. By comparison, *iso ΔBaraA*
853 flies that used the same vector for mutant generation live as wild-type. ATM^{8} flies
854 suffer precocious neurodegeneration and are included as short-lived controls
855 (Petersen et al. 2013). **I)** *ΔBaraB*^{LC1} crossed to the genomic deficiency line (*Df(9063)*)
856 supports a partial-lethal effect of *BaraB* mutation.
857

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.

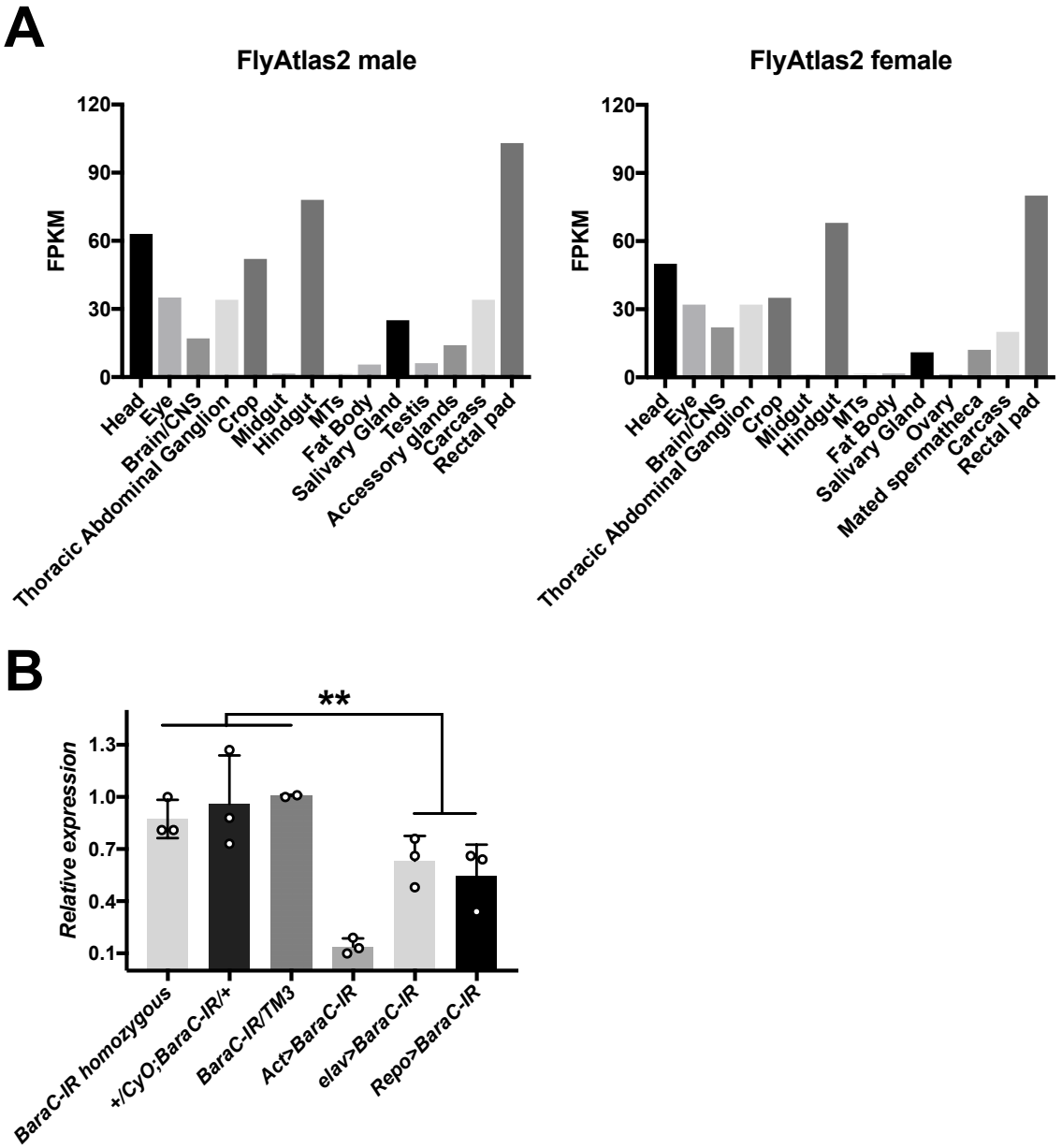


Figure S5: *BaraC* is expressed in the nervous system, but also the hindgut and rectal pads. A) FlyAtlas2 expression data for *BaraC*. **B)** RT-qPCR of *BaraC* in whole flies using different Gal4 drivers to express *BaraC* RNAi. *BaraC* is knocked down by both the *elav-Gal4* and *Repo-Gal4* nervous 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.

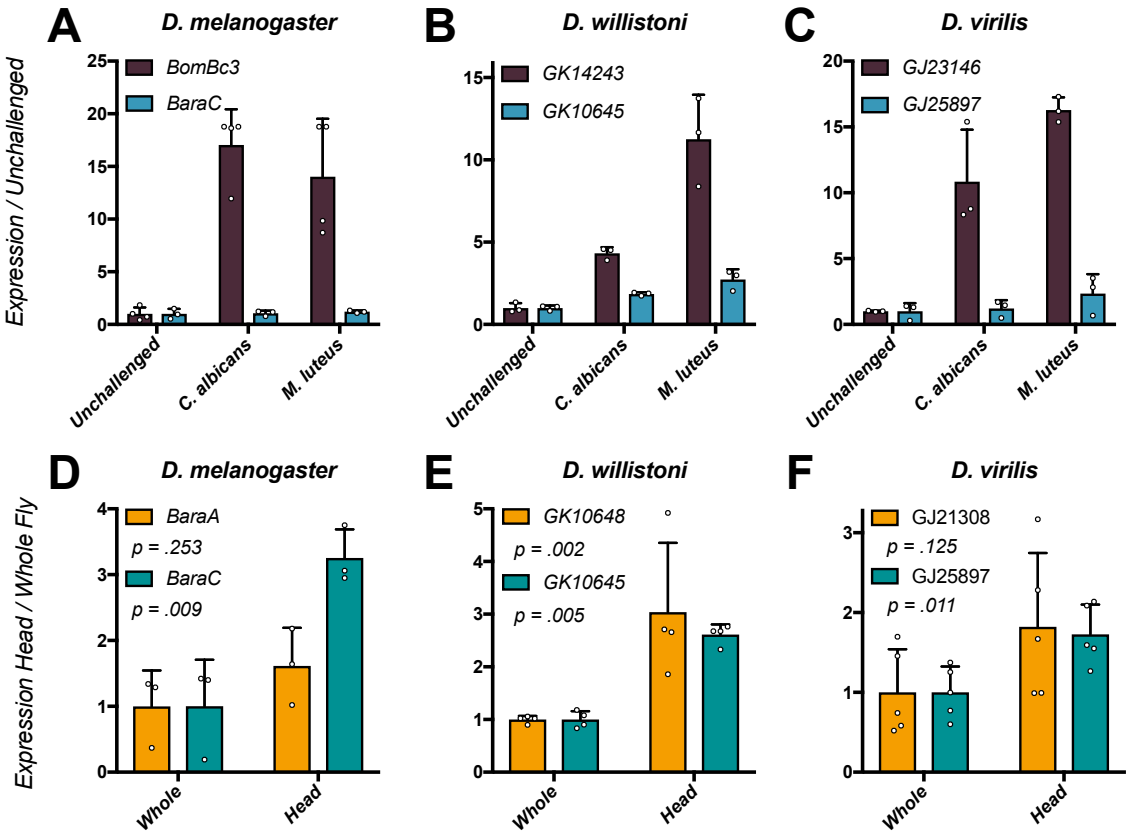


Figure S6: RT-qPCR of *Baramicin* genes in diverse species. A-C) Independent IM24-specific genes in *D. melanogaster* (A), *D. willistoni* (B), and *D. virilis* (C) are not induced by infection. *BomBc3* is included as an immune-induced control. **D-F)** The independent IM24-specific genes (blue) of *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.

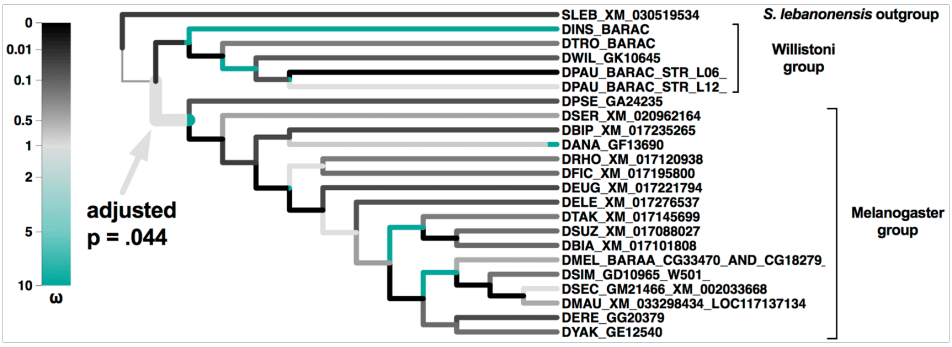


Figure S7: Adaptive branch site relative-effects likelihood (aBSREL) analysis of the Sophophora ATP8A locus (primarily *BaraA* genes) for episodic diversifying selection. One branch along the tree had a significant dN/dS distinguishing Willistoni group *Baramicins* from *D. melanogaster* and relatives (p-adj = .044).

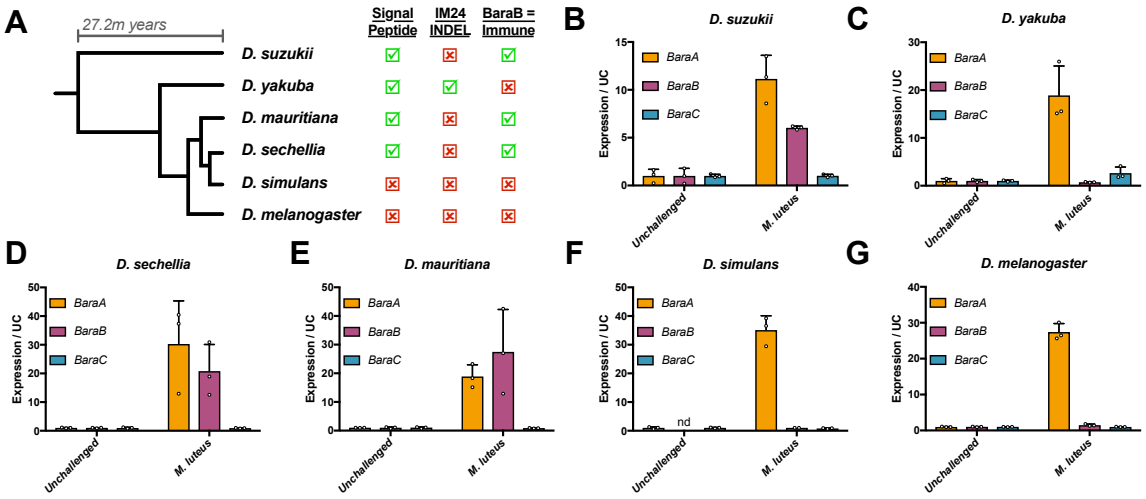


Figure S8: The *D. melanogaster BaraB* gene acquired its non-immune role only recently. A) Phylogeny of the Melanogaster species group, with divergence times from (Kumar et al. 2017). The presence of a functional signal peptide (Fig. 6A), and the disruption of the *D. yakuba* IM24 peptide by an in-frame 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* and *D. mauritiana*. *BaraB* is also induced *D. sechellia* but not in the sister species *D. simulans*, nor in *D. melanogaster,* which lack signal peptide structures. *Drosophila yakuba BaraB* is not immune-induced (C), has an insertion event in its IM24 peptide (Fig. 6A), and its sister species *D. erecta* has pseudogenized its *BaraB* orthologue (Fig. 6B), suggesting pseudogenization may explain the lack of immune induction in *D. yakuba BaraB*.

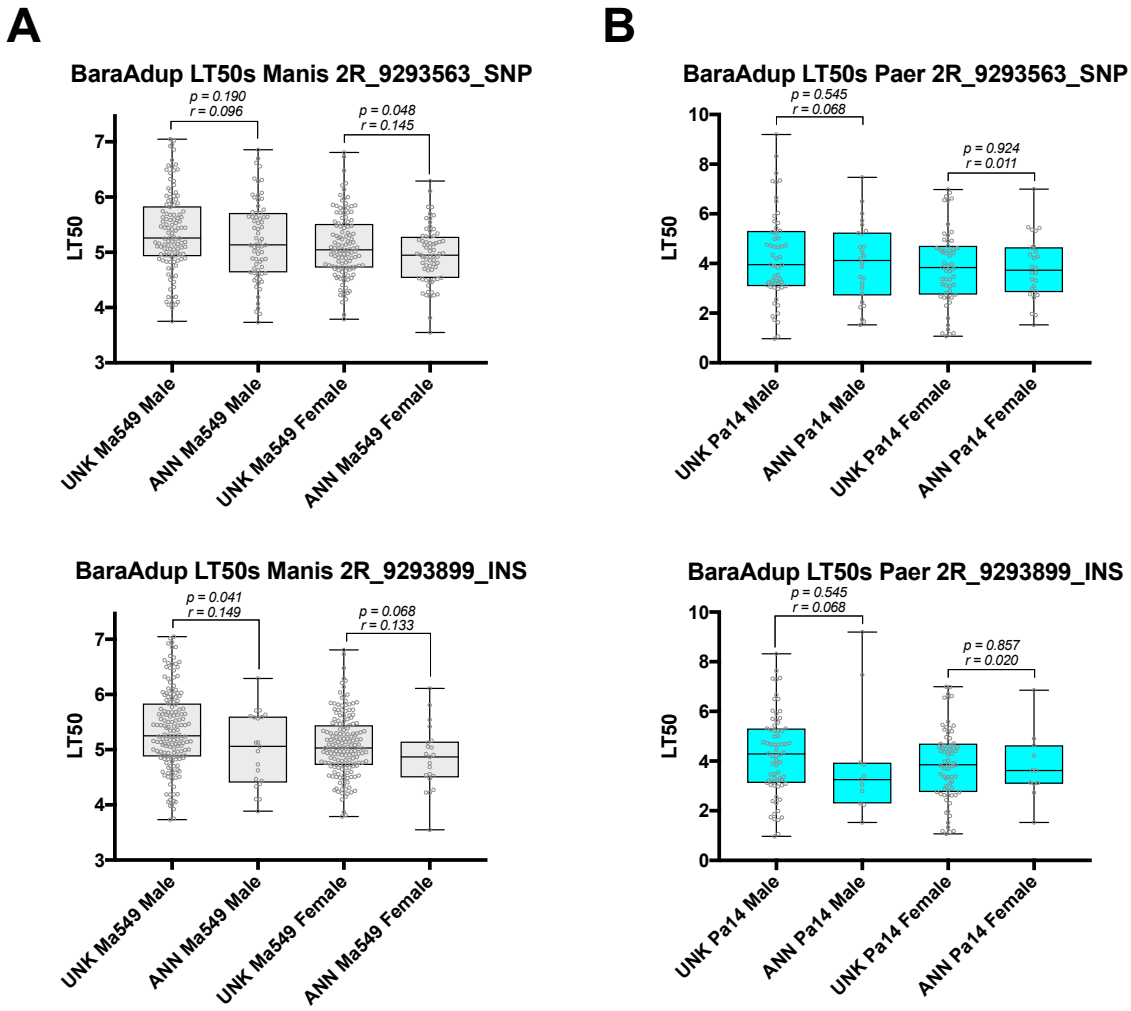


Figure S9: Variant call status at the *BaraA* locus predicts defence specifically against fungal infection in data from Wang et al. (Wang et al. 2017). Marginal and significant correlations (0.09 < |*rho*| < 0.16) were found between unknown status and *BaraA* variant sites in defence against **A)** *Metarhizium anisopliae* fungal infection (0.03 < p < 0.10, except males*2R_9293563_SNP p = 0.196), but not **B)** *Pseudomonas aeruginosa* bacterial infection. Pearson's rho and p-values come from Pearson's correlation matrix implemented in Prism 7. LT50 reflects the time taken until 50% of flies perished, and so higher LT50 indicates longer survival. Example variant site identifiers are given above each plot. Boxplots here show underlying data for two representative sites split by sex and *BaraA* variant call status.

918 **References:**
919

920 Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak
921 S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide
922 predictions using deep neural networks. *Nat. Biotechnol.* 37:420–423.

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

925 Barrett AJ, Rawlings ND, Woessner JF eds. 2004. Handbook of proteolytic enzymes.
926 2nd ed. Amsterdam ; Boston: Elsvier Academic Press

927 Brogden KA. 2005. Antimicrobial peptides: Pore formers or metabolic inhibitors in
928 bacteria?

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

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

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

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

942 Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. 2010. Datamonkey 2010: A
943 suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*
944 26:2455–2457.

945 Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, Nguyen M,
946 Haditsch U, Raha D, Griffin C, et al. 2019. Porphyromonas gingivalis in
947 Alzheimer's disease brains: Evidence for disease causation and treatment
948 with small-molecule inhibitors. *Science Advances* 5.

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

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

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

958 Gosztyla ML, Brothers HM, Robinson SR. 2018. Alzheimer's Amyloid-β is an
959 Antimicrobial Peptide: A Review of the Evidence.

960 Halldórsdóttir K, Árnason E. 2015. Trans-species polymorphism at antimicrobial
961 innate immunity cathelicidin genes of Atlantic cod and related species. *PeerJ*
962 3:e976.

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

966 Hanson MA, Cohen LB, Marra A, Iatsenko I, Wasserman SA, Lemaitre B. 2020. The
967 Drosophila Baramicin polypeptide gene protects against fungal infection.
968 bioRxiv Available from:
969 http://biorxiv.org/lookup/doi/10.1101/2020.11.23.394148

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

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

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

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

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

984 Huang J, Lou Y, Liu J, Bulet P, Jiao R, Hoffmann JA, Liegeois S, Li Z, Ferrandon D. 2020.
985 The BaramicinA gene is required at several steps of the host defense against
986 Enterococcus faecalis and Metarhizium robertsii in a septic wound infection

987 model in Drosophila melanogaster. bioRxiv Available from:
988 http://biorxiv.org/lookup/doi/10.1101/2020.11.23.394809

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

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

995 Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web
996 portal for protein modeling, prediction and analysis. *Nature protocols*
997 10:845–858.

998 Kim BY, Wang J, Miller DE, Barmina O, Delaney EK, Thompson A, Comeault AA,
999 Peede D, D'Agostino ERR, Pelaez J, et al. 2021. Highly contiguous assemblies
1000 of 101 drosophilid genomes. *eLife* 10:e66405.

1001 Kobler JM, Rodriguez Jimenez FJ, Petcu I, Grunwald Kadow IC. 2020. Immune
1002 Receptor Signaling and the Mushroom Body Mediate Post-ingestion Pathogen
1003 Avoidance. *Curr Biol* 30:4693-4709.e3.

1004 Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane
1005 protein topology with a hidden Markov model: application to complete
1006 genomes. *J Mol Biol* 305:567–580.

1007 Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for
1008 Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution*
1009 34:1812–1819.

1010 Lazzaro BP, Zasloff M, Rolff J. 2020. Antimicrobial peptides: Application informed by
1011 evolution. *Science* 368.

1012 Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. 2018. FlyAtlas 2: A new version
1013 of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq
1014 and sex-specific data. *Nucleic Acids Research* 46:D809–D815.

1015 Lee M, Shi X, Barron AE, McGeer E, McGeer PL. 2015. Human antimicrobial peptide
1016 LL-37 induces glial-mediated neuroinflammation. *Biochemical Pharmacology*
1017 94:130–141.

1018 Lemaitre B, Reichhart JM, Hoffmann JA. 1997. Drosophila host defense: differential
1019 induction of antimicrobial peptide genes after infection by various classes of
1020 microorganisms. *Proceedings of the National Academy of Sciences of the*
1021 *United States of America* 94:14614–14619.

1022 Levy F, Rabel D, Charlet M, Bulet P, Hoffmann JA, Ehret-Sabatier L. 2004. Peptidomic
1023 and proteomic analyses of the systemic immune response of Drosophila.
1024 *Biochimie* 86:607–616.

1025 Lezi E, Zhou T, Koh S, Chuang M, Sharma R, Pujol N, Chisholm AD, Eroglu C,
1026 Matsunami H, Yan D. 2018. An Antimicrobial Peptide and Its Neuronal
1027 Receptor Regulate Dendrite Degeneration in Aging and Infection. *Neuron*
1028 97:125-138.e5.

1029 Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y,
1030 Magwire MM, Cridland JM, et al. 2012. The Drosophila melanogaster Genetic
1031 Reference Panel. *Nature* 482:173–178.

1032 Masuzzo A, Manière G, Viallat-Lieutaud A, Avazeri É, Zugasti O, Grosjean Y, Kurz CL,
1033 Royet J. 2019. Peptidoglycan-dependent NF-κB activation in a small subset of
1034 brain octopaminergic neurons controls female oviposition. *eLife* 8:e50559.

1035 Moir RD, Lathe R, Tanzi RE. 2018. The antimicrobial protection hypothesis of
1036 Alzheimer's disease. *Alzheimer's & Dementia* 14:1602–1614.

1037 Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner
1038 T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic
1039 selection. *Mol Biol Evol* 32:1365–1371.

1040 Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, Griffin RS, Belfer I,
1041 Dai F, Smith SB, et al. 2010. A Genome-wide Drosophila screen for heat
1042 nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell*
1043 143:628–638.

1044 Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time
1045 RT-PCR. *Nucleic Acids Res.* 29:e45.

1046 Pineda A, Burré J. 2017. Modulating membrane binding of α-synuclein as a
1047 therapeutic strategy. *Proceedings of the National Academy of Sciences*
1048 114:1223–1225.

1049 Sackton TB, Lazzaro BP, Clark AG, Wittkopp P. 2017. Rapid expansion of immune-
1050 related gene families in the house fly, musca domestica. *Molecular Biology*
1051 *and Evolution*.

1052 Sackton TB, Lazzaro BP, Schlenke T a, Evans JD, Hultmark D, Clark AG. 2007.
1053 Dynamic evolution of the innate immune system in Drosophila. *Nature*
1054 *genetics* 39:1461–1468.

1055 Semple F, Dorin JR. 2012. ß-Defensins: Multifunctional Modulators of Infection,
1056 Inflammation and More? *J Innate Immun* 4:337–348.

1057 Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. 2021. Innate Immunity
1058 Promotes Sleep through Epidermal Antimicrobial Peptides. *Current Biology*
1059 31:564-577.e12.

1060 Tennessen JA. 2005. Molecular evolution of animal antimicrobial peptides:
1061 Widespread moderate positive selection.

1062 Toda H, Williams JA, Gulledge M. 2019. A sleep-inducing gene, nemuri , links sleep
1063 and immune function in Drosophila. 515:509–515.

1064 Unckless RL, Howick VM, Lazzaro BP. 2016. Convergent Balancing Selection on an
1065 Antimicrobial Peptide in Drosophila. *Current Biology* 26:257–262.

1066 Unckless RL, Lazzaro BP. 2016. The potential for adaptive maintenance of diversity
1067 in insect antimicrobial peptides. *Philosophical Transactions of the Royal*
1068 *Society B: Biological Sciences* 371:20150291.

1069 Vilcinskas A, Mukherjee K, Vogel H. 2013. Expansion of the antimicrobial peptide
1070 repertoire in the invasive ladybird Harmonia axyridis. *Proceedings of the*
1071 *Royal Society B: Biological Sciences*.

1072 Wang JB, Lu H-L, St Leger RJ. 2017. The genetic basis for variation in resistance to
1073 infection in the Drosophila melanogaster genetic reference panel. *PLoS*
1074 *Pathog* 13:e1006260.

1075 Wang Y, Zhu S. 2011. The defensin gene family expansion in the tick Ixodes
1076 scapularis. *Developmental and Comparative Immunology*.

1077 Zanetti M. 2005. The role of cathelicidins in the innate host defenses of mammals.
1078 *Curr Issues Mol Biol* 7:179–196.

1079 Zhang W, Yang J, He B, Walker SE, Zhang H, Govindarajoo B, Virtanen J, Xue Z, Shen
1080 H-B, Zhang Y. 2016. Integration of QUARK and I-TASSER for Ab Initio Protein
1081 Structure Prediction in CASP11. *Proteins* 84 Suppl 1:76–86.

1082