# *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 [1–3]. AMP genes frequently experience events of duplication and loss [4–7] and undergo rapid evolution at the sequence level [8,9]. This rapid evolution is driven by positive selection, and acts to either maintain genetic variation (balancing selection), or differentiate AMP sequence (diversifying selection) [10–15]. The selective pressures that drive these evolutionary outcomes are likely the consequence of host-pathogen

interactions. One striking example is the maintenance of two alleles of *Diptericin A* in *Drosophila* that predict defence against infection by *Providencia rettgeri* bacteria [16]. However AMPs and AMP-like genes in various species have also recently been implicated in non-immune roles as regulators of sleep [17,18], memory [19], and neurodegeneration [20–24] in both flies and nematodes. These new contexts suggest that the evolutionary forces acting on AMP genes may not be driven strictly by trade-offs in host defence, but rather by conflicts between roles in immunity and other non-immune functions.

[This part in yellow is too much]

For instance, Diptericins are membrane-disrupting antimicrobial peptides of flies (Diptera) that are required for defence against infection [4,16,25]. It was therefore surprising that the *D. melanogaster* gene *Diptericin B (DptB)* affects memory processes [19]. In this study, *DptB* derived from the fly fat body (analogous to the mammalian liver) regulated the ability of the fly to form long-term memory associations [19]. Another study in nematodes showed that an immune-induced polypeptide (NLP-29) binds to a G-protein coupled receptor (NPR-12) triggering neurodegeneration through activation of the NPR-12-dependent autophagy pathway [20]. The two *Drosophila* AMP-like effectors *GNBP-like3* and *IBIN* have also been shown to affect neuronal processes derived from expression within the nervous system [19,26]. In humans, the *Cathelicidin* gene encodes the AMP LL-37, which is implicated in glia-mediated neuroinflammation and Alzheimer's disease [24,27], alongside evidence of Alzheimer's as being stimulated by pathogen-derived inflammation [21]. Thus, immune effectors best known for their roles in the systemic immune response can have surprising interactions with host neurology, derived either from expression within neuronal tissue or through production via immune-responsive tissue.

We recently described a novel antifungal peptide gene of *Drosophila melanogaster* that we named *Baramicin A* (*BaraA*) [28]. A unique aspect of *BaraA* is its precursor protein structure, which encodes a polypeptide cleaved into multiple mature products by interspersed furin cleavage sites. The use of furin cleavage sites to produce two mature peptides from a single polypeptide precursor is common in

73 animal AMP genes [1,3]. However, *BaraA* represents an exceptional case as multiple
74 tandem repeat peptides are produced from the translation of a single coding
75 sequence, effectively resembling a "protein-based operon"; this tandem repeat
76 structure has also been found in two other AMPs of flies and bees [15,29]. The
77 immature precursor protein of *D. melanogaster BaraA* encodes three types of
78 domains: an IM24 domain, three tandem repeats of IM10-like domains, and an IM22
79 domain (**Fig.** 1A). *BaraA* mutants are susceptible to infection by fungi, and
80 preliminary *in vitro* experiments suggest the *BaraA* IM10-like peptides have
81 antifungal activity [28,30]. The remaining *Baramicin* domains encoding IM22 and
82 IM24 remain uncharacterized. Curiously, *BaraA* deficient flies also display an erect
83 wing behavioural phenotype upon infection, suggesting that *BaraA* could have
84 additional roles beyond immunity [28].
85 In this study, we describe the evolution of the immune-induced *BaraA* locus.
86 We then characterize two other *Baramicin* genes in the genome of *D. melanogaster*
87 (*BaraB* and *BaraC)*, providing insight on the function of each *Baramicin* domain.
88 Surprisingly, only *BaraA* is immune-induced, while *BaraB* and *BaraC* are enriched in
89 the nervous system. Moreover independent *Baramicin*s in other *Drosophila* have
90 convergently evolved to resemble *BaraC*, both in expression and gene structure. We
91 resolved the history of *Baramicin* genomic translocations, revealing properties of
92 each protein domain. We suggest that distinct *Baramicin* domains perform different
93 functions with either host protein targets or antimicrobial activity. By characterizing
94 the distinct roles of sub-peptides in an AMP gene, we provide a case study for how
95 other polypeptide AMP-encoding genes (such as fly *DptB* or human *Cathelicidin*)
96 could play dual roles in both immune and non-immune processes.
97

98 # Results and Discussion

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

100 The *D. melanogaster BaraA* gene encodes three key types of mature peptide
101 (**Fig.** 1A): one IM24 peptide, three IM10-like peptides, and one IM22 peptide that is
102 structurally related to the IM10-like peptides [28]. We recently found that *BaraA*

103 contributes to the *Drosophila* defence against filamentous fungi [28]. During this
104 investigation, we found that while the *BaraA* locus in the *D. melanogaster* R6
105 reference genome encodes two copies of *BaraA* (*CG33470* and *CG18279*, *BaraA1* and
106 *BaraA2* respectively), other fly strains carry only one *BaraA* gene (**Fig.** 1B). Several
107 lines of evidence indicate that this duplication is a very recent evolutionary event.
108 First: in the Dmel_R6 genome assembly, 3,380bp of contiguous sequence in the
109 *BaraA1* and *BaraA2* loci are 100% identical, including the regions upstream
110 (1,140bp) and downstream (892bp) of the 5' and 3' UTRs. Second: this duplication
111 is not found in close relatives such as *Drosophila simulans* and *Drosophila sechellia*.
112 And third: this duplication is not fixed in laboratory (**Table** S1) or wild flies (below).
113 The *Drosophila* Genetic Reference Panel (DGRP) is a genome-sequenced
114 collection of wild-caught flies from the USA [31]. As the *D. melanogaster* R6 genome
115 assembly encodes two copies of *BaraA*, we reasoned that mapping the genomes of
116 DGRP strains with variable copy number to this reference genome would result in a
117 failure to call genetic variants ("call = unknown") in the *BaraA* gene region. Indeed, a
118 sharp spike in unknown variant calls occurs just upstream of *BaraA2* in the region
119 that is identical to *BaraA1* (**Fig.** S1A and predictive region in **Fig.** 1B). We performed
120 a PCR-based screen of 132 DGRP strains using primers specific to the duplication
121 (PCR amplicon in **Fig.** 1B), finding 18/132 flies were PCR-positive for the
122 duplication (~14%). Next, we used Pearson's correlation analysis to ask if genetic
123 variant status (annotated or unknown) correlated with PCR-positive genotypes
124 from our screen. We recovered a significant correlation between *BaraA* PCR status
125 and a string of nine genetic variants from 2R_9293471_SNP to 2R_9293576_SNP
126 (Pearson's correlation: $1e^{-4} < p < 5e^{-3}$ at all nine sites). Specifically, flies with
127 annotated alleles were always PCR-negative (53/53), while 17/18 PCR-positive
128 lines had unknown annotations throughout these nine sites. However 61/114 DGRP
129 strains had unknown calls despite also being PCR-negative for the duplication. Thus,
130 unknown call status is an imperfect but informative predictor of the duplication
131 event. Finally we asked if *BaraA* copy number might inform on infection outcome in
132 data from Wang et al. [32]. Agreeing with an antifungal role for *BaraA*, we recovered
133 striking correlations between these same nine variant sites and susceptibility to

134 infection by *Metarhizium anisopliae* fungi (.09 < Pearson's r < .16, and .03 < p-value
135 < .10; supplementary information and supplementary data file 1). Thus, considering
136 variation at the *BaraA* locus can improve interpretation of DGRP responses to
137 infection.
138 Here we annotate *BaraA* copy number in both lab strains and wild-caught
139 flies. Given this region's poor resolution, cryptic SNPs or indels could have
140 interfered with our PCR assay. Nevertheless, we reveal an unusual pattern in the
141 DGRP that provides insight into how *BaraA* copy number seemingly affects
142 antifungal defence. Importantly, our results highlight how lack of information (i.e.
143 unknown variant status) can have predictive value. It will be important to consider
144 *BaraA* copy number in infection studies, or perhaps even behavioural responses
145 where AMP genes have previously been implicated [19,26,28,33].
146
147 ***Baramicin A is an ancestral immune effector with convergent polymorphisms***
148 In *D. melanogaster, BaraA* is regulated by the Toll immune pathway [28,30].
149 Using BLAST, we recovered *BaraA*-like genes encoding each Baramicin peptide
150 (IM24, IM10-like, and IM22) across the genus *Drosophila* and in the outgroup
151 *Scaptodrosophila lebanonensis*. We next performed infection experiments to confirm
152 that these *BaraA*-like genes were immune-inducible. We infected the diverse species
153 *D. melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, and *D. neotesteacea* (last
154 common ancestor ~50mya [34]) with *Micrococcus luteus* and *Candida albicans*, two
155 microbes that stimulate the Toll pathway (**Fig.** 2A). In all five species, *BaraA*-like
156 genes were immune-induced, confirming that the ancestral *Baramicin* was an
157 immune-induced gene (**Fig.** 2B-F).
158 The predominant peptide products of *D. melanogaster BaraA* are the three
159 IM10-like peptides IM10, IM12, and IM13, which can inhibit fungal growth in vitro
160 [28]. Antimicrobial peptides can display polymorphisms found in different species,
161 suggesting selective pressure (e.g. balancing selection [16]) that favours one allele
162 or the other in a context-dependent manner. Such polymorphisms are commonly
163 found in Dipteran AMPs [4,12,35,36]. We found that the 23-residue IM10-like
164 peptides of *Baramicin* genes display signatures of convergent evolution, similar to

165 other fly AMPs (**Fig.** S2 and supplementary information). Of particular note, an
166 $^{6}$RPXR$^{9}$ motif universally encodes aspartate (D) or glycine (G) at residue 8, which
167 has convergently evolved multiple times (**Fig.** S2C). Future studies of these peptides
168 should incorporate variation of IM10-like peptide residues into activity assays and
169 interpretations of host-pathogen dynamics.
170
171 ***Identification of BaraB and BaraC, two non-immune Baramicin paralogs***
172 While investigating *Baramicin* evolution across species, we also recovered
173 two paralogous *Baramicin* genes in *D. melanogaster* through reciprocal BLAST
174 searches: *CG13749* and *CG30285*, which we name *BaraB* and *BaraC* respectively. The
175 three *Baramicin* genes are scattered on the right arm of chromosome II at
176 cytological positions 44F9 (*BaraB*), 50A5 (*BaraA*), and 57F8 (*BaraC*). While *BaraA*-
177 like genes are conserved throughout the genus *Drosophila*, *BaraB* is conserved only
178 in Melanogaster group flies, and *BaraC* is found only in Melanogaster and Obscura
179 group flies, indicating that both paralogs stem from duplication events of a *BaraA*-
180 like ancestor (a summary of the *Baramicin* loci is shown later). In the *D.*
181 *melanogaster* reference genome, these paralogous *Baramicin* genes are united by
182 the presence of the IM24 domain (**Fig. 3**A). In the case of *BaraB*, we additionally
183 recovered a frameshift mutation (2R_4821599_INS) causing a premature stop
184 segregating in the DGRP leading to the loss of IM13 and IM22 relative to the *BaraA*
185 gene structure (**Fig. 3**A); this truncation is present in the Dmel_R6 genome
186 assembly, but many DGRP strains encode a CDS with either a standard (e.g.
187 DGRP38) or extended (e.g. DGRP101) IM22 domain (a DGRP *BaraB* alignment is
188 provided in supplementary data file 2). Moreover, in contrast to BaraA, the initial
189 IM10-like peptide of BaraB no longer follows a furin cleavage site (**Fig. 3**A), and
190 encodes a serine (RSXR) in its IM10-like motif instead of the proline (RPXR) of
191 *BaraA*-like genes. Each of these mutations prevents the secretion of classical IM10-
192 like and IM22 peptides. Finally, *BaraC* encodes only IM24 tailed by a
193 transmembrane domain at the C terminus (TMHMM v2.0 [37]), and thus lacks both
194 the IM10-like peptides and IM22 (**Fig. 3**A).

195 *BaraA* is strongly induced following microbial challenge, being
196 predominantly regulated by the Toll pathway with a minor input from the Immune
197 Deficiency (Imd) pathway [28,30]. We assayed the expression of *BaraB* and *BaraC* in
198 wild-type flies, and also flies with defective Toll ($spz^{rm7}$) or Imd ($Rel^{E20}$) signalling to
199 see if their basal expression relied on these pathways. Surprisingly, neither gene
200 was induced upon infection regardless of microbial challenge (**Fig. 3**B and **Fig.** S3A-
201 B). However *BaraC* levels were reduced in $spz^{rm7}$ mutants regardless of treatment
202 (cumulative data in **Fig.** S3C, p = .005), suggesting that *BaraC* basal expression is
203 affected by Toll signalling. We next generated a novel time course of development
204 from egg to adult to monitor the expression of the three *Baramicin* genes. We found
205 that expression of all genes increased over development and reached their highest
206 level in young adults (**Fig.** 3C). Of note, *BaraB* expression approached the lower limit
207 of our assay's detection sensitivity at early life stages. However *BaraB* was robustly
208 detected beginning at the pupal stage, indicating it is expressed during
209 metamorphosis. *BaraC* expression also increased markedly between the L3 larval
210 stage and pupal stage.
211 Here we reveal that *BaraA* is part of a larger gene family. While the *BaraA*
212 gene was first described as an immune effector, the two more phylogenetically-
213 restricted *Baramicin* paralogs *BaraB* and *BaraC* are not induced by infection in *D.*
214 *melanogaster*. Both *BaraB* and *BaraC* first see increased expression during pupation,
215 and are ultimately expressed at their highest levels in adults.
216
217 ***Dmel\BaraB is required in the nervous system over the course of development***
218 A simple interpretation of the truncated gene structure and low levels of
219 *BaraB* expression is that this gene is undergoing pseudogenization. To explore
220 *BaraB* function, we used two mutations for *BaraB* ($\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$,
221 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 $\Delta BaraB^{LC1}$ and $\Delta BaraB^{LC4}$ mutations differ in their ultimate effect,
224 as $\Delta BaraB^{LC1}$ 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 **S4**A-B). The *ΔBaraB*$^{LC4}$ mutation however deletes the entire locus, leading to *BaraB*
227 null flies (**Fig. S4**A). We further introgressed both mutations into the DrosDel
228 isogenic background (referred to as *iso*) for seven generations. *ΔBaraB*$^{LC4}$
229 homozygotes are lethal during larval development, whereas the hypomorphic
230 *ΔBaraB*$^{LC1}$ flies allow for homozygous adults to emerge (**Fig. S4**C). Thus we used
231 *ΔBaraB*$^{LC1}$ for experiments involving adults. Isogenization was performed as
232 described in Ferreira et al. [38]. We assessed *ΔBaraB*$^{LC1}$ mutant viability using
233 crosses between *ΔBaraB*$^{LC1}$*/CyO* heterozygous females and *ΔBaraB*$^{LC1}$ homozygous
234 males. Under mendelian ratios of inheritance, this cross is expected to result in a 1:1
235 ratio of *ΔBaraB*$^{LC1}$/CyO and *ΔBaraB* $^{LC1}$ homozygous progeny. However, *BaraB*
236 deficient flies showed reduced viability using both mutations independent of genetic
237 background (**Fig. S4**C). This reduced viability was exacerbated by rearing at 29°C,
238 reaching a lethality rate over 50% (**Fig. 3**D). Using a *CyO-GFP* reporter to track
239 genotype in larvae revealed that the lethal phase occurs primarily in the late larval
240 and pupal stages (**Fig. S4**D-G), agreeing with a role for *BaraB* in larvae/pupae
241 previously suggested by increased expression at this stage. Some emergent flies also
242 exhibited a nubbin-like wing phenotype (FlyBase: FBrf0220532 and e.g. in **Fig. 3**E),
243 where the wings were stuck in a shrivelled state for the remainder of the fly's
244 lifespan. However, a plurality of *ΔBaraB*$^{LC1}$ homozygotes successfully emerged, and
245 unlike their siblings, had no immediate morphological or locomotory defects. The
246 lifespan of morphologically normal *iso ΔBaraB*$^{LC1}$ adults is nevertheless significantly
247 shorter compared to wild-type flies and *iso ΔBaraB*$^{LC1}$*/CyO* siblings (**Fig. S4**H). We
248 confirmed these developmental effects using ubiquitous gene silencing with
249 *Actin5C-Gal4 (Act-Gal4)* to drive two *BaraB* RNAi constructs (*TRiP-IR* and *KK-IR*).
250 Both constructs resulted in significant lethality and occurrence of nubbin-like wings
251 (**Table** S2). Additionally, we crossed *iso ΔBaraB*$^{LC1}$ to a genomic deficiency line
252 (BDSC9063) at both 25°C and 29°C. These crosses resulted in significantly reduced
253 numbers of eclosing *ΔBaraB*$^{LC1}$*/Df(9063)* flies at 25°C ($n = 114$, $\chi^2$ $p < .001$) and 29°C
254 ($n = 63$, $\chi^2$ $p < .001$) (**Fig. S4**I).

255 These data demonstrate a significant consequence of *BaraB* disruption on
256 fitness. While whole-fly *BaraB* expression is low, these results suggest that *BaraB* is
257 not pseudogenized, and instead performs an integral function during development.
258 The fact that there is a bimodal outcome in hypomorph-like $\Delta BaraB^{LC1}$ adults (either
259 severe defects or generally healthy) suggests *BaraB* is involved in passing some
260 checkpoint during larval/pupal development. Flies deficient for *BaraB* are more
261 likely to fail at this developmental checkpoint, resulting in either lethality or
262 developmental defects.
263
264 ***The Baramicin paralogs BaraB and BaraC are expressed in the nervous system***
265 Given that whole-fly *BaraB* depletion produced lethality and nubbin-like
266 wings, we next sought to determine in which tissue(s) *BaraB* is required. *BaraB*
267 mutant locomotory defects led us to suspect an involvement in the nervous system.
268 We silenced *BaraB* in the nervous system at 25°C or at 29°C for greater efficiency,
269 using a pan-neural *elav-Gal4* driver, and for *BaraB-IR* used both the *TRiP-IR* and *KK-*
270 *IR* lines. We additionally combined this approach with *UAS-Dicer2* (*Dcr2*) to further
271 strengthen gene silencing [39]. We used the *TRiP-IR* line with a *CyO* balancer
272 chromosome (*elav-Gal4 * TRiP-IR/CyO*), allowing the same simple mendelian
273 approach to assess lethality as used previously. Both *elav>TRiP-IR* and *elav>Dcr2,*
274 *TRiP-IR* caused partial lethality and occasional nubbin-like wings ( $\chi^2$ $p < .02$, **Table**
275 S2). For *KK-IR*, we instead crossed homozygous flies and so we did not assess
276 lethality using mendelian inheritance. However, no adults emerged when *elav>Dcr2,*
277 *KK-IR* flies were reared at 29°C. Rare emergents ($n^{total} = 11$ after three experiments)
278 occurred at 25°C, all of which bore nubbin-like wings. Using *elav-Gal4* at 29°C
279 without *Dcr2*, we observed greater numbers of emerging adults, but 100% of flies
280 had nubbin-like wings (**Fig. 3**E, **Table** S2). Finally, *elav>KK-IR* flies at 25°C suffered
281 both partial lethality and nubbin-like wings, but normal-winged flies began
282 emerging ( $\chi^2$ $p < .001$, **Table** S2). This analysis indicates that *BaraB* is expressed in
283 the nervous system, and this expression readily explains both the lethality and
284 nubbin-like wing phenotypes. Moreover, we observed a consistent spectrum of

285 developmental defects using *elav-Gal4>BaraB-IR* wherein strength of gene silencing
286 correlates with severity of defects.
287 *BaraB* is not the only *Baramicin* gene now specialized for expression in the
288 nervous system. Tissue-specific transcriptomic data indicate that *BaraC* is expressed
289 in various neural tissues including the eye, brain, and the thoracic abdominal
290 ganglion (**Fig. S5**A) [40]. However *BaraC* is also highly enriched in the hindgut and
291 rectal pads, pointing to a complex expression pattern [40,41]. We next searched
292 FlyBrainAtlas (http://scope.aertslab.org, [42]) to narrow down which neural
293 subtypes *BaraC* was expressed in. Strikingly, *BaraC* was expressed in all glial cell
294 types, fully overlapping the glia marker *Repo* (**Fig. 3**F). To confirm these
295 observations, we compared the effects of *BaraC* RNA silencing (*BaraC-IR*) using *Act-*
296 *Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia) drivers on *BaraC*
297 expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that of control flies
298 (**Fig. S5**B). By comparison *elav-Gal4* reduced *BaraC* expression to ~63% that of
299 controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls (overall
300 controls vs. neural/glia-IR, p = .002). Collectively, our results support the notion
301 that *BaraC* is expressed in the nervous system, and are consistent with *BaraC*
302 expression being most localized to glial cells.
303
304 ***Extensive genomic turnover of the Baramicin gene family***
305 The results thus far show that *BaraA*-like genes are consistently immune-
306 induced in all *Drosophila* species (**Fig.** 2), and the *BaraA* antifungal IM10-like
307 peptides evolve rapidly (**Fig.** S2). However, the two phylogenetically-restricted
308 paralogs *Dmel\BaraB* and *Dmel\BaraC* are not immune-induced, and are truncated
309 in a fashion that deletes some or all of the IM10-like peptides (**Fig. 3**A). These two
310 *Baramicins* are now enriched in the nervous system (**Fig. 3**E-F). In the case of *BaraB*,
311 a role in the nervous system is evidenced by severe defects recapitulated using pan-
312 neural RNA silencing. In the case of *BaraC*, nervous system expression is evidenced
313 by a clear overlap with *Repo*-expressing cells.
314 The three *D. melanogaster Baramicin* genes are united by their IM24 domain,
315 and are scattered across different loci along chromosome 2R, suggesting a series of

316 complex genetic events. To determine the ancestry of each *D. melanogaster*
317 *Baramicin* gene, we traced their evolutionary history by analyzing genomic synteny.
318 This allowed us to determine ancestry beyond sequence similarity-based
319 approaches. Using this strategy, we traced a series of genomic duplications,
320 inversions, and locus translocation events that ultimately resulted in the three
321 *Baramicin* loci of *D. melanogaster*. Our analysis reveals that these three loci
322 ultimately stem from a single-locus ancestor encoding only one *Baramicin* gene that
323 resembled *Dmel\BaraA* (**Fig. 4**A). This is evidenced by the presence of only a single
324 *BaraA*-like gene in the outgroup *S. lebanonensis*, and also in multiple lineages of the
325 subgenus Drosophila (**Fig. 4**B). Indeed, the general *BaraA* gene structure encoding
326 IM24, tandem repeats of IM10-like peptides, and IM22 is conserved in *S.*
327 *lebanonensis* and all *Drosophila* species (**Fig. 4**C). On the other hand, the *Dmel\BaraC*
328 gene comes from an ancient duplication restricted to the subgenus Sophophora, and
329 *Dmel\BaraB* resulted from a more recent duplication found only in the
330 Melanogaster group (**Fig. 4**B).
331 This genomic synteny analysis revealed a striking observation: the *D.*
332 *willistoni* gene *Dwil\GK10648* is syntenic with the *Dmel\BaraC* locus (**Fig. 4**A), yet it
333 is immune-induced (**Fig.** 2D) and retains a *BaraA*-like gene structure (**Fig. 4**C).
334 Meanwhile, the other *D. willistoni Baramicin Dwil\GK10645* is found at the locus
335 syntenic with *BaraA*, yet this gene has undergone an independent truncation to
336 encode just an IM24 peptide (similar to *Dmel\BaraC*). Thus these two *D. willistoni*
337 genes have evolved similar to *D. melanogaster BaraA/BaraC*, but in a vice versa
338 fashion. This suggests a pattern of convergent evolution with two key points: **i)** the
339 duplication event producing *Dmel\BaraA* and *Dmel\BaraC* originally copied a full-
340 length *BaraA*-like gene to both loci, and **ii)** the derivation of an IM24-specific gene
341 structure has occurred more than once (*Dmel\BaraC* and *Dwil\GK10645*). Indeed,
342 another independent IM24-specific *Baramicin* gene is present in *D. virilis*
343 (*Dvir\GJ25897*), which is a direct sister of the *BaraA*-like gene *Dvir\GJ21309* (**Fig.**
344 **4**C). We evaluated the expression of these truncated *Baramicins* encoding only IM24
345 in both *D. willistoni* and *D. virilis* upon infection. As was the case for *Dmel\BaraC*,
346 neither gene is immune-induced (**Fig. S6**A-C). Given the glial expression of

347 *Dmel\BaraC,* we reasoned that the heads of adult flies (rich in nerve tissue) should
348 be enriched in *BaraC* compared to whole animals. Indeed we saw a significant
349 enrichment of *BaraC* in the heads of *D. melanogaster* males compared to whole flies,
350 which was not the case for *BaraA* (**Fig. S6**D). We further saw a consistent and
351 significant enrichment in the head for the IM24-specific genes *Dwil\GK10645* and
352 *Dvir\GJ25897*, while *BaraA*-like genes were more variable in expression (**Fig. S6**E-
353 F).
354 Thus, multiple independent IM24-specific *Baramicins* apparently do not play
355 an immune role, and are enriched in the head. In the case of *Dmel\BaraC*, this is
356 likely due to expression in glia. Strikingly, we observe an independent but parallel
357 evolution of expression pattern and gene structure in *Baramicins* of *D. willistoni* and
358 *D. virilis*. A summary of these expression data is included in **Fig. 4**C. Phylogenetic
359 and genomic synteny analysis reveals that the gene structure and immune
360 expression pattern of *BaraA* reflect the ancestral state, while *Dmel\BaraB* and
361 *Dmel\BaraC* are paralogs derived from independent duplication events. These two
362 genes, which are truncated and delete both IM10-like and IM22 domains, have also
363 lost immune function.
364
365 ***Evolution of IM24 in lineage-specific fashions***
366 Multiple independent *Baramicin* genes have lost both IM10-like and IM22
367 peptides, and also converge on loss of immune induction. Taken together, these
368 truncations and expression patterns suggest that the IM10-like peptides and IM22
369 are strictly useful during the immune response, consistent with a recently described
370 antifungal role for IM10-like peptides [28]. Inversely, non-immune *Baramicin* genes
371 have repeatedly and independently truncated to encode primarily IM24. This
372 suggests that IM24 is adaptable to different contexts – a protein domain with utility
373 extending beyond the immune response.
374 Thus far, we have discussed IM24 primarily as a unifying feature of the
375 *Baramicin* gene family. To understand how this universal domain differs across gene
376 lineages, we next investigated the evolution of the IM24 domain at the sequence
377 level. We analyzed IM24 domain evolution via multiple statistical approaches

378 (including Likelihood (FEL), Bayesian (FUBAR), and Count-based (SLAC) analyses)
379 using the HyPhy package implemented in Datamonkey.org, [43] with separate codon
380 alignments of *BaraA, BaraB,* or *BaraC* IM24 domains beginning at their conserved
381 $Q^1$ starting residue. We collected additional genomic sequence data from recently-
382 sequenced fly species [44] to better inform on the sequence evolution of these genes
383 and loci. We employed analyses of each lineage using locus-specific alignments (e.g.
384 genes at the *stum* locus in **Fig. 4**B were all analyzed together). Globally, FEL, FUBAR,
385 and SLAC site-specific analyses suggest strong purifying selection across the IM24
386 domain (data in **supplementary data file 3**), agreeing with the general protein
387 structure of IM24 being broadly conserved. However one site (residue 29) was
388 consistently highlighted as evolving under positive selection using each type of
389 statistical approach for those genes located at the Sophophora *ATP8A* locus (*BaraA*
390 genes and *Dwil\GK10645*: p-adj < .05; **Fig. 5**A). This site is universally Proline in
391 *Baramicin* genes located at the stum locus (*BaraC*-like), but is variable in both the
392 *BaraA* and *BaraB* lineages (**Fig. 5**A). We also note that two sites on either side of site
393 29 (site 27 and site 31) similarly diverge by lineage in an otherwise highly
394 conserved region of the IM24 domain. FUBAR analysis (but not FEL or SLAC) also
395 found evidence of positive selection at site 31 (p-adj = .026), suggesting this
396 neighbouring site could also be evolving in a non-random fashion. Similar analyses
397 of the *BaraB* and *stum* loci *Baramicins* did not find evidence of site-specific positive
398 selection.
399 We next assessed the potential for diversifying selection across the
400 phylogeny using two branch-site methods (aBSREL and BUSTED [45]). Both branch-
401 site analyses found evidence of episodic diversifying selection in different lineages.
402 For the *ATP8A* locus *Baramicins (BaraA* genes and *Dwil\GK10645)*, BUSTED whole-
403 phylogeny analysis found evidence of diversifying selection in at least one branch
404 across the phylogeny (p-adj = .008). aBSREL analysis also finds evidence of
405 diversifying selection specifically at the node delineating the Willistoni group from
406 other Sophophora (**Fig.** S7). This node corresponds to the parallel and opposite
407 evolution of BaraA-like and BaraC-like protein structures at this locus in *D.*
408 *melanogaster* compared to *D. willistoni* (**Fig. 4**C), indicating both gene structure and

409 sequence evolved rapidly in these branches. Finally one other string of residues is
410 also striking: a specific motif with lineage-restricted patterns occurs at sites 40-48,
411 numbered according to the Q[1] starting residue. In *Dmel\BaraB*, this motif encodes
412 the residues [40]HHASSPAD[48]. The three C-terminal residues of this motif are
413 diagnostic for each gene lineage (*BaraA, BaraB,* and *BaraC* have RGE, PXE, or
414 (S/N)GQ respectively; **Fig. 5**A). We also recovered a signature of diversifying
415 selection in the stum locus *Baramicin* gene of *Drosophila paulistorum* using recent
416 genomic sequence for strains L06 and L12 (see [44] and supplementary
417 information), suggesting population-level data may reveal additional sites of
418 interest.
419 Globally, *Baramicin* IM24 peptides have experienced positive selection. Site-
420 specific analyses highlight residue 29 in the *ATP8A* locus lineage. We further found
421 evidence that, alongside major structural changes specific to gene lineages, the IM24
422 sequence itself evolved under diversifying selection. Thus, not only are IM10-like
423 and IM22 peptides independently lost in *Dmel\BaraC* and *Dwil\GK10645*, alongside
424 parallel loss of immune induction, but their IM24 domains also evolved rapidly
425 following the speciation event distinguishing these lineages. Another 8-residue
426 motif encoding "HHASSPAD" in Dmel\BaraB is diagnostic of *Baramicin* lineages. We
427 could not generate a reasonable model of the IM24 peptide conformation using
428 Phyre2 [46], QUARK, or TASSER protein modelling methodologies [47]. However
429 the pattern of residues at site 29 and neighbouring residues at site 27 and site 31
430 may suggest this region has some importance in interactions with targets of IM24.
431 The Dmel\BaraB HHASSPAD motif also shows striking patterns in different gene
432 lineages, though we did not find evidence of positive selection at this motif.
433
434 ***A timeframe for BaraB neofunctionalization***
435 Selection analyses support the notion that IM24 evolves rapidly in different
436 lineages. One other striking feature of the Dmel\BaraB protein is the absence of a
437 signal peptide structure (SignalP 5.0 [48]). Signal peptide sequence is conserved in
438 all *Baramicin* lineages, except in *BaraB* of *D. melanogaster* and also *D. simulans* (last
439 common ancestor 5.9mya [34]); these species encode a conserved region of parallel

440 length (**Fig.** 5A). This is surprising, as *D. simulans* is phylogenetically more closely
441 related to *D. sechellia* (last common ancestor 3.3mya) and *D. mauritiana* (last
442 common ancestor 4.5mya), yet *D. sechellia* and *D. mauritiana* retain signal peptide
443 structures in their *BaraB* genes; we note that the *D. sechellia* BaraB signal peptide is
444 divergent due to an insertion event (**Fig.** 5A), but is still predicted to enable protein
445 secretion (SignalP 5.0 [48]). We also previously noted that a genetic variant
446 truncating the *BaraB* CDS is segregating in wild *D. melanogaster* (Fig. 3A and
447 supplementary data file 2). However we recovered only full-length *Baramicin* gene
448 structures encoding IM24, IM10-like peptides, and IM22 in *BaraB* orthologues of
449 other species (except *D. erecta* where the *BaraB* orthologue appears pseudogenized
450 by the loss of the *BaraB* start codon and multiple indel events).
451 We reason that the *D. simulans BaraB* N-terminus – which mirrors that of *D.*
452 *melanogaster* – was either introgressed from one species to the other prior to the
453 complete development of hybrid inviability, or reflects incomplete lineage sorting of
454 this locus in the Melanogaster species complex. In either scenario, two major events
455 affect the *D. melanogaster* BaraB protein structure: i) loss of signal peptide
456 sequence, and ii) premature stop variants truncating the precursor protein. Patterns
457 of residue conservation at IM24 site 29 or in the BaraB $^{40}$HHASSPAD$^{48}$ domain may
458 also determine different species' *BaraB* gene function. We were curious if any of
459 these events correlated with the loss of an immune role, presumably alongside
460 neofunctionalization of *BaraB* for its specific role in the nervous system. To this end,
461 we designed qPCR assays for *Baramicin* genes in phylogenetically diverse members
462 of the Melanogaster species group including *D. suzukii, D. yakuba, D. mauritiana, D.*
463 *sechellia,* and *D. simulans* (last common ancestor 27.2mya [34]). We then infected
464 these flies with *M. luteus* bacteria, known to activate the Toll pathway. We found
465 that the *BaraB* orthologues of *D. suzukii*, *D. mauritiana* and *D. sechellia* all remain
466 immune-induced, indicating that the ancestral *BaraB* gene was still immune-
467 regulated following its genomic translocation (**Fig.** 5B). However like *Dmel\BaraB*,
468 importantly the *BaraB* gene of *D. simulans* was not immune-induced and is
469 expressed at a very low level despite direct sister species having immune-induced
470 *BaraB* genes.

These observations provide a precise phylogenetic timeframe confirming that the last point where the *Dmel\BaraB* lineage encoded a functional signal peptide was in the last common ancestor of the Melanogaster species complex *5*.9mya. The *D. melanogaster* and *D. simulans BaraB* genes appear to have lost immune regulation only recently, as *BaraB* is robustly expressed and immune-induced in the Melanogaster sister species *D. mauritiana* and *D. sechellia*. Taken together, we demonstrate not only that this AMP-encoding gene has repeatedly evolved non-immune roles in multiple species, but that one such event unfolded very recently in the Melanogaster species complex with significant implications on fitness during development. We therefore characterize the precise timeframe wherein an AMP-encoding gene transitioned to perform a novel function in the nervous system, and highlight key sequence and structural changes correlated with this neofunctionalization event.

## Conclusion and Future Directions

The *Baramicin* gene family has undergone a varied history of genomic duplication, translocation, and complex sequence evolution. The *BaraA* duplication may be especially relevant to immunity studies given it is variable in common lab stocks. Our initial exploration using data from a previous DGRP infection screen suggests that *BaraA* copy number can explain a great deal of DGRP strain response to infection (also see supplementary information). As *BaraA* underlies a sexually dimorphic "erect wing" response to infection [28], and other short immune peptides have recently been implicated in behavioural regulation [18,19,26,33], *BaraA* copy number variation may be an important consideration beyond immune function.

The ancestral *BaraA*-like gene structure is comprised of multiple distinct peptide domains (i.e. IM24, IM10-like peptides, IM22). Two recent studies suggest the *BaraA* gene and specifically the IM10-like peptides have antifungal properties [28,30]. We found a striking convergent polymorphism in the IM10-like peptides (**Fig.** S2) similar to polymorphisms found in other AMPs that can have drastic effects on host-pathogen interactions [16]. We also discovered multiple non-immune

501 *Baramicin* paralogs in various *Drosophila* species that have convergently lost IM10-
502 like peptides, reinforcing a role for the IM10-like peptides specifically in the immune
503 response. Development of these peptides or other AMPs as therapeutics should
504 consider the consequences of such polymorphisms on antimicrobial activity.

505 The IM24 domain unites the *Baramicin* gene family. It was therefore
506 surprising that multiple *Baramicin* genes retain this domain, but have lost immune
507 induction. In the case of *D. melanogaster BaraB* and *BaraC* we found enrichment in
508 the nervous system, where *BaraB* knockdown by the pan-neural driver *elav-Gal4*
509 phenocopies both the lethality and nubbin-like wing phenotype. We additionally
510 investigated the effect of *BaraB* RNAi using Gal4 drivers in non-neural tissues
511 including the fat body (*c564-Gal4*), hemocytes (*hml-Gal4*), the gut (esg-Gal4),
512 malpighian tubules (MyO-Gal4), the wing disc (*nubbin-Gal4*), and in myocytes (*mef2-*
513 *gal4*) to no effect. We also screened neural drivers specific for glia (*Repo-Gal4*),
514 motor neurons (*D42-*, *VGMN-*, and *OK6-Gal4*), and a recently-made *BaraA-Gal4*
515 driver that is expressed in the larval ventral nervous system [28]. However all these
516 *Gal4>BaraB-IR* flies were viable and never exhibited overt morphological defects.
517 While we did not identify the specific neural cell type that expresses *BaraB*, our Gal4
518 approach rules out various options including tissue overlapping in expression with
519 *BaraA*. Future study will benefit from systematically knocking down *BaraB* in
520 neuronal subtypes, and perhaps exploring *BaraB* evolution in the Melanogaster
521 species complex.

522 Adding to the complex patterns in IM24 evolution and gene function, the
523 cellular location of each Baramicin peptide is unclear. The Dmel\BaraB protein has
524 lost the otherwise conserved signal peptide structure of the *Baramicin* gene family, a
525 mutation that prevents secretion. Certainly *BaraA* IM24 is secreted, as this peptide
526 was first described for its robust extracellular signature as a component in immune-
527 induced hemolymph [28,30,49]. Likewise, *Dmel\BaraC* bears a couple key
528 signatures indicating it is likely to reach the cell membrane: i) it encodes a signal
529 peptide that should enable entry into the secretory pathway, and ii) its IM24 domain
530 is preceded by an "SA" dipeptidyl peptidase site. XA or XP motifs are cleaved by
531 dipeptidyl peptidases at the cell membrane to maturate secreted peptides [50], and

532 indeed XA or XP motifs are seen preceding the $Q^1$ residue in all IM24 domains of
533 BaraA and BaraC proteins (**Fig. 5**A), and an "SP" motif precedes each of the *D.*
534 *melanogaster* IM10-like peptides [28]. Given IM24 conservation in genes with totally
535 different expression profiles, the general IM24 protein is not strictly tied to
536 immunity. In this regard, AMPs and neuropeptides share many features including:
537 small size, cationic net charge, and amphipathicity [51]. Indeed Dmel\BaraA IM24 is
538 both relatively short (96 AA) and cationic (+2.4 at pH=7). However, while IM24
539 sequence is broadly conserved (**Fig.** 5A), changes present in Dmel\BaraC make it
540 strikingly anionic by comparison: the net charge of the putative extracellular part is
541 -4.6 at pH=7 assuming the same N-terminal pGlu modification found in BaraA [37].
542 The glia-expressed *Dmel\BaraC* encodes a transmembrane tail that is found
543 throughout the Melanogaster and Obscura groups. This domain should enable
544 insertion into either the endogenous cell membrane, or the membrane of a
545 neighbouring cell. The latter possibility is intriguing, as glia envelope and regulate
546 neurons. Discovering the binding partner(s) of any of *Baramicin* IM24 proteins
547 should be exciting, as this will highlight a protein scaffold with either immune
548 and/or neural functional domains.
549 While *Baramicins* are restricted to drosophilids, they are not the only AMP
550 gene family united by a domain not necessarily functioning in the immune response.
551 One human AMP recently implicated in chronic neuroinflammatory disease is the
552 Cathelicidin LL-37 [24,27,52]. The *Cathelicidin* gene family is unified by its N-
553 terminal "Cathelin" domain, however this region has no antimicrobial activity in
554 vitro [53]. Instead, the Cathelicidins encode their mature AMPs (e.g. LL-37) at the
555 protein C-terminus. The role of innate immunity and specifically antimicrobial
556 peptides in neurological disorders is a budding and exciting field of exploration [17–
557 20,22,26,54]. The *Baramicins* add to this growing list of immune-neural interactions
558 by providing an evolutionary case study demonstrating not all peptides of AMP-
559 encoding genes are immune-specific. Indeed, most AMP studies have a peculiar
560 blind spot in this regard, as *in vivo* experiments typically disrupt entire genes, while
561 ex vivo or in vitro approaches typically manipulate only the canonical mature AMP.
562 Future studies on AMPs relations to disease processes should benefit greatly from

563 dissecting the functions of AMPs at the gene level from the specific functions of their
564 encoded peptides, which could act in totally separate tissues and contexts.
565 In summary, we characterize how an ancestral AMP-encoding gene has
566 evolved under a number of unique evolutionary scenarios. In light of the repeated
567 derivation of IM24-specific *Baramicin* genes – including at least two specifically
568 involved in neural processes – it will be interesting to consider the functions of AMP
569 genes in neural processes not simply at the level of the gene, but at the level of the
570 mature peptides produced by that gene. The distinct functional roles of the IM24
571 domain in neural processes and IM10-like peptides in immunity suggest that
572 selection on AMP genes may go beyond host-pathogen interactions. In the case of
573 *Baramicin* and other recently-characterized invertebrate AMPs, natural selection
574 may impose trade-offs that contrast pressures imposed by essential roles in host
575 processes (e.g. in neural tissues) with maintenance of a competent host defence.
576 Given the many commonalities between AMPs and neuropeptides, these canonical
577 immune effectors might be adapted for non-immune functions more commonly than
578 appreciated.
579
580 **Acknowledgements**
581 We would like to thank Maria Litovchenko for advice, Ana Marija Jakšić for
582 generously providing DGRP flies, Rob Unckless for stimulating discussion, Huang et
583 al. [30] for collaborative cooperation, and Florent Masson, Hannah Westlake, and
584 the anonymous reviewers for a thorough reading of the manuscript and appreciated
585 feedback. This research was supported by Sinergia grant CRSII5_186397 awarded to
586 Bruno Lemaitre. The *BaraB*$^{LC1}$ and *BaraB*$^{LC4}$ mutations were graciously provided by
587 Steven Wasserman and generated by Lianne Cohen, who we also thank for their
588 critical involvement in characterizing *Baramicin A* [28].
589
590 # Materials and Methods
591 ***DGRP population screening and bioinformatics analyses***
592 Genomic sequence data were downloaded from GenBank default reference
593 assemblies and Kim et al. [44], and DGRP sequence data from
594 http://dgrp2.gnets.ncsu.edu/ [31]. Sequence comparisons and alignment figures
595 were prepared using Geneious R10 [55], Prism 7, and Inkscape. Alignments were
596 performed using MUSCLE or MAFFT followed by manual curation, and phylogenetic
597 analyses were performed to validate sequence patterns using the Neighbour Joining,
598 PhyML, RaxML, and MrBayes plugins in Geneious. Figure S1B was generated using
599 the UCSC *D. melanogaster* DGRP2 genome browser. Selection analyses were

600 performed using the HyPhy package implemented in datamonkey.org [43]. Codon
601 alignments of the IM24 domain used in **Fig. 5**A is included as a .fasta file in
602 **supplementary data file 3** alongside outputs from FEL, FUBAR, SLAC, and aBSREL
603 selection analyses.
604
605 ***Fly genetics***
606 The *BaraB*$^{LC1}$ and *BaraB*$^{LC4}$ mutations were generated using CRISPR with two
607 gRNAs and an HDR vector by cloning 5' and 3' region-homologous arms into the
608 pHD-DsRed vector, and consequently *ΔBaraB* flies express DsRed in their eyes,
609 ocelli, and abdomen. The following PAM sites were used for CRISPR bordering the
610 *BaraB* region. Slashes indicate the cut site: 5': GCGGGCAACAGATGTGTTCA/GGG
611 3': GTCCATTGCTTATTCAAAAA/TGG. These mutants were generated in the
612 laboratory of Steve Wasserman by Lianne Cohen, who graciously allowed their use
613 in this study. All fly stocks including Gal4 and RNAi lines are listed in
614 **supplementary data file 4**. Experiments were performed at 25°C unless otherwise
615 indicated. When possible, genetic crosses of 6-8 males and 6-8 females were
616 performed in both directions to test for an effect of the X or Y chromosomes on
617 *BaraB*-mediated lethality; crosses in both directions yielded similar results in all
618 cases and reported data are pooled results. Fly diet consisted of a nutrient-rich lab
619 standard food: 3.72g agar, 35.28g cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL
620 propionic acid, 15.9mL moldex, and $H_2O$ to 600mL.
621
622 ***Infection experiments***
623 Bacteria and yeast were grown to mid-log phase shaking at 200rpm in their
624 respective growth media (LB, BHI, or YPG) and temperature conditions, and then
625 pelleted by centrifugation to concentrate microbes. Resulting cultures were diluted
626 to OD = 200 at 600nm before infections to measure gene expression. The following
627 microbes were grown at 37°C: *Escherichia coli strain 1106* (LB) and *Candida albicans*
628 (YPG). *Micrococcus luteus* was grown at 29°C in LB. For **Fig.** 2 and S6A, pooled fly
629 samples were collected either 6 hours post-infection (*E. coli*) or 24 hours post-
630 infection (*C. albicans, M. luteus*) prior to RNA extraction on pools of 5 adult males.
631 These timepoints correspond to the maximal expression inputs of the Imd (6hpi) or
632 Toll (24hpi) NF-κB signalling pathways, which are most specifically induced by
633 Gram-negative bacteria (Imd) or Gram-positive bacteria or fungi (Toll) [56]. Flies
634 were pricked in the thorax as described in [25].
635 RNA extractions were performed using TRIzol™, Ambion DNAse treatment,
636 and PrimeScript RT according to manufacturer's protocols. RT-qPCR was performed
637 using PowerUP SYBR Green master mix with primers listed in **supplementary data**
638 **file 5**. Gene expression differences were analyzed using the PFAFFL method [57].
639 For gene expression experiments requiring dissection of heads, pools of 20 males
640 were used for either whole flies or heads dissected in ice-cold PBS and transferred
641 immediately to a tube kept on dry ice.
642

643 ## Figure captions
644
645 **Figure 1: The *BaraA* duplication is not fixed in wild flies. A)** Gene structure of
646 *BaraA*, including its three key domain types: IM24 (orange), IM10-like (green), and
647 IM22 (teal). SP = signal peptide. Purple triangles indicate furin cleavage sites. **B)**
648 Depiction of the *BaraA* duplication. The PCR amplicon used to detect presence of the
649 *BaraA* duplication event, and the region in the DGRP of unknown variant calls
650 correlated with the duplication are annotated (detailed in Fig. S1).
651
652 **Figure 2: The ancestral *BaraA* gene was immune-induced. A)** Cladogram of
653 species used in B-F. **B-F)** Expression of *BomBc3* control genes (brown) or *BaraA-like*
654 genes (orange) in diverse *Drosophila* species upon infection. In all cases, both
655 *BomBc3* and *BaraA-like* genes are induced upon infection by either *C. albicans* yeast
656 or *M. luteus* bacteria.
657
658 **Figure 3: The *D. melanogaster Baramicin* gene family. A)** Two *Baramicin*
659 paralogs in addition to *BaraA* are present in the Dmel_R6 reference genome, united
660 by the presence of an IM24 peptide. Relative to *BaraA*, two frameshift mutations in
661 *BaraB* are segregating in the DGRP. One results in the same mature protein
662 annotated in the Dmel_R6 reference genome (coloured domains), while the other
663 elongates the IM22 peptide (e.g. *DGRP101*). Most strains carry a *BaraB* gene of
664 similar length to *BaraA* (e.g. DGRP38). Additionally, the *BaraB* gene does not encode
665 a functional signal peptide. **B)** Only *BaraA* is immune-induced. *BaraB* and *BaraC* do
666 not respond to infection, though basal *BaraC* expression relies on Toll signalling
667 (**Fig.** S3C). **C)** Time series of whole animal *Baramicin* expression over the course of
668 development. Expression values are normalized within each gene, with expression
669 in the adult set to 100% (left panel). For context, normalizing each gene to *BaraA* in
670 adults shows that *BaraC* and especially *BaraB* expression is much lower at a whole
671 fly level (right panel). **D)** The $\Delta BaraB^{LC1}$ mutation is partial lethal, an effect that is
672 exaggerated by rearing flies at 29°C. **E)** Example of a nubbin-like wing, which is
673 phenocopied by *BaraB* gene silencing using the neural driver *elav-Gal4*. **F)** The
674 *BaraC* gene almost perfectly matches the expression pattern of glia specific *Repo*-
675 expressing cells in single cell RNA sequencing of the adult fly brain [42].
676
677 **Figure 4: *Baramicin* evolutionary history. A)** Detailed map of genomic
678 neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D.*
679 *melanogaster*, detailing inferred duplication, inversion, and translocation events.
680 Gene names are given as found in *D. melanogaster*. **B)** Cladogram and genomic loci
681 further detailing the series of events leading to the extant *Baramicin* loci of
682 *Drosophila* species. Loci in *S. lebanonensis* and flies in the subgenus *Drosophila*
683 encode only one *Baramicin* gene, indicating the ancestral drosophilid likely encoded
684 only one *Baramicin*. **C)** IM24-specific *Baramicins* arose from convergent evolution
685 both in gene structure and expression profile. Genomic loci are described here as
686 *ATP8A* or *stum* reflecting prominent genes neighbouring the *Baramicins* (see **Fig.**
687 5A). Expression dynamics relating to immune-induction or enrichment in the head

688 (checked boxes) are shown in **Fig.** 2 and **Fig.** S6. The *Baramicin* loci in *D. willistoni*
689 are syntenic with *D. melanogaster*, but evolved in a vice versa fashion (purple
690 arrow). The *D. virilis Baramicins GJ21309* and *GJ25897* are direct sister genes (100%
691 identity at N-terminus).
692
693 **Figure 5: *BaraB* and IM24 rapid evolution. A)** Evolution of the signal peptide and
694 core IM24 domain. Residue highlighting indicates agreement with *Dmel\BaraB*.
695 Arrows indicate site 29 and the *D. melanogaster* $^{40}$HHASSPAD$^{48}$ domain. Insertion
696 events in *D. yakuba* and *D. sechellia BaraB* are denoted as XX to save space. The *D.*
697 *sechellia* signal peptide is predicted to remain functional (SignalP 5.0). The
698 cladogram on the left shows genomic relatedness (by speciation and locus)
699 independent of sequence similarity. Background colouring is included to show birth
700 of novel *Baramicin* loci/lineages. **B)** Proposed timeframe for gain of novel *BaraB*
701 function. The *D. erecta BaraB* gene is pseudogenized, and the *D. yakuba* gene is not
702 immune-induced and encodes a 9-residue insertion in the IM24 peptide bordering
703 the HHASSPAD domain (see **A**). However the *BaraB* genes of *D. suzukii*, and both *D.*
704 *mauritiana* and *D. sechellia* remain inducible by infection, while the *BaraB* genes of
705 *D. simulans* and *D. melanogaster* are not, and are expressed at very low levels (Fig.
706 S8). This pattern suggests that *BaraB* of *D. melanogaster* (and *D. simulans*) acquired
707 its non-immune role only recently, and is correlated with the loss of the *BaraB* signal
708 peptide. Speciation timeframes adapted from Timetree.org [34].
709
710

711 # Supplementary figures and tables
712
713 **Figure S1: The *BaraA* locus is poorly resolved in DGRP genome assemblies. A)**
714 The number of unknown variant calls rises immediately after the start of the
715 identical promoter region shared by *BaraA1* and *BaraA2,* stabilizing at ~125/205
716 DGRP lines. Approaching the limits of Illumina resolution (~300bp), a second spike
717 in unknown calls stabilizes at ~174/205 lines corresponding to the limits of 454
718 sequencing (~450bp), after which unknown call frequency seems more randomly
719 distributed. **B)** The *BaraA1* and *BaraA2* gene regions are totally devoid of mapped
720 variants (dashed boxes). We speculate this is due to an artefact during genomic
721 assembly, where reads mapping equally to the two identical *BaraA* genes were
722 discarded as non-specific.
723
724 **Table S1: the *BaraA* duplication is variable in common lab stocks.** Inferred
725 *BaraA* copy numbers are given for six different lab stocks including *Oregon R* (*OR-R*),
726 *iso DrosDel* (*iso* $w^{1118}$), *Exelexis*, a $w^{1118}$ line from the VDRC, a *yellow white* (*yw*) line
727 used previously by our group, and *Canton S*. All samples were additionally screened
728 for *CG30059* as a DNA extraction positive control.
729
730 **Figure S2: Convergent evolution maintains a polymorphism in Baramicin**
731 **IM10-like peptides. A)** Alignment of IM10-like peptides from diverse *Drosophila*
732 species. The G/D polymorphic site is highlighted in lavender (residue 8). Broadly
733 conserved polymorphisms are also seen at sites 3 (H/V/Q), 4 (V/I), 5 (A/E/S), and 9
734 (R/H). **B)** The IM10-like polymorphic residues confer different hydrophobicity and
735 isoelectric points (PI). The three *D. melanogaster* IM10-like peptides are given as
736 examples. Bar height indicates greater hydrophobicity/PI. **C)** The syntenic codon of
737 IM12 in Melanogaster group flies encodes either a D or G residue owing to
738 convergent G-A nucleotide transitions. The D residue is far more common in
739 additional outgroups (**Fig.** S2E), suggesting the ancestral state encoded D, and G was
740 derived at least twice. **D)** IM22 encodes a unique peptide with a domain similar to
741 the IM10-like $^{6}$RP(G/D)R$^{9}$ motif, except both D and G are encoded in tandem. There
742 is a furin cleavage site upstream of IM22 in all *BaraA-like* genes. In *S. lebanonensis*
743 there is an SP dipeptidyl peptidase site that is also predicted to be cleaved off. In *D.*
744 *melanogaster BaraA* the GIND motif is also cleaved off by an unknown process [28].
745 In subgenus Drosophila flies, the mature IM22 peptide begins immediately following
746 the furin cleavage site. **E)** Presentation of the residue 8 G/D polymorphism in all
747 IM10-like peptides (i.e. IM10, IM12, IM13) of Melanogaster group flies. 41/48
748 encode an Aspartic acid, while 7/48 encode a Glycine. However, in the Quinaria
749 species group, only Glycine residues are observed (**Fig.** S2A), suggesting G/D
750 frequency is lineage-specific.
751
752 **Figure S3: Additional assays of Baramicin expression upon infection. A)**
753 Neither *BaraB* nor *BaraC* are regulated by the Imd pathway, which is specifically
754 stimulated by *E. coli* infection. **B)** Neither *BaraB* nor *BaraC* are induced after
755 infection by *M. luteus*. **C)** *BaraC* levels were consistently depressed in $spz^{rm7}$ flies in

756 the unchallenged condition (UC) or upon infection with *C. albicans* (**Fig.** 3B) or *M.*
757 *luteus* (**Fig.** S4B). Data here are pooled for *iso* wild type or *iso spz*$^{rm7}$ flies without
758 regard for infection treatment (student's t, p = .005).
759
760 **Figure S4: *BaraB* mutation is highly deleterious and penetrance of effect**
761 **depends on genetic background. A)** Diagram of *BaraB* mutant loci. The *ΔBaraB*$^{LC1}$
762 mutation encodes an unintended insertion in the *BaraB* promoter, and both a SNP in
763 the promoter and frameshift in protein product (asterisks). The *ΔBaraB*$^{LC4}$ mutation
764 encodes the intended knock-in of a DsRed cassette in place of the *BaraB* gene
765 producing a full gene knockout. **B)** *ΔBaraB*$^{LC1}$ is a *BaraB* hypomorph. Under normal
766 qPCR assay conditions, *BaraB* expression is not detected. However using highly
767 concentrated cDNA beyond our assay's valid range (100ng/10μL reaction), we could
768 detect *BaraB* transcript in *ΔBaraB*$^{LC1}$ flies. Quantification shown here is intended
769 only to show that *BaraB* transcript can be recovered from *ΔBaraB*$^{LC1}$ homozygotes,
770 and to give a sense of relative whole-fly expression levels. **C)** Emergent frequencies
771 of *ΔBaraB* flies in a variety of genetic backgrounds. *iso ΔBaraB*$^{LC1}$ flies were
772 backcrossed for seven generations. *w; ΔBaraB/CyO* flies represent the original
773 background the mutation was generated in, while *ΔBaraB/CyO-GFP* flies have had
774 additional genetic variation introduced in their 1$^{st}$ and 3$^{rd}$ chromosomes. Rare
775 homozygous L3 larvae were observed in *ΔBaraB*$^{LC4}$/CyO-GFP vials (<5% of L3
776 larvae), but these larvae never successfully eclosed. **D)** Aborted pupae (yellow
777 arrows) are a common occurrence in *ΔBaraB* vials, and sometimes contain fully-
778 developed adults that simply never eclosed. In **E-G:** ns = not significant, * = p < .05,
779 *** p = < .001. **E)** The ratio of *ΔBaraB*$^{LC1}$/CyO-GFP to *ΔBaraB*$^{LC1}$ homozygous larvae
780 drops between the S2 and S3 larval stages ($\chi^2$, p = .515 and p = .012 respectively). **F)**
781 Frequency of successfully eclosing adults using *BaraB*$^{LC4}$/CyO-GFP flies. **G)**
782 Frequency of successfully eclosing adults using *BaraB*$^{LC1}$/CyO-GFP flies. **H)** *BaraB*
783 mutation negatively affects lifespan. *iso ΔBaraB*$^{LC1}$ homozygotes suffer reduced
784 lifespan even relative to their *iso ΔBaraB*$^{LC1}$*/CyO* siblings. By comparison, *iso ΔBaraA*
785 flies that used the same vector for mutant generation live as wild-type. ATM$^{8}$ flies
786 suffer precocious neurodegeneration and are included as short-lived controls [58].
787 **I)** *ΔBaraB*$^{LC1}$ crossed to the genomic deficiency line (*Df(9063)*) supports a partial-
788 lethal effect of *BaraB* mutation.
789
790 **Table S2: *BaraB* RNAi summary statistics.** Crosses used either the TRiP or KK
791 *BaraB-IR* lines, driven by either *Actin5C-Gal4* or *elav-Gal4,* sometimes including *UAS-*
792 *Dcr2*. Rearing at 29°C and inclusion of *UAS-Dcr2* increases the strength of RNA
793 silencing.
794
795 **Figure S5: *BaraC* is expressed in the nervous system, but also the hindgut and**
796 **rectal pads. A)** FlyAtlas2 expression data for *BaraC*. **B)** RT-qPCR of *BaraC* in whole
797 flies using different Gal4 drivers to express *BaraC* RNAi. *BaraC* is knocked down by
798 both the *elav-Gal4* and *Repo-Gal4* nervous system drivers. Cumulatively, nervous
799 system drivers significantly depress *BaraC* expression compared to *BaraC-IR*
800 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 [34]. 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. [32].** 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.

844 **References:**
845


846 1. Hanson MA, Lemaitre B. New insights on Drosophila antimicrobial peptide
847 function in host defense and beyond. Curr Opin Immunol. 2020;62: 22–30.
848 doi:10.1016/j.coi.2019.11.008

849 2. Semple F, Dorin JR. ß-Defensins: Multifunctional Modulators of Infection,
850 Inflammation and More? J Innate Immun. 2012;4: 337–348.
851 doi:10.1159/000336619

852 3. Gerdol M, Schmitt P, Venier P, Rocha G, Rosa RD, Destoumieux-Garzón D.
853 Functional Insights From the Evolutionary Diversification of Big Defensins.
854 Front Immunol. 2020;11: 758. doi:10.3389/fimmu.2020.00758

855 4. Hanson MA, Lemaitre B, Unckless RL. Dynamic Evolution of Antimicrobial
856 Peptides Underscores Trade-Offs Between Immunity and Ecological Fitness.
857 Frontiers in Immunology. 2019;10: 2620. doi:doi: 10.3389/fimmu.2019.02620

858 5. Sackton TB, Lazzaro BP, Clark AG, Wittkopp P. Rapid expansion of immune-
859 related gene families in the house fly, musca domestica. Molecular Biology and
860 Evolution. 2017. doi:10.1093/molbev/msw285

861 6. Vilcinskas A, Mukherjee K, Vogel H. Expansion of the antimicrobial peptide
862 repertoire in the invasive ladybird Harmonia axyridis. Proceedings of the Royal
863 Society B: Biological Sciences. 2013. doi:10.1098/rspb.2012.2113

864 7. Wang Y, Zhu S. The defensin gene family expansion in the tick Ixodes
865 scapularis. Developmental and Comparative Immunology. 2011.
866 doi:10.1016/j.dci.2011.03.030

867 8. Sackton TB, Lazzaro BP, Schlenke T a, Evans JD, Hultmark D, Clark AG. Dynamic
868 evolution of the innate immune system in Drosophila. Nature genetics.
869 2007;39: 1461–1468. doi:10.1038/ng.2007.60

870 9. Lazzaro BP, Zasloff M, Rolff J. Antimicrobial peptides: Application informed by
871 evolution. Science. 2020;368. doi:10.1126/science.aau5480

872 10. Halldórsdóttir K, Árnason E. Trans-species polymorphism at antimicrobial
873 innate immunity cathelicidin genes of Atlantic cod and related species. PeerJ.
874 2015;3: e976. doi:10.7717/peerj.976

875 11. Hellgren O, Sheldon BC, Buckling A. In vitro tests of natural allelic variation of
876 innate immune genes (avian ??-defensins) reveal functional differences in
877 microbial inhibition. Journal of Evolutionary Biology. 2010;23: 2726–2730.
878 doi:10.1111/j.1420-9101.2010.02115.x

879 12. Chapman JR, Hill T, Unckless RL. Balancing selection drives maintenance of
880 genetic variation in Drosophila antimicrobial peptides. Genome Biology and
881 Evolution. 2019;11: 2691–2701. doi:https://doi.org/10.1093/gbe/evz191

882 13. Jiggins FM, Kim KW. A screen for immunity genes evolving under positive
883 selection in Drosophila. Journal of Evolutionary Biology. 2007;20: 965–970.
884 doi:10.1111/j.1420-9101.2007.01305.x

885 14. Tennessen JA. Molecular evolution of animal antimicrobial peptides:
886 Widespread moderate positive selection. 2005. doi:10.1111/j.1420-
887 9101.2005.00925.x

888 15. Hanson MA, Hamilton PT, Perlman SJ. Immune genes and divergent
889 antimicrobial peptides in flies of the subgenus Drosophila. BMC evolutionary
890 biology. 2016;16: 228. doi:10.1186/s12862-016-0805-y

891 16. Unckless RL, Howick VM, Lazzaro BP. Convergent Balancing Selection on an
892 Antimicrobial Peptide in Drosophila. Current Biology. 2016;26: 257–262.
893 doi:10.1016/j.cub.2015.11.063

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

896 18. Sinner MP, Masurat F, Ewbank JJ, Pujol N, Bringmann H. Innate Immunity
897 Promotes Sleep through Epidermal Antimicrobial Peptides. Current Biology.
898 2021;31: 564-577.e12. doi:10.1016/j.cub.2020.10.076

899 19. Barajas-azpeleta R, Wu J, Gill J, Welte R. Antimicrobial peptides modulate long-
900 term memory. PLoS Genetics. 2018; 1–26. doi:10.1371/journal.pgen.1007440

901 20. Lezi E, Zhou T, Koh S, Chuang M, Sharma R, Pujol N, et al. An Antimicrobial
902 Peptide and Its Neuronal Receptor Regulate Dendrite Degeneration in Aging
903 and Infection. Neuron. 2018;97: 125-138.e5. doi:10.1016/j.neuron.2017.12.001

904 21. Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, et al.
905 Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease
906 causation and treatment with small-molecule inhibitors. Science Advances.
907 2019;5. doi:10.1126/sciadv.aau3333

908 22. Gosztyla ML, Brothers HM, Robinson SR. Alzheimer's Amyloid-β is an
909 Antimicrobial Peptide: A Review of the Evidence. 2018. doi:10.3233/JAD-
910 171133

911 23. Pineda A, Burré J. Modulating membrane binding of α-synuclein as a
912 therapeutic strategy. Proceedings of the National Academy of Sciences.
913 2017;114: 1223–1225. doi:10.1073/pnas.1620159114

914 24. De Lorenzi E, Chiari M, Colombo R, Cretich M, Sola L, Vanna R, et al. Evidence
915 that the human innate immune peptide LL-37 may be a binding partner of
916 amyloid-β and inhibitor of fibril assembly. Journal of Alzheimer's Disease.
917 2017;59: 1213–1226. doi:10.3233/JAD-170223

918 25. Hanson MA, Dostálová A, Ceroni C, Poidevin M, Kondo S, Lemaitre B. Synergy
919 and remarkable specificity of antimicrobial peptides in vivo using a systematic
920 knockout approach. eLife. 2019;8. doi:10.7554/elife.44341

921 26. Ebrahim SAM, Talross GJS, Carlson JR. Sight of parasitoid wasps accelerates
922 sexual behavior and upregulates a micropeptide gene in Drosophila. Nat
923 Commun. 2021;12: 2453. doi:10.1038/s41467-021-22712-0

924 27. Lee M, Shi X, Barron AE, McGeer E, McGeer PL. Human antimicrobial peptide
925 LL-37 induces glial-mediated neuroinflammation. Biochemical Pharmacology.
926 2015;94: 130–141. doi:10.1016/j.bcp.2015.02.003

927 28. Hanson MA, Cohen LB, Marra A, Iatsenko I, Wasserman SA, Lemaitre B. The
928 Drosophila Baramicin polypeptide gene protects against fungal infection.
929 bioRxiv; 2020 Nov. doi:10.1101/2020.11.23.394148

930 29. Casteels-Josson K, Capaci T, Casteels P, Tempst P. Apidaecin multipeptide
931 precursor structure: a putative mechanism for amplification of the insect
932 antibacterial response. The EMBO journal. 1993;12: 1569–78.
933 doi:10.1002/j.1460-2075.1993.tb05801.x

934 30. Huang J, Lou Y, Liu J, Bulet P, Jiao R, Hoffmann JA, et al. The BaramicinA gene is
935 required at several steps of the host defense against Enterococcus faecalis and
936 Metarhizium robertsii in a septic wound infection model in Drosophila
937 melanogaster. bioRxiv; 2020 Nov. doi:10.1101/2020.11.23.394809

938 31. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, et al. The
939 Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482: 173–8.
940 doi:10.1038/nature10811

941 32. Wang JB, Lu H-L, St Leger RJ. The genetic basis for variation in resistance to
942 infection in the Drosophila melanogaster genetic reference panel. PLoS Pathog.
943 2017;13: e1006260. doi:10.1371/journal.ppat.1006260

944 33. Kobler JM, Rodriguez Jimenez FJ, Petcu I, Grunwald Kadow IC. Immune
945 Receptor Signaling and the Mushroom Body Mediate Post-ingestion Pathogen
946 Avoidance. Curr Biol. 2020;30: 4693-4709.e3. doi:10.1016/j.cub.2020.09.022

947 34. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines,
948 Timetrees, and Divergence Times. Molecular biology and evolution. 2017;34:
949 1812–1819. doi:10.1093/molbev/msx116

950 35. Unckless RL, Lazzaro BP. The potential for adaptive maintenance of diversity in
951 insect antimicrobial peptides. Philosophical Transactions of the Royal Society
952 B: Biological Sciences. 2016;371: 20150291. doi:10.1098/rstb.2015.0291

953 36. Lazzaro BP, Clark AG. Molecular population genetics of inducible antibacterial
954 peptide genes in Drosophila melanogaster. Mol Biol Evol. 2003;20: 914–923.
955 doi:10.1093/molbev/msg109

956 37. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane
957 protein topology with a hidden Markov model: application to complete
958 genomes. J Mol Biol. 2001;305: 567–580. doi:10.1006/jmbi.2000.4315

959 38. Ferreira ÁG, Naylor H, Esteves SS, Pais IS, Martins NE, Teixeira L. The Toll-
960 dorsal pathway is required for resistance to viral oral infection in Drosophila.
961 PLoS Pathog. 2014;10: e1004507. doi:10.1371/journal.ppat.1004507

962 39. Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, et al. A Genome-
963 wide Drosophila screen for heat nociception identifies α2δ3 as an
964 evolutionarily conserved pain gene. Cell. 2010;143: 628–638.
965 doi:10.1016/j.cell.2010.09.047

966 40. Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. FlyAtlas 2: A new version
967 of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and
968 sex-specific data. Nucleic Acids Research. 2018;46: D809–D815.
969 doi:10.1093/nar/gkx976

970 41. Hammonds AS, Bristow CA, Fisher WW, Weiszmann R, Wu S, Hartenstein V, et
971 al. Spatial expression of transcription factors in Drosophila embryonic organ
972 development. Genome Biol. 2013;14: R140. doi:10.1186/gb-2013-14-12-r140

973 42. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, et al. A Single-
974 Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell. 2018;174: 982-
975 998.e20. doi:10.1016/j.cell.2018.05.057

976 43. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: A
977 suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics.
978 2010;26: 2455–2457. doi:10.1093/bioinformatics/btq429

979 44. Kim BY, Wang J, Miller DE, Barmina O, Delaney EK, Thompson A, et al. Highly
980 contiguous assemblies of 101 drosophilid genomes. Evolutionary Biology; 2020
981 Dec. doi:10.1101/2020.12.14.422775

982 45. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-
983 wide identification of episodic selection. Mol Biol Evol. 2015;32: 1365–1371.
984 doi:10.1093/molbev/msv035

985 46. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web
986 portal for protein modeling, prediction and analysis. Nature protocols.
987 2015;10: 845–58. doi:10.1038/nprot.2015.053

988 47. Zhang W, Yang J, He B, Walker SE, Zhang H, Govindarajoo B, et al. Integration of
989 QUARK and I-TASSER for Ab Initio Protein Structure Prediction in CASP11.
990 Proteins. 2016;84 Suppl 1: 76–86. doi:10.1002/prot.24930

991 48. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O,
992 Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep
993 neural networks. Nat Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-
994 0036-z

995 49. Levy F, Rabel D, Charlet M, Bulet P, Hoffmann JA, Ehret-Sabatier L. Peptidomic
996 and proteomic analyses of the systemic immune response of Drosophila.
997 Biochimie. 2004;86: 607–616. doi:10.1016/j.biochi.2004.07.007

998 50. Barrett AJ, Rawlings ND, Woessner JF, editors. Handbook of proteolytic
999 enzymes. 2nd ed. Amsterdam ; Boston: Elsvier Academic Press; 2004.

1000 51. Brogden KA. Antimicrobial peptides: Pore formers or metabolic inhibitors in
1001 bacteria? 2005. doi:10.1038/nrmicro1098

1002 52. Moir RD, Lathe R, Tanzi RE. The antimicrobial protection hypothesis of
1003 Alzheimer's disease. Alzheimer's & Dementia. 2018;14: 1602–1614.
1004 doi:10.1016/j.jalz.2018.06.3040

1005 53. Zanetti M. The role of cathelicidins in the innate host defenses of mammals.
1006 Curr Issues Mol Biol. 2005;7: 179–196.

1007 54. Masuzzo A, Manière G, Viallat-Lieutaud A, Avazeri É, Zugasti O, Grosjean Y, et al.
1008 Peptidoglycan-dependent NF-κB activation in a small subset of brain
1009 octopaminergic neurons controls female oviposition. eLife. 2019;8: e50559.
1010 doi:10.7554/eLife.50559

1011 55. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al.
1012 Geneious Basic: An integrated and extendable desktop software platform for
1013 the organization and analysis of sequence data. Bioinformatics. 2012.
1014 doi:10.1093/bioinformatics/bts199

1015 56. Lemaitre B, Reichhart JM, Hoffmann JA. Drosophila host defense: differential
1016 induction of antimicrobial peptide genes after infection by various classes of
1017 microorganisms. Proceedings of the National Academy of Sciences of the United
1018 States of America. 1997;94: 14614–9. doi:10.1073/pnas.94.26.14614

1019 57. Pfaffl MW. A new mathematical model for relative quantification in real-time
1020 RT-PCR. Nucleic Acids Res. 2001;29: e45. doi:10.1093/nar/29.9.e45

1021 58. Petersen AJ, Katzenberger RJ, Wassarman DA. The innate immune response
1022 transcription factor relish is necessary for neurodegeneration in a Drosophila
1023 model of ataxia-telangiectasia. Genetics. 2013;194: 133–142.
1024 doi:10.1534/genetics.113.150854

1025

Figure 1

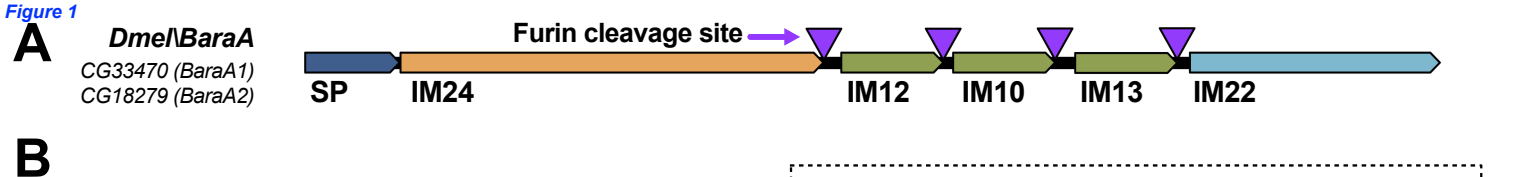

A
Dmel\BaraA
CG33470 (BaraA1)
CG18279 (BaraA2)
Furin cleavage site
SP
IM24
IM12
IM10
IM13
IM22
B
50A gene region
CG33470
CG30059
ATP8A
PCR amplicon
CG18279
CG33470
CG18278
CR45470
CG30059
ATP8A
DGRP predictive region

Figure 2

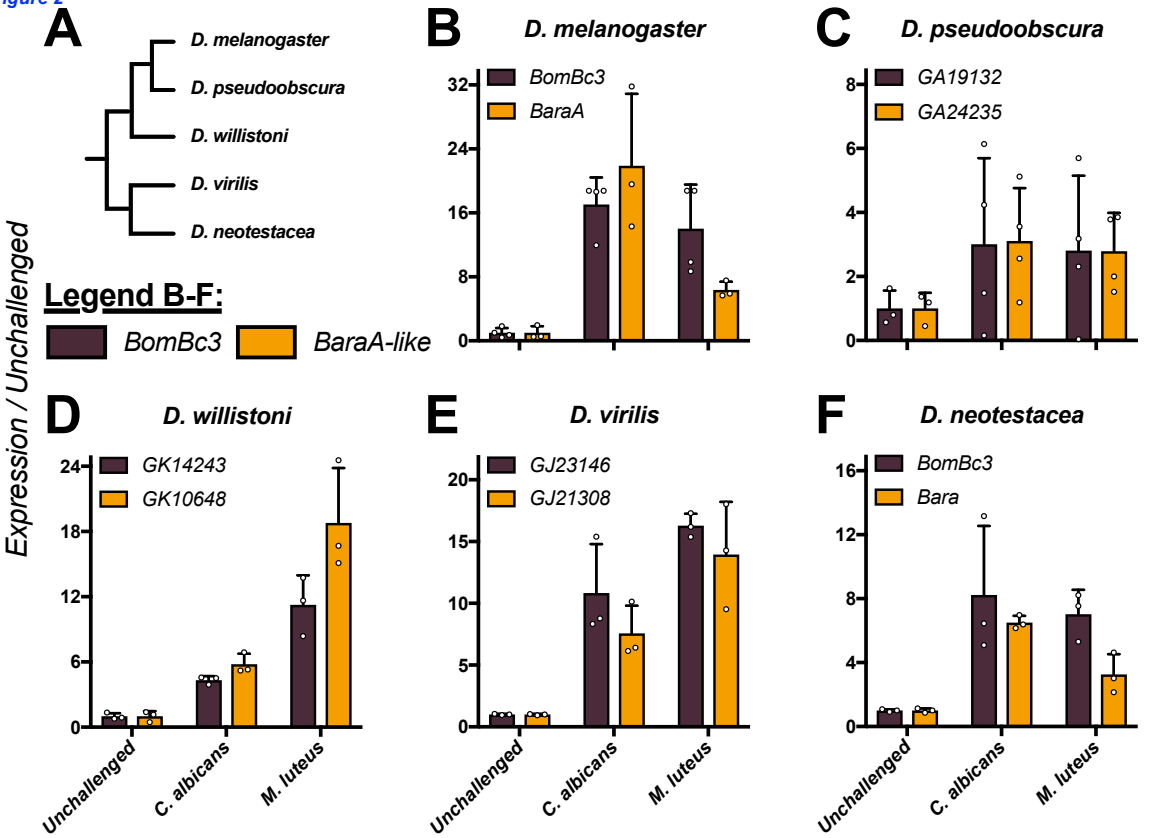

A
D. melanogaster
D. pseudoobscura
D. willistoni
D. virilis
D. neotestacea
Legend B-F:
BomBc3
BaraA-like
Expression / Unchallenged
B
D. melanogaster
BomBc3
BaraA
C
D. pseudoobscura
GA19132
GA24235
D
D. willistoni
GK14243
GK10648
E
D. virilis
GJ23146
GJ21308
F
D. neotestacea
BomBc3
Bara
Unchallenged
C. albicans
M. luteus

Figure 3

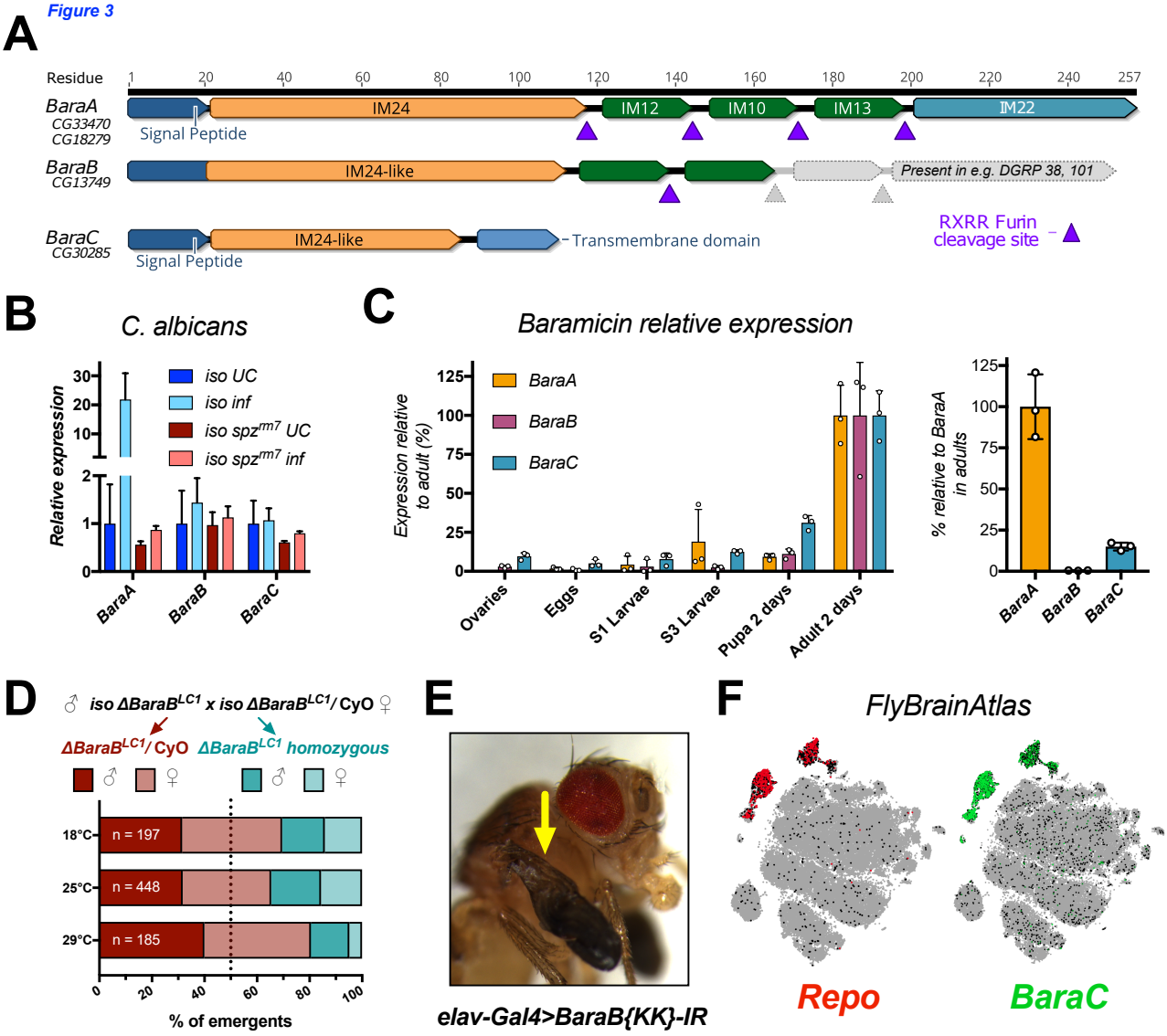
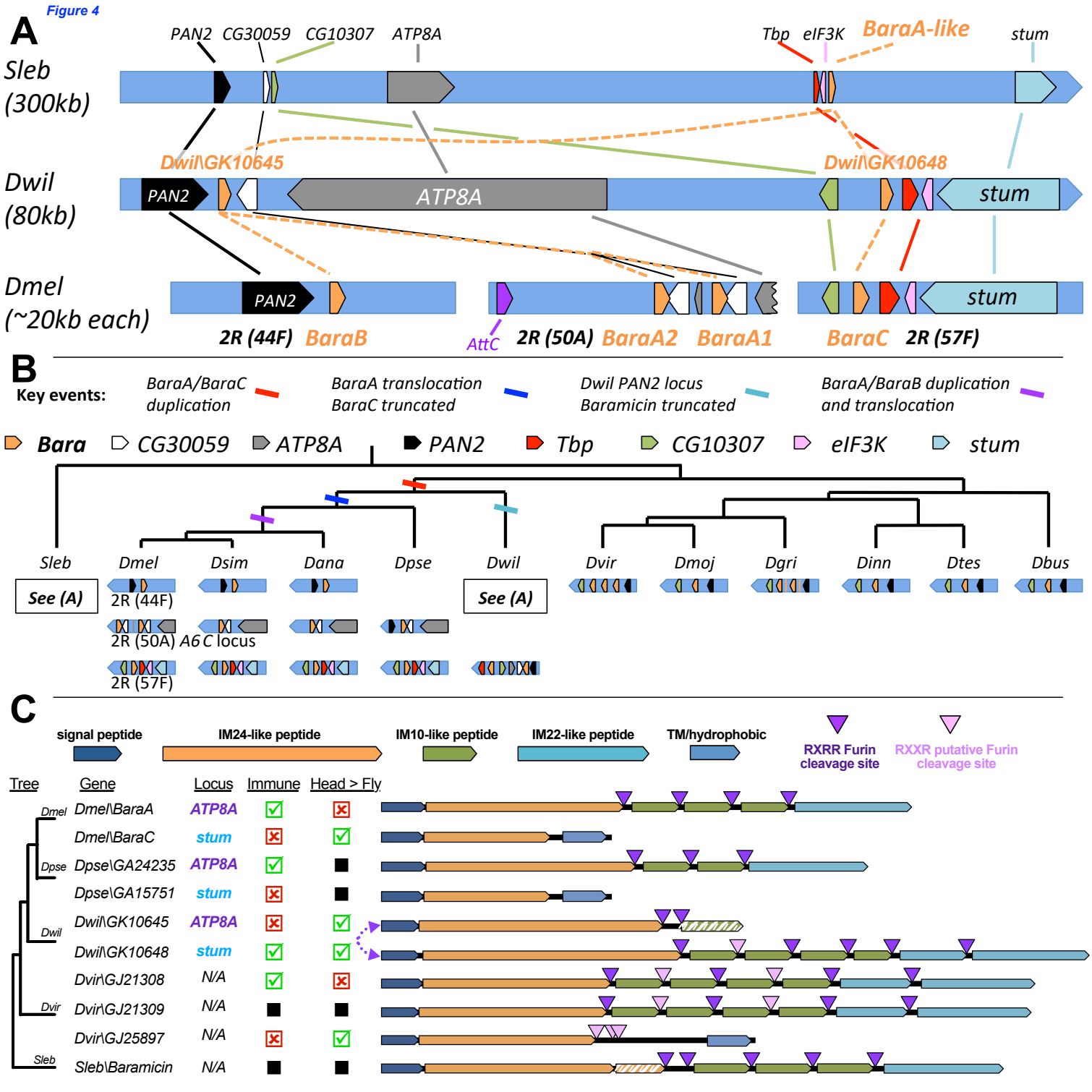

A
Residue
1
20
40
60
80
100
120
140
160
180
200
220
240
257
BaraA
CG33470
CG18279
Signal Peptide
IM24
IM12
IM10
IM13
IM22
BaraB
CG13749
IM24-like
Present in e.g. DGRP 38, 101
BaraC
CG30285
Signal Peptide
IM24-like
Transmembrane domain
RXRR Furin cleavage site
B
C. albicans
Relative expression
iso UC
iso inf
iso spz^rm7 UC
iso spz^rm7 inf
BaraA
BaraB
BaraC
C
Baramicin relative expression
Expression relative to adult (%)
BaraA
BaraB
BaraC
Ovaries
Eggs
S1 Larvae
S3 Larvae
Pupa 2 days
Adult 2 days
% relative to BaraA in adults
D
♂ iso ΔBaraB^LC1 x iso ΔBaraB^LC1/ CyO ♀
ΔBaraB^LC1/ CyO
ΔBaraB^LC1 homozygous
18°C
25°C
29°C
n = 197
n = 448
n = 185
% of emergents
E
elav-Gal4>BaraB{KK}-IR
F
FlyBrainAtlas
Repo
BaraC

Figure 4


A
Sleb (300kb)
PAN2
CG30059
CG10307
ATP8A
Tbp
eIF3K
BaraA-like
stum
Dwil (80kb)
Dwil\GK10645
Dwil\GK10648
PAN2
ATP8A
stum
Dmel (~20kb each)
PAN2
stum
2R (44F)
BaraB
AttC
2R (50A)
BaraA2
BaraA1
BaraC
2R (57F)
B
Key events:
BaraA/BaraC duplication
BaraA translocation BaraC truncated
Dwil PAN2 locus Baramicin truncated
BaraA/BaraB duplication and translocation
Bara
CG30059
ATP8A
PAN2
Tbp
CG10307
eIF3K
stum
Sleb
Dmel
Dsim
Dana
Dpse
Dwil
Dvir
Dmoj
Dgri
Dinn
Dtes
Dbus
See (A)
See (A)
2R (44F)
2R (50A)
A6C locus
2R (57F)
C
signal peptide
IM24-like peptide
IM10-like peptide
IM22-like peptide
TM/hydrophobic
RXRR Furin cleavage site
RXXR putative Furin cleavage site
Tree
Gene
Locus
Immune
Head > Fly
Dmel
Dmel\BaraA
ATP8A
Dmel\BaraC
stum
Dpse
Dpse\GA24235
ATP8A
Dpse\GA15751
stum
Dwil
Dwil\GK10645
ATP8A
Dwil\GK10648
stum
Dvir
Dvir\GJ21308
N/A
Dvir\GJ21309
N/A
Dvir\GJ25897
N/A
Sleb
Sleb\Baramicin
N/A

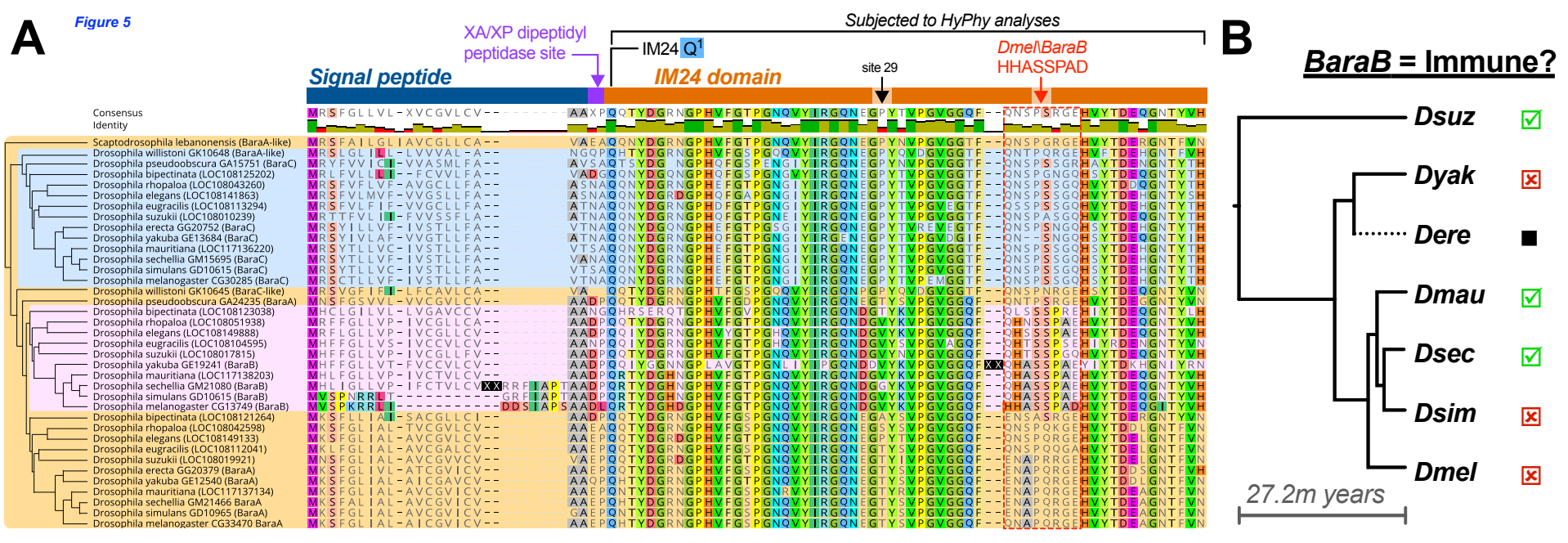
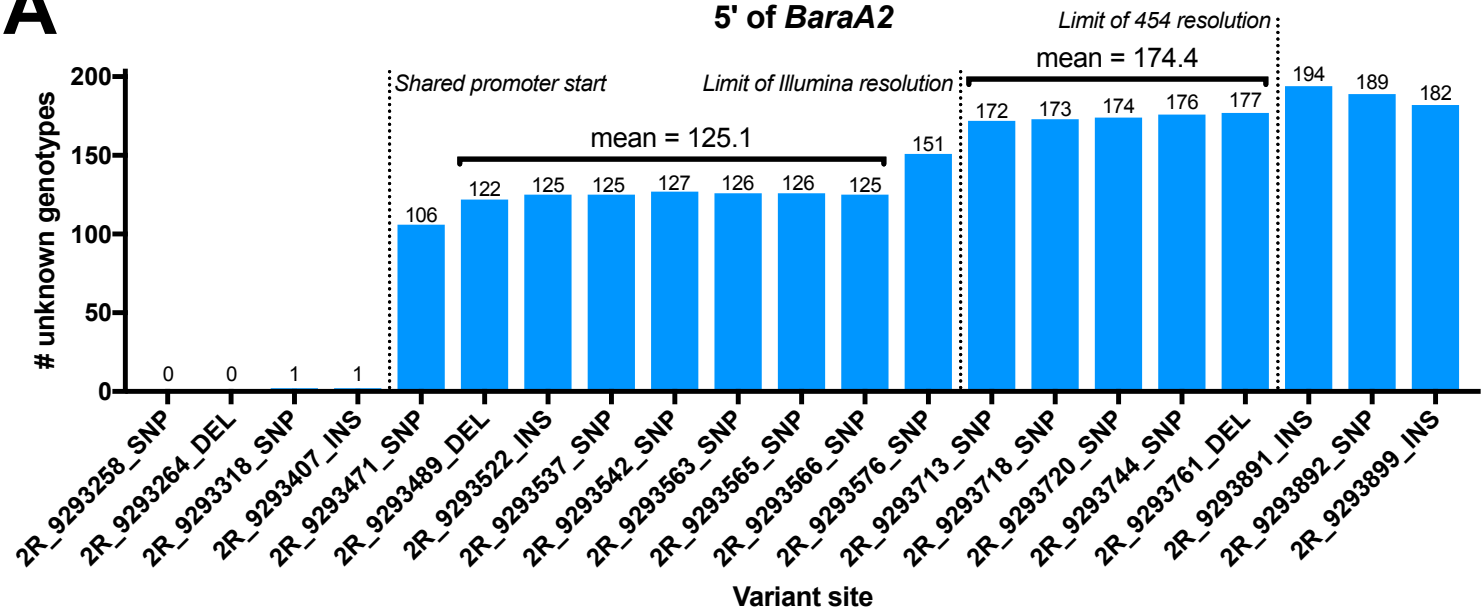

A
Figure 5
XA/XP dipeptidyl peptidase site
Subjected to HyPhy analyses
IM24 Q1
Signal peptide
IM24 domain
site 29
Dmel\BaraB HHASSPAD
Consensus
Identity
Scaptodrosophila lebanonensis (BaraA-like)
Drosophila willistoni GK10648 (BaraA-like)
Drosophila pseudoobscura GA15751 (BaraC)
Drosophila bipectinata (LOC108125202)
Drosophila rhopaloa (LOC108043260)
Drosophila elegans (LOC108141863)
Drosophila eugracilis (LOC108113294)
Drosophila suzukii (LOC108010239)
Drosophila erecta GG20752 (BaraC)
Drosophila yakuba GE13684 (BaraC)
Drosophila mauritiana (LOC117136220)
Drosophila sechellia GM15695 (BaraC)
Drosophila simulans GD10615 (BaraC)
Drosophila melanogaster CG30285 (BaraC)
Drosophila willistoni GK10645 (BaraC-like)
Drosophila pseudoobscura GA24235 (BaraA)
Drosophila bipectinata (LOC108123038)
Drosophila rhopaloa (LOC108051938)
Drosophila elegans (LOC108149888)
Drosophila eugracilis (LOC108104595)
Drosophila suzukii (LOC108017815)
Drosophila yakuba GE19241 (BaraB)
Drosophila mauritiana (LOC117138203)
Drosophila sechellia GM21080 (BaraB)
Drosophila simulans GD10615 (BaraB)
Drosophila melanogaster CG13749 (BaraB)
Drosophila bipectinata (LOC108121264)
Drosophila rhopaloa (LOC108042598)
Drosophila elegans (LOC108149133)
Drosophila eugracilis (LOC108112041)
Drosophila suzukii (LOC108019921)
Drosophila erecta GG20379 (BaraA)
Drosophila yakuba GE12540 (BaraA)
Drosophila mauritiana (LOC117137134)
Drosophila sechellia GM21466 BaraA
Drosophila simulans GD10965 (BaraA)
Drosophila melanogaster CG33470 BaraA
B
BaraB = Immune?
Dsuz
Dyak
Dere
Dmau
Dsec
Dsim
Dmel
27.2m years

*Figure S1*

**A**


5' of BaraA2
Shared promoter start
Limit of Illumina resolution
Limit of 454 resolution
mean = 125.1
mean = 174.4
# unknown genotypes
200
150
100
50
0
0
0
1
1
106
122
125
125
127
126
126
125
151
172
173
174
176
177
194
189
182
2R_9293258_SNP
2R_9293264_DEL
2R_9293318_SNP
2R_9293407_INS
2R_9293471_SNP
2R_9293489_DEL
2R_9293522_INS
2R_9293537_SNP
2R_9293542_SNP
2R_9293563_SNP
2R_9293565_SNP
2R_9293566_SNP
2R_9293576_SNP
2R_9293713_SNP
2R_9293718_SNP
2R_9293720_SNP
2R_9293744_SNP
2R_9293761_DEL
2R_9293891_INS
2R_9293892_SNP
2R_9293899_INS
Variant site

**B**

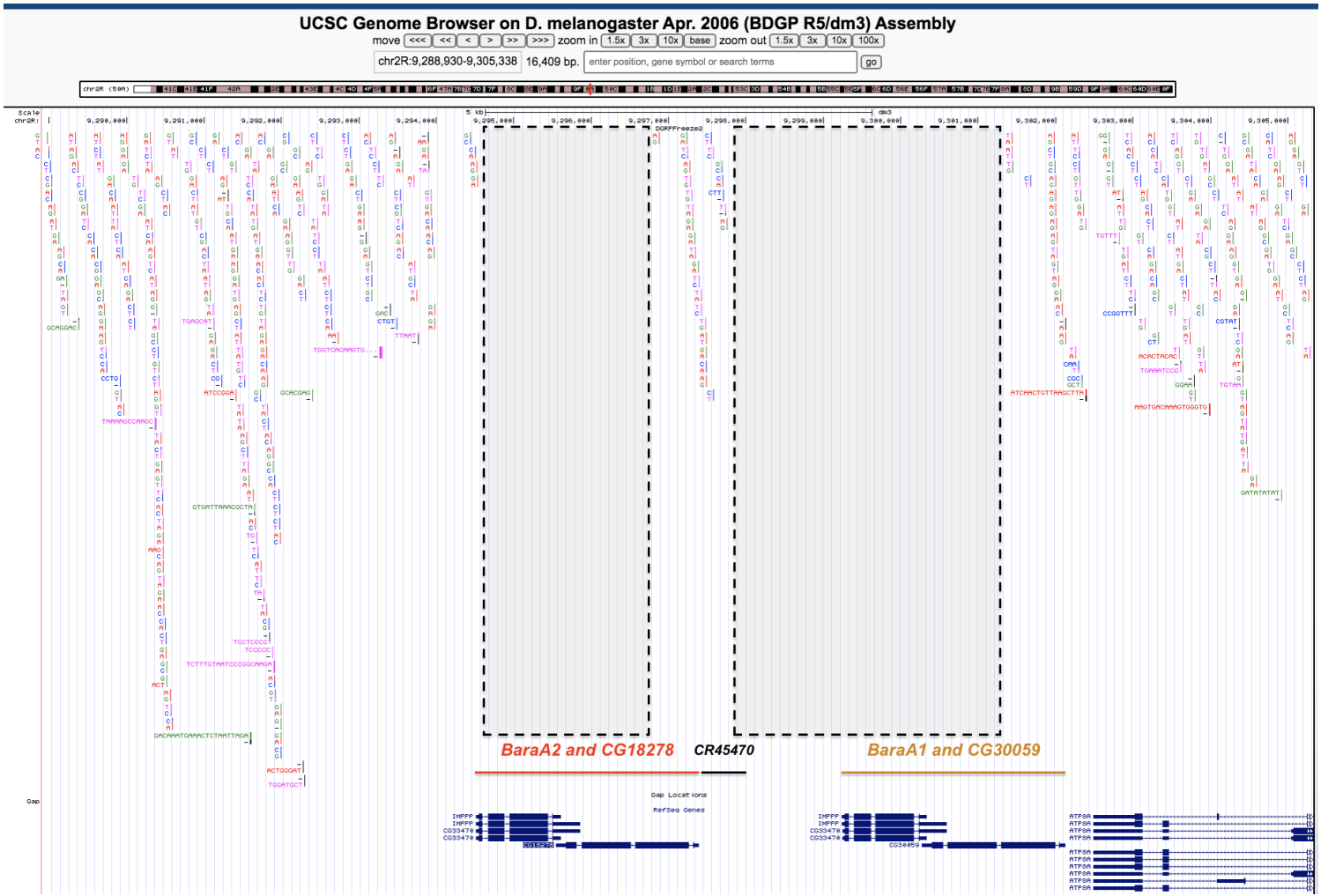

UCSC Genome Browser on D. melanogaster Apr. 2006 (BDGP R5/dm3) Assembly
move
zoom in
zoom out
chr2R:9,288,930-9,305,338
16,409 bp.
go
BaraA2 and CG18278
CR45470
BaraA1 and CG30059
Gap Locations
RefSeq Genes

*Figure S2*

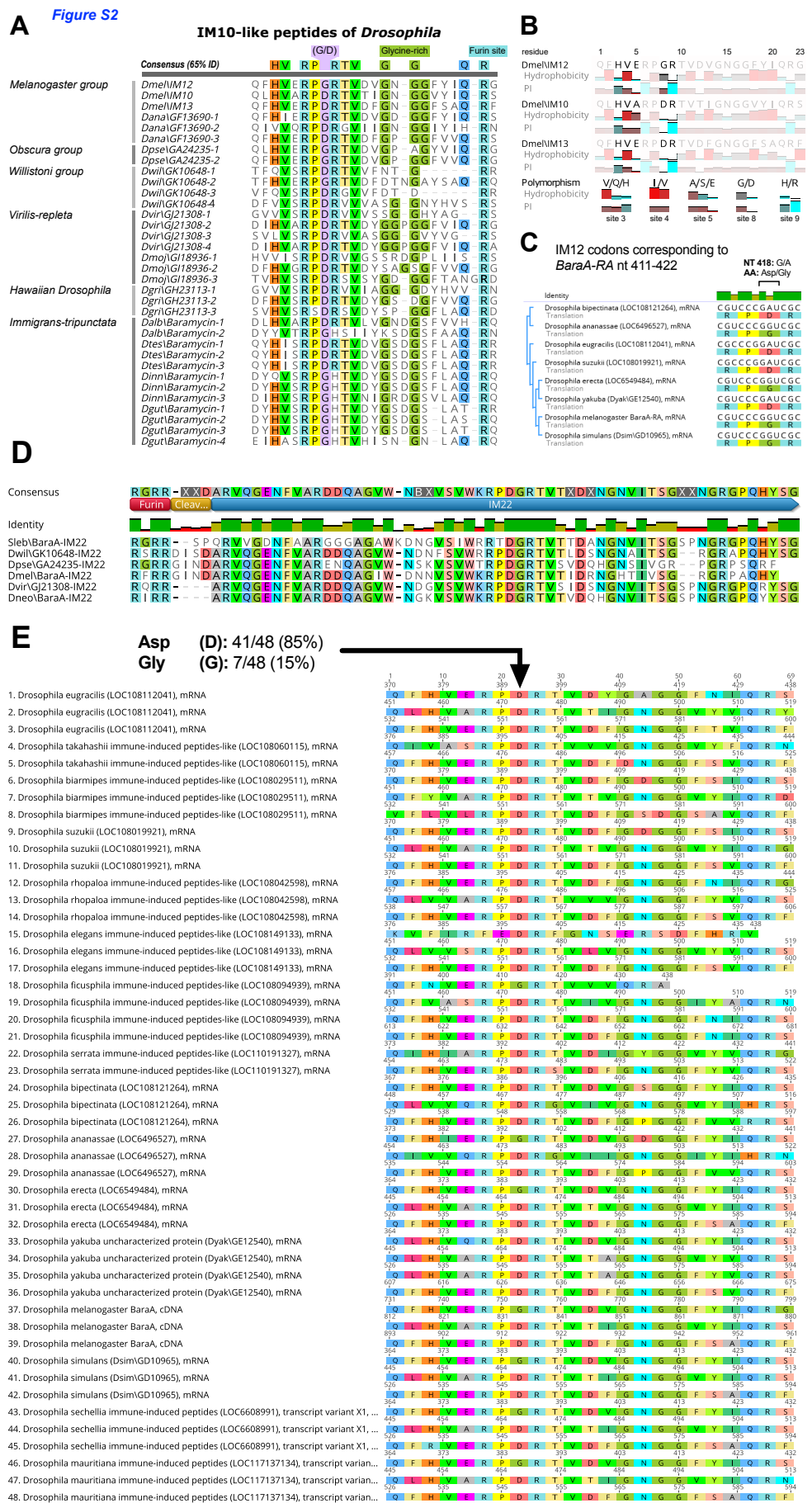

A
IM10-like peptides of Drosophila
(G/D)
Glycine-rich
Furin site
Consensus (65% ID)
Melanogaster group
Dme\IM12
Dme\IM10
Dme\IM13
Dana\GF13690-1
Dana\GF13690-2
Dana\GF13690-3
Obscura group
Dpse\GA24235-1
Dpse\GA24235-2
Willistoni group
Dwil\GK10648-1
Dwil\GK10648-2
Dwil\GK10648-3
Dwil\GK10648-4
Virilis-repleta
Dvir\GJ21308-1
Dvir\GJ21308-2
Dvir\GJ21308-3
Dvir\GJ21308-4
Dmoj\GI18936-1
Dmoj\GI18936-2
Dmoj\GI18936-3
Hawaiian Drosophila
Dgri\GH23113-1
Dgri\GH23113-2
Dgri\GH23113-3
Immigrans-tripunctata
Dalb\Baramycin-1
Dalb\Baramycin-2
Dtes\Baramycin-1
Dtes\Baramycin-2
Dtes\Baramycin-3
Dinn\Baramycin-1
Dinn\Baramycin-2
Dinn\Baramycin-3
Dgut\Baramycin-1
Dgut\Baramycin-2
Dgut\Baramycin-3
Dgut\Baramycin-4
B
residue
Dme\IM12
Dme\IM10
Dme\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...

Figure S3

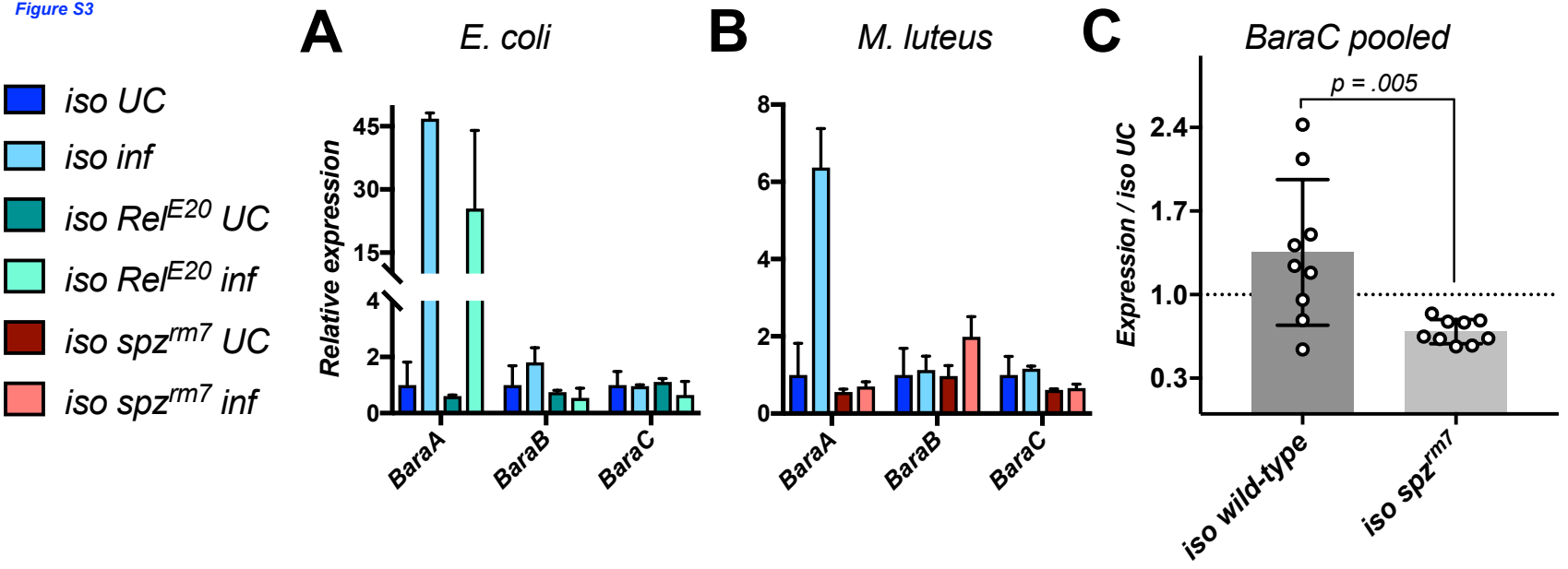

iso UC
iso inf
iso Rel$^{E20}$ UC
iso Rel$^{E20}$ inf
iso spz$^{rm7}$ UC
iso spz$^{rm7}$ inf
A
E. coli
Relative expression
45
30
15
4
2
0
BaraA
BaraB
BaraC
B
M. luteus
8
6
4
2
0
BaraA
BaraB
BaraC
C
BaraC pooled
p = .005
Expression / iso UC
2.4
1.7
1.0
0.3
iso wild-type
iso spz$^{rm7}$

Figure S4

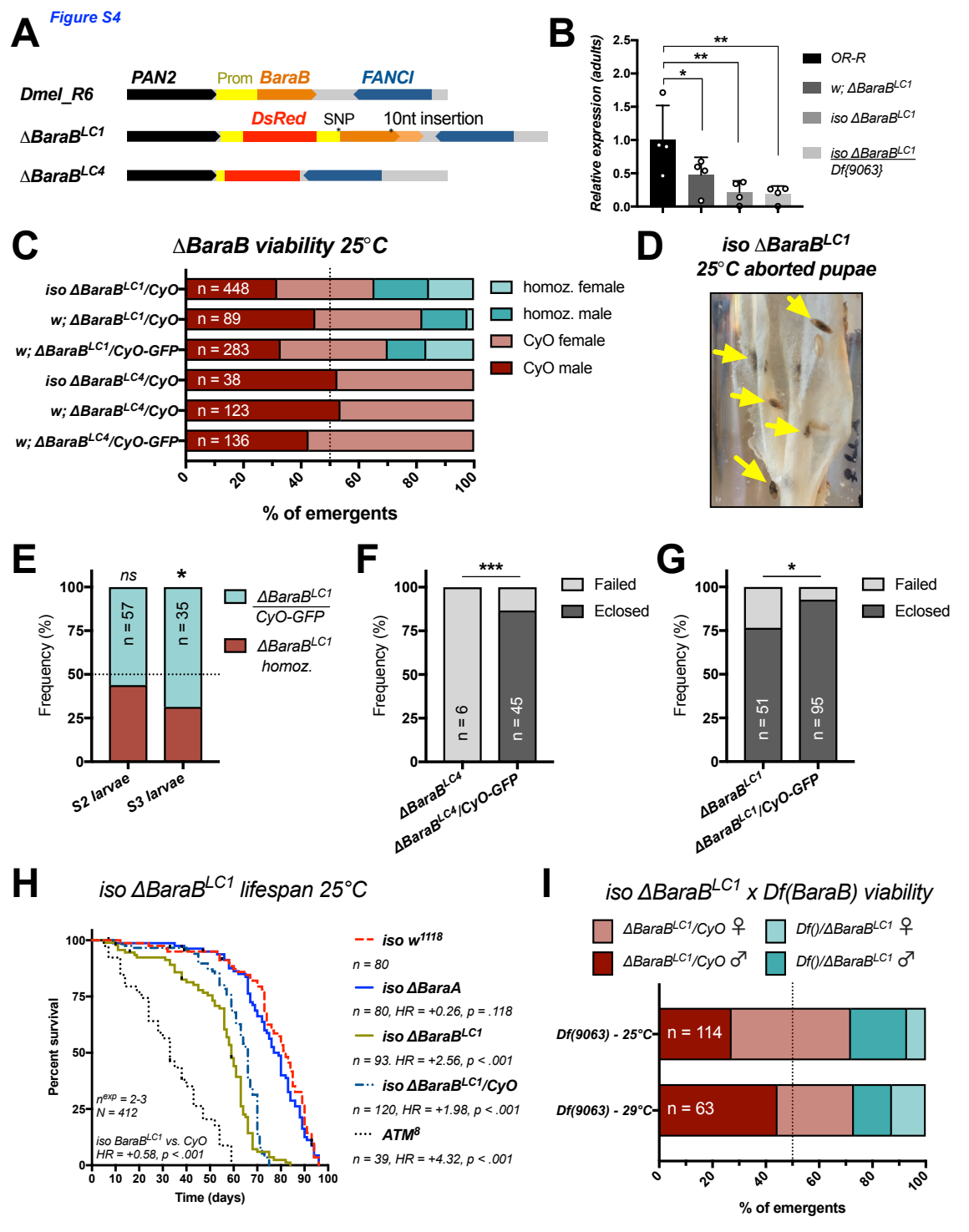
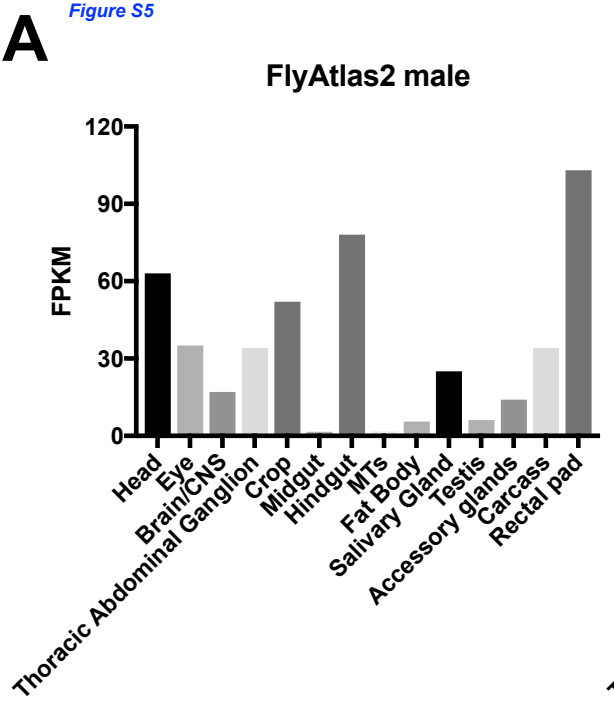
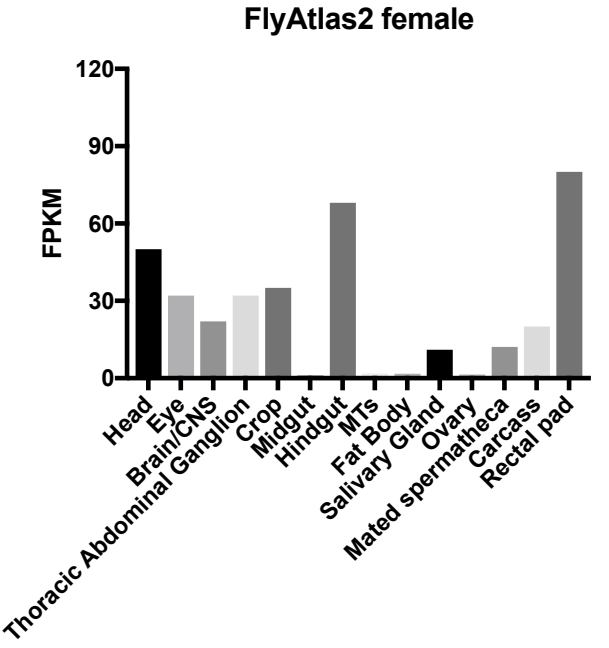

A
PAN2
Prom
BaraB
FANCI
Dmel_R6
DsRed
SNP
10nt insertion
ΔBaraB^LC1
ΔBaraB^LC4
B
Relative expression (adults)
OR-R
w; ΔBaraB^LC1
iso ΔBaraB^LC1
iso ΔBaraB^LC1 Df{9063}
C
ΔBaraB viability 25°C
iso ΔBaraB^LC1/CyO
w; ΔBaraB^LC1/CyO
w; ΔBaraB^LC1/CyO-GFP
iso ΔBaraB^LC4/CyO
w; ΔBaraB^LC4/CyO
w; ΔBaraB^LC4/CyO-GFP
n = 448
n = 89
n = 283
n = 38
n = 123
n = 136
homoz. female
homoz. male
CyO female
CyO male
% of emergents
D
iso ΔBaraB^LC1
25°C aborted pupae
E
ns
*
n = 57
n = 35
ΔBaraB^LC1/CyO-GFP
ΔBaraB^LC1 homoz.
Frequency (%)
S2 larvae
S3 larvae
F
***
Failed
Eclosed
n = 6
n = 45
ΔBaraB^LC4
ΔBaraB^LC4/CyO-GFP
G
*
n = 51
n = 95
ΔBaraB^LC1
ΔBaraB^LC1/CyO-GFP
H
iso ΔBaraB^LC1 lifespan 25°C
Percent survival
Time (days)
n^exp = 2-3
N = 412
iso BaraB^LC1 vs. CyO
HR = +0.58, p < .001
iso w^1118
n = 80
iso ΔBaraA
n = 80, HR = +0.26, p = .118
iso ΔBaraB^LC1
n = 93. HR = +2.56, p < .001
iso ΔBaraB^LC1/CyO
n = 120, HR = +1.98, p < .001
ATM^8
n = 39, HR = +4.32, p < .001
I
iso ΔBaraB^LC1 x Df(BaraB) viability
ΔBaraB^LC1/CyO ♀
ΔBaraB^LC1/CyO ♂
Df()/ΔBaraB^LC1 ♀
Df()/ΔBaraB^LC1 ♂
Df(9063) - 25°C
Df(9063) - 29°C
n = 114
n = 63
% of emergents

*Figure S5*


A
FlyAtlas2 male
FPKM
120
90
60
30
0
Head
Eye
Brain/CNS
Thoracic Abdominal Ganglion
Crop
Midgut
Hindgut
MTs
Fat Body
Salivary Gland
Testis
Accessory glands
Carcass
Rectal pad


FlyAtlas2 female
FPKM
120
90
60
30
0
Head
Eye
Brain/CNS
Thoracic Abdominal Ganglion
Crop
Midgut
Hindgut
MTs
Fat Body
Salivary Gland
Ovary
Mated spermatheca
Carcass
Rectal pad

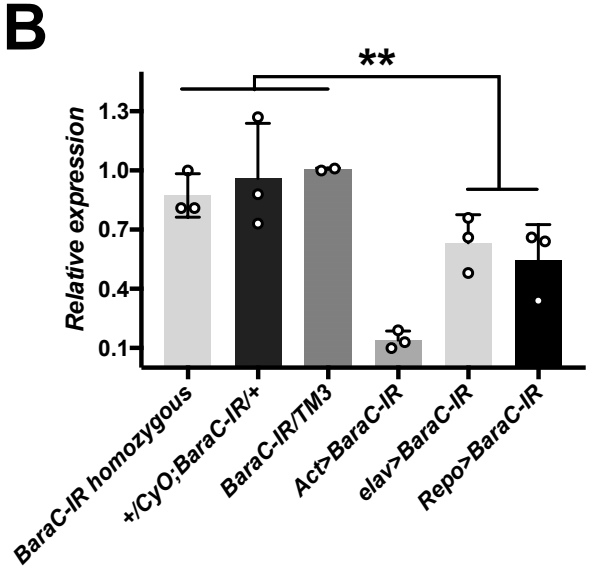

B
**
Relative expression
1.3
1.0
0.7
0.4
0.1
BaraC-IR homozygous
+/CyO;BaraC-IR/+
BaraC-IR/TM3
Act>BaraC-IR
elav>BaraC-IR
Repo>BaraC-IR

Figure S6

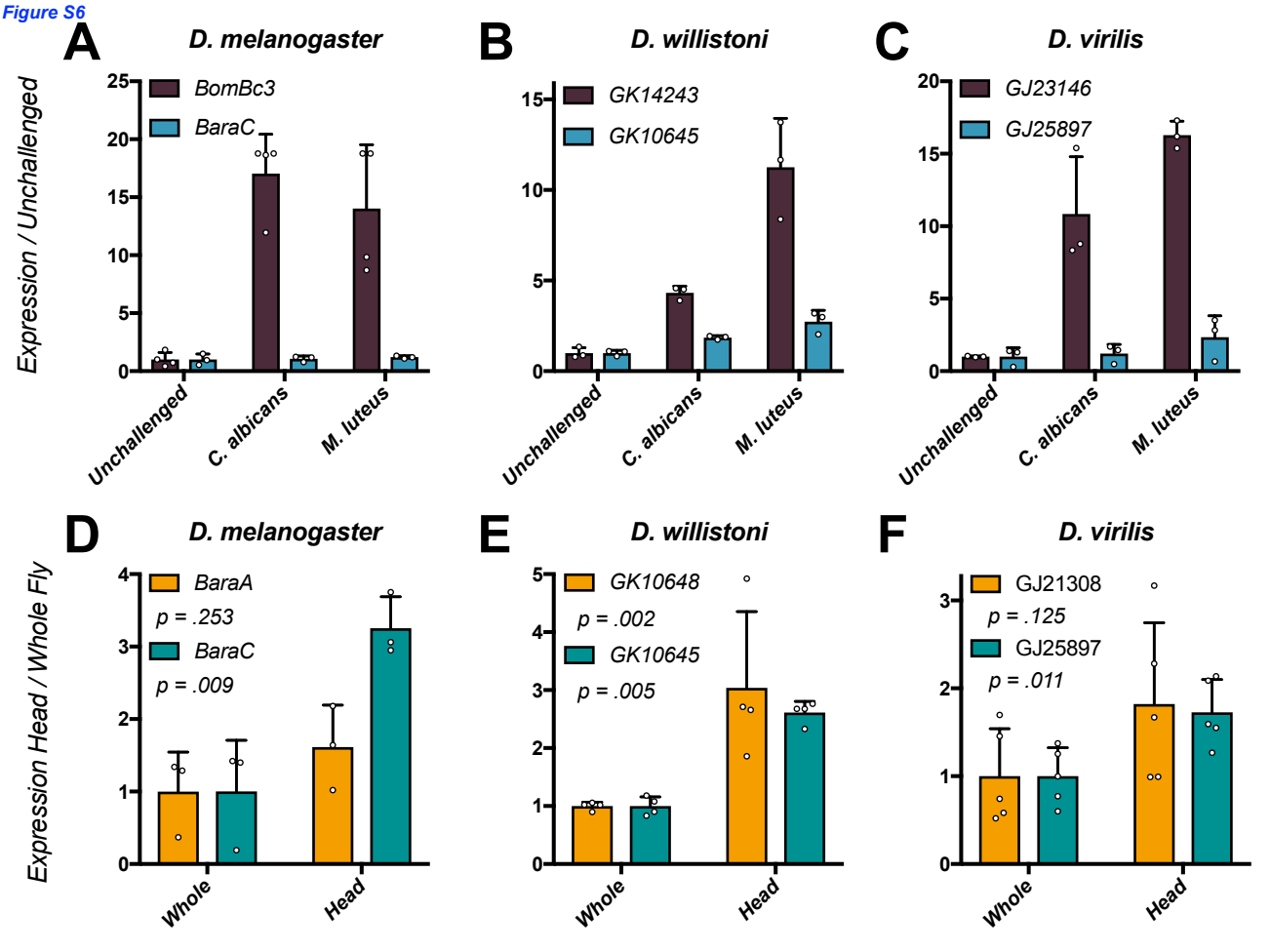

A
D. melanogaster
BomBc3
BaraC
Expression / Unchallenged
Unchallenged
C. albicans
M. luteus
B
D. willistoni
GK14243
GK10645
C
D. virilis
GJ23146
GJ25897
D
D. melanogaster
BaraA
p = .253
BaraC
p = .009
Expression Head / Whole Fly
Whole
Head
E
D. willistoni
GK10648
p = .002
GK10645
p = .005
F
D. virilis
GJ21308
p = .125
GJ25897
p = .011

*Figure S7*

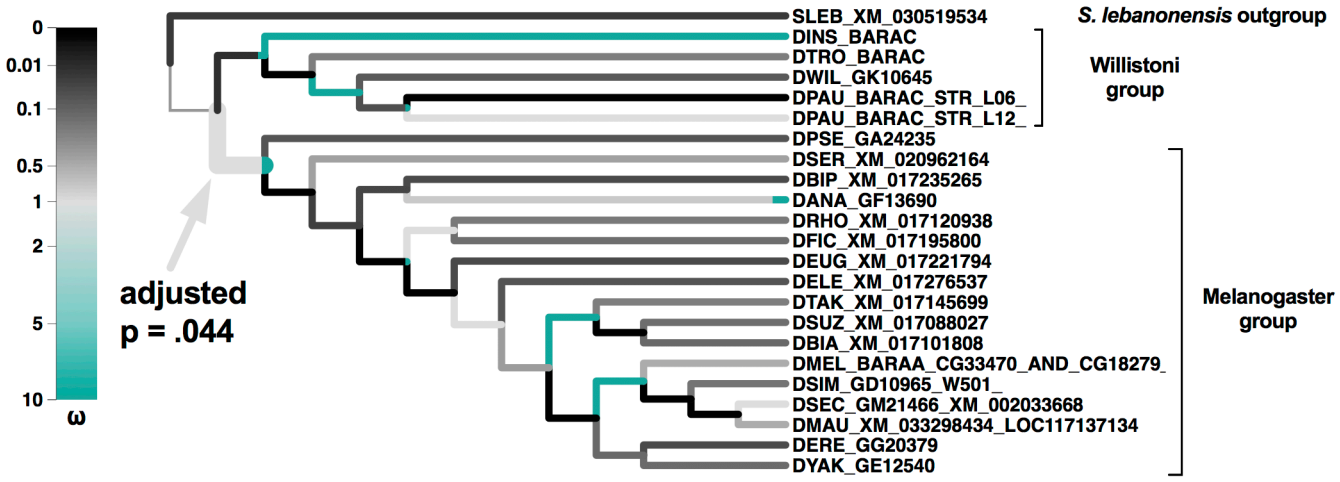
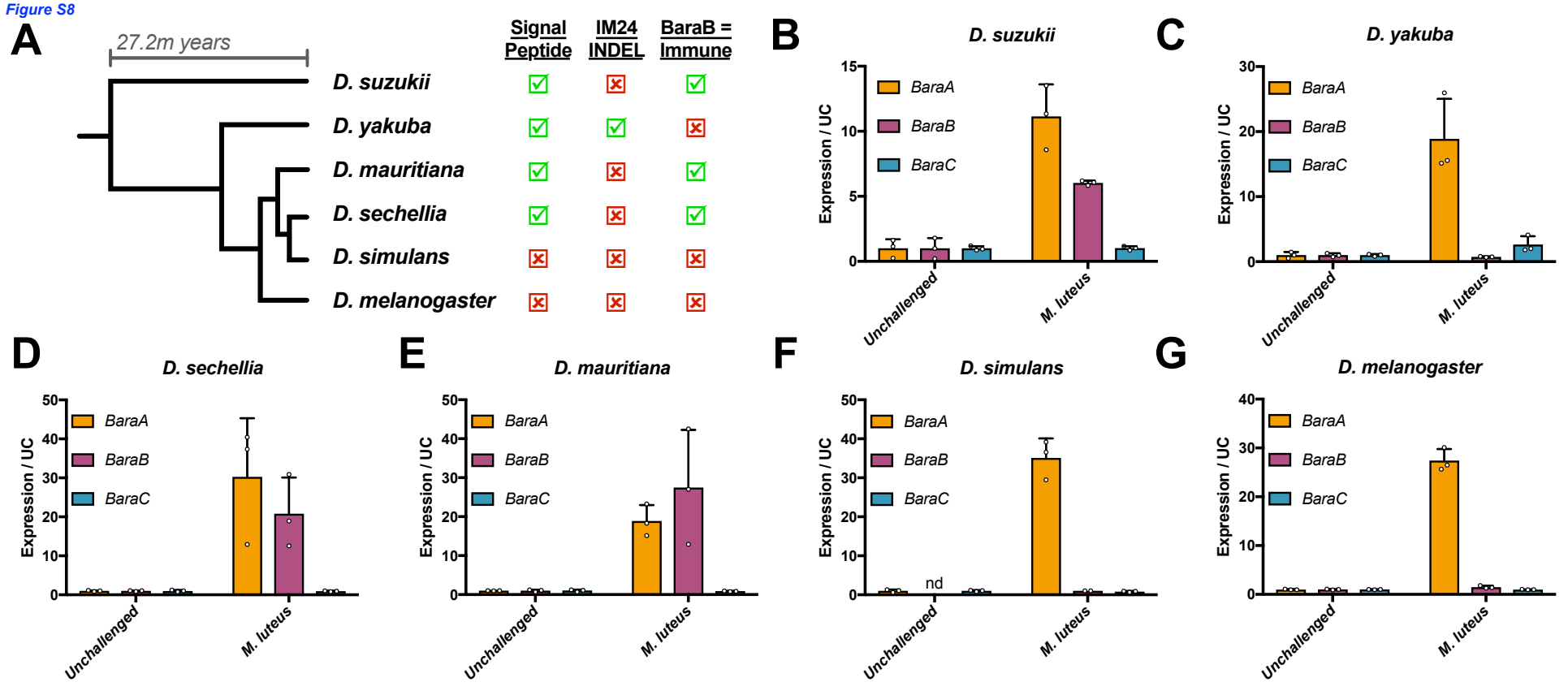
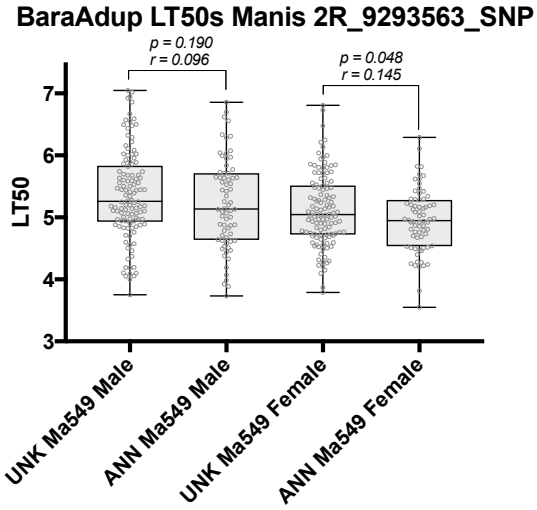

0
0.01
0.1
0.5
1
2
5
10
ω
adjusted
p = .044
SLEB_XM_030519534
DINS_BARAC
DTRO_BARAC
DWIL_GK10645
DPAU_BARAC_STR_L06_
DPAU_BARAC_STR_L12_
DPSE_GA24235
DSER_XM_020962164
DBIP_XM_017235265
DANA_GF13690
DRHO_XM_017120938
DFIC_XM_017195800
DEUG_XM_017221794
DELE_XM_017276537
DTAK_XM_017145699
DSUZ_XM_017088027
DBIA_XM_017101808
DMEL_BARAA_CG33470_AND_CG18279_
DSIM_GD10965_W501_
DSEC_GM21466_XM_002033668
DMAU_XM_033298434_LOC117137134
DERE_GG20379
DYAK_GE12540
S. lebanonensis outgroup
Willistoni group
Melanogaster group

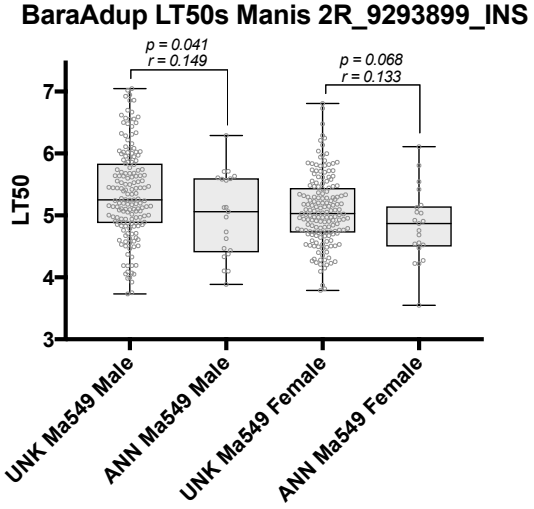
Figure S8


A
27.2m years
Signal Peptide
IM24 INDEL
BaraB = Immune
D. suzukii
D. yakuba
D. mauritiana
D. sechellia
D. simulans
D. melanogaster
B
D. suzukii
C
D. yakuba
D
D. sechellia
E
D. mauritiana
F
D. simulans
G
D. melanogaster
BaraA
BaraB
BaraC
Expression / UC
Unchallenged
M. luteus
nd

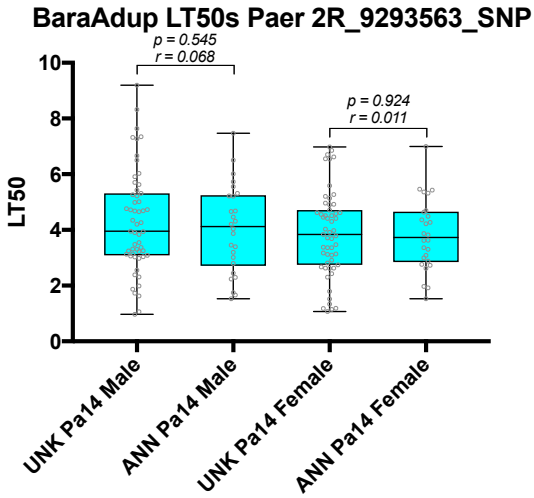
A

*Figure S9*


BaraAdup LT50s Manis 2R_9293563_SNP
p = 0.190
r = 0.096
p = 0.048
r = 0.145
LT50
7
6
5
4
3
UNK Ma549 Male
ANN Ma549 Male
UNK Ma549 Female
ANN Ma549 Female


BaraAdup LT50s Manis 2R_9293899_INS
p = 0.041
r = 0.149
p = 0.068
r = 0.133
LT50
7
6
5
4
3
UNK Ma549 Male
ANN Ma549 Male
UNK Ma549 Female
ANN Ma549 Female

B


BaraAdup LT50s Paer 2R_9293563_SNP
p = 0.545
r = 0.068
p = 0.924
r = 0.011
LT50
10
8
6
4
2
0
UNK Pa14 Male
ANN Pa14 Male
UNK Pa14 Female
ANN Pa14 Female

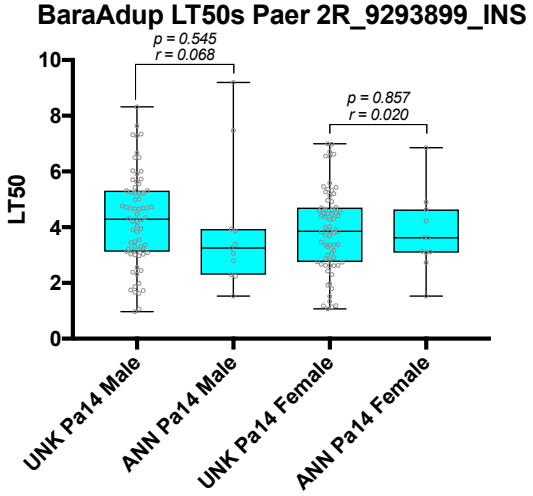

BaraAdup LT50s Paer 2R_9293899_INS
p = 0.545
r = 0.068
p = 0.857
r = 0.020
LT50
10
8
6
4
2
0
UNK Pa14 Male
ANN Pa14 Male
UNK Pa14 Female
ANN Pa14 Female

*Table S1*

| 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 S2*

| 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 | | |