Extensive duplication and convergent sequence evolution of an antimicrobial peptide 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 trans-species 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

41 evolution at the sequence level [8,9]. This rapid evolution is driven by positive
42 selection, and acts to either maintain genetic variation (balancing selection), or
43 differentiate AMP sequence (diversifying selection) [10–15]. The selective pressures
44 that drive these evolutionary outcomes are likely the consequence of host-pathogen
45 interactions. One striking example is the maintenance of two alleles of *Diptericin A*
46 in *Drosophila* that predict defence against infection by *Providencia rettgeri* bacteria
47 [16]. However AMPs and AMP-like genes in various species have also recently been
48 implicated in non-immune roles as regulators of sleep [17], memory [18], and
49 neurodegeneration [19–23]. These new contexts suggest that the evolutionary
50 forces acting on AMP genes may not be driven strictly by trade-offs in host defence,
51 but rather by conflicts between roles in immunity and other non-immune functions.
52
53 We recently described a novel antifungal peptide gene of *Drosophila*
54 *melanogaster* that we named *Baramicin A* (*BaraA*) [24]. A unique aspect of *BaraA* is
55 its precursor protein structure, which encodes a polypeptide cleaved into multiple
56 mature products by interspersed furin cleavage sites. The use of furin cleavage sites
57 to produce two mature peptides from a single polypeptide precursor is common in
58 animal AMP genes [1,3]. However, *BaraA* represents an exceptional case as multiple
59 tandem repeat peptides are produced from the translation of a single coding
60 sequence, effectively resembling a "protein-based operon." The immature precursor
61 protein of *D. melanogaster BaraA* encodes three types of domains: an IM24 domain,
62 three tandem repeats of IM10-like domains, and an IM22 domain (**Fig.** 1A). *BaraA*
63 mutants are susceptible to infection by fungi, and preliminary *in vitro* experiments
64 suggest the *BaraA* IM10-like peptides have antifungal activity [24,25]. The
65 remaining Baramicin domains encoding IM22 and IM24 remain uncharacterized.
66 Curiously, *BaraA* deficient flies also display an erect wing behavioural phenotype
67 upon infection [24], suggesting that *BaraA* could have additional roles beyond
68 immunity.

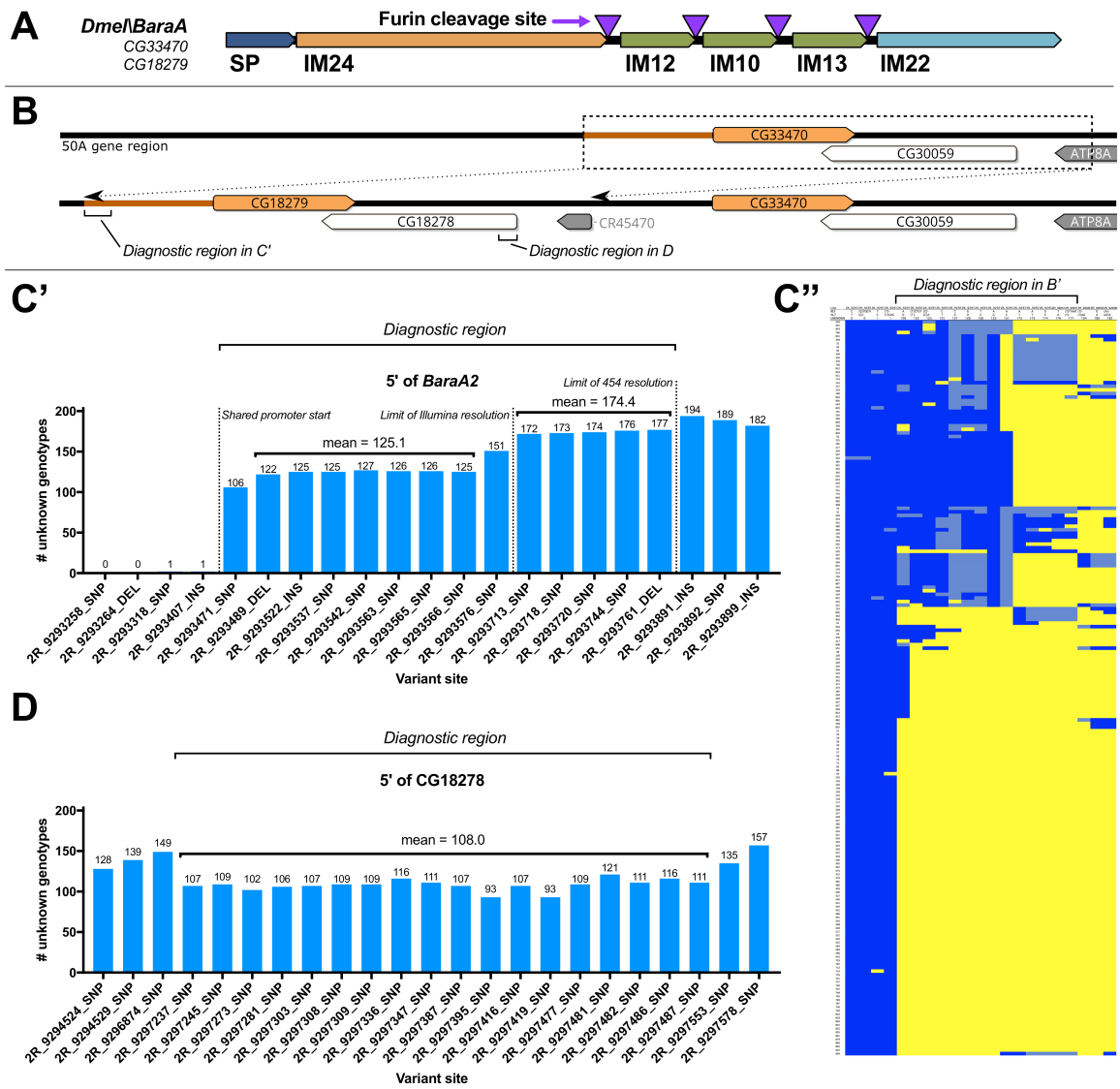


Figure 1: The *BaraA* duplication is not fixed in wild flies. A) Gene structure of *BaraA*, including its three key domain types: IM24 (orange), IM10-like (green), and IM22 (teal). SP = signal peptide. Purple triangles indicate furin cleavage sites. **B)** Depiction of the *BaraA* duplication. Diagnostic regions used in (C) and (D) are annotated. **C)** The number of unknown variant calls rises immediately after the start of the identical promoter region shared by *BaraA1* and *BaraA2,* stabilizing at 61% (~125/205) of the DGRP lines (**C'**). Upon reaching the limit of Illumina resolution, a second spike in unknown calls stabilizes at ~174/205 lines. **C''** shows a heatmap of DGRP strains where mapped variants are listed in blue (light blue = alternate allele), while yellow indicates unknown variant calls. **D)** The 5' region of the duplicated sulfatase *CG18278* shows a stable unknown variant call frequency at ~53% (~108/205) of the DGRP lines.

81 In this study, we characterize a duplication of the *BaraA* locus, and also
82 describe two additional *Baramicin* genes encoded in the genome of *D. melanogaster*:
83 *BaraB* and *BaraC*. Surprisingly, only *BaraA* is immune-induced, while *BaraB* and
84 *BaraC* are specifically enriched in the nervous system. Moreover, independent
85 *Baramicin* genes in other *Drosophila* species have convergently evolved to resemble
86 *BaraC*, both in expression and gene structure. We resolved the evolutionary history
87 of *Baramicin* across species, revealing properties that are unique to each of the
88 *Baramicin* domains. By doing so, we provide a case study for how an AMP-encoding
89 gene can play dual roles in both immune and non-immune processes.
90

91 # Results and Discussion

92

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

94

95 The *D. melanogaster BaraA* gene encodes three key types of mature peptide
96 (**Fig.** 1A): one IM24 peptide, three IM10-like peptides, and one IM22 peptide that is
97 structurally related to the IM10-like peptides [24]. We recently found that the
98 immune induced *BaraA* gene contributes to the *Drosophila* defence against
99 filamentous fungi [24]. During this investigation, we found that while the *BaraA*
100 locus in the *D. melanogaster* R6 reference genome encodes two copies of *BaraA*
101 (*CG33470* and *CG18279*, *BaraA1* and *BaraA2* respectively), other fly strains carry
102 only one *BaraA* gene (**Fig.** 1B). Several lines of evidence indicate that this
103 duplication is a very recent evolutionary event. First: in the Dmel_R6 genome,
104 3,380bp of contiguous sequence in the *BaraA1* and *BaraA2* loci are 100% identical,
105 including the regions upstream (1,140bp) and downstream (892bp) of the 5' and 3'
106 UTRs. Second: this duplication is not found in *D. melanogaster* sister species such as
107 *Drosophila simulans* and *Drosophila sechellia*. And third: this duplication is not fixed
108 in wild-type flies (**Fig.** 1).
109

110 As the reference genome encodes two copies of *BaraA*, we reasoned that
111 mapping the genome of a strain with only one copy to the reference genome (two-
112 copies) would result in multiple genetic variant calls of "unknown" in *BaraA* as those

113 reads would map to only one of the two *BaraA* gene regions in the reference
114 genome. Using this logic, we searched for the *BaraA* duplication in the Drosophila
115 Genetic Reference Panel (DGRP), a panel of wild-type flies collected in the USA
116 whose genomes were assembled using a mix of Illumina and 454 sequencing. These
117 two assembly methods differed in read length: ~300bp (Illumina) and 450-550bp
118 (454 sequencing) [26]. As a consequence, we expected to see regions of unknown
119 variant calling that were ~300bp long, and then an additional ~150-250 bp
120 specifically for genomes that used 454 sequencing. As predicted, a sharp spike in
121 unknown variant calls occurs upstream of *BaraA2*, which begins at the variant site
122 2R_9293471_INS (**Fig.** 1C). This signal remains relatively stable at ~61% over the
123 next ~300bp, which includes seven genetic variant sites. Following this region, there
124 is a second spike in unknown variant calls, which again settles at ~61% frequency
125 spanning the next 178bp. An alternate estimate of the duplication frequency can be
126 made using sequence in the 5' gene region of *CG18728*, which is also duplicated in
127 lines carrying two copies of *BaraA* (**Fig.** 1B). In this region, there is a relatively
128 steady number of unknown variant calls with a frequency of ~53% (**Fig.** 1D). This
129 suggests that approximately 53-61% of DGRP strains have only one *BaraA* gene. We
130 validated *BaraA* copy number in a subset of DGRP strains using PCR. This
131 independent validation supports the use of both loci in **Fig.** 1C-D in interpreting the
132 DGRP sequencing data as neither region perfectly predicts copy number, but a
133 decision tree using information from both loci correctly predicted *BaraA* copy
134 number in 7/7 cases (**Table** S1). Other regions of the *BaraA* locus show less
135 consistent variant calling patterns (**Fig.** S1A). Lastly, we found that *BaraA* copy
136 number varies in common *D. melanogaster* lab strains (**Table** S1).
137
138 This suggests that only 39-47% of DGRP strains carry the *BaraA* duplication
139 found in the Dmel_R6 reference genome. Due to the difficulty in mapping highly
140 similar duplicate genes, it is not clear whether the 100% similarity of *BaraA1* and
141 *BaraA2* found in the reference genome is true of all DGRP strains. Indeed, there are
142 variant calls in the *BaraA2* promoter region, but no variant calls in the *BaraA1*
143 promoter region (**Fig.** S1B), which may be an artefact caused by variants in *BaraA1*

144 erroneously mapping to *BaraA2*. Regardless, this analysis clearly shows that the
145 *BaraA* duplication is not fixed in *D. melanogaster*, and it is common for wild flies and
146 lab strains to carry either one or two copies of *BaraA*.
147
148 ***Baramicin A is an ancestral immune effector***
149
150 In *D. melanogaster, BaraA* is regulated by the Toll immune signalling pathway
151 [24,25]. Using BLAST, we further recovered *BaraA*-like genes that encode all three
152 kinds of Baramicin peptides (IM24, IM10-like, and IM22) across the genus
153 *Drosophila* and in the outgroup *Scaptodrosophila lebanonensis*. We next performed
154 infection experiments in different species to confirm that their *BaraA*-like genes
155 were immune-inducible. We infected the phylogenetically diverse species *D.*
156 *melanogaster*, *D. pseudoobscura*, *D. willistoni*, *D. virilis*, and *D. neotesteacea* with
157 *Micrococcus luteus* and *Candida albicans*, and included orthologues of the *Drosophila*
158 Toll effector *Dmel\BomBc3* as an immune-induced control (**Fig.** 2A). In all five
159 species, *BaraA*-like genes were immune-induced, confirming that the ancestral
160 *Baramicin* was an immune-induced gene (**Fig.** 2B-F). We therefore demonstrate that
161 *BaraA*-like genes are consistently immune induced in all species tested.
162
163 ***The BaraA antifungal IM10*-like *peptides display trans-species polymorphisms***
164
165 Trans-species polymorphisms are variable residues that appear repeatedly
166 in orthologous genes of different species. Such polymorphisms imply balancing
167 selection that favours one allele or the other in a context-dependent manner, and
168 are commonly found in Dipteran AMPs [4,12,27]. The predominant peptide
169 products of *D. melanogaster BaraA* are the three IM10-like peptides IM10, IM12, and
170 IM13, which have been shown to inhibit fungal growth in vitro [24]. We next
171 investigated the degree of conservation in *Drosophila* species for these 23-residue
172 peptides. Strikingly, we found that IM10-like peptides across the *Drosophila* genus
173 have a recurrent polymorphic site in an otherwise conserved ^{6}RPXR^{9} motif, at
174 residue 8 (G/D) (**Fig.** 3A). We also note three other polymorphic residues at site 3

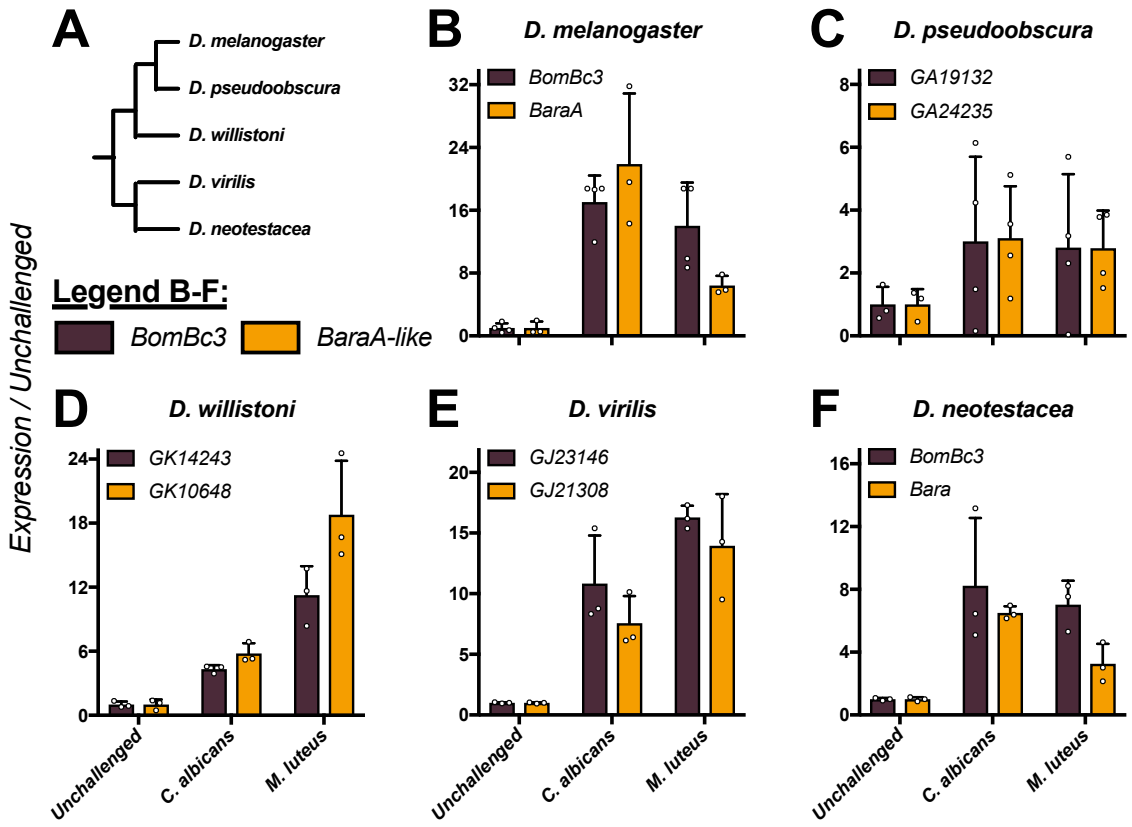


Figure 2: The ancestral *BaraA* gene was immune-induced. A) Cladogram of species used in B-F. **B-F)** Expression of *BomBc3* controls (brown) or *BaraA-like* genes (orange) in diverse *Drosophila* species upon infection. In all cases, both *BomBc3* and *BaraA-like* genes are induced upon infection by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

(H/V/Q), site 4 (V/I), and site 5 (A/E/S) that are broadly observed across the genus. Furthermore, site 9 universally encodes an Arginine, except in flies of the Immigrans-Tripunctata radiation where a Histidine is observed instead (**Fig. 3A**). These polymorphisms impact the hydrophobicity and isoelectric point of these peptides (**Fig.** 3B). Notably, the IM22 peptide encoded at the C-terminus of *BaraA* also encodes an IM10-like motif, except the residue 8 G/D site instead encodes both D and G in tandem (i.e. ^{27}RPDGR^{31}), which is conserved across *Drosophila* species (**Fig.** S2).

The number of IM10-like repeats within the *BaraA*-like gene structure is variable across species (e.g. in **Fig.** 5C), making it difficult to assess the synteny of

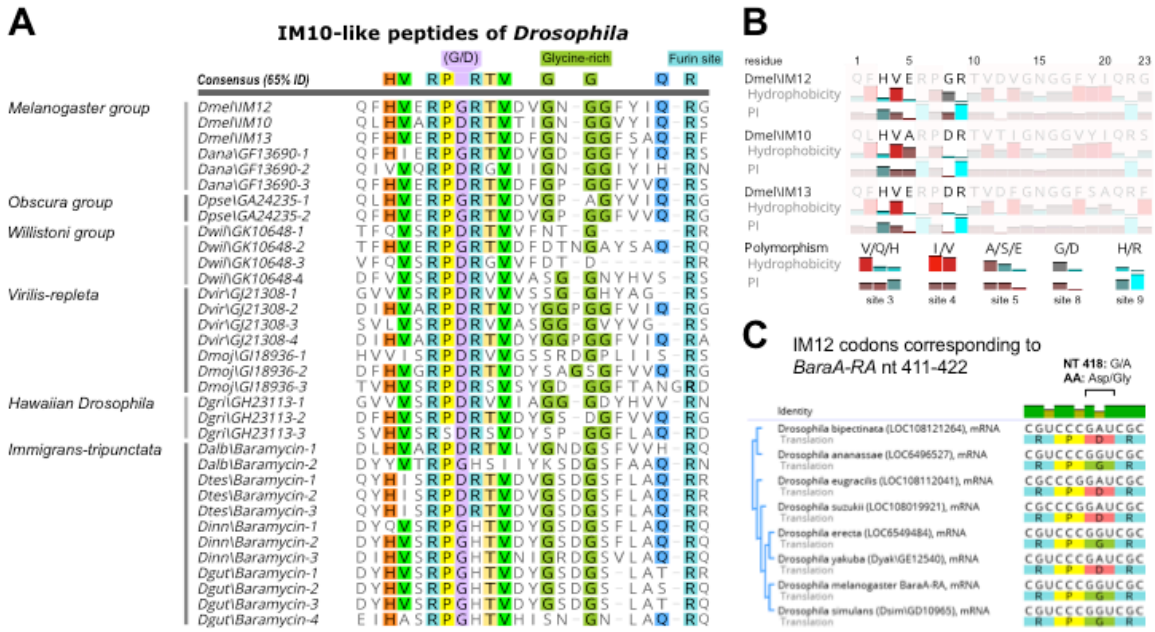


Figure 3: Convergent evolution maintains a polymorphism in Baramicin IM10-like peptides. A) Alignment of IM10-like peptides from diverse *Drosophila* species. The G/D polymorphic site is highlighted in lavender (residue 8). Broadly conserved polymorphisms are also seen at sites 3 (H/V/Q), 4 (V/I), 5 (A/E/S), and 9 (R/H). **B)** The IM10-like polymorphic residues confer different hydrophobicity and isoelectric points (PI). The three *D. melanogaster* IM10-like peptides are given as examples. Bar height indicates greater hydrophobicity/PI. **C)** The syntenic codon of IM12 in Melanogaster group flies encodes either a D or G residue owing to convergent G-A nucleotide transitions. The D residue is far more common in additional outgroups (**Fig.** S3), suggesting the ancestral state encoded D, and G was derived at least twice.

codons in distantly related species. However among closely related species of the Melanogaster group, we nevertheless recovered G/D trans-species polymorphisms at syntenic codons. Focusing on *BaraA*-like genes in the Melanogaster group, the first IM10-like peptide repeat is IM12. In the "RPXR" motif of IM12, RPGR is found in *D. ananassae, D. erecta, D. simulans,* and *D. melanogaster.* Meanwhile an RPDR motif is found in *D. bipectinata, D. eugracilis, D. suzukii,* and *D. yakuba* (**Fig.** 3C). At the RNA level these polymorphic sites are encoded as either GGU (G) or GAU (D). Thus there is a split between G/D in IM12 in two sets of sister lineages: *D. ananassae* vs. *D. bipectinata,* and *D. erecta* vs. *D. yakuba*. Furthermore, while *D. melanogaster* encodes RPGR, *D. eugracilis* and *D. suzukii* outgroups instead encode RPDR. These examples demonstrate that the G/D trans-species polymorphism of IM10-like peptides arose independently in different lineages, and thus derive from recurrent mutations found

216 in the reference genomes of given species. Overall, the frequency of either residue
217 across all IM10-like peptides in the Melanogaster species group is: RPDR (~85%) or
218 RPGR (~15%) (**Fig.** S3). However, the G/D frequency differs drastically in other
219 *Drosophila* lineages (**Fig.** 3A). As such, the IM10-like G/D frequency in Melanogaster
220 group flies should not be generalized to how advantageous one allele might be in
221 another species.
222
223 Thus trans-species polymorphisms in IM10-like peptides are found across
224 the genus *Drosophila*. In particular, the residue 8 G/D polymorphism is unlikely to
225 be a product of simple inheritance, as this would require multiple speciation events
226 to maintain both alleles of the polymorphism. Moreover G/D frequency differs in a
227 lineage-specific manner, suggesting selection acts to promote one allele or the other
228 depending on life history. These IM10-like polymorphisms confer distinct
229 biochemical properties (i.e. hydrophobicity, isoelectric point), which are likely to
230 modify antimicrobial activity. We conclude that the IM10-like peptides encoded by
231 *BaraA* evolve rapidly in lineage-specific fashions.
232
233 ***Identification of BaraB and BaraC, two non-immune Baramicin paralogs in D.***
234 ***melanogaster***
235
236 While investigating *Baramicin* evolution across species, we also recovered
237 two paralogous *Baramicin* genes in *D. melanogaster* through reciprocal BLAST
238 searches: *CG13749* and *CG30285*, which we name *BaraB* and *BaraC* respectively. The
239 three *Baramicin* genes are scattered on the right arm of chromosome II at
240 cytological positions 44F9(*BaraB*), 50A5 (*BaraA*), and 57F8 (*BaraC*). While *BaraA*-
241 like genes are conserved throughout the genus *Drosophila*, *BaraB* is conserved only
242 in Melanogaster group flies, and *BaraC* is found only in Melanogaster and Obscura
243 group flies, indicating that both paralogs stem from duplication events of a *BaraA*-
244 like ancestor (see **Fig.** 5 for a summary of the *Baramicin* loci). In the *D. melanogaster*
245 reference genome, these paralogous *Baramicin* genes are united by the presence of
246 the IM24 domain (**Fig.** 4A). In the case of *BaraB*, we additionally recovered a

247 frameshift mutation causing a premature stop segregating in the DGRP, leading to
248 the loss of IM13 and IM22 relative to the *BaraA* gene structure (**Fig.** 4A); this
249 truncation is present in the Dmel_R6 reference genome, but many DGRP strains
250 encode a CDS including either a standard (e.g. DGRP38) or extended (e.g. DGRP101)
251 IM22 domain. Moreover, in contrast to BaraA, the initial IM10-like peptide of BaraB
252 no longer follows a furin cleavage site (**Fig.** 4A), and now also encodes a serine
253 (RSXR) in its IM10-like motif instead of the proline (RPXR) of *BaraA*-like genes. Each
254 of these mutations prevents the secretion of classical IM10-like and IM22 peptides.
255 Finally, BaraC encodes only IM24 tailed by a transmembrane domain at the C
256 terminus (TMHMM v2.0 [28]), and thus lacks both the IM10-like peptides and IM22
257 (**Fig.** 4A).
258
259 *BaraA* is strongly induced following microbial challenge, being
260 predominantly regulated by the Toll pathway with minor input from the Immune
261 Deficiency (Imd) pathway [24]. We assayed the expression of *BaraB* and *BaraC* in
262 wild-type flies, and also flies with defective Toll (spz^{rm7}) or Imd (Rel^{E20}) signalling to
263 see if their basal expression relied on these pathways. Surprisingly, neither gene
264 was induced upon infection regardless of microbial challenge (**Fig.** 4B and **Fig.** S4).
265 *BaraC* levels were reduced in spz^{rm7} mutants regardless of treatment (cumulative
266 data in **Fig.** S4C, p = .005), suggesting that *BaraC* basal expression might be affected
267 by Toll signalling. We next monitored the expression of the three *Baramicin* genes
268 over the course of development from egg to adult. We found that expression of all
269 genes increased over development and reached their highest level in young adults
270 (**Fig.** 4C). Of note, *BaraB* expression approached the lower limit of our assay's
271 detection sensitivity at early life stages. However *BaraB* was robustly detected
272 beginning at the pupal stage, indicating it is expressed during metamorphosis. *BaraC*
273 expression also increased markedly between the S3 larval stage and pupal stage.

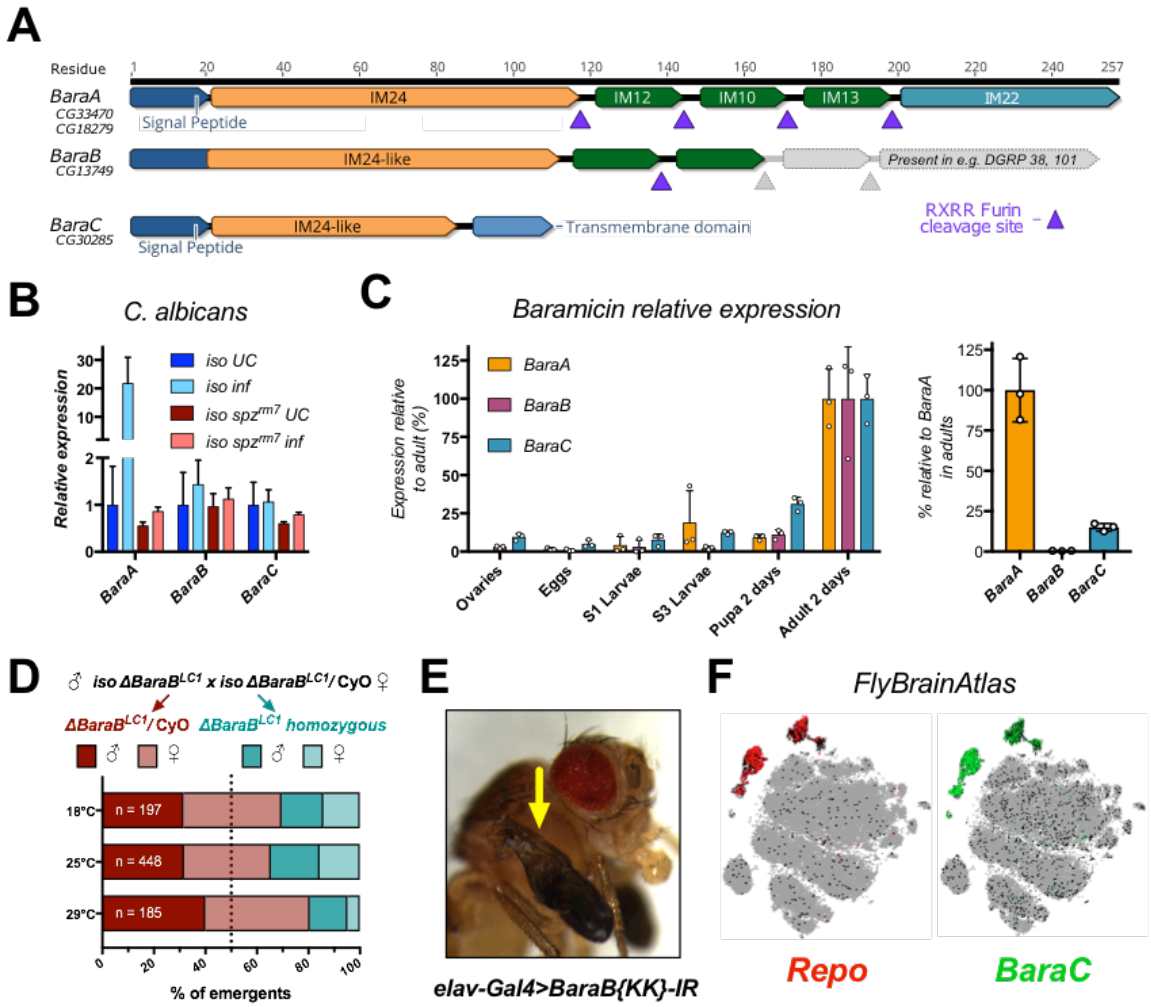


Figure 4: The *D. melanogaster Baramicin* gene family. A) Two *Baramicin* paralogs in addition to *BaraA* are present in the Dmel_R6 reference genome, united by the presence of an IM24 peptide. Relative to *BaraA*, two frameshift mutations in *BaraB* are segregating in the DGRP. One results in the same mature protein annotated in the Dmel_R6 reference genome (coloured domains), while the other elongates the IM22 peptide (e.g. *DGRP101*). Most strains carry a *BaraB* gene of similar length to *BaraA* (e.g. DGRP38). Additionally, the *BaraB* gene does not encode a functional signal peptide. **B)** Only *BaraA* is immune-induced. *BaraB* and *BaraC* do not respond to infection, though basal *BaraC* expression depends on Toll signalling (**Fig.** S4C). **C)** Time series of whole animal *Baramicin* expression over the course of development. Expression values are normalized within each gene, with expression in the adult set to 100% (left panel). For context, normalizing each gene to *BaraA* in adults shows that *BaraC* and especially *BaraB* expression is much lower at a whole fly level (right panel). **D)** The $\Delta BaraB^{LC1}$ mutation is partial lethal, an effect that is exaggerated by rearing flies at 29°C. **E)** Example of a nubbin-like wing, which is phenocopied by *BaraB* gene silencing using the neural driver *elav-Gal4*. **F)** The *BaraC* gene almost perfectly matches the expression pattern of glia specific *Repo*-expressing cells in single cell RNA sequencing of the adult fly brain [29].

293 Here we reveal that *BaraA* is part of a larger gene family. While the *BaraA*
294 gene was first described as an immune effector, the two more phylogenetically-
295 restricted *Baramicin* paralogs *BaraB* and *BaraC* are not induced by infection in *D.*
296 *melanogaster*. Both *BaraB* and *BaraC* first see increased expression during pupation,
297 and are ultimately expressed at their highest levels in adults.
298
299 ***The Dmel\BaraB gene is required in the nervous system over the course of***
300 ***development***
301
302 A simple interpretation of the truncated gene structure and low levels of
303 *BaraB* expression is that this gene is undergoing pseudogenization. To explore
304 *BaraB* function, we used two null mutations for *BaraB* (*ΔBaraB*^{LC1} *and ΔBaraB*^{LC4},
305 generously gifted by S.A. Wasserman). These mutations were made using a CRISPR
306 double gRNA approach to replace the *BaraB* locus with sequence from the pHD-
307 DsRed vector. We further introgressed the *ΔBaraB*^{LC1} mutation into the DrosDel
308 isogenic background (referred to as *iso*) as described in Ferreira et al. [30]. We
309 assessed *BaraB* mutant viability using crosses between *ΔBaraB/CyO* heterozygous
310 females and *ΔBaraB* homozygous males. Under mendelian ratios of inheritance, this
311 cross is expected to result in a 1:1 ratio of *ΔBaraB*/CyO and *ΔBaraB* homozygous
312 progeny. However, *BaraB* deficient flies exhibited reduced viability regardless of
313 mutation or genetic background (**Fig.** 4D and S5A). This reduced viability was
314 exacerbated by rearing at 29°C, reaching a lethality rate over 50%. Using a *CyO-GFP*
315 reporter to track genotype in larvae revealed that the lethal phase occurs primarily
316 in the late larval and pupal stages (**Fig.** S5B-E), agreeing with a role for *BaraB* in
317 pupae previously suggested by increased expression at this stage (**Fig.** 4C). Many
318 *ΔBaraB*^{LC1} flies never emerge from their pupal cases (**Fig.** S5B). Dissections
319 confirmed that some adult flies fully developed inside the puparium even to the
320 point of abdominal pigmentation maturing, indicating they did not perish but rather
321 failed to eclose in spite of developing to adulthood. Beyond overt lethality, many
322 emerging homozygotes had locomotory issues and were frequently found
323 floundering in the food medium, even when given tissue paper to facilitate climbing.

324 Some emergents also exhibited a nubbin-like wing phenotype (FlyBase:
325 FBrf0220532 and e.g. **Fig.** 4E), where the wings were stuck in a shrivelled state for
326 the remainder of the fly's lifespan. However, a plurality of homozygotes successfully
327 emerged, and unlike their siblings, had no immediate morphological or locomotory
328 defects. The lifespan of otherwise normal *iso ΔBaraB*^{LC1} adults is significantly
329 shorter compared to wild-type flies and even *iso ΔBaraB*^{LC1}*/CyO* siblings (**Fig.** S5F).
330 We confirmed these developmental effects using ubiquitous gene silencing with
331 *Actin5C-Gal4 (Act-Gal4)* to drive two *BaraB* RNAi constructs (*TRiP-IR* and *KK-IR*).
332 Both constructs resulted in significant lethality and occurrence of nubbin-like wings
333 (**Table** S2). Additionally, we crossed *iso ΔBaraB*^{LC1} to a genomic deficiency line
334 (BDSC9063) at both 25°C and 29°C. These crosses resulted in significantly reduced
335 numbers of eclosing *ΔBaraB*^{LC1}*/Df(9063)* flies at 25°C ($n = 114$, χ^2 $p < .001$) and 29°C
336 ($n = 63$, χ^2 $p < .001$) (**Fig.** S5G)
337
338 These data demonstrate a significant consequence of *BaraB* mutation on
339 fitness. While whole-fly *BaraB* expression is low, these results suggest that *BaraB* is
340 not pseudogenized, and instead performs an integral function during development.
341 The fact that there is a bimodal outcome in adults (either severe defects or generally
342 healthy) suggests *BaraB* is involved in passing some checkpoint during larval/pupal
343 development. Flies lacking *BaraB* are more likely to fail at this developmental
344 checkpoint, resulting in either lethality or developmental defects.
345
346 ***The Baramicin paralogs BaraB and BaraC are expressed in the nervous system***
347
348 Given that whole-fly *BaraB* depletion produced lethality and nubbin-like
349 wings, we next sought to determine in which tissue(s) *BaraB* is required. *BaraB*
350 mutant locomotory defects led us to suspect an involvement in the nervous system.
351 We silenced *BaraB* in the nervous system at 25°C or at 29°C for greater efficiency,
352 using a pan-neural *elav-Gal4* driver, and for *BaraB-IR* used both the *TRiP-IR* and *KK-*
353 *IR* lines. We additionally combined this approach with *UAS-Dicer2* (*Dcr2*) to further
354 strengthen gene silencing [31]. We used the *TRiP-IR* line with a *CyO* balancer

355 chromosome (*elav-Gal4* * *TRiP-IR/CyO*), allowing the same simple mendelian
356 approach to assess lethality as used previously. Both *elav>TRiP-IR* and *elav>Dcr2,*
357 *TRiP-IR* caused partial lethality and occasional nubbin-like wings (χ^2 p < .02, **Table**
358 S2). For *KK-IR*, we instead crossed homozygous flies and so we did not assess
359 lethality using mendelian inheritance. However, no adults emerged when *elav>Dcr2,*
360 *KK-IR* flies were reared at 29°C. Rare emergents (n^{total} = 11 after three experiments)
361 occurred at 25°C, all of which bore nubbin-like wings. Using *elav-Gal4* at 29°C
362 without *Dcr2*, we observed greater numbers of emerging adults, but 100% of flies
363 had nubbin-like wings (**Fig.** 4E, **Table** S2). Finally, *elav>KK-IR* flies at 25°C suffered
364 both partial lethality and nubbin-like wings, but normal-winged flies began
365 emerging (χ^2 p < .001, **Table** S2). This analysis indicates that *BaraB* is expressed in
366 the nervous system, and this expression readily explains both the lethality and
367 nubbin-like wing phenotypes. Moreover, we observed a consistent spectrum of
368 developmental defects using *elav-Gal4>BaraB-IR* wherein strength of gene silencing
369 correlates with severity of defects.
370
371 *BaraB* is not the only *Baramicin* gene now specialized for expression in the
372 nervous system. Tissue-specific transcriptomic data indicate that *BaraC* is expressed
373 in various neural tissues including the eye, brain, and the thoracic abdominal
374 ganglion (**Fig.** S6A) [32]. However *BaraC* is also highly enriched in the hindgut and
375 rectal pads, pointing to a complex expression pattern [32,33]. We next searched
376 FlyBrainAtlas (http://scope.aertslab.org, [29]) to narrow down which neural
377 subtypes *BaraC* was expressed in. Strikingly, *BaraC* was expressed in all glial cell
378 types, fully overlapping the glia marker *Repo* (**Fig.** 4F). To confirm these
379 observations, we compared the effects of *BaraC* RNA silencing (*BaraC-IR*) using *Act-*
380 *Gal4* (ubiquitous), *elav-Gal4* (neural) and *Repo-Gal4* (glia) drivers on *BaraC*
381 expression. *Act-Gal4* reduced *BaraC* expression to just ~14% that of control flies
382 (**Fig.** S6B). By comparison *elav-Gal4* reduced *BaraC* expression to ~63% that of
383 controls, while *Repo-Gal4* led to *BaraC* levels only 57% that of controls (overall
384 controls vs. neural/glia-IR, p = .002). Collectively, our results support the notion

385 that *BaraC* is expressed in the nervous system, and are consistent with *BaraC*
386 expression being most localized to glial cells.
387
388 ***Extensive genomic turnover of the Baramicin gene family***
389
390 The results thus far show that *BaraA*-like genes are consistently immune-
391 induced in all *Drosophila* species (**Fig.** 2), and the *BaraA* antifungal IM10-like
392 peptides evolve rapidly (**Fig.** 3). However, the two phylogenetically-restricted
393 paralogs *Dmel\BaraB* and *Dmel\BaraC* are not immune-induced, and are truncated
394 in a fashion that deletes some or all of the IM10-like peptides (**Fig.** 4A). These two
395 *Baramicins* are now enriched in the nervous system (**Fig.** 4E-F). In the case of *BaraB*,
396 a role in the nervous system is evidenced by severe defects recapitulated using pan-
397 neural RNA silencing. In the case of *BaraC*, this is evidenced by a clear overlap with
398 *Repo*-expressing cells.
399
400 The three *D. melanogaster Baramicin* genes are united by their IM24 domain,
401 and are scattered across different loci along chromosome 2R, suggesting a series of
402 complex genetic events. To determine the ancestry of each *D. melanogaster*
403 *Baramicin* gene, we traced their evolutionary history by analyzing genomic synteny.
404 This allowed us to determine ancestry beyond sequence similarity-based
405 approaches. Using this strategy, we traced a series of genomic duplications,
406 inversions, and locus translocation events that ultimately resulted in the three
407 *Baramicin* loci of *D. melanogaster*. Our analysis reveals that these three loci
408 ultimately stem from a single-locus ancestor encoding only one *Baramicin* gene that
409 resembled *Dmel\BaraA* (**Fig.** 5A). This is evidenced by the presence of only a single
410 *BaraA*-like gene in the outgroup *S. lebanonensis*, and also in multiple lineages of the
411 subgenus Drosophila (**Fig.** 5B). Indeed, the general *BaraA* gene structure encoding
412 IM24, tandem repeats of IM10-like peptides, and IM22 is conserved in *S.*
413 *lebanonensis* and all *Drosophila* species (**Fig.** 5C). On the other hand, the *Dmel\BaraC*
414 gene is an ancient duplication restricted to the subgenus Sophophora, and
415 *Dmel\BaraB* resulted from a more recent duplication found only in the
416 Melanogaster group (**Fig.** 5B).

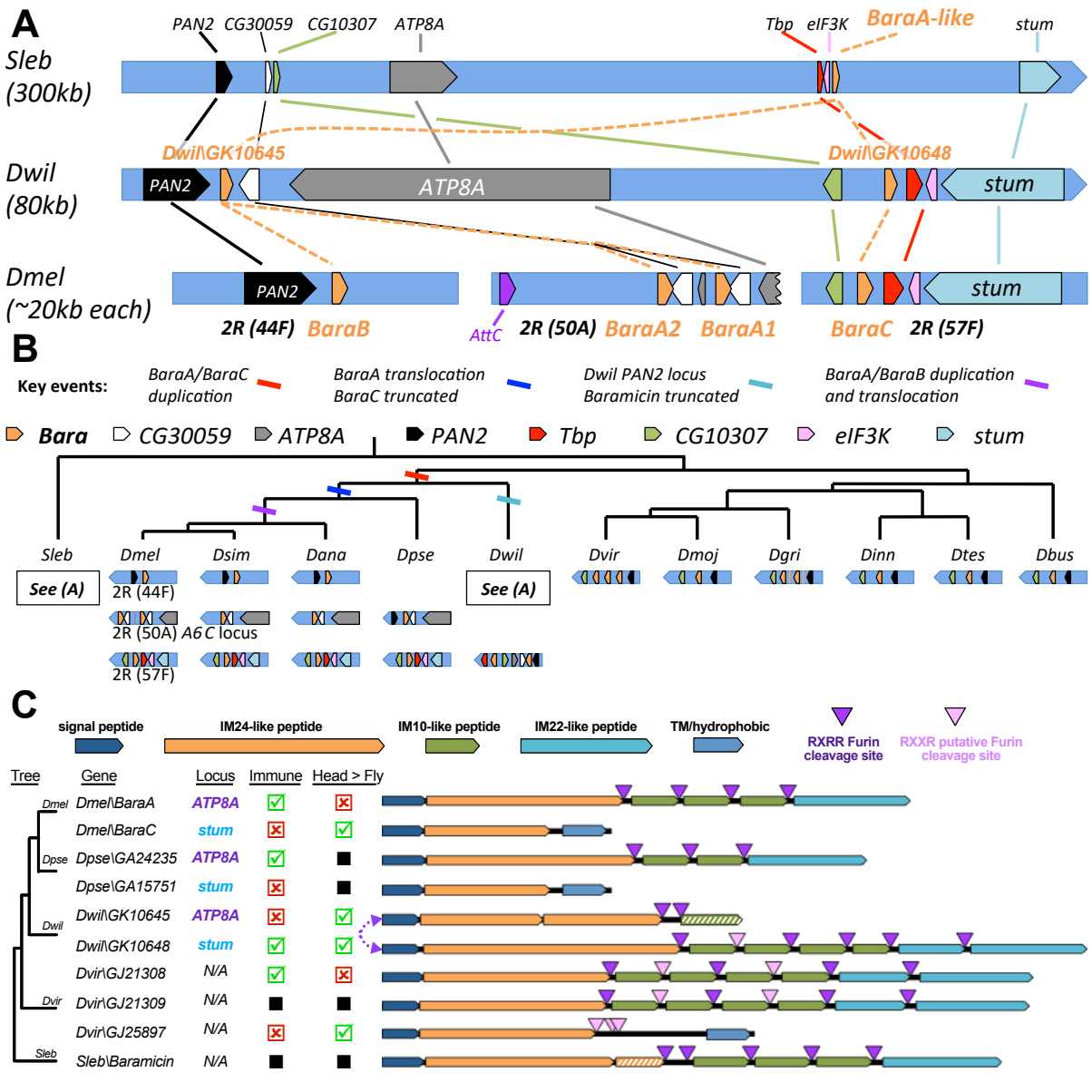


Figure 5: *Baramicin* evolutionary history. A) Map of genomic neighbourhoods in the outgroup drosophilids *S. lebanonensis, D. willistoni,* and *D. melanogaster*, detailing inferred duplication, inversion, and translocation events. Gene names are given as found in *D. melanogaster*. **B)** Cladogram and genomic loci further detailing the series of events leading to the extant *Baramicin* loci of *Drosophila* species. Loci in *S. lebanonensis* and flies in the subgenus *Drosophila* encode only one *Baramicin* gene, indicating the ancestral drosophilid likely encoded only one *Baramicin*. **C)** IM24-specific *Baramicins* arose from convergent evolution both in gene structure and expression profile. Genomic loci are described here as *ATP8A* or *stum* reflecting prominent genes neighbouring the *Baramicins* (see **Fig.** 5A). Expression dynamics relating to immune-induction or enrichment in the head (checked boxes) are shown in **Fig.** 2 and **Fig.** S7. The *Baramicin* loci in *D. willistoni* are syntenic with *D. melanogaster*, but evolved in a vice versa fashion (purple arrow). The *D. virilis Baramicins GJ21309* and *GJ25897* are direct sister genes (100% identity at N-terminus).

432 This genomic synteny analysis revealed a striking observation: the *D.*
433 *willistoni* gene *Dwil\GK10648* is syntenic with the *Dmel\BaraC* locus (**Fig.** 5A), yet it
434 is immune-induced (**Fig.** 2D) and retains a *BaraA*-like gene structure (**Fig.** 5C).
435 Meanwhile, the other *D. willistoni Baramicin Dwil\GK10645* is found at the locus
436 syntenic with *BaraA/BaraB*, yet this gene has undergone an independent truncation
437 to encode just an IM24 peptide (similar to *Dmel\BaraC*). Thus these two *D. willistoni*
438 genes have evolved similar to *D. melanogaster BaraA/BaraC*, but in a vice versa
439 fashion. This suggests a pattern of convergent evolution with two key points: **i)** the
440 duplication event producing *Dmel\BaraA* and *Dmel\BaraC* originally copied a full-
441 length *BaraA*-like gene to both loci, and **ii)** the derivation of an IM24-specific gene
442 structure has occurred more than once (*Dmel\BaraC* and *Dwil\GK10645*). Indeed,
443 another independent IM24-specific *Baramicin* gene is present in *D. virilis*
444 (*Dvir\GJ25897*), which is a direct sister of the *BaraA*-like gene *Dvir\GJ21309* (**Fig.**
445 5C). We evaluated the expression of these truncated *Baramicins* encoding only IM24
446 in both *D. willistoni* and *D. virilis* upon infection. As was the case for *Dmel\BaraC*,
447 neither gene is immune-induced (**Fig.** S7A-C). Given the glial expression of
448 *Dmel\BaraC,* we reasoned that the heads (rich in nerve tissue) of adult flies should
449 be enriched in *BaraC* compared to whole animals. Indeed we saw a significant
450 enrichment of *BaraC* in the heads of *D. melanogaster* males compared to whole flies,
451 which was not the case for *BaraA* (**Fig.** S7D). We further saw a consistent and
452 significant enrichment in the head for the IM24-specific genes *Dwil\GK10645* and
453 *Dvir\GJ25897*, while *BaraA*-like genes were more variable in expression (**Fig.** S7E-F).
454
455 Thus, multiple independent IM24-specific *Baramicins* apparently do not play
456 an immune role, and are enriched in the head. In the case of *Dmel\BaraC*, this is
457 likely due to expression in glia. Strikingly, we observe an independent but parallel
458 evolution of expression pattern and gene structure in *Baramicins* of *D. willistoni* and
459 *D. virilis*. A summary of these expression data is included in **Fig.** 5C. Phylogenetic
460 and genomic synteny analysis reveals that the gene structure and immune
461 expression pattern of *BaraA* reflect the ancestral state, while *Dmel\BaraB* and
462 *Dmel\BaraC* are paralogs derived from independent duplication events. These two

463 genes, which are truncated and delete both IM10-like and IM22 domains, have also
464 lost immune function.
465
466 ***Evolution of IM24 in lineage-specific fashions***
467
468 Multiple independent *Baramicin* genes have lost both IM10-like and IM22
469 peptides, and also converge on loss of immune induction. Taken together, these
470 truncations and expression patterns suggest that the IM10-like peptides and IM22
471 are strictly useful during the immune response, consistent with their recently
472 described antifungal role [24]. Inversely, non-immune *Baramicin* genes have
473 repeatedly and independently truncated to encode primarily IM24. This suggests
474 that IM24 is adaptable to different contexts – a protein domain with utility
475 extending beyond the immune response.
476
477 Thus far, we have discussed IM24 primarily as a unifying feature of the
478 *Baramicin* gene family. To understand how this universal domain differs across gene
479 lineages, we next investigated the evolution of the IM24 domain at the sequence
480 level. We analyzed IM24 domain evolution via multiple approaches using the HyPhy
481 package (implemented in Datamonkey.org, [34]) with a codon alignment of *BaraA,*
482 *BaraB,* and *BaraC* IM24 domains beginning at their conserved Q^1 starting residue.
483 Globally, FEL, FUBAR, and SLAC site-specific analyses suggest strong purifying
484 selection across the IM24 domain, agreeing with the general protein structure of
485 IM24 being broadly conserved (analysis outputs in **supplementary data file 1**).
486 However alignment of Baramicin IM24 domains revealed a specific motif with
487 lineage-restricted patterns. In *Dmel\BaraB*, this motif encodes the residues
488 ^{40}HHASSPAD^{48}, numbered according to the Q^1 starting residue. This HHASSPAD
489 domain shows lineage-restricted residues in *BaraA, BaraB,* and *BaraC* (**Fig.** 6A).
490 Such lineage-restricted residues suggest that selection on Baramicin IM24 coincided
491 with rapid evolution of this motif following gene duplication events, and then
492 subsequently experienced purifying selection within each lineage (as suggested by
493 FEL, FUBAR, and SLAC). This pattern could mask signals of positive selection,

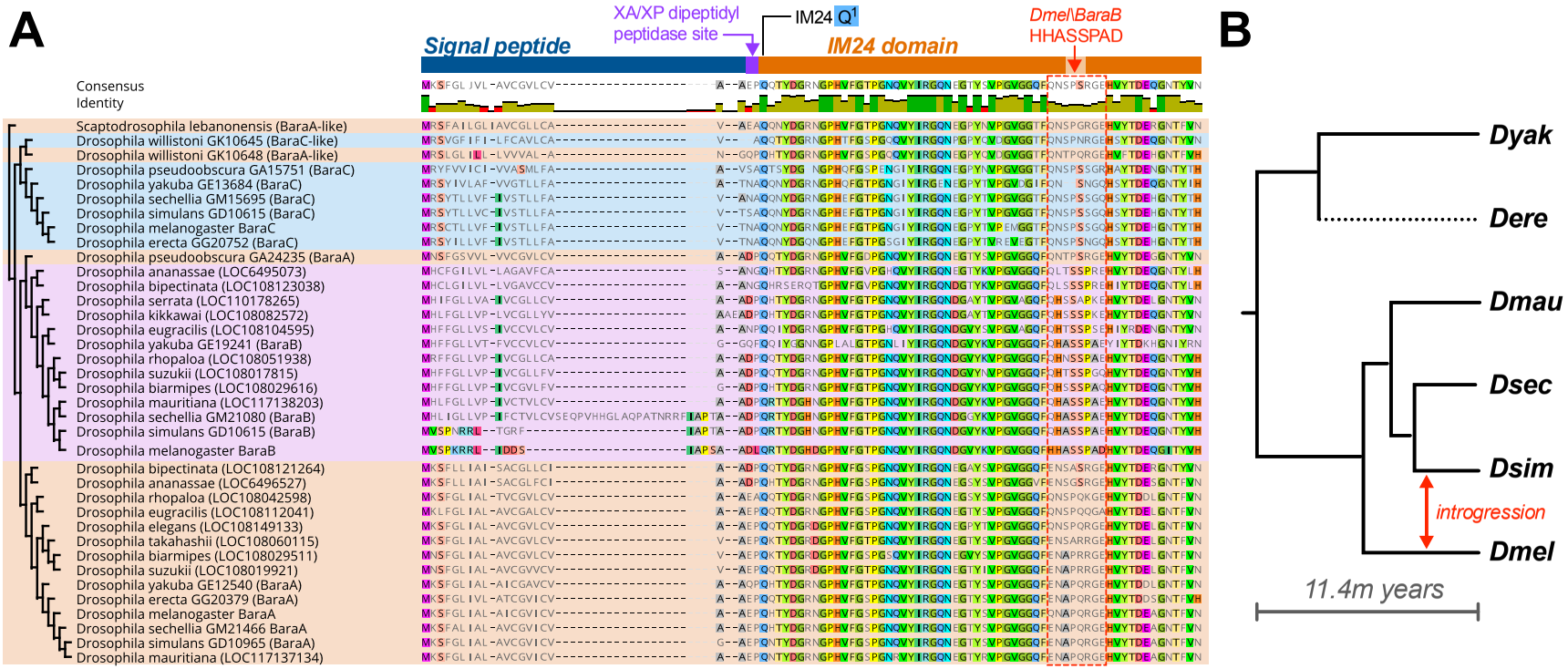


Figure 6: *BaraB* and IM24 rapid evolution. A) Evolution of the signal peptide and core IM24 domain with a focus on *Baramicins* of the Melanogaster group. Residue highlighting indicates agreement with *Dmel\BaraB*. The *D. melanogaster* ^{40}HHASSPAD^{48} domain is bordered by red dashed lines, where unique residues are found restricted within each lineage. For example: the residues at the end of this domain are ^{46}RGE^{48} in *BaraA*-like genes and also *Dwil\GK10645*, ^{46}(S/N)GQ^{48} in *BaraC* genes, and ^{46}PXE^{48} in *BaraB* genes (except in *D. melanogaster*). The signal peptide region, dipeptidyl peptidase cleavage site, and IM24 domain including the Q^{1} site are indicated. The cladogram along the left shows a neighbour-joining tree derived from the residues shown (1000 bootstraps). Background colouring indicates *Baramicin* clade. **B)** Proposed timeframe for the introgression of the extant *Dmel\BaraB* protein structure between *D. melanogaster* and *D. simulans*. The *D. erecta BaraB* gene is pseudogenized. Speciation timeframes adapted from Timetree.org.

494
495

496 particularly given the short 61-residue region of homologous coding sequence (**Fig.**
497 6A). To test this hypothesis, we next used adaptive branch site relative-effects
498 likelihood (aBSREL) to screen for episodic diversifying selection, a more sensitive
499 method to detect punctuated instances of positive selection [34]. aBSREL analysis of
500 the IM24 domain found one instance of diversifying selection along the branch
501 distinguishing the subgenus Drosophila and *D. willistoni* (subgenus Sophophora) as
502 outgroups from other Sophophora species (**Fig.** S8). This branch corresponds to the
503 evolutionary event distinguishing the *Dmel\BaraA* and *Dmel\BaraC* lineages from
504 the *Baramicins* found in *D. willistoni*. This positive selection event therefore
505 accompanies the parallel and vice versa evolution of *BaraA*-like and IM24-specific
506 gene structures and expression patterns in *D. melanogaster* and *D. willistoni* (**Fig.**
507 5C).
508
509 These findings highlight that not only are IM10-like and IM22 peptides lost in
510 e.g. *Dmel\BaraC* and *Dwil\GK10645*, but the IM24 domain in these genes also
511 evolved rapidly following these events. An 8-residue motif ("HHASSPAD" in
512 Dmel\BaraB) appears to be a highly variable motif amongst different genes, but the
513 sequence within is conserved in each gene lineage (**Fig.** 6A). Moreover, an ancient
514 instance of diversifying selection distinguishes the *Baramicins* of *D. melanogaster*
515 and *D. willistoni*, which evolved towards similar gene structures independently.
516
517 ***Introgression and loss: a timeframe for BaraB neofunctionalization***
518
519 Another striking evolutionary pattern in Dmel\BaraB is the loss of the N-
520 terminal signal peptide (**Fig.** 4). Intriguingly, a *BaraA*-like signal peptide is
521 conserved across species in all *Baramicin* lineages, except in *BaraB* of *D.*
522 *melanogaster* and also *D. simulans* (last common ancestor 5.9mya [35]); these
523 species encode a conserved region of parallel length that lacks a signal peptide
524 structure (**Fig.** 6A, SignalP 5.0 [36]). This is surprising, as *D. simulans* is
525 phylogenetically more closely related to *D. sechellia* (last common ancestor 3.3mya)
526 and *D. mauritiana* (last common ancestor 4.5mya), yet *D. sechellia* and *D. mauritiana*

527 retain signal peptide structures in their *BaraB* genes; we note that the *D. sechellia*
528 BaraB signal peptide is divergent due to a separate insertion event (**Fig.** 6A), but is
529 still predicted to enable protein secretion (SignalP 5.0 [36]).
530
531 We reason that the *D. simulans BaraB* N-terminus – which mirrors that of *D.*
532 *melanogaster* – was introgressed from one species to the other prior to the complete
533 development of hybrid inviability between these lineages. Taken together, these
534 observations provide a precise phylogenetic timeframe that confirms that the last
535 point where the *Dmel\BaraB* lineage encoded a functional signal peptide was in the
536 last common ancestor of *D. melanogaster, D. simulans, D. sechellia,* and *D.mauritiana*
537 5.9mya. This overt change in protein structure prevents BaraB secretion by a
538 standard signal peptide mechanism. Even more precisely, the introgression of the
539 *Dmel\BaraB*-like gene structure into *D. melanogaster/D. simulans* pinpoints that the
540 current *BaraB* gene structure existed prior to the development of hybrid inviability
541 between these two species, but after reproductive isolation between *D. simulans* and
542 *D. sechellia* began (**Fig.** 6B). It is tempting to speculate that this evolutionary event
543 correlates with the neofunctionalization of *BaraB* for some developmental process
544 in the nervous system. Perhaps supporting this notion, the *BaraB* gene of the
545 outgroup *D. erecta* is pseudogenized (**Fig.** 6B): there is a mutation in the ATG start
546 codon (now ATT), alongside frameshifts in the IM24 domain leading to premature
547 stops. It seems unlikely that *BaraB* in the last common ancestor of *D. melanogaster*
548 and *D. erecta* (11.4mya) functioned as it does now in *D. melanogaster* given that
549 *Dmel\BaraB* mutation is highly deleterious.
550
551 We speculate that loss of a signal peptide correlates with Dmel\BaraB
552 function diverging from its sister gene Dmel\BaraA, though we do not to invoke this
553 loss as the cause of *Dmel\BaraB* derived function. However, using this signal as a
554 proxy for derived function, Timetree.org [35], and confirmed Melanogaster group
555 hybrid inviabilities [37], we estimate the timeframe of *D. melanogaster-D. simulans*
556 *BaraB* introgression at between ~2.6 and 3.3mya, which dates the extant
557 *Dmel\BaraB* gene structure to between 5.9mya and 2.6mya. Assuming loss of the

558 *Baramicin* signal peptide correlates with derivation of novel function, we suggest
559 that the developmental role of *Dmel\BaraB* was only recently derived between 2.6-
560 5.9 million years ago.
561
562 ## Conclusion and Future Directions
563
564 The *Baramicin* gene family has undergone a varied history of genomic
565 duplication, translocation, and complex sequence evolution. Here we describe an
566 actively segregating duplication of the immune-induced *BaraA*, which may be
567 relevant to immunity studies given that the duplication is variable in common lab
568 stocks. The ancestral *BaraA*-like gene structure is comprised of multiple distinct
569 peptide domains (i.e. IM24, IM10-like peptides, IM22). The loss of IM10-like
570 peptides in non-immune *Baramicin* paralogs and inter-species polymorphisms in
571 the IM10-like peptides of the *BaraA*-like genes reinforces the notion that these
572 peptides are likely key in host-pathogen interactions. Two recent studies suggest
573 that the *BaraA* gene and specifically the IM10-like peptides have antifungal
574 properties [24,25]. Future research investigating these peptides as potential
575 therapeutics would benefit from explicitly testing the consequences of the G/D
576 polymorphism on microbicidal activity. Broadly conserved polymorphisms across
577 IM10-like peptides at sites 3 (H/Q/V), 4 (V/I), 5 (A/E/S) and 9 (R/H) should also be
578 of interest to potential IM10-like peptide activity screens.
579
580 The IM24 domain unites the *Baramicin* gene family. As a consequence, the
581 evolution of *Baramicin* gene structure, pattern of expression, and IM24 domain
582 sequence divergence, all yield insight into the contexts in which IM24 acts. Notably,
583 multiple IM24-specialized genes are not immune induced, and are enriched in the
584 head; for *D. melanogaster BaraB* and *BaraC* we determine this head enrichment
585 derives from the nervous system. *BaraB* knockdown by the pan-neural driver *elav-*
586 *Gal4* phenocopies both the lethality and nubbin-like wing phenotype. We
587 additionally investigated the effect of *BaraB* RNAi using Gal4 drivers specific for glia
588 (*Repo-Gal4*), motor neurons (*D42-, VGMN-,* and *OK6-Gal4*), the wing disc (*nubbin-*

589 *Gal4*), myocytes (*mef2-Gal4*), and with a recently-made *BaraA-Gal4* driver [24].
590 However all these *Gal4>BaraB-IR* flies were viable and never exhibited overt
591 morphological defects. While we did not identify the specific neural cell type that
592 expresses *BaraB*, our Gal4 approach rules out a number of options including: glia,
593 the wing disc, motor neurons, myocytes, or neural tissue overlapping with
594 expression of *BaraA* in the nervous system. Future study will benefit from
595 systematically knocking down *BaraB* in neuronal subtypes. Furthermore, *BaraB*
596 gene knockout in the Melanogaster group (i.e. in *D. simulans, D. sechellia,* and/or *D.*
597 *mauritiana*) may also validate whether the important role of *BaraB* in development
598 correlates with the loss of a signal peptide structure.
599
600 While we do not identify the exact tissue of *BaraB* expression, we do
601 highlight two evolutionary events that are associated specifically with the
602 *Dmel\BaraB* lineage. First, the global picture informed by analyzing the IM24 region
603 of *BaraA, BaraB,* and *BaraC* reveals that the residues aligned with the Dmel\BaraB
604 HHASSPAD domain are highly variable, bounded by otherwise conserved IM24
605 sequence (**Fig.** 6A). Second, the Dmel\BaraB protein has lost the otherwise
606 conserved signal peptide structure of the *Baramicin* gene family, suggesting its
607 cellular localization differs from that of other *Baramicins*. Certainly BaraA IM24 is
608 secreted, as this peptide is robustly detected in immune-induced hemolymph
609 [24,25,38]. Likewise, Dmel\BaraC bears a couple key signatures indicating it is likely
610 to reach the plasma membrane: i) it encodes a signal peptide that should enable
611 entry into the secretory pathway, and ii) the IM24 domain is preceded by an "SA"
612 dipeptidyl peptidase site; XA or XP motifs are cleaved by dipeptidyl peptidases at
613 the cell membrane to maturate secreted peptides [39]. Indeed XA or XP motifs are
614 seen preceding the Q^1 residue in all IM24 domains of BaraA and BaraC proteins in
615 **Fig.** 6A. These patterns suggest *BaraC* encodes an initially secreted protein.
616
617 *BaraC* is unique as a tool to understand the activity of IM24, owing to its
618 IM24-specific structure. As the BaraC IM24 domain is tailed by a transmembrane
619 helix, the BaraC C-terminus should insert itself into the membrane of either the

620 BaraC-producing cell, or a neighbouring cell. In either scenario, the IM24 domain of
621 BaraC would likely be exposed to the extracellular space, suggesting a role for the
622 BaraC IM24 domain in receiving some humoral signal, or interacting with other
623 molecules embedded in the membrane. Identifying the binding partner(s) of either
624 the secreted BaraA IM24, the putatively extracellular BaraC IM24 domain, or BaraB
625 IM24 whose cellular localization is unknown, should be informative to this domain's
626 general function. Further genetic and biochemical approaches using tagged IM24
627 constructs should elucidate BaraA, BaraB, and BaraC IM24 localization patterns.
628 Manipulating the variable region of different IM24 peptides should also confirm if
629 this motif underlies the specificity of different IM24 domains to their paralog-
630 specific functions, including *Baramicin* phenotypes such as *BaraB*-dependent
631 nubbin-like wings, or the erect wing phenotype *BaraA* mutants display upon
632 infection [24].
633
634 It is clear from our comparative approach that the function of the IM24
635 protein is not strictly related to immune processes. Nevertheless it is possible that
636 IM24 also possesses antimicrobial activity. Indeed AMPs and neuropeptides share
637 many common features including: small size, cationic net charge, and
638 amphipathicity [40], and BaraA IM24 is both relatively short (96 AA) and cationic
639 (+2.4 at pH=7). However *Baramicins* are not the only AMP gene family united by a
640 domain that does not necessarily function as an immune effector. One human AMP
641 recently implicated in chronic neuroinflammatory disease is the Cathelicidin LL-37
642 [23,41,42]. The *Cathelicidin* gene family is unified by its N-terminal "Cathelin"
643 domain, however this region has no antimicrobial activity in vitro [43]. Instead, the
644 Cathelicidins encode their mature AMPs (e.g. LL-37) at the protein C-terminus. The
645 role of innate immunity and specifically antimicrobial peptides in neurological
646 disorders is a budding and exciting field of exploration [17–19,21,44]. The
647 *Baramicins* add to the growing list of immune-neural interactions by providing an
648 evolutionary case study that demonstrates that not all peptidic domains of AMP-
649 encoding genes are immune-specific. Indeed, most AMP studies have a peculiar
650 blind spot in this regard, as *in vivo* experiments typically disrupt entire genes, while

651 ex vivo or in vitro approaches typically manipulate only the canonical mature AMP
652 domain. Future studies on AMP contributions to various disease processes should
653 benefit greatly from dissecting the functions of AMPs at the gene level from the
654 specific functions of their encoded peptides, which could act in totally separate
655 tissues and contexts.
656
657 In summary, we characterize how an ancestral AMP-encoding gene has
658 evolved under a number of unique evolutionary scenarios. In light of the repeated
659 derivation of IM24-specific *Baramicin* genes – including at least two specifically
660 involved in neural processes – it will be interesting to consider the functions of AMP
661 genes in neural processes not simply at the level of the gene, but at the level of the
662 mature peptides produced by that gene. The distinct functional roles of the IM24
663 domain in neural processes and IM10-like peptides in immunity suggest that
664 selection on AMP genes may go beyond host-pathogen interactions. In the case of
665 *Baramicin* and other recently-characterized invertebrate AMPs, natural selection
666 may impose trade-offs that contrast pressures imposed by essential roles in host
667 processes (e.g. in neural tissues) with maintenance of a competent host defence.
668 Given the many commonalities between AMPs and neuropeptides, these canonical
669 immune effectors might be adapted for non-immune functions more commonly than
670 appreciated.
671
672 **Acknowledgements**
673
674 We would like to thank Maria Litovchenko for assistance with DGRP
675 screening, Rob Unckless for stimulating discussion, Huang et al. [25] for
676 collaborative cooperation, and Florent Masson and Hannah Westlake for a thorough
677 reading of the manuscript and much appreciated advice. This research was
678 supported by Sinergia grant CRSII5_186397 awarded to Bruno Lemaitre. The
679 *BaraB*^{LC1} and *BaraB*^{LC4} mutations were graciously provided by Steven Wasserman
680 and generated by Lianne Cohen, who we also thank for their critical involvement in
681 characterizing *Baramicin A* [24].
682
683

684 **Materials and Methods (507 words)**
685
686 *DGRP population screening and bioinformatics analyses*
687
688 Sequence data from the DGRP were downloaded from
689 http://dgrp2.gnets.ncsu.edu/ [26]. Sequence comparisons and alignment figures
690 were prepared using Geneious R10 [45], Prism 7, and Inkscape. Alignments were
691 performed using MUSCLE or MAFFT followed by manual curation, and phylogenetic
692 analyses were performed to validate sequence patterns using the Neighbour Joining,
693 PhyML, RaxML, and MrBayes plugins in Geneious. Figure S1B was generated using
694 the UCSC *D. melanogaster* DGRP2 genome browser. Selection analyses were
695 performed using the HyPhy package implemented in datamonkey.org [34]. Codon
696 alignment of the IM24 domain used in **Fig.** 6A is included as a .fasta file in
697 **supplementary data file 1** alongside outputs from FEL, FUBAR, SLAC, and aBSREL
698 selection analyses.
699
700 *Fly genetics*
701
702 The *BaraB*^{LC1} and *BaraB*^{LC4} mutations were generated using CRISPR with two
703 gRNAs and an HDR vector by cloning 5' and 3' region-homologous arms into the
704 pHD-DsRed vector, and consequently *ΔBaraB* flies express DsRed in their eyes,
705 ocelli, and abdomen. The following PAM sites were used for CRISPR bordering the
706 *BaraB* region. Slashes indicate the cut site: 5': GCGGGCAACAGATGTGTTCA/GGG
707 3': GTCCATTGCTTATTCAAAAA/TGG. These mutants were generated in the
708 laboratory of Steve Wasserman by Lianne Cohen, who graciously allowed their use
709 in this study. All fly stocks including Gal4 and RNAi lines are listed in
710 **supplementary data file 1**. Experiments were performed at 25°C unless otherwise
711 indicated. When possible, genetic crosses of 6-8 males and 6-8 females were
712 performed in both directions to test for an effect of the X or Y chromosomes on
713 *BaraB*-mediated lethality; crosses in both directions yielded similar results in all
714 cases. Fly diet consisted of a nutrient-rich lab standard food: 3.72g agar, 35.28g
715 cornmeal, 35.28g yeast, 36mL grape juice, 2.9mL propionic acid, 15.9mL moldex,
716 and H_2O to 600mL.
717
718 *Infection experiments*
719
720 Bacteria and yeast were grown to mid-log phase shaking at 200rpm in their
721 respective growth media (LB, BHI, or YPG) and temperature conditions, and then
722 pelleted by centrifugation to concentrate microbes. Resulting cultures were diluted
723 to OD = 200 at 600nm before infections to measure gene expression. The following
724 microbes were grown at 37°C: *Escherichia coli strain 1106* (LB) and *Candida albicans*
725 (YPG). *Micrococcus luteus* was grown at 29°C in LB. For **Fig.** 2 and S6A, pooled fly
726 samples were collected either 6 hours post-infection (*E. coli*) or 24 hours post-
727 infection (*C. albicans, M. luteus*) prior to RNA extraction on pools of 5 adult males.
728 These timepoints correspond to the maximal expression inputs of the Imd (6hpi) or

729 Toll (24hpi) NF-κB signalling pathways, which are most specifically induced by
730 Gram-negative bacteria (Imd) or Gram-positive bacteria or fungi (Toll) [46]. Flies
731 were pricked in the thorax as described in [47].
732
733 RNA extractions were performed using TRIzol™, Ambion DNAse treatment,
734 and PrimeScript RT according to manufacturer's protocols. RT-qPCR was performed
735 using PowerUP SYBR Green master mix with primers listed in **supplementary data**
736 **file 1**. Gene expression differences were analyzed using the PFAFFL method [48].
737 For gene expression experiments requiring dissection of heads, pools of 20 males
738 were used for either whole flies or heads dissected in ice-cold PBS and transferred
739 immediately to a tube kept on dry ice.
740
741 **References:**
742

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

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

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

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

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

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

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

764 8. Sackton TB, Lazzaro BP, Schlenke T a, Evans JD, Hultmark D, Clark AG. Dynamic
765 evolution of the innate immune system in Drosophila. Nature genetics.
766 2007;39: 1461–1468. doi:10.1038/ng.2007.60

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

769 10. Halldórsdóttir K, Árnason E. Trans-species polymorphism at antimicrobial
770 innate immunity cathelicidin genes of Atlantic cod and related species. PeerJ.
771 2015;3: e976. doi:10.7717/peerj.976

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

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

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

782 14. Tennessen JA. Molecular evolution of animal antimicrobial peptides:
783 Widespread moderate positive selection. 2005. doi:10.1111/j.1420-
784 9101.2005.00925.x

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

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

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

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

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

798 20. Dominy SS, Lynch C, Ermini F, Benedyk M, Marczyk A, Konradi A, et al.
799 Porphyromonas gingivalis in Alzheimer's disease brains: Evidence for disease
800 causation and treatment with small-molecule inhibitors. Science Advances.
801 2019;5. doi:10.1126/sciadv.aau3333

802 21. Gosztyla ML, Brothers HM, Robinson SR. Alzheimer's Amyloid-β is an
803 Antimicrobial Peptide: A Review of the Evidence. 2018. doi:10.3233/JAD-
804 171133

805 22. Pineda A, Burré J. Modulating membrane binding of α-synuclein as a
806 therapeutic strategy. Proceedings of the National Academy of Sciences.
807 2017;114: 1223–1225. doi:10.1073/pnas.1620159114

808 23. De Lorenzi E, Chiari M, Colombo R, Cretich M, Sola L, Vanna R, et al. Evidence
809 that the human innate immune peptide LL-37 may be a binding partner of
810 amyloid-β and inhibitor of fibril assembly. Journal of Alzheimer's Disease.
811 2017;59: 1213–1226. doi:10.3233/JAD-170223

812 24. Hanson MA, Cohen LB, Marra A, Iatsenko I, Wasserman SA, Lemaitre B. The
813 Drosophila Baramicin polypeptide gene protects against fungal infection.
814 Immunology; 2020 Nov. doi:10.1101/2020.11.23.394148

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

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

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

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

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

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

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

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

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

845 34. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: A
846 suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics.
847 2010;26: 2455–2457. doi:10.1093/bioinformatics/btq429

848 35. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines,
849 Timetrees, and Divergence Times. Molecular biology and evolution. 2017;34:
850 1812–1819. doi:10.1093/molbev/msx116

851 36. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O,
852 Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep
853 neural networks. Nat Biotechnol. 2019;37: 420–423. doi:10.1038/s41587-019-
854 0036-z

855 37. Lachaise D, David JR, Lemeunier F, Tsacas L, Ashburner M. The Reproductive
856 Relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D.
857 melanogaster from the Afrotropical Region. Evolution. 1986;40: 262.
858 doi:10.2307/2408806

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

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

864 40. Brogden KA. Antimicrobial peptides: Pore formers or metabolic inhibitors in
865 bacteria? 2005. doi:10.1038/nrmicro1098

866 41. Lee M, Shi X, Barron AE, McGeer E, McGeer PL. Human antimicrobial peptide
867 LL-37 induces glial-mediated neuroinflammation. Biochemical Pharmacology.
868 2015;94: 130–141. doi:10.1016/j.bcp.2015.02.003

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

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

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

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

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

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

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

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

895
896

Supplementary Information

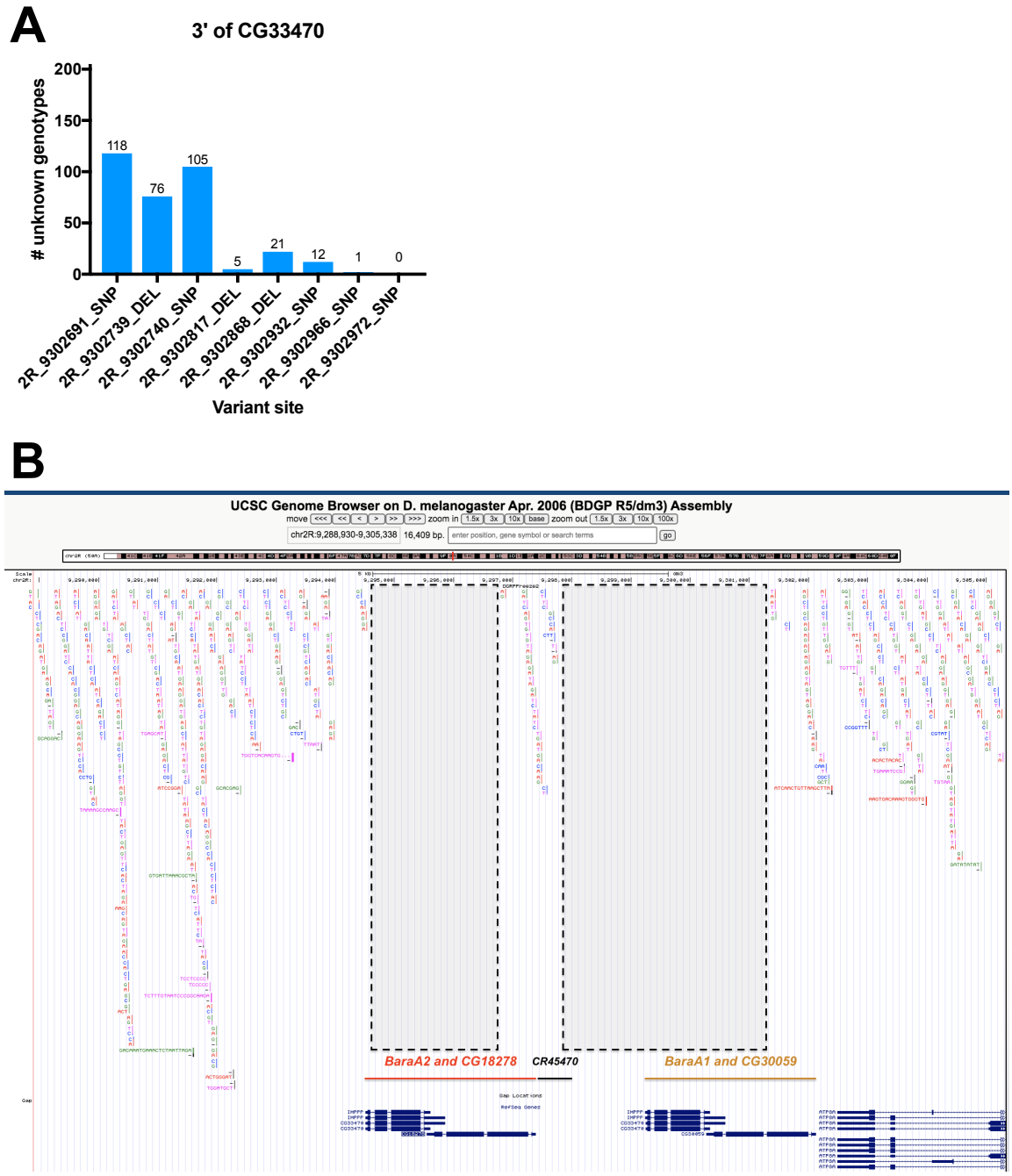


Figure S1: Other regions of the *BaraA* locus in the DGRP are less informative regarding the *BaraA* duplication. A) The 3' region of *CG33470 (BaraA1)* appears less informative in determining the frequency of the *BaraA* duplication in the DGRP, but also generally supports a duplication frequency near 50% in the DGRP. **B)** The *BaraA1* and *BaraA2* gene regions are totally devoid of mapped variants (dashed boxes). We speculate this is due to an artefact during genomic assembly, where reads mapping to either of the two identical *BaraA* genes were discarded as non-specific read mapping.

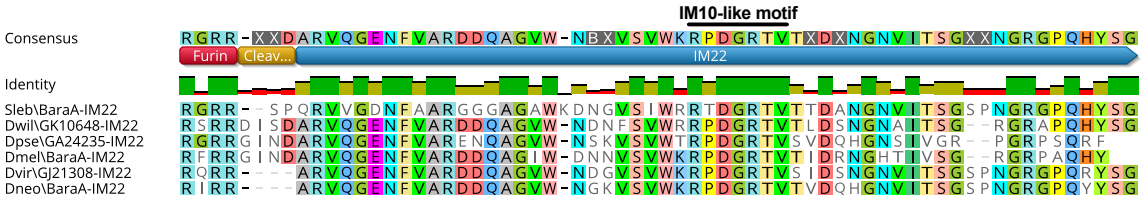


Figure S2: The IM22 domain encodes an IM10-like motif. IM22 encodes a unique peptide with a domain similar to the IM10-like ^{6}RP(G/D)R^{9} motif, except both D and G are encoded in tandem. There is a furin cleavage site upstream of IM22 in all *BaraA-like* genes. In *S. lebanonensis* there is an SP dipeptidyl peptidase site that is also predicted to be cleaved off. In *D. melanogaster BaraA* the GIND motif is also cleaved off by an unknown process [24]. In subgenus Drosophila flies, the mature IM22 peptide begins immediately following the furin cleavage site.

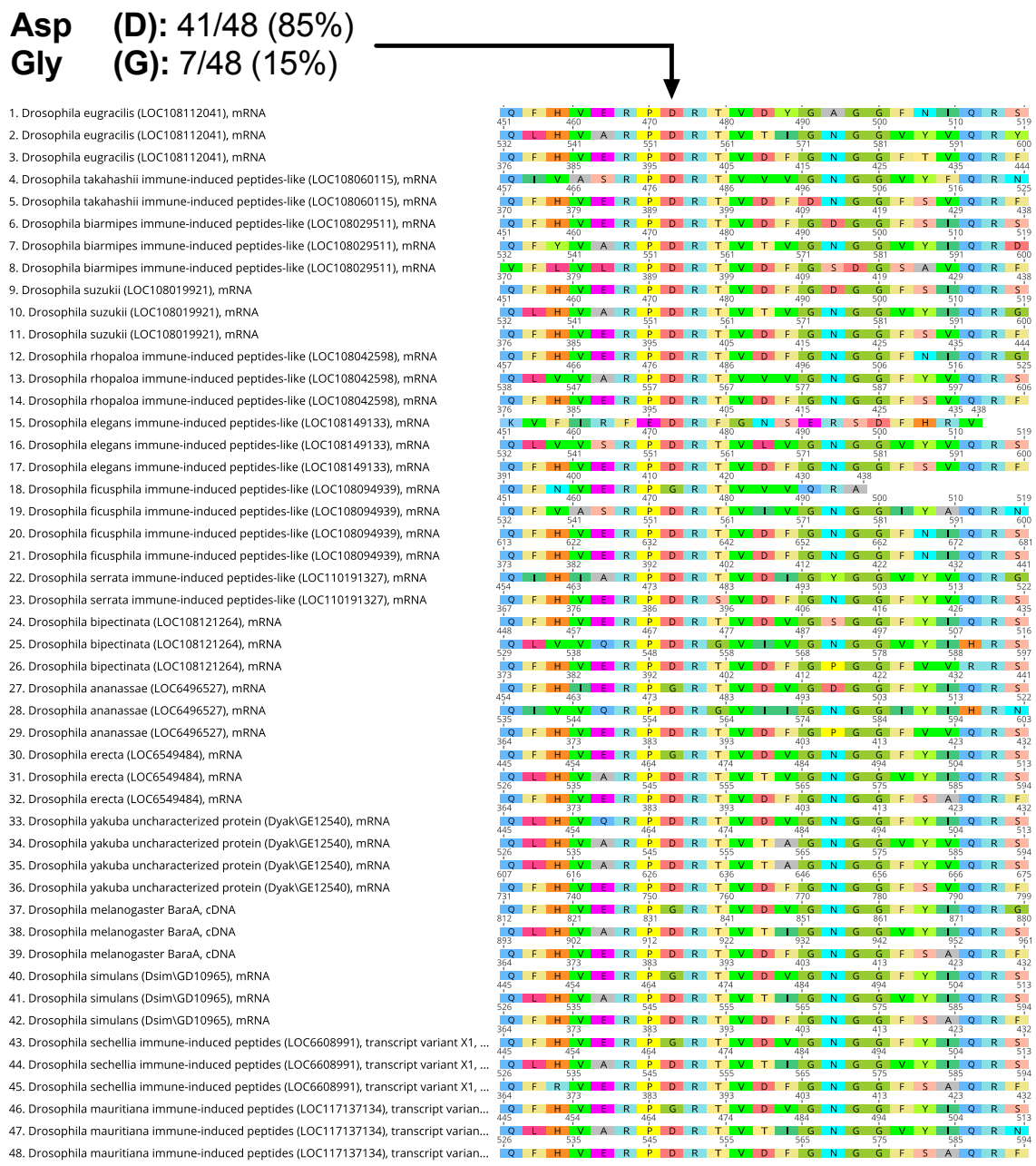


Figure S3: G/D polymorphism frequency in Melanogaster group flies. Presentation of the residue 8 G/D polymorphism in all IM10-like peptides (i.e. IM10, IM12, IM13) of Melanogaster group flies. 41/48 encode an Aspartic acid, while 7/48 encode a Glycine. However, in the Quinaria species group, only Glycine residues are observed (**Fig.** 2A), suggesting G/D frequency is lineage-specific.

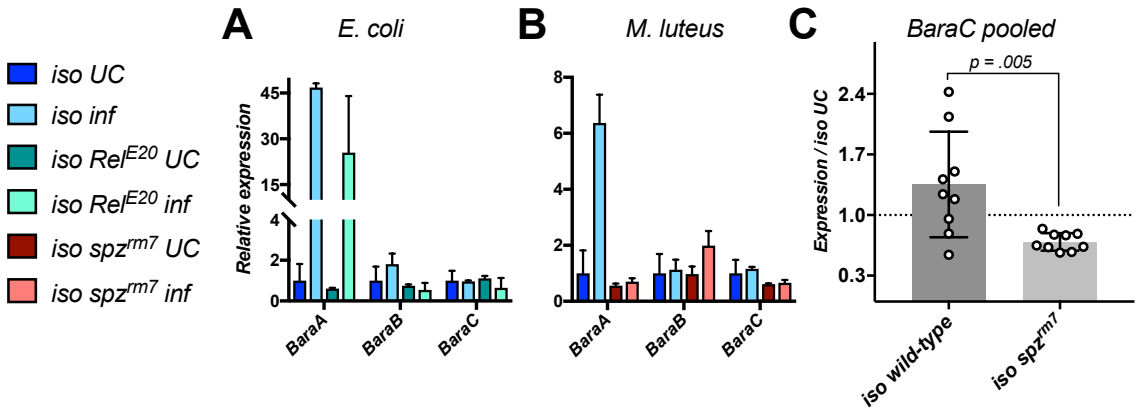


Figure S4: Additional assays of Baramicin expression upon infection. A) Neither *BaraB* nor *BaraC* are regulated by the Imd pathway, which is specifically stimulated by *E. coli* infection. **B)** Neither *BaraB* nor *BaraC* are induced after infection by *M. luteus*. **C)** *BaraC* levels were consistently depressed in *spz^{rm7}* flies in the unchallenged condition (UC) or upon infection with *C. albicans* (**Fig.** 3B) or *M. luteus* (**Fig.** S4B). Data here are pooled for *iso* wild type or *iso spz^{rm7}* flies without regard for infection treatment (student's t, p = .005).

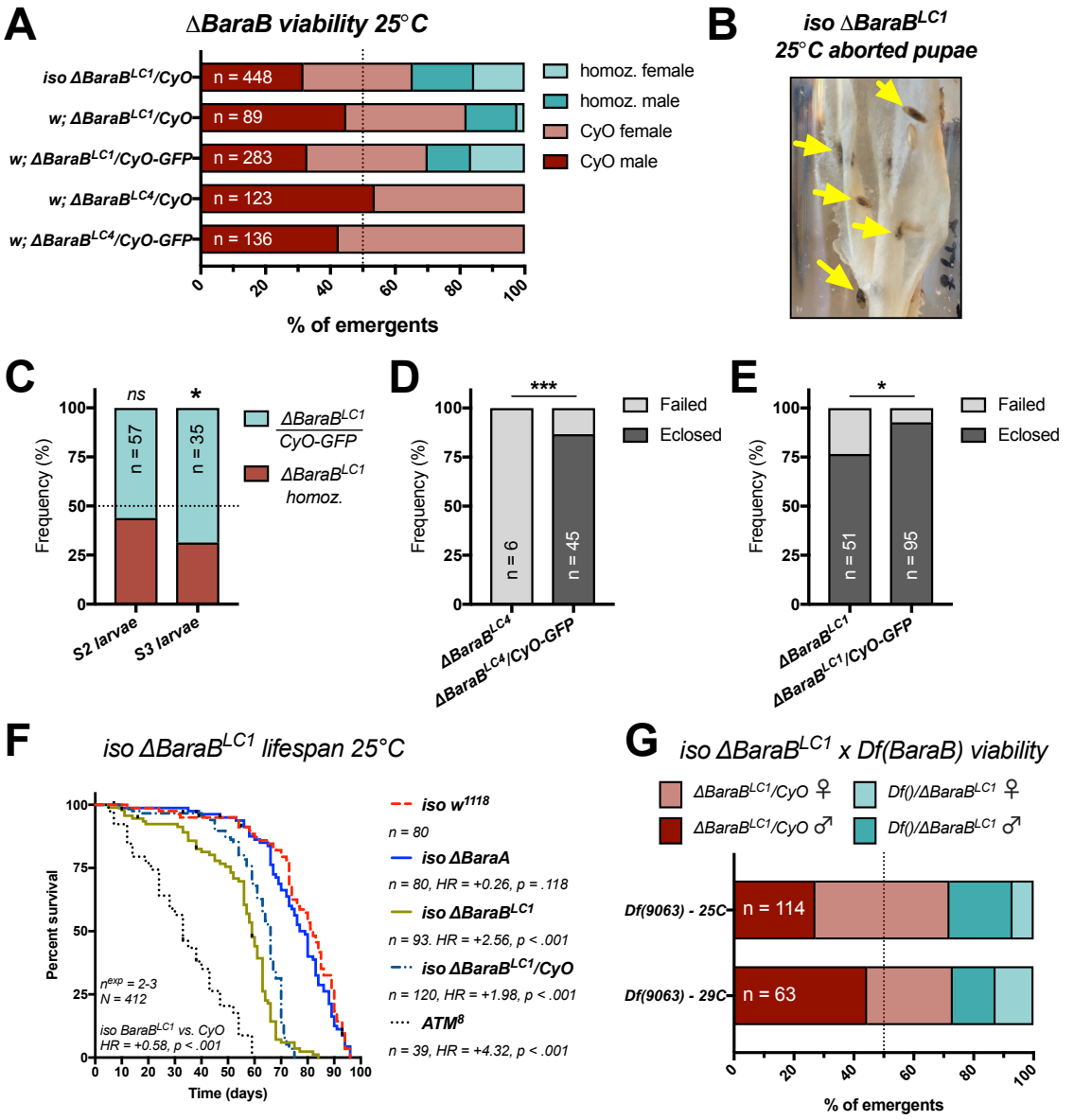
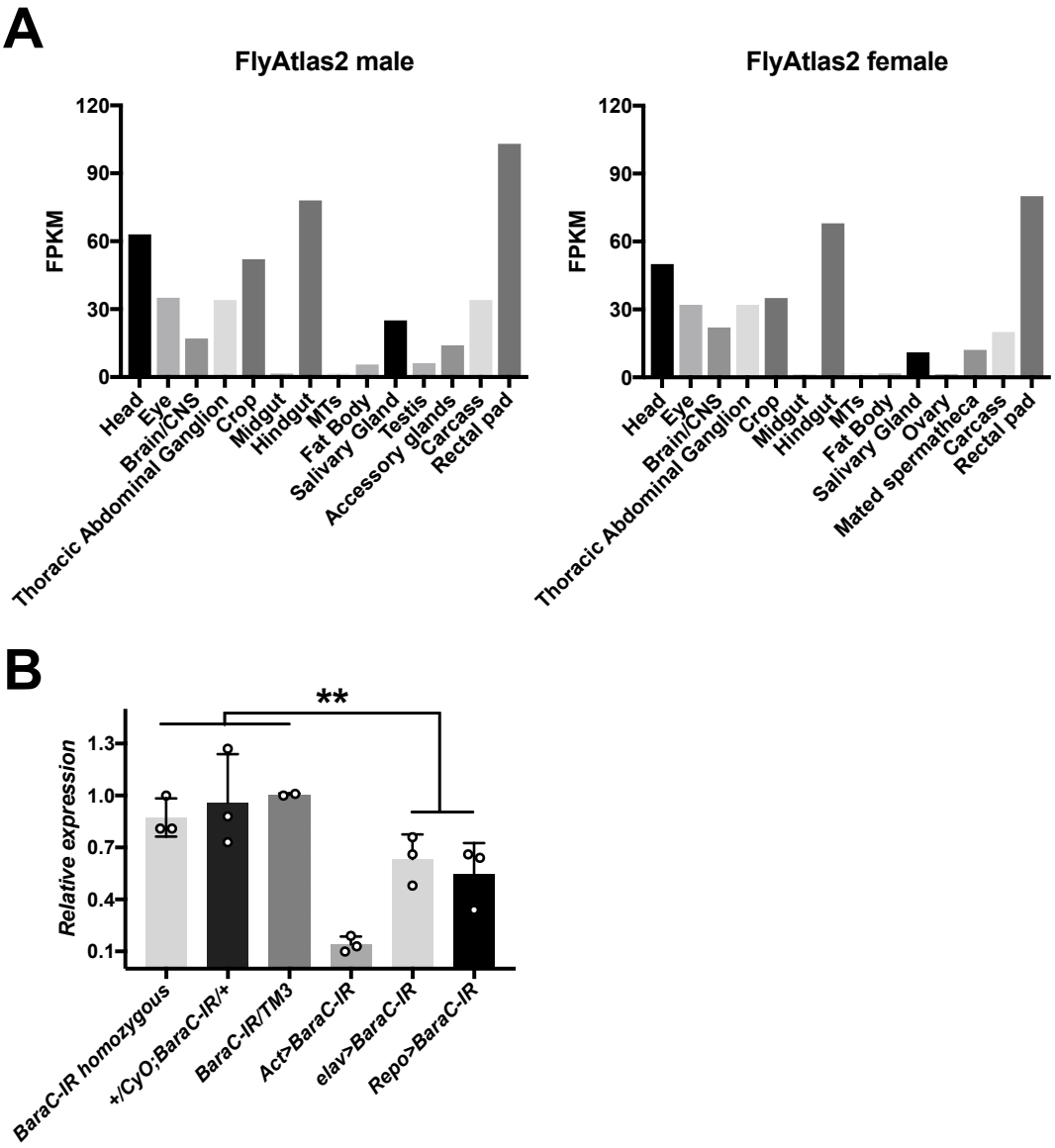


Figure S5: *BaraB* mutation is highly deleterious and penetrance of effect depends on genetic background. A) Emergent frequencies of *ΔBaraB* flies in a variety of genetic backgrounds. *iso ΔBaraB*^{LC1} flies were backcrossed for seven generations. *w; ΔBaraB/CyO* flies represent the original background the mutation was generated in, while *ΔBaraB/CyO-GFP* flies have had additional genetic variation introduced in their 1st and 3rd chromosomes. Rare homozygous L3 larvae were observed in *BaraB*^{LC4}/CyO-GFP vials (anecdotally a rate of <2% of L3 larvae), but these larvae never successfully eclosed. **B)** Aborted pupae (yellow arrows) are a common occurrence in *ΔBaraB* vials, and sometimes contain fully-developed adults that simply never eclosed. **C-E:** ns = not significant, * = p < .05, *** = p < .001. **C)** The ratio of *ΔBaraB*^{LC1}/CyO-GFP to *ΔBaraB*^{LC1} homozygous larvae drops between the S2 and S3 larval stages (χ^2, p = .515 and p = .012 respectively). **D)** Frequency of successfully eclosing adults using *BaraB*^{LC4}/CyO-GFP flies. **E)** Frequency of successfully eclosing adults using *BaraB*^{LC1}/CyO-GFP flies. **F)** *BaraB* mutation negatively affects lifespan. *iso ΔBaraB*^{LC1} homozygotes suffer reduced

946 lifespan even relative to their *iso ΔBaraB*^{LC1}*/CyO* siblings. By comparison, *iso ΔBaraA*
947 flies that used the same vector for mutant generation live as wild-type. ATM^{8} flies suffer
948 precocious neurodegeneration and are included as short-lived controls [49]. **G)** Crosses
949 using the genomic deficiency line *Df(9063)* support a partial-lethal effect of *BaraB*
950 mutation.
951



952
953 **Figure S6: *BaraC* is expressed in the nervous system, but also the hindgut and**
954 **rectal pads. A)** FlyAtlas2 expression data for *BaraC*. **B)** RT-qPCR of *BaraC* in whole flies
955 using different Gal4 drivers to express *BaraC* RNAi. *BaraC* is knocked down by both the
956 *elav-Gal4* and *Repo-Gal4* nervous system drivers. Cumulatively, nervous system drivers
957 significantly depress *BaraC* expression compared to *BaraC-IR* controls (student's t, $p <$
958 $.01$). Ubiquitous knockdown using *Act>BaraC-IR* provides a comparative knockdown to
959 better understand the strength of nervous system-specific knockdowns at the whole fly
960 level.

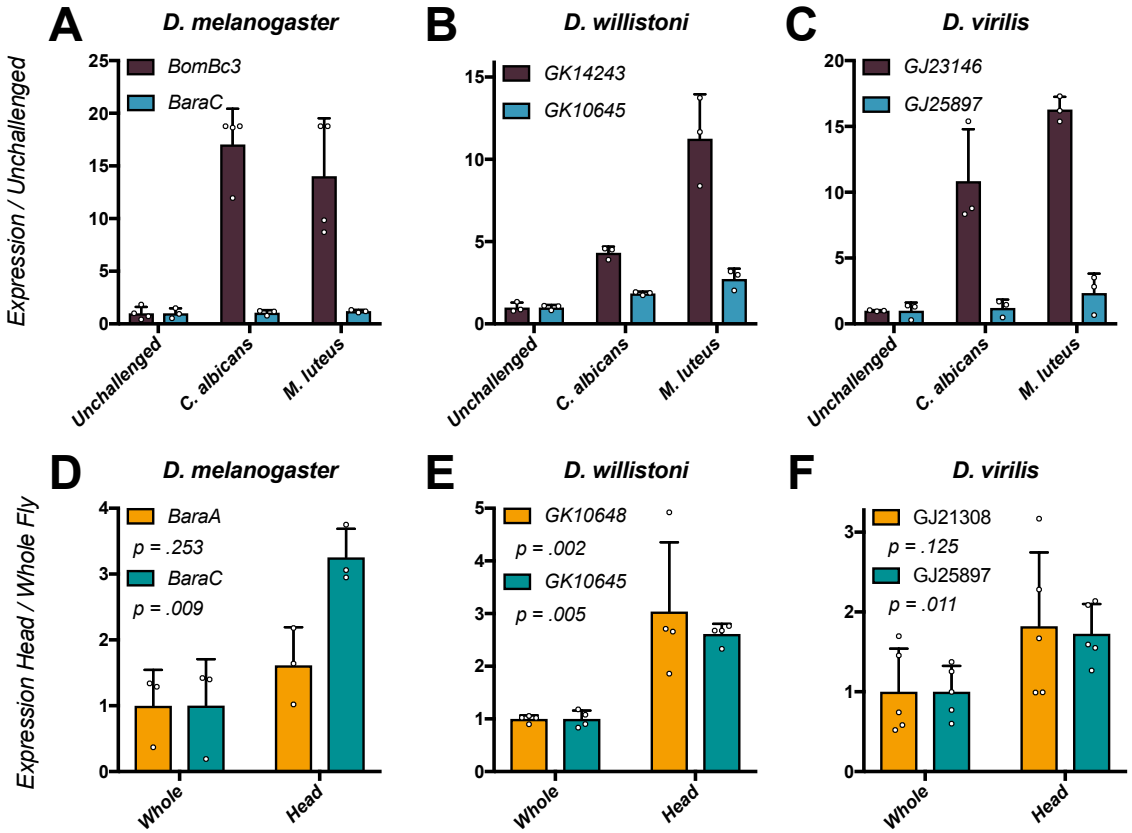


Figure S7: RT-qPCR of *Baramicin* genes in diverse species. A-C) Independent IM24-specific genes in *D. melanogaster* (A), *D. willistoni* (B), and *D. virilis* (C) are not induced by infection. *BomBc3* is included as an immune-induced control. **D-F)** The independent IM24-specific genes (blue) of *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.

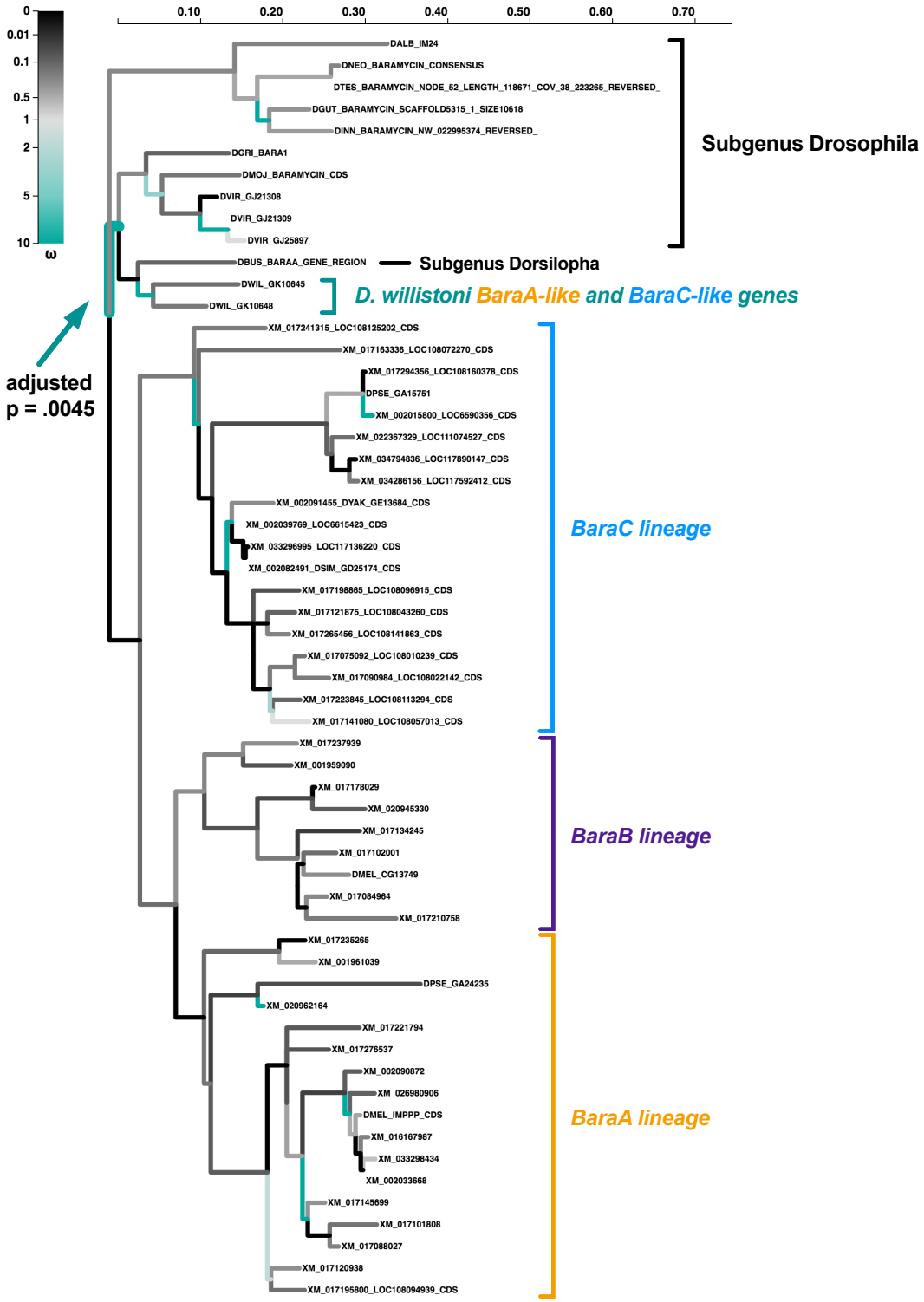


Figure S8: Adaptive branch site relative-effects likelihood (aBSREL) analysis of the IM24 domain for episodic diversifying selection. One branch along the tree had a significant dN/dS distinguishing *D. willistoni Baramicins* from *D. melanogaster* and other outgroups. A number of other branches also have high ω values, though these were not statistically significant (e.g. the branch distinguishing *Dvir\GJ21309* and *Dvir\GJ25897* from their immune-induced sibling *Dvir\GJ21308*).

Genotype	Inferred copy #	Gtype	Junction PCR band	CG30059 band	5' BaraA2 prediction	5' CG18278 prediction	5' BaraA2 correct?	5' CG18278 correct?
DGRP	**2**	*DGRP-38*	**YES**	**YES**	**1**	**2**	No	**Yes**
	2	*DGRP-88*	**YES**	**YES**	**1**	**2**	No	**Yes**
	1	*DGRP-91*	NO	**YES**	**1**	*unclear*	**Yes**	*unclear*
	1	*DGRP-158*	NO	**YES**	**1**	*unclear*	**Yes**	*unclear*
	2	*DGRP-318*	**YES**	**YES**	**2**	*unclear*	**Yes**	*unclear*
	1	*DGRP-822*	NO	**YES**	**1**	*unclear*	**Yes**	*unclear*
	1	*DGRP-832*	NO	**YES**	**1**	*unclear*	**Yes**	*unclear*
Wild-type	**2**	*OR-R*	**YES**	**YES**				
	2	*iso w1118*	**YES**	**YES**				
	2	*Exelexis*	**YES**	**YES**				
	1	*w VDRC*	NO	**YES**				
	1	*yw*	NO	**YES**				
	1	*Canton S*	NO	**YES**				

Table S1: the *BaraA* duplication is variable in common lab stocks. Inferred *BaraA* copy numbers are given for various DGRP and wild-type strains following PCR diagnosis using primers specific to the duplication event. All samples were additionally screened for *CG30059* as a DNA extraction positive control. We used unknown variant call patterns to predict *BaraA* copy number in the DGRP using both the 5' region of *BaraA2,* and the 5' region of *CG18278*. We were successful in predicting *BaraA* copy number for 5/7 DGRP strains using just the 5' region of *BaraA2* alone. We could not discern clearly defined pattern groups in the 5' region of *CG18278* that allowed us to specifically predict lines with only one copy of *BaraA*. However some strains called variants across the entirety of the *CG18278* gene region, and we presumed these to be clear indicators of having the duplication. This assumption was correct in 2/2 cases; notably in the two where existing unknown variants in the 5' of *BaraA2* predicted incorrectly. We suggest that using a decision tree to predict *BaraA* copy number will provide best results. This decision tree is informed by the 5' region of the *BaraA2* promoter, but gives veto power to the 5' region of *CG18278* if variants are called in *CG18278* throughout. This post-hoc approach successfully predicts *BaraA* copy number in the DGRP in 7/7 strains.

979
980
981

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	37	32.133	**p < .001**
elav-Gal4> ;; female		f	47		
25°C	nubbin-like wing	m	31		
		f	5		
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		

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