# Functional divergence of the *bag of marbles* gene in the *Drosophila melanogaster* species group

Jaclyn E. Bubnell[1], Cynthia K.S. Ulbing[1], Paula Fernandez-Begne[1], and Charles F. Aquadro[1*]

[1] Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA

*Correspondence: Jaclyn Bubnell, jeb486@cornell.edu and Charles Aquadro, cfa1@cornell.edu

Keywords: Adaptive evolution, CRISPR/Cas9, germline stem cell, gametogenesis

Authors performed the following: JEB - conceptualization, investigation, validation, formal analysis, and writing. CKSU and PFB - investigation, and validation, CFA - conceptualization, funding acquisition, supervision, and writing.

## Abstract

The combination of recent advances in both genomic and gene editing technologies have opened up new possibilities for assessing the functional consequences and drivers of positive selection. In *Drosophila melanogaster,* a key germline stem cell differentiation factor, *bag of marbles* (*bam*) shows rapid bursts of amino acid fixations between its *D. melanogaster* and its sibling species *D. simulans*, but not in the outgroup species *D. ananassae*. We previously hypothesized that a genetic conflict with the maternally inherited, intracellular bacteria *W. pipientis* could be driving the adaptive evolution of *bam* as *W. pipientis* increases the fertility of a *bam* partial loss of function mutant. However, we have not been able to further test this hypothesis by assessing *bam* variation in other *Drosophila* lineages and their interactions with *W. pipientis* because *bam* function has not been examined in non-*melanogaster Drosophila* species. Since *bam* is rapidly evolving at the protein level, its function may not be conserved between species, and therefore different evolutionary pressures may be shaping *bam* in individual lineages. Here, we ask if *bam* is necessary for GSC daughter differentiation in five *Drosophila* species in the *melanogaster* species group that span approximately 15 million years of divergence and show different patterns of nucleotide sequence evolution at *bam*. We find that *bam* function is not fully conserved across these species, and that *bam* function may change on a relatively short time scale. Ultimately, we conclude that a simple gain in function as the germline stem cell differentiation factor alone does not explain our population genetic and functional genetic results we have observed. Our findings provide a foundation on which to explore the evolution of *bam* as a GSC differentiation factor and its interactions with *W. pipientis* in specific lineages.

## Introduction

In most sexually reproducing animals, egg and sperm production begins with the differentiation of germline stem cells. These germline stem cells (GSCs) can divide both to self-renew to maintain the germline and/or to produce daughter cells that differentiate to produce gametes [1]. The cellular processes of daughter cell differentiation and stem cell self-renewal are the most fundamental functions of stem cells and are therefore highly regulated and conserved. In the germline the mis-regulation of either of these processes leads to sterility [2]. However, many of the genes that control GSC maintenance and GSC daughter differentiation in *D. melanogaster* are rapidly evolving due to positive selection for amino acid diversification [3–5].

One such gene is *bag-of-marbles* (*bam*) which exhibits a rapid burst of amino acid substitutions between the sister species *D. melanogaster* and *D. simulans*, acts as the key switch for GSC daughter cell differentiation in *D. melanogaster* females, and is necessary for terminal differentiation of spermatogonia in males [3,6][7]. In females, *bam* is repressed in GSCs, and upon its expression promotes differentiation by binding to the protein product of *benign gonial cell neoplasm* (*bgcn)* as well as a series of other protein partners to repress the translation of self-renewal factors *nanos* and *eIF4a* [8,9]. The GSC daughter cell differentiates into a cystoblast, which then goes through four rounds of mitotic divisions while *bam* localizes to the fusome, a cellular structure that interconnects the resulting dividing cells that make up the cyst. In these cells, Bam functions with Bgcn to regulate mitotic synchrony [10]. Before meiosis begins, *bam* expression is repressed. In males, *bam* is expressed in GSCs and its expression

68 increases as differentiation progresses [7]. In early dividing spermatogonia, Bam protein
69 must reach a threshold to bind with partners Bgcn and *tumorous testis* (*tut*) to repress
70 *mei-P26* and switch the cellular program from proliferation to terminal differentiation to
71 begin meiosis [11]. After this switch, *bam* expression is rapidly repressed. Loss of *bam*
72 function in either sex daughter cell differentiation and results in the over-proliferation of
73 either GSCs in females or spermatogonia in males causing tumors and sterility [6]. The
74 intriguing observation that *bam* is both rapidly evolving due to positive selection and
75 functions as a switch for GSC daughter differentiation has led us to investigate what
76 could be driving its adaptive evolution [12].
77
78 Many genes involved in reproduction show signals of adaptive evolution and are
79 involved in forms of evolutionary conflict. In *Drosophila,* many genes that function in
80 later stages of reproduction (relative to GSC genes), such as in the female reproductive
81 tract and in seminal fluid are critical for fertility and fitness but also show signals of
82 adaptive evolution [13–18]. Combinations of population, quantitative, and functional
83 genetic studies have established that sexual selection and sexual conflicts both
84 between males and females and between males likely explain many of the signals of
85 adaptation in later stages of reproduction. However, the evolutionary pressures that
86 affect post-mating reproductive genes are not likely to affect GSC genes, as GSC genes
87 are expressed at the very first stages of gametogenesis [3,5]. However, there are other
88 conflicts in the germline that may shape the evolution of GSC genes. For example, the
89 GSC genes *piwi* and *fs(1)Yb* show signals of adaptive evolution in *D. melanogaster* and
90 play important roles in regulating transposable element expression in the germline

91 [5,19,20]. Additionally, natural variation in the GSC gene *Bruno* is associated with
92 tolerance to P-element activity [21]. Therefore, interactions with transposable elements
93 may drive the evolution of GSC genes involved in their regulation. However, *bam* has
94 no known role in regulating transposable elements and lacks any domains predicted to
95 bind DNA or RNA so it is unlikely that conflicts with transposable elements shape its
96 evolution.
97
98 We previously identified a genetic interaction between *bam* and the endosymbiotic,
99 maternally inherited, bacterium *W. pipientis* [22]. Infection with the *w*Mel strain of *W.*
100 *pipientis* partially rescued the fertility and cellular GSC daughter differentiation defects
101 of a *bam* partial loss-of-function mutant (hypomorph) in *D. melanogaster* females [12].
102 We also found evidence that the consequences of genetic divergence between *D.*
103 *melanogaster* and *D. simulans bam* only functionally affect females. A full copy of the *D.*
104 *simulans bam* expressed as a transgene in a *D. melanogaster bam* null background did
105 not fully complement female *D. melanogaster bam* function, but it did fully rescue male
106 *bam* function [12]. These two pieces of evidence, that *W. pipientis* rescues a *bam*
107 hypomorph in females, and that functional consequences of *bam* divergence between
108 *D. melanogaster* and *D. simulans* is female specific, are consistent with the possibility
109 that interactions between the *Drosophila* germline and *W. pipientis* drives positive
110 selection at *bam*. As *W. pipientis* utilizes female host cues such as actin formation in
111 order to infect the germline for its own propagation [23,24], *W. pipientis* may be in
112 conflict with its host for control of oogenesis [25]. Our lab sought to further test this
113 hypothesis by asking if other *Drosophila* species with known *W. pipientis* infection

114 histories within the time frame of those in *D. melanogaster* and *D. simulans* also show
115 evidence of positive selection at *bam* [26]. However, we reported that in *D. ananassae*
116 which is also infected with *W. pipientis*, there is no evidence of positive selection at *bam*
117 [26].
118
119 While the finding in *D. ananassae* that *bam* is not evolving under positive selection
120 despite the species being currently infected with *W. pipientis* seems to refute the
121 hypothesis that *W. pipientis* drives the adaptive evolution of *bam,* two notable caveats
122 must be considered. First, we do not know the entire *W. pipientis* infection history of a
123 lineage and therefore conflict between *bam* and *W. pipientis* may have occurred at any
124 time in that lineage. A signal of positive selection that we observe in a population
125 sample today may be the result of a *W. pipientis* infection in the past. Second, we do
126 not know if *bam* function is conserved between *D. melanogaster* and *D. ananassae*. Our
127 expectation that *W. pipientis* infection status should predict patterns of selection at *bam*
128 if *W. pipientis* drives *bam* evolution also assumes that *bam* function is conserved in the
129 lineages under comparison. However, although *bam* has predicted orthologs throughout
130 the *Drosophila* genus, there is reason to believe that *bam* function may not be
131 conserved. Across *Drosophila*, *bam* sequence is highly divergent, making it even
132 difficult to confidently align between *D. melanogaster* and *D. ananassae* [3].
133 Additionally, *bam* is predicted to be a novel gene to a dipteran lineage containing the
134 genus *Drosophila* and a few other closely related genera [27].
135

136 The *bam* sequence itself has many properties of novel genes: it does not contain any
137 known DNA or RNA binding motifs, it binds with multiple protein partners, it contains
138 disordered regions, and it is expressed in a tight spatiotemporal manner [28–32]. Novel
139 genes that become fixed often gain critical roles in conserved gene networks which may
140 drive their evolution by providing functional flexibility for conserved processes [30]. It is
141 possible that *bam* was a novel gene to the common ancestor of *Drosophila* (and
142 possibly some other dipterans) and has functionally diversified across the *Drosophila*
143 genus, only gaining its role in germline stem cell differentiation in the lineage leading to
144 *D. melanogaster* and *D. simulans*, thereby resulting in adaptive evolution for this
145 function. Additionally, it is possible that *bam* quickly evolved a role in GSC daughter
146 differentiation in *Drosophila,* but that function has been shaped by germline conflicts in
147 specific lineages. Conflict with germline agents such as *W. pipientis* may even drive the
148 fixation of novel genes in new roles that successfully evade their control.
149
150 Here, we ask if *bam* is necessary for gametogenesis in diverse species within the
151 *Drosophila* genus that show different patterns of *bam* evolution. We ask this question to
152 understand two concepts fundamental to making predictions about drivers of positive
153 selection at *bam.* First, we need to know if the signals of positive selection we observed
154 in *D. melanogaster* and *D. simulans* can be explained by a new role for *bam* as the
155 GSC daughter differentiation switch. Second, if we know in which lineages *bam* is
156 necessary for GSC daughter differentiation, we can identify appropriate lineages to use
157 for further hypothesis testing regarding factors that drive the adaptive evolution of *bam*.
158

159 In this study we analyzed additional population samples of species mainly within the
160 *melanogaster* species group and a few more distant lineages for nucleotide sequence
161 variation at *bam.* We find further evidence of lineage-specific signals of adaptation
162 within the *melanogaster* species subgroup, and no signals observed outside of the
163 *melanogaster* species subgroup. We generated *bam* loss-of-function alleles in five
164 diverse *Drosophila* species in the *melanogaster* group and find that *bam* is not
165 necessary for gametogenesis in one or both sexes in lineages that do not have
166 evidence of positive selection at *bam*. Together, our data suggest that functional
167 diversification alone does not likely explain the signals of adaptive evolution at *bam,* and
168 a complex evolutionary history between GSC function and lineage-specific pressures
169 may explain the population genetic and functional genetic patterns at *bam*.
170

171 ## Materials and Methods

172 **Fly stocks and rearing:**

173 We raised fly stocks on standard cornmeal molasses food at room temperature. We
174 used yeast glucose food for fertility assays. The *D. melanogaster* $w^{1118}$ isogenic line and
175 $w^{1118}$ balancer lines were gifts from Luis Teixeira, and the *D. simulans* $w^{501}$ line (stock
176 14021-0251.011), the *D. yakuba* white eyed line [14021-226.03] and the *D. ananassae*
177 reference genome (14024-0371.14) lines are from the Drosophila species stock center
178 (http://blogs.cornell.edu/drosophila/). The *W. pipientis*-free *D. ananassae* line Nou83
179 was a gift from Artyom Kopp. The *D. teissieri* line (*teissieri* syn) was a gift from Daniel
180 Matute [33].
181

182 We tested all lines for *W. pipientis* infection status by qPCR (NEB Luna Universal qPCR
183 kit) using primer sets for *wsp* and *arm* (S1 Table). Lines that tested positive (*D. teissieri*,
184 *D. ananassae*, *D. simulans*) were treated with tetracycline (200 µg/mL) for three
185 generations to remove *W. pipientis* [34]. We allowed these lines to recover for at least
186 three generations before using them in any analyses. We were unable to fully clear the
187 *D. ananassae* reference line of *W. pipientis* with tetracycline, so to obtain a *W. pipientis*
188 free stock, we crossed males of our *bam* null lines to females from the Nou83 line,
189 which we confirmed to be *W. pipientis*-free by qPCR.
190
191 **Population samples:**
192 The *D. yakuba*, *D. santomea*, and *D. teissieri* population sample lines were gifted by
193 Brandon Cooper and Daniel Matute [35]. The *D. yakuba* and *D. teissieri* lines were
194 collected in Bioko in 2013, the *D. santomea* sample was collected in Sao Tome in 2015.
195 The *D. affinis* population sample lines were gifted by Robert Unckless (S2 Table). The
196 *D. mojavensis* population sample is from Organ Pipe National Monument, Arizona and
197 DNA from this sample was gifted by Erin Kelleher and Luciano Matzkin. The *D.*
198 *pseudoobscura* population sample is from Apple Hill, California, as described in [36].
199 We used previously prepared DNA from the *D. pseudoobscura* population sample as
200 described in [36].
201
202 **Cloning for bam null constructs**:
203 We obtained the nucleotide sequence information for designing constructs in *D.*
204 *melanogaster, D. simulans*, *D. yakuba*, and *D. ananassae* from FlyBase [27]. We

205 obtained sequence information for the *D. teissieri* sequences from Daniel Matute. We
206 used Geneious for all cloning design.
207
208 We used the NEB Q5 High Fidelity 2X master mix to generate all PCR products. We gel
209 extracted and purified PCR products using the NEB Monarch DNA gel extraction kit or
210 Qiagen MinElute gel extraction kit. IDT primers were used for PCR, sequencing, and
211 cloning. We generated the donor plasmids for the 3xP3-DsRed and 3xP3-YFP bam
212 disruption lines using the NEB HiFi Assembly Cloning kit into the pHD-attP-DsRed
213 vector from flyCRISPR [37] as follows: we amplified 1.5 kb homology arms from
214 genomic DNA of the appropriate species stock flanking the insertion site for 3xP3-
215 DsRed or 3xP3-YFP. 3xP3-DsRed was amplified from the pHD-attP-DsRed plasmid
216 and 3xP3-YFP was amplified from the *D. simulans* nos-Cas9 line, a gift from David
217 Stern. We then gel extracted the two homology arms and the appropriate 3xP3 marker,
218 purified them, and assembled them into the pHD vector backbone using the
219 manufacturer's protocol (S3 Table).
220
221 We used the U6:3 plasmid from CRISPRflydesign [38] to express gRNAs for generating
222 the *D. yakuba* lines. The gRNA target sequence was generated by annealing primers
223 with the gRNA sequence and BbsI site overhangs, and ligating into the BbsI ligated
224 U6:3 vector (T4 ligase NEB, BbsI NEB) (S3 Table). We used NEBalpha competent cells
225 for transformations.
226

227 We prepped and purified all plasmids for embryo injections with the Qiagen plasmid
228 plus midi-prep kit. We additionally purified plasmids for injections with phenol-chloroform
229 extraction to remove residual RNases. We confirmed plasmid sequences with Sanger
230 sequencing (Cornell BRC Genomics Facility).
231
232 **CRISPR/Cas9:**
233 gRNA selection:
234 We chose gRNAs using the flyCRISPR target finder [37] for *D. melanogaster*, *D.*
235 *simulans*, *D. yakuba*, and *D. ananassae*. For *D. teissieri*, we used one completely
236 conserved gRNA from *D. yakuba.* We also chose an additional gRNA from *D. yakuba*
237 that had two divergent sites between *D. yakuba* and *D. teissieri,* thus we modified them
238 to the *D. teissieri* sites. We selected gRNAs that had no predicted off- targets in the
239 reference genome (S4 Table & S5 Table ). For *D. yakuba*, we used the U6:3 gRNA
240 plasmid from CRISPRfly design as described above. For the remaining species (*D.*
241 *melanogaster*, *D. simulans*, *D. teissieri*, and *D. ananassae*), we used synthetic gRNAs
242 (sgRNAs) from Synthego. We used 1-3 gRNAs per injection to increase the chances of
243 a successful CRISPR event. We used multiple gRNAs if there were suitable options for
244 that site to increase efficiency (multiple gRNAs within 50 bps with 0 predicted off-
245 targets).
246
247 Injections:
248 All CRISPR/Cas9 injections were carried out by Genetivision. Lines were injected with
249 the plasmid donor, gRNAs (synthetic or plasmid), Cas9 protein (Synthego), and an

250 siRNA for Lig4 (IDT DNA). The same siRNA was used for *D. melanogaster*, *D.*
251 *simulans*, *D. yakuba*, and *D. teissieri* (pers communication with Daniel Barbash and
252 David Stern S6 Table). We created a *D. ananassae* specific siRNA due to sequence
253 divergence (S6 Table).
254
255 For the *bam* disruption lines, we screened for eye color in F1s in house using a
256 Nightsea fluorescence system with YFP (cyan) and DsRed (green) filters. We
257 backcrossed all CRISPR/Cas9 mutants to the stock line we used for injections for three
258 generations. All mutants are maintained as heterozygous stocks. We confirmed all
259 CRISPR insertions by Sanger sequencing (Cornell BRC Genomics Core Facility).
260
261 Genotypes for generating bam disruption homozygotes:
262 We generated two *bam* disruption lines in each species in order to create the genotypes
263 for the comparative *bam* null analyses. We generated one by disrupting *bam* with 3xP3-
264 DsRed and the other with 3xP3-YFP. We selected DsRed or YFP positive flies and
265 crossed them to generate homozygous *bam* disruption flies ($bam^{3xp3\text{-}DsRed}/bam^{3xP3\text{-}YFP}$).
266 This cross also results in DsRed or YFP positive heterozygous *bam* null progeny as well
267 as *bam* wildtype progeny without any fluorescent eye marker.
268
269 **Fertility assays:**
270 Female fertility:
271 All female fertility assays were performed following the approach of Flores et al. [12].
272 Specifically, we collected virgin females and aged them 2-3 days (*D. melanogaster*, *D.*

273 *simulans*, *D. yakuba,* and *D. teissieri*) or 5-7 days (*D. ananassae*). We collected all
274 genotypes from each bottle to control for bottle effects. We collected virgin males of the
275 wildtype genotype for each species and aged them for the same timeframe as the
276 females of that species. We evenly distributed males collected from different bottles
277 across the female genotypes to control for any bottle effects. We crossed single females
278 collected within 48 hours of each other to two virgin males. We allowed the trio to mate
279 for 9 days, and then flipped them to new vials. Except for *D. ananassae*, as fertility was
280 very low, we cleared the trio after day 9 and did not continue further. For all other
281 species crosses, after an additional 9 days, we cleared the trio. Then, we counted the
282 progeny for each trio every other day to get the total adult progeny per female.
283
284 Male fertility:
285 All male fertility assays were performed as described for the female fertility assays
286 except that one tester male was crossed to two virgin females.
287
288 Food and conditions for fertility assays:
289 For *D. melanogaster*, *D. simulans*, *D. ananassae*, and *D. teissieri*, we performed all
290 fertility assays on yeast-glucose food. For *D. yakuba*, we performed fertility assays on
291 cornmeal-molasses food, as this line did not thrive on the yeast-glucose food. We kept
292 all crosses and fertility experiments in an incubator at 25°C with a 12-hour light-dark
293 cycle.
294
295

296 **Statistics**:

297 For the fertility assays we used estimation statistics to assess the mean difference
298 (effect size) of adult progeny between the wildtype *bam* genotype and each individual
299 additional *bam* genotype. For generating the estimation statistics and plots [39], we
300 used the dabest package in Python (v. 0.3.1) with 5,000 bootstrap resamples.
301 Estimation statistics provide a non-parametric alternative to other statistical tests of the
302 difference of the mean (for example, ANOVA), and allow us to understand the
303 magnitude of the effect of *bam* genotype on the fertility phenotype. In text, we report
304 significance as a mean difference (effect size) outside the 95% bootstrap confidence
305 interval. While one benefit of these statistics is that they are less biased than reporting
306 traditional P-values, we also report P-values for approximate permutation tests with
307 5,000 permutations.

308

309 **Immunostaining:**

310 We used the following primary antibodies, anti-vasa antibody from Santa Cruz
311 biologicals (anti rabbit 1:200), anti-Hts-1B1 from Developmental Studies Hybridoma
312 bank (anti-mouse 1:4 for serum, 1:40 for concentrate item), and anti-Fas3 from DSHB
313 (anti-mouse 1:50). We used the following secondary antibodies: Alexaflour 488, 568
314 (Invitrogen) at 1:500.

315

316 We performed immunostaining as described in Aruna et al. [38] and Flores et al.
317 [12,40]. Briefly, we dissected ovaries and testes in ice-cold 1X PBS and pipetted ovaries
318 up and down to improve antibody permeability. We fixed tissues in 4%

319 paraformaldehyde, washed with PBST, blocked in PBTA (Alfa Aesar, with 0.01% Triton
320 X added), and then incubated in the appropriate primary antibody in PBTA overnight.
321 We then washed (PBST), blocked (PBTA), and incubated the tissue in the appropriate
322 secondary antibody for 2 hours, then washed (PBST) and mounted in mounting media
323 with DAPI (Vecta shield or Prolong Glass with NucBlue) for imaging.
324
325 **Confocal Microscopy:**
326 We imaged ovaries and testes on a Zeiss i880 confocal microscope with 405nm, 488
327 nm, and 568 nm laser lines at 40X (Plan-Apochromat 1.4 NA, oil) (Cornell BRC Imaging
328 Core Facility). We analyzed and edited images using Fiji (ImageJ).
329
330 **RNA extraction:**
331 For RT-qPCR experiments we dissected ovaries from 10 one-day old females per
332 biological replicate or testes from 15 one-day old males in ice cold 1x PBS. The tissue
333 was immediately placed in NEB DNA/RNA stabilization buffer, homogenized with a
334 bead homogenizer, and then stored at -80°C until RNA extraction. We used the NEB
335 Total RNA mini-prep kit for all RNA extractions according to the manufacturer's
336 protocol and included the on column DNAse treatment. We assessed RNA
337 concentration and quality with a Qubit 4 fluorometer and a NanoDrop.
338
339 **RT-qPCR:**
340 For all RT-qPCR analyses, we used the Luna One Step RT-qPCR kit from NEB
341 according to the manufacturer's instructions at 10µl total volume in 384 well plates. We

342 used a Viia7 qPCR machine with the appropriate cycle settings as described in the
343 Luna One Step RT-qPCR kit.
344
345 For the analysis of *bam* expression in *D. ananassae*, we used primers to *D. ananassae*
346 *bam* with *D. ananassae Rp49* as a control (S1 Table). To assess the possibility that the
347 *bam* disruption allele induced exon skipping of the disrupted exon 1 in *D. teissieri*, we
348 designed a primer pair that spanned exon 1 to exon 2 and another pair that spanned
349 exon 2 to exon 3 (S1 Table). We used the exon 2 - exon 3 target as the control. For
350 both the *D. teissieri* and *D. ananassae* assays described above, we included a standard
351 curve for each target on every plate at a 1:2 dilution and then used the standard curve
352 method to calculate the relative quantity of *bam* (to either *Rp49* for *D. ananassae* or
353 *bam* exon 2-exon 3 for *D. teissieri*) with the QuantStudio software using the wildtype
354 *bam* genotype as a control for each species.
355
356 For the allele specific analysis in *D. teissieri,* we designed one primer pair specific to
357 the *bam* 3xP3-DsRed and 3xP3-YFP insertion alleles and another primer pair specific
358 to the wildtype allele. We then used these two targets to measure the relative
359 expression of the insertion allele to the wildtype allele in a single heterozygous sample.
360 The primer pair for the *bam* insertion alleles target the sequence between the SV40-
361 polyA tail of the 3xP3-DsRed and 3xP3-YFP insertion and the *bam* sequence
362 downstream of the insertion (S1 Table). The primer pair for the *bam* wildtype allele
363 target *bam* upstream and downstream of the 3xP3-DsRed and 3xP3-YFP insertion,
364 and therefore will not amplify the large insertion allele under RT-qPCR conditions (S1

365 Table). We included a standard curve for each target on every plate at a 1:2 dilution
366 and then used the standard curve method in the QuantStudio software to calculate the
367 Ct and Ct standard error for each target and each sample. We report the relative
368 quantity of the 3xP3-DsRed or 3xP3-YFP allele using the wildtype allele in the same
369 sample as a control by calculating $2^{-(nullCt-wildtypeCt)}$.
370
371 **Genomic analysis of *bam* copy number in *D. teissieri:***
372 While *bam* is reported as a 1:1 syntenic ortholog on FlyBase for *D. melanogaster, D.*
373 *simulans, D. yakuba,* and *D. ananassae,* data for *D. teissieri* is not available. To ask if
374 *bam* has been recently duplicated in *D. teissieri,* we generated a blast database (blastn
375 v 2.9.0) from a new de novo assembled *D. teissieri* reference genome (accession
376 GCA_016746235.1) and blasted *D. teissieri bam* from the resequencing described
377 above.
378
379 Additionally, to assess read coverage for *bam* compared to the surrounding loci and the
380 entire chromosome 3R, we used publicly available short read WGS data for *D. teissieri*
381 (SRR13202235) and used bwa-mem (v 0.7.17) to map reads to a new de novo
382 assembled *D. yakuba* reference genome (accession GCA_016746365.1). We visualized
383 the alignment and assessed read coverage using Geneious.
384
385
386
387

388 **Population genetic analyses:**

389 **Resequencing:**

390 We prepared genomic DNA from *D. affinis, D. yakuba*, *D. santomea*, and *D. teissieri*
391 samples using the Qiagen PureGene kit. We used Primer3 to generate primers to
392 amplify *bam* from *D. affinis, D. mojavensis, D. pseudoobscura, D. yakuba*, *D. santomea*,
393 and *D. teissieri* population samples (S7 Table). We used existing polymorphism data
394 (Daniel Matute, personal communication) to generate primers for *D. yakuba*, *D.*
395 *santomea*, and *D. teissieri*. We resequenced *bam* using Sanger sequencing from
396 5'UTR-3'UTR for *D. yakuba*, *D. santomea*, and *D. teissieri* and *bam* exons and introns
397 for *D. affinis* (Cornell BRC Genomics Facility). We used Geneious for aligning the
398 sequences and calling polymorphisms. We generated fasta files for each individual. We
399 used DnaSP to phase individuals with heterozygous sites, and then randomly picked
400 one allele for further analyses [3]. We will deposit all resequencing data described here
401 in Genbank.

402

403 **Existing polymorphism data:**

404 We used polymorphism data for *D. melanogaster* from the Zambia population sample
405 from the Drosophila genome Nexus [41], as this is the largest population sample with
406 intact *bam* sequence and represents a population in the *D. melanogaster* ancestral
407 range. We used polymorphism data for *D. simulans* from the California population
408 sample [42]. For the *D. melanogaster* Zambia population sample, we removed a single
409 individual that contained missing nucleotide calls in the *bam* coding sequence (ZI117).
410 As we filtered polymorphisms at < 5% frequency, omitting this single individual would

411 not affect our downstream analyses. In the *D. simulans* CA population sample there
412 were no individuals with missing data.
413
414 *D. ananassae bam* polymorphism, and single *D. atripex, D. bipectinata bam* sequences
415 are from Choi et al. [26]. We obtained single *bam* sequences for species used for
416 outgroups from FlyBase (*D. erecta*), and Genbank (*D. eugracilis, D. arizonae, D.*
417 *navojoa, D. miranda, D. subobscura).*
418
419 **Alignments and ancestral sequence reconstruction:**
420 We performed all multiple sequence alignments for the *bam* coding sequences using
421 PRANK (v.170427) with the -codon and -F parameters and using the PRANK generated
422 guide tree. As *bam* is highly divergent across the *Drosophila* genus, we generated
423 multiple sequence alignments only among closely related clusters of species to ensure
424 alignment quality. We generated a multiple sequence alignment of the *melanogaster*
425 subgroup species (*D. melanogaster, D. simulans, D. yakuba, D. santomea, D. teissieri,*
426 *D. erecta, D. eugracilis*), a multiple sequence alignment of *D. ananassae, D. atripex,*
427 and *D. bipectinata*, a multiple sequence alignment of *D. mojavensis, D. arizonae, and D.*
428 *navajo*, and a multiple sequence alignment of *D. affinis, D. pseudoobscura, D. miranda,*
429 and *D. subobscura*.
430
431 We used only unique sequences for input to codeml. We used the codeml package from
432 PAML (v. 4.9) to generate the predicted common ancestor sequences to calculate
433 lineage-specific divergence for *bam* with the McDonald-Kreitman test. We used the

434 PRANK alignments and their corresponding gene trees as input to codeml with the
435 following control file parameters (noisy = 9, verbose = 2, runmode = 0, seqtype = 1,
436 CodonFreq = 2, clock = 0, aaDist = 0, model = 0, NSsites = 0, icode = 0, getSE = 0,
437 RateAncestor = 1, Small_Diff = .5e-6, cleandata = 0, method = 1).
438
439 We used the http://mkt.uab.cat/mkt/mkt.asp webtool to perform the standard McDonald-
440 Kreitman test (MKT) comparing nonsynonymous vs. synonymous changes [43]. We
441 excluded polymorphic sites at < 5% frequency, as these are likely slightly deleterious
442 alleles that have not yet been purged by purifying selection [44]. For each species
443 where we had polymorphism data, we performed the MKT using the specified predicted
444 common ancestral sequence to calculate lineage-specific divergence. We report the
445 values of the contingency table, the P-value of the Chi-Square test, as well as alpha, the
446 proportion of fixations predicted to be due to positive selection [45].
447
448 **Results**
449 ***bam* shows lineage-specific bursts of positive selection in the *yakuba* complex,**
450 **and no evidence of positive selection outside of the *melanogaster* species**
451 **subgroup**
452 Previously, our group found that *bam* is evolving under strong positive selection
453 detected by the McDonald-Kreitman test in *D. melanogaster* and *D. simulans*, but not in
454 *D. ananassae* [3,26]. As these three species span about 15 million years of divergence,
455 we wanted to know if there was evidence of adaptation in lineages more closely related
456 to *D. melanogaster* and *D. simulans*, or if this signal was unique to the individual *D.*

457 *melanogaster* and *D. simulans* lineages. We expanded the species sampled to include
458 three closely related species from the *yakuba* complex, *D. yakuba, D. santomea,* and *D.*
459 *teissieri*. We also took advantage of published population genomic data from large
460 inbred lines to expand our assessment of *bam* variation in an additional population
461 sample each for *D. melanogaster* and *D. simulans*.
462
463 In addition to further sampling more closely related species to *D. melanogaster* and *D.*
464 *simulans,* we also wanted to ask if there was any evidence of adaptive evolution at *bam*
465 in lineages outside of the melanogaster group. We resequenced *bam* from population
466 samples of three more distantly related *Drosophila* species: *D. pseudoobscura*, *D.*
467 *affinis*, and *D. mojavensis,* these species span approximately 40 million years of
468 divergence [46].
469
470 To ask if any population samples showed departures from selective neutrality consistent
471 with positive selection at *bam* similar to what we previously observed for *D.*
472 *melanogaster* and *D. simulans*, we used a lineage-specific McDonald-Kreitman test
473 (MKT) [47], specifically measuring nonsynonymous and synonymous changes. The null
474 hypothesis of the MKT is that the ratio of nonsynonymous to synonymous polymorphism
475 within a species is equal to the ratio of nonsynonymous to synonymous divergence
476 between species (or in our case, a predicted common ancestral sequence). Previously,
477 we have detected signatures of positive selection at *bam* as an excess of
478 nonsynonymous fixations in the *D. melanogaster* and *D. simulans* lineages [3]. As *bam*
479 appears to be experiencing episodic signals of adaptive evolution restricted to particular

480 lineages, we measured lineage-specific divergence. We used codeml to generate
481 reconstructed ancestral sequences at the node of interest by maximum likelihood. We
482 then used this sequence to calculate nonsynonymous and synonymous divergences for
483 each population sample.
484
485 For *D. melanogaster*, we utilized the Zambia population sample from the *Drosophila*
486 genome Nexus to measure *bam* variation [41]. The Zambia sample contains the largest
487 sample size of intact *bam* sequence (few missing calls), and represents variation found
488 in the ancestral range of *D. melanogaster*. We used the predicted common ancestor of
489 *D. melanogaster* and *D. simulans* and found similar patterns of polymorphism and
490 divergence to the smaller samples of *D. melanogaster* from the USA and Zimbabwe
491 previously reported by Bauer DuMont et al. [3] (Fig 1). The *D. melanogaster* lineage
492 shows a strong signal of adaptive evolution (alpha = 0.912, P<0.000) (Fig 1).
493
494 For *D. simulans bam*, we utilized the CA, USA population sample from Signor et al [42].
495 The previous *D. simulans bam* sample our lab reported was also from the USA, but
496 from North Carolina [3,48]. We used the same common ancestral sequence as we did
497 for the *D. melanogaster* analysis to calculate divergence. However, we did not observe
498 the same patterns of divergence and polymorphism in this sample as we previously
499 observed in the NC sample (Fig 1). In the CA *bam* sample, we cannot reject neutrality
500 (alpha = 0.539, P=0.084), in contrast to our previous finding in the NC sample
501 (alpha=0.631, P=0.049). It is notable that although the ratios were not significantly
502 different, the trend in this lineage also shows a high alpha value and higher

503 nonsynonymous divergences. The difference between the two samples seems to be
504 driven by higher nonsynonymous polymorphism in the CA sample compared to the NC
505 sample. The likely explanation for this discrepancy is the observed genome-wide
506 excess of intermediate frequency polymorphisms in the CA population sample that is
507 consistent with a combination of population contraction and possibly soft sweeps on
508 standing genetic variation within the past 100 years in this CA *D. simulans* population
509 [42]. The introduction of the *w*Ri strain of *W. pipientis* into southern CA around the
510 1980s led to the spread of this strain through CA with initially strong Cytoplasmic
511 Incompatibility [49] that would have also resulted in a severe bottleneck particularly in
512 the female effective population size (e.g. [50]). A population bottleneck can bring slightly
513 deleterious polymorphisms to higher frequency. While the frequency of these
514 deleterious polymorphisms will then slowly decline, during this time their frequency in
515 the population is thus an overestimation of their probability of their fixation and thus
516 weaken power of the MKT and underestimate alpha (e.g. [51])
517
518 We used the predicted common ancestor of the *yakuba* complex (*D. yakuba*, *D.*
519 *santomea*, *D. teissieri*) to estimate divergence for *bam* in each of the species population
520 samples. For *D. yakuba* and *D. santomea* we observed signals of positive selection for
521 *bam* (*D. yakuba*: alpha=0.723, P=0.006), (*D. santomea*: alpha=0.726, P=0.024) (Fig 1).
522 In contrast, we could not reject neutrality for *D. teissieri bam* (alpha=-0.011, P=0.983)
523 (Fig 1).
524

525 In lineages outside of the *melanogaster* species subgroup, we did not reject neutrality at
526 *bam* for any of the species we sampled. For *D. ananassae*, we reassessed divergence
527 with the predicted common ancestor of *D. atripex* and *D. ananassae* to further polarize
528 the patterns of divergence in the *D. ananassae* lineage as Choi et al. [26] reported
529 divergence between *D. ananassae* and *D. atripex*. Our findings are consistent with the
530 previous study (alpha = 0.816, P= 0.089) [26]. The alpha value is high here although not
531 significant, which suggests it may be informative to sample *bam* from additional *D.*
532 *ananassae* populations with larger sample sizes (n=14 here). We calculated divergence
533 for *bam* from population samples of *D. pseudoobscura* and *D. affinis* from their
534 predicted common ancestor and observed similar patterns of polymorphism and
535 divergence (*D. pseudoobscura*: alpha=0.250, P=1), (*D. affinis*: alpha=0.487, P=0.360),
536 neither of which were close to rejecting neutral expectations (Fig 1). We used the
537 predicted common ancestor of *D. mojavensis* and *D. arizonae bam* to calculate
538 divergence in the *D. mojavensis* lineage and again observed no trend towards an
539 excess of nonsynonymous divergence (alpha=0.250, P=1) (Fig 1).
540
541 **Generating *bam* null alleles for non-*melanogaster* species**
542 We believe that there are three main possibilities that could explain the patterns of
543 positive selection we have observed at *bam* across the *Drosophila* genus. The first is
544 that *bam* function is conserved across all of the species sampled and the only lineages
545 that show signals of adaptation are those that are experiencing an evolutionary
546 pressure to which *bam* is adapting. The second possibility is that *bam* function is not
547 conserved across the *Drosophila* genus, and a novel role for *bam* in GSC daughter

548 differentiation is under positive selection in those lineages. The third possibility
549 combines aspects of the first two, *bam's* function in GSC daughter differentiation is
550 novel to the *melanogaster* group, but adaptation is driven by lineage-specific factors to
551 which *bam* is responding.
552
553 As *bam* sequence is rapidly evolving across the *Drosophila* genus but we only observed
554 signals of adaptive evolution in four species all within the melanogaster group (spanning
555 approximately eight million years of divergence [46]), we wanted to distinguish between
556 the possibilities that *bam* is adapting to lineage-specific factors (i.e., endosymbionts, or
557 changes in life history), and that *bam* adapting due to acquiring a role in GSC daughter
558 differentiation. Importantly, before we can generate hypotheses about lineage-specific
559 factors that would drive the adaptation of a GSC daughter differentiation factor, we need
560 to ask if *bam* is a GSC daughter differentiation factor in that lineage. While it is difficult
561 to assess if all of *bam's* known functions in *D. melanogaster* are conserved in diverse
562 *Drosophila* species due to the lack of genetic tools in non-*melanogaster* species, we
563 can ask if ask if *bam*'s core function is conserved in other species by assessing if *bam*
564 is necessary for GSC daughter differentiation in females and males.
565
566 To ask if *bam* is rapidly evolving due to a novel role as the GSC daughter differentiation
567 switch and simultaneously ask if the *bam* phenotype is conserved to test hypotheses
568 about lineage-specific selection pressures, we designed and generated a series of *bam*
569 null alleles in diverse species in the melanogaster species group to assay for GSC
570 daughter differentiation (Fig 2A). To do this, we used CRISPR/Cas9 to introduce a

571 3xP3-YFP or a 3xP3-DsRed gene cassette into the first exon of *bam*, thereby disrupting
572 the *bam* coding sequence and introducing a premature termination codon (Fig 2B). This
573 also generates an allele that is trackable by eye color, which is necessary for the non-
574 *melanogaster* species of which we do not have balancer chromosomes to maintain
575 alleles that cause sterility. Being able to track the null allele is a concern in this design if
576 *bam* is necessary for GSC daughter differentiation in the species tested, because a loss
577 of function allele will cause sterility (as it does in *D. melanogaster* males and females).
578 To generate the homozygous *bam* null genotype, we cross the 3xP3-YFP line to the
579 3xP3-DsRed line and select flies with both DsRed and YFP positive eyes. This scheme
580 also allows us to use the same cross to assay the heterozygous and wildtype *bam*
581 siblings (Table 1).
582
583 However, we were unable to obtain a *D. simulans* 3xP3-DsRed 1$^{st}$ exon disruption line,
584 but as we did find that a *D. simulans* line with a 3xP3-DsRed insertion in its 2$^{nd}$ intron
585 resulted in homozygous sterile females and males with tumorous ovaries and testes,
586 while heterozygous females and males were fertile with wildtype ovary and testes
587 phenotypes. As this edit was specific to the *bam* locus, we presumed it was an mRNA
588 null allele and we used this line for our analyses. Thus, if the phenotype of the 3xP3-
589 DsRed 2$^{nd}$ intron allele over the 3xP3-YFP 1$^{st}$ exon allele is the same as the
590 homozygous 3xP3-DsRed 2$^{nd}$ intron allele, this is consistent with classification as a
591 genetic null.
592

593 We used *D. melanogaster* as a control to ensure our method would generate a loss of
594 function *bam* phenotype, as the *bam* null phenotype has been well characterized in this
595 species (Table 1). We also generated *bam* nulls in two other species that exhibited
596 signals of positive selection – *D. simulans* and *D. yakuba*, and two of which we did not
597 have evidence of positive selection – *D. teissieri* and *D. ananassae* (Fig 2A).
598

599 ***Bam* is necessary for female and male fertility in *D. melanogaster*, *D. simulans*,**
600 **and *D. yakuba*, but only in *D. ananassae* females, and in neither *D. teissieri* males**
601 **nor females.**

602 In *D. melanogaster*, both male and female *bam* nulls are sterile as early germline cells
603 fail to differentiate and proceed through gametogenesis. In classic null alleles of *D.*
604 *melanogaster bam* ($bam^{\Delta 59}$ and $bam^{\Delta 86}$), one copy of functional *bam* is sufficient to
605 rescue the *bam* null sterility phenotype. We first asked if our novel *bam* null alleles
606 behaved this way in *D. melanogaster*. We found that in both males and females,
607 heterozygotes of each 3xP3-DsRed or 3xP3-YFP disruption line showed similar fertility
608 to wildtype *bam* and that the *bam* null genotype was completely sterile (Fig 3, Fig S1).
609 We observed this same pattern for *D. simulans* and *D. yakuba* (Fig 3, Fig S1). However,
610 when we assessed fertility of the *D. teissieri bam* null, we found that there was no
611 difference in fertility between any of the genotypes tested for females or males (Fig 3,
612 Fig S1). The *bam* null genotype did not result in sterility, or even reduced fertility. In *D.*
613 *ananassae*, *bam* null females were sterile, and one copy of the null allele of each 3xP3-
614 DsRed or 3xP3-YFP disruption did not significantly reduce fertility, but in *D. ananassae*
615 males, *bam* nulls were fertile, with no significant mean difference in progeny across the

616 tested genotypes (Fig 3, Fig S1). Therefore, *bam* is not necessary for fertility in all
617 species in the *melanogaster* species group, or even all species within the *melanogaster*
618 species subgroup.
619
620 **Consistent with the fertility results, *bam* is also necessary for germ cell**
621 **differentiation in *D. simulans* and *D. yakuba* females and males, but not in *D.***
622 ***teissieri* males or females, and only *D. ananassae* females.**
623 In *D. melanogaster* females, *bam* null ovaries feature over-proliferating germline stem
624 cells that fail to differentiate into cystoblasts [6,52]. This results in the germarium and
625 developing downstream cysts accumulating small, GSC-like cells resembling tumors
626 (Fig 4A). In *D. melanogaster bam* null males, spermatogonia differentiation is blocked
627 which results in testes with over-proliferating spermatogonia that fail to develop into
628 terminally differentiated spermatocytes (Fig 4A) [7,53].
629
630 We imaged ovaries from 2-5 day old *D. melanogaster* females from each *bam*
631 genotype. We assayed for the characteristic *bam* null tumor phenotype using antibodies
632 to Vasa to visualize the morphology of the germline and Hts-1B1 to visualize the
633 fusome, which fails to fully form in *bam* mutants [54,55]. We found that the *bam* null
634 genotype resulted in the characteristic *bag of marbles* phenotype consistent with classic
635 *bam* null alleles, with tumorous ovaries filled with undifferentiated GSC-like cells (Fig
636 4B). One copy of wildtype *bam* was sufficient to rescue the *bam* null phenotype, as has
637 been previously observed for other *bam* null alleles (Fig 4B) [12,54]. Therefore, our
638 3xP3-DsRed and 3xP3-YFP disruptions successfully recapitulate the *bam* null

639 phenotype in females. We imaged testes from 3-5 day old *D. melanogaster* males and
640 also immunostained with antibodies to Vasa to label the germline and Hts-1B1 to label
641 the fusome. We found that the *bam* null genotype resulted in tumorous testes filled with
642 small over-proliferating cells that fail to differentiate into spermatocytes (Fig 4B). We
643 found that one copy of wildtype *bam* was sufficient to rescue the *bam* null tumor
644 phenotype, as has been previously reported for classic *bam* null alleles [56,57](Fig 4B).
645 Therefore, our *bam* null alleles also recapitulate the previously described *bam* null
646 phenotype in *D. melanogaster* males.
647
648 We repeated the same analysis for the other species in the study. We found for *D.*
649 *simulans* and *D. yakuba* both female and male homozygous *bam* null genotypes
650 exhibited the characteristic ovarian and testes tumors as described in *D. melanogaster*
651 (Fig 4B). In these two species for both males and females, one copy of *bam* was also
652 sufficient to rescue the *bam* null phenotype as observed in *D. melanogaster* (Fig 4B).
653 However, in the *D. teissieri bam* null genotype, we observed no evidence of GSC over-
654 proliferation in females or spermatogonia over-proliferation in males. We observe the
655 same phenotype for the heterozygous alleles in both males and females, as well as
656 wildtype (Fig 4B). This indicates that *bam* is not necessary for GSC daughter
657 differentiation in *D. teissieri* females nor is *bam* necessary for spermatogonia
658 differentiation in males. In *D. ananassae bam* null females, we observed the
659 characteristic *bam* null phenotype observed in *D. melanogaster* with over-proliferating
660 GSC-like cells in the germarium and cysts and undeveloped fusomes (Fig 4B). One
661 copy of *bam* was sufficient to rescue the *bam* null tumorous phenotype (Fig 4B).

662 Therefore the *D. ananassae* female *bam* null phenotype is consistent with that of *D.*
663 *melanogaster*, *D. simulans*, and *D. yakuba*. In contrast, in *D. ananassae* males of the
664 *bam* null genotype we did not observe over-proliferating spermatogonia (Fig 4B). The
665 homozygous *bam* null genotype resulted in the same phenotype as the heterozygous
666 bam null genotypes and wildtype indicating that *bam* is not necessary for the switch
667 from proliferating spermatogonia to differentiated spermatocytes in *D. ananassae*
668 males. Therefore, these cytological data are fully consistent with our fertility assay
669 observations.
670
671 **Confirming the loss-of-function status of *bam* in the *bam* null disruption design**
672 However, as the only available Bam antibody is weak in *D. melanogaster* and *D.*
673 *simulans* we cannot confirm the absence of Bam protein in all the species tested.
674 Although we used the same design to generate the *bam* null alleles across this study
675 and therefore expect the 3xP3-DsRed and 3xP3-YFP disruptions to have the same
676 effect on *bam* in all five species, we further confirmed the loss-of-function status of the
677 *bam* disruption alleles in *D. ananassae* and *D. teissieri* given their functional divergence.
678
679 First, we asked if *bam* had undergone duplication events in *D. ananassae* and *D.*
680 *teissieri* using publicly available sequence data and found no evidence of *bam*
681 duplication in either species (details described in S1 File, S8 Table).
682
683 Next, since the 3xP3-DsRed and 3xP3-YFP insertions introduce a premature
684 termination codon, we expect the mRNA of the disrupted *bam* allele to be degraded

through nonsense mediated mRNA decay (NMD), but if the premature termination codon induced exon skipping, the disrupted *bam* exon could be alternatively spliced out resulting in expression of a truncated *bam* allele, which may be sufficient for function [58–60]. We used RT-qPCR to ask if patterns of expression in the *bam* null alleles in *D. ananassae* males and *D. teissieri* females and males were consistent with expectations for NMD. We found that for *D. ananassae* and *D. teissieri*, the expression of the *bam* null disruption alleles were reduced relative to the *bam* wildtype allele to levels consistent with NMD (~1-50%, details in S1 File, Fig S2). Therefore, the mRNA of the *bam* disruption alleles are likely targeted for degradation. Additionally, since both *bam* null *D. teissieri* males and females show wildtype phenotypes, we asked if the exon disrupted by 3xP3-YFP or 3xP3-DsRed in the *bam* null genotypes was expressed in proportion to the non-disrupted exons relative to the wildtype genotype to look for evidence of exon skipping. We found that the ratio of exon 1 to exon 2 expression was not significantly different between *D. teissieri* nulls and *D. teissieri* wildtype genotypes for males and females (details in S1 File, FigS4).

Although *bam* null *D. ananassae* males and *D. teissieri* males and females exhibited wildtype phenotypes, together the genomic and RT-qPCR analyses indicate that the *bam* null alleles are loss-of-function alleles. This evidence therefore supports our conclusion that *bam* is not required for spermatogenesis in *D. ananassae* and *D. teissieri* nor oogenesis in *D. teissieri.*

708 ***D. ananassae bam* null males show evidence of germline defects**

709 As we found that *bam* was necessary for oogenesis in *D. ananassae* but not for
710 spermatogenesis, we assayed for more subtle evidence of germline defects in *D.*
711 *ananassae* males by immunostaining testes with an antibody to Fas3 which is
712 expressed in epithelial cells in the hub, the testes microenvironment that houses the
713 somatic and germline stem cell population. In *D. melanogaster bam* null males, Fas3 is
714 mis-expressed outside of the hub, as the spermatogonia over proliferate (Fig 5A&B). In
715 *D. ananassae bam* null males, we also observed misexpression of Fas3 outside of the
716 hub, although not nearly as severe as in *D. melanogaster bam* null males (Fig 5C&D).
717 This indicates that while *bam* is not necessary for differentiation in *D. ananassae* males,
718 it may be playing a role in maintenance of the stem cell population.

719

720 **Discussion**

721 **A novel role in GSC daughter differentiation is not likely the sole driver of**
722 **adaptive evolution at *bam***

723 Using representative *Drosophila* species including the *melanogaster* species group, we
724 combined population genetic analyses and functional genetic experiments to evaluate
725 the evolutionary dynamics of genetic variation at *bam* and the functional consequences
726 of its evolution. We observe additional lineages within the *melanogaster* species
727 subgroup with evidence of positive selection for amino acid diversification. Our
728 comparative assessment of *bam* requirement in gametogenesis across diverse
729 *Drosophila* species indicates that the functional evolution of *bam* is complex. The two
730 species for which we found no signal of positive selection at *bam* (*D. teissieri* and *D.*

731 *ananassae*) also do not require *bam* for gametogenesis in one or both sexes. This
732 pattern is not fully consistent with the hypothesis that positive selection is driven alone
733 by a gain in *bam* function as a GSC daughter differentiation factor. *D. ananassae* is an
734 outgroup to all of the other species tested, and thus for this hypothesis to fit, *bam* would
735 have gained and lost its function multiple times in the *melanogaster* species group. The
736 most parsimonious explanation of our functional data is that *bam* was required for
737 oogenesis and spermatogenesis in the common ancestor of the species tested (Fig 2A),
738 then *bam* function was lost in *D. ananassae* males, and then lost again independently in
739 *D. teissieri* males and females.
740
741 Therefore, we cannot fully distinguish between the possibility that the signal of adaptive
742 evolution at *bam* is driven by functional divergence and the possibility that positive
743 selection is due to conflict between *bam*'s GSC daughter differentiation function and
744 lineage-specific agents such as *W. pipientis.* However, it does seem unlikely that
745 positive selection is solely driven by a gain in GSC function. In the second scenario,
746 *bam* must be necessary for GSC function for there to be an evolutionary conflict. The
747 finding that the requirement for *bam* is lineage-specific may also indicate that lineage-
748 specific conflicts could drive functional diversification of GSC genes. One illuminating
749 test would be to ask if *D. teissieri bam* is sufficient for GSC daughter differentiation in *D.*
750 *yakuba* using the reciprocal hemizygosity test [61], or in *D. melanogaster* using
751 CRISPR/Cas9. If it is, it would be further evidence that *bam* functioned as a necessary
752 GSC daughter differentiation factor in the common ancestor of the *melanogaster*
753 species subgroup. If *D. teissieri bam* is not sufficient for gametogenesis, it is possible

754 that either its function has been completely lost due to drift in *D. teissieri*, or that it was
755 not required in the common ancestor of the *melanogaster* species group. Similar
756 experiments could be done with *D. ananassae bam* in order to further refine the timeline
757 for the emergence of *bam's* role as the switch for GSC daughter differentiation.
758
759 Additionally, our findings illustrate that *bam* ortholog function is be conserved, even for
760 though it has a critical function for fitness. As most functional genetic studies have been
761 performed only in single representative model organisms, we often do not have
762 functional confirmation for all orthologs in a comparative study. This fact is especially
763 important to consider in making and evaluating hypotheses about drivers of genetic
764 signals of positive selection. If function is not conserved across orthologs, then it is not
765 likely for that gene to be experiencing the same evolutionary pressures in each species.
766
767 **Episodic signals of positive selection at *bam*, but only in the *melanogaster* group**
768 It is notable that we have only observed signals of adaptive evolution at *bam* within the
769 melanogaster subgroup which spans approximately 10 million years of divergence, and
770 only in four lineages within this group (*D. melanogaster, D. simulans, D. santomea* and
771 *D. yakuba*) [46]. This pattern combined with our finding that *bam* function is not
772 conserved in all species tested implies that *bam* may not necessary for GSC daughter
773 differentiation in more divergent species outside of the *melanogaster* species group.
774 Therefore, testing hypotheses about the presence or absence of *W. pipientis* or other
775 factors to explain the episodic amino acid diversification at *bam* should not include
776 comparisons with these divergent species.

We currently have the most evidence in the *D. melanogaster* lineage for pervasive positive selection at *bam*, as we detect departures from neutrality for multiple population samples from different continents, whereas in *D. simulans* we observe a significant signal of positive selection in the NC population, but not in the CA population. As discussed above, this seems likely to be due an inflation of polymorphism, including nonsynonymous variants, associated with strongly nonequilibrium dynamics caused by a recent bottleneck in CA. This pattern was not evident in the NC population. It remains possible that differences in sample sizes and sequencing methodologies may also contribute to the different results obtained (NC n=10 Sanger sequences, CA n=182 low-coverage genome sequences). To better understand how *bam* is evolving in *D. simulans,* it will also be beneficial to sample *bam* from additional populations in the *D. simulans* ancestral range.

For the remaining species in our study (other than *D. melanogaster*), we currently only have samples from single populations. It will be interesting to further sample *bam* and its functional interacting network in additional populations within the species described here to further understand lineage-specific functional evolution at *bam*. This may be particularly interesting for the species in the *yakuba* complex, where we see functional divergence of *bam* on a short time scale (approximately 5 million years). Additionally, there is existing sampling data for populations in these species over time in which *W. pipientis* infection frequencies increased [35]. Are there changes in allele frequencies

799 over time in these populations that can help explain regions of *bam* under selection in
800 *D. yakuba* but not in *D. teissieri*?
801
802 **The function of a key stem cell daughter differentiation factor is divergent in**
803 **closely related species**
804
805 The results of past studies of GSC evolution [3,4,26,62] and our findings here pose the
806 question of whether GSC developmental functions are frequent targets of adaptation.
807 There is increasing evidence that this is so. Selfish transposable elements are regulated
808 by piRNAs to protect the germline genome. There is natural variation in the gene *Bruno*
809 that provides tolerance to deleterious P-element expression in the genome [21]. In a
810 scRNA-seq study of male *D. melanogaster* testes, Witt et al. [63] report many fixed de
811 novo genes expressed in GSCs and early spermatogonia (where *bam* is expressed) at
812 similar levels to testis specific genes. In contrast, segregating de novo genes show
813 lower expression in GSCs and early spermatogonia implying that cell type expression
814 may impact the likelihood for a de novo gene to become fixed. As de novo gene
815 evolution is increasingly recognized as a critical mode of adaptation, especially in
816 reproduction, GSCs/early spermatogonia may be generating necessary variation for
817 adaptation [63]. A recent finding that the insect gene *oskar,* which is critical for germ
818 cell formation and oogenesis, arose *de novo* through a horizontal gene transfer event
819 from a prokaryote, resulting in a fusion between prokaryotic and eukaryotic sequence
820 illustrates yet another mechanism by which novel genes can take on key developmental
821 roles in reproduction [64,65].

822

823 GSCs and early differentiating cells exhibit high gene expression both in number of
824 genes and transcript abundance which likely contributes to novel transcript expression
825 [63]. Novel transcripts in the early germline could provide the substrate for evolutionary
826 responses to intracellular agents like *W. pipientis*, TEs, and viruses. For example, a
827 novel gene product may physically interact with the intracellular agent to block its
828 function. Or, as many novel genes evolve connections with conserved networks, a novel
829 gene that integrates into a functional network manipulated by *W. pipientis* may eliminate
830 or reduce the ability of *W. pipientis* to recognize and manipulate that function. In either
831 case, we may predict that gene to evolve in an arms race with the infectious intracellular
832 agents, resulting in positive selection for alleles that evade manipulation. Another
833 consequence of these interactions could be the evolution of sex-biased gene function.
834 While many of the genes involved in GSC daughter differentiation and maintenance are
835 similar between females and males, many have sex-specific interactions, functions, or
836 expression patterns [66]. If a gene is evolving in conflict with an intracellular pathogen
837 that differentially affects females and males, this could lead to sex-biased function.
838

839 In addition to *bam,* there is other evidence that essential GSC genes may be gained
840 and lost. For example, unpublished data from our lab indicates that the *fs(1)Yb* ortholog
841 is missing in the *D. pseudoobscura, D. persimilis* and *D. miranda* lineage (Jae Young
842 Choi, personal communication). This gene is expressed in cap cells which adhere to the
843 GSCs in the germarium to maintain self-renewal [67]. In the cap cells, *fs(1)Yb* functions
844 in GSC maintenance as well as in the regulation of transposable elements [19]. In the

845 *D. melanogaster* and *D. simulans* lineages, *fs(1)Yb* shows a strong signal of amino acid
846 diversification [12], and in *D. melanogaster*, females null for *fs(1)Yb* are sterile [68].
847 Additionally, in *D. melanogaster*, *bam* and *ovarian tumor* (*otu*) share some sequence
848 homology, suggesting they may have evolved from a single locus, or perhaps the
849 physical properties of these proteins are common in novel proteins that gain roles in
850 GSC regulation [6].
851
852 It certainly seems plausible that genetic conflict is one way for new genes to evolve
853 required functions in conserved processes such as GSC daughter differentiation. The *D.*
854 *melanogaster* testes and the accessory gland show enriched expression of novel
855 genes, many of which show signals of adaptive evolution, and some of which are
856 predicted to be driven by sexual conflict [63,69–73].
857
858 We believe it is worth further exploring if stem cells that function in other developmental
859 processes outside of reproduction that also show signals of adaptive evolution are
860 functionally divergent, as Choi et al. [4] also found levels of adaptive evolution in neural
861 stem cell genes were similar to what we observe in germline stem cell genes. Perhaps
862 the genes that regulate differentiation and maintenance of stem cells provide adaptive
863 potential for functional innovation in critical developmental processes. In fact, two
864 studies report that adaptively evolving neural stem cells are enriched for roles in
865 neurogenesis, the process of neural stem cell differentiation into neurons and glia
866 [4,74]. *D. melanogaster* intestinal stem cells are a growing model for understanding
867 epithelial stem cell regulation, and studies have shown that the midgut responds to

868 nutritional changes by altering modes of stem cell division to regulate midgut growth
869 [75]. As the *Drosophila* gut is in contact with the environment and documented to
870 respond to changes in food, microbiota, and pathogens, intestinal stem cell regulatory
871 genes may also play roles in adaptive responses.
872
873
874 **Acknowledgements:**
875 We thank Santiago Herrera Alvarez for resequencing *bam* from *D. affinis* and Jae
876 Young Choi for resequencing *bam* from *D. pseudoobscura* and *D. mojavensis.* We
877 thank Daniel Matute, Brandon Cooper, Robert Unckless, Erin Kelleher, and Luciano
878 Matzkin for the population samples they provided for this study. Additionally, we thank
879 David Stern for his advice and guidance in designing the *bam* null alleles. We are
880 grateful to Mariana F. Wolfner, Andrew G. Clark, Miwa Wenzel and Catherine
881 Kagemann for valuable input and discussion of these experiments and results and
882 helpful comments on this manuscript.

883 REFERENCES


884 1. Kahney EW, Snedeker JC, Chen X. Regulation of Drosophila germline stem cells.
885 Curr Opin Cell Biol. 2019;60: 27–35. doi:10.1016/j.ceb.2019.03.008
886 2. Gleason RJ, Anand A, Kai T, Chen X. Protecting and diversifying the germline.
887 Genetics. 2018;208: 435–471. doi:10.1534/genetics.117.300208
888 3. Dumont VLB, Flores HA, Wright MH, Aquadro CF. Recurrent Positive Selection at
889 Bgcn, a Key Determinant of Germ Line Differentiation, Does Not Appear to be
890 Driven by Simple Coevolution with Its Partner Protein Bam. [cited 14 Apr 2020].
891 doi:10.1093/molbev/msl141
892 4. Choi JY, Aquadro CF. Molecular Evolution of Drosophila Germline Stem Cell and
893 Neural Stem Cell Regulating Genes. [cited 14 Apr 2020]. doi:10.1093/gbe/evv207
894 5. Flores HA, DuMont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, et al. Adaptive
895 evolution of genes involved in the regulation of germline stem cells in Drosophila
896 melanogaster and D. simulans. G3 Genes, Genomes, Genet. 2015;5: 583–592.
897 doi:10.1534/g3.114.015875
898 6. McKearin DM, Spradling AC. Bag-of-marbles: A Drosophila gene required to
899 initiate both male and female gametogenesis. Genes Dev. 1990;4: 2242–2251.
900 doi:10.1101/gad.4.12b.2242
901 7. Insco ML, Leon A, Tam CH, McKearin DM, Fuller MT. Accumulation of a
902 differentiation regulator specifies transit amplifying division number in an adult
903 stem cell lineage. Proc Natl Acad Sci U S A. 2009. doi:10.1073/pnas.0912454106
904 8. Li Y, Minor NT, Park JK, McKearin DM, Maines JZ. Bam and Bgcn antagonize
905 Nanos-dependent germ-line stem cell maintenance. Proc Natl Acad Sci U S A.

906 2009;106: 9304–9309. doi:10.1073/pnas.0901452106
907 9. Ting X. Control of germline stem cell self-renewal and differentiation in the
908 *Drosophila* ovary: concerted actions of niche signals and intrinsic factors. Wiley
909 Interdiscip Rev Dev Biol. 2013;2: 261–273. doi:10.1002/wdev.60
910 10. Chen D, McKearin D. Dpp Signaling Silences bam Transcription Directly to
911 Establish Asymmetric Divisions of Germline Stem Cells. Curr Biol. 2003.
912 doi:10.1016/j.cub.2003.09.033
913 11. Chen D, Wu C, Zhao S, Geng Q, Gao Y, Li X, et al. Three RNA Binding Proteins
914 Form a Complex to Promote Differentiation of Germline Stem Cell Lineage in
915 Drosophila. PLoS Genet. 2014. doi:10.1371/journal.pgen.1004797
916 12. Flores HA, Bubnell JE, Aquadro CF, Barbash DA. The Drosophila bag of marbles
917 Gene Interacts Genetically with Wolbachia and Shows Female-Specific Effects of
918 Divergence. 2015. doi:10.1371/journal.pgen.1005453
919 13. Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. Molecular population
920 genetics of male accessory gland proteins in Drosophila. Genetics. 2000.
921 14. Prokupek A, Hoffmann F, Eyun S Il, Moriyama E, Zhou M, Harshman L. An
922 evolutionary expressed sequence tag analysis of drosophila spermatheca genes.
923 Evolution (N Y). 2008. doi:10.1111/j.1558-5646.2008.00493.x
924 15. Kelleher ES, Markow TA. Duplication, selection and gene conversion in a
925 Drosophila mojavensis female reproductive protein family. Genetics. 2009.
926 doi:10.1534/genetics.108.099044
927 16. Tsaur SC, Ting CT, Wu CI. Positive selection driving the evolution of a gene of
928 male reproduction, Acp26Aa, of Drosophila: II. Divergence versus polymorphism.

929 Mol Biol Evol. 1998. doi:10.1093/oxfordjournals.molbev.a026002
930 17. Wigby S, Brown NC, Allen SE, Misra S, Sitnik JL, Sepil I, et al. The Drosophila
931 seminal proteome and its role in postcopulatory sexual selection. Philos Trans R
932 Soc Lond B Biol Sci. 2020. doi:10.1098/rstb.2020.0072
933 18. Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. Evolutionary
934 EST analysis identifies rapidly evolving male reproductive proteins in Drosophila.
935 Proc Natl Acad Sci. 2001;98: 7375–7379. doi:10.1073/pnas.131568198
936 19. Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. An in vivo
937 RNAi assay identifies major genetic and cellular requirements for primary piRNA
938 biogenesis in Drosophila. EMBO J. 2010. doi:10.1038/emboj.2010.212
939 20. Fablet M, Akkouche A, Braman V, Vieira C. Variable expression levels detected in
940 the Drosophila effectors of piRNA biogenesis. Gene. 2014.
941 doi:10.1016/j.gene.2013.11.095
942 21. Kelleher ES, Jaweria J, Akoma U, Ortega L, Tang W. QTL mapping of natural
943 variation reveals that the developmental regulator bruno reduces tolerance to P-
944 element transposition in the Drosophila female germline. PLoS Biol. 2018.
945 doi:10.1371/journal.pbio.2006040
946 22. Werren JH, Baldo L, Clark ME. Wolbachia: Master manipulators of invertebrate
947 biology. Nature Reviews Microbiology. 2008. doi:10.1038/nrmicro1969
948 23. Newton ILG, Savytskyy O, Sheehan KB. Wolbachia Utilize Host Actin for Efficient
949 Maternal Transmission in Drosophila melanogaster. PLoS Pathog. 2015.
950 doi:10.1371/journal.ppat.1004798
951 24. Ote M, Yamamoto D. Impact of Wolbachia infection on Drosophila female

952 germline stem cells. Current Opinion in Insect Science Elsevier Inc.; Feb 1, 2020
953 pp. 8–15. doi:10.1016/j.cois.2019.10.001
954 25. Serbus LR, Sullivan W. A cellular basis for Wolbachia recruitment to the host
955 germline. PLoS Pathog. 2007. doi:10.1371/journal.ppat.0030190
956 26. Choi JY, Aquadro CF. The coevolutionary period of Wolbachia pipientis infecting
957 Drosophila ananassae and its impact on the evolution of the host germline stem
958 cell regulating genes. Mol Biol Evol. 2014. doi:10.1093/molbev/msu204
959 27. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al.
960 FlyBase 2.0: The next generation. Nucleic Acids Res. 2019.
961 doi:10.1093/nar/gky1003
962 28. Wilson BA, Foy SG, Neme R, Masel J. Young genes are highly disordered as
963 predicted by the preadaptation hypothesis of de novo gene birth. Nat Ecol Evol.
964 2017. doi:10.1038/s41559-017-0146
965 29. Jiang X, Assis R. Natural Selection Drives Rapid Functional Evolution of Young
966 Drosophila Duplicate Genes. [cited 7 May 2020]. doi:10.1093/molbev/msx230
967 30. Chen S, Zhang YE, Long M. New Genes in Drosophila Quickly Become Essential.
968 Science (80- ). 2010;330: 1682–1685. doi:10.1126/science.1196380
969 31. Vakirlis N, McLysaght A. Computational Prediction of De Novo Emerged Protein-
970 Coding Genes. Methods in Molecular Biology. Humana Press Inc.; 2019. pp. 63–
971 81. doi:10.1007/978-1-4939-8736-8_4
972 32. Klasberg S, Bitard-Feildel T, Callebaut I, Bornberg-Bauer E. Origins and structural
973 properties of novel and *de novo* protein domains during insect evolution. FEBS J.
974 2018;285: 2605–2625. doi:10.1111/febs.14504

975 33. Turissini DA, Liu G, David JR, Matute DR. The evolution of reproductive isolation
976 in the Drosophila yakuba complex of species. J Evol Biol. 2015.
977 doi:10.1111/jeb.12588
978 34. Glover DM, Raff J, Karr TL, O'Neill SL, Lin H, Wolfner MF. Parasites in Drosophila
979 embryos. Nature. 1990. doi:10.1038/348117a0
980 35. Cooper BS, Ginsberg PS, Turelli M, Matute DR. Wolbachia in the Drosophila
981 yakuba complex: Pervasive frequency variation and weak cytoplasmic
982 incompatibility, but no apparent effect on reproductive isolation. Genetics. 2017.
983 doi:10.1534/genetics.116.196238
984 36. Schaeffer SW, Miller EL. Nucleotide sequence analysis of Adh genes estimates
985 the time of geographic isolation of the Bogota population of Drosophila
986 pseudoobscura. Proc Natl Acad Sci U S A. 1991. doi:10.1073/pnas.88.14.6097
987 37. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et
988 al. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair
989 in Drosophila. Genetics. 2014;196: 961–971. doi:10.1534/genetics.113.160713
990 38. Port F, Chen HM, Lee T, Bullock SL. Optimized CRISPR/Cas tools for efficient
991 germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S
992 A. 2014. doi:10.1073/pnas.1405500111
993 39. Ho J, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. Moving beyond P values:
994 data analysis with estimation graphics. Nature Methods. 2019.
995 doi:10.1038/s41592-019-0470-3
996 40. Aruna S, Flores HA, Barbash DA. Reduced fertility of Drosophila melanogaster
997 Hybrid male rescue (Hmr) mutant females is partially complemented by Hmr

998 orthologs from sibling species. Genetics. 2009. doi:10.1534/genetics.108.100057
999 41. Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, et
1000 al. The drosophila genome nexus: A population genomic resource of 623
1001 Drosophila melanogaster genomes, including 197 from a single ancestral range
1002 population. Genetics. 2015. doi:10.1534/genetics.115.174664
1003 42. Signor SA, New FN, Nuzhdin S. A large panel of drosophila simulans reveals an
1004 abundance of common variants. Genome Biol Evol. 2018.
1005 doi:10.1093/gbe/evx262
1006 43. Egea R, Casillas S, Barbadilla A. Standard and generalized McDonald-Kreitman
1007 test: a website to detect selection by comparing different classes of DNA sites.
1008 Nucleic Acids Res. 2008. doi:10.1093/nar/gkn337
1009 44. Charlesworth J, Eyre-Walker A. The McDonald-Kreitman test and slightly
1010 deleterious mutations. Mol Biol Evol. 2008. doi:10.1093/molbev/msn005
1011 45. Eyre-Walker A. The genomic rate of adaptive evolution. Trends in Ecology and
1012 Evolution. 2006. doi:10.1016/j.tree.2006.06.015
1013 46. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al.
1014 Evolution of genes and genomes on the Drosophila phylogeny. Nature. 2007.
1015 doi:10.1038/nature06341
1016 47. Siddiq MA, Loehlin DW, Montooth KL, Thornton JW. Experimental test and
1017 refutation of a classic case of molecular adaptation in Drosophila melanogaster.
1018 Nat Ecol Evol. 2017;1. doi:10.1038/s41559-016-0025
1019 48. Aquadro CF, Lado KM, Noon WA. The rosy region of Drosophila melanogaster
1020 and Drosophila simulans. I. Contrasting levels of naturally occurring DNA

1021 restriction map variation and divergence. Genetics. 1988.
1022 49. Hoffmann AA, Turelli M, Simmons GM. Unidirectional incompatibility between
1023 populations of Drosophila simulans. Evolution (N Y). 1986. doi:10.1111/j.1558-
1024 5646.1986.tb00531.x
1025 50. Cariou M, Duret L, Charlat S. The global impact of Wolbachia on mitochondrial
1026 diversity and evolution. J Evol Biol. 2017. doi:10.1111/jeb.13186
1027 51. Messer PW, Petrov DA. Frequent adaptation and the McDonald-Kreitman test.
1028 Proc Natl Acad Sci U S A. 2013. doi:10.1073/pnas.1220835110
1029 52. Ting X. Control of germline stem cell self-renewal and differentiation in the
1030 *Drosophila* ovary: concerted actions of niche signals and intrinsic factors. Wiley
1031 Interdiscip Rev Dev Biol. 2013;2: 261–273. doi:10.1002/wdev.60
1032 53. Insco ML, Bailey AS, Kim J, Olivares GH, Wapinski OL, Tam CH, et al. A self-
1033 limiting switch based on translational control regulates the transition from
1034 proliferation to differentiation in an adult stem cell lineage. Cell Stem Cell. 2012.
1035 doi:10.1016/j.stem.2012.08.012
1036 54. Lavoie CA, Ohlstein B, McKearin DM. Localization and function of bam protein
1037 require the benign gonial cell neoplasm gene product. Dev Biol. 1999.
1038 doi:10.1006/dbio.1999.9346
1039 55. Hinnant TD, Merkle JA, Ables ET. Coordinating Proliferation, Polarity, and Cell
1040 Fate in the Drosophila Female Germline. Frontiers in Cell and Developmental
1041 Biology. Frontiers Media S.A.; 2020. p. 19. doi:10.3389/fcell.2020.00019
1042 56. Gönczy P, Matunis E, DiNardo S. bag-of-marbles and benign genial cell
1043 neoplasm act in the germline to restrict proliferation during Drosophila

1044 spermatogenesis. Development. 1997.
1045 57. Shivdasani AA, Ingham PW. Regulation of Stem Cell Maintenance and Transit
1046 Amplifying Cell Proliferation by TGF-β Signaling in Drosophila Spermatogenesis.
1047 Curr Biol. 2003. doi:10.1016/j.cub.2003.10.063
1048 58. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nature
1049 Structural and Molecular Biology. 2009. doi:10.1038/nsmb.1550
1050 59. Mou H, Smith JL, Peng L, Yin H, Moore J, Zhang XO, et al. CRISPR/Cas9-
1051 mediated genome editing induces exon skipping by alternative splicing or exon
1052 deletion. Genome Biol. 2017. doi:10.1186/s13059-017-1237-8
1053 60. Chen D, Tang JX, Li B, Hou L, Wang X, Kang L. CRISPR/Cas9-mediated genome
1054 editing induces exon skipping by complete or stochastic altering splicing in the
1055 migratory locust. BMC Biotechnol. 2018. doi:10.1186/s12896-018-0465-7
1056 61. Stern DL. Identification of loci that cause phenotypic variation in diverse species
1057 with the reciprocal hemizygosity test. Trends in Genetics. 2014.
1058 doi:10.1016/j.tig.2014.09.006
1059 62. Flores HA, Dumont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, et al. Adaptive
1060 Evolution of Genes Involved in the Regulation of Germline Stem Cells in
1061 Drosophila melanogaster and D. simulans. 2015 [cited 14 Apr 2020].
1062 doi:10.1534/g3.114.015875
1063 63. Witt E, Benjamin S, Svetec N, Zhao L. Testis single-cell RNA-seq reveals the
1064 dynamics of de novo gene transcription and germline mutational bias in
1065 drosophila. Elife. 2019. doi:10.7554/eLife.47138
1066 64. Jenny A, Hachet O, Závorszky P, Cyrklaff A, Weston MDJ, St Johnston D, et al. A

1067 translation-independent role of oskar RNA in early Drosophila oogenesis.
1068 Development. 2006;133: 2827–2833. doi:10.1242/dev.02456
1069 65. Blondel L, Jones TEM, Extavour CG. Bacterial contribution to genesis of the novel
1070 germ line determinant Oskar. Elife. 2020. doi:10.7554/eLife.45539
1071 66. Cash AC, Andrews J. Fine scale analysis of gene expression in Drosophila
1072 melanogaster gonads reveals Programmed cell death 4 promotes the
1073 differentiation of female germline stem cells. BMC Dev Biol. 2012.
1074 doi:10.1186/1471-213X-12-4
1075 67. Kirilly D, Xie T. The Drosophila ovary: An active stem cell community. Cell
1076 Research. 2007. doi:10.1038/sj.cr.7310123
1077 68. Swan A, Hijal S, Hilfiker A, Suter B. Identification of new x-chromosomal genes
1078 required for Drosophila oogenesis and novel roles for fs(1)Yb, brainiac and dunce
1079 [4]. Genome Research. 2001. doi:10.1101/gr.156001
1080 69. Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. Novel genes derived
1081 from noncoding DNA in Drosophila melanogaster are frequently X-linked and
1082 exhibit testis-biased expression. PNAS. 2006. Available:
1083 www.pnas.orgcgidoi10.1073pnas.0509809103
1084 70. Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously
1085 unannotated, rapidly evolving seminal fluid genes in Drosophila. Genome Res.
1086 2009;19: 886–896. doi:10.1101/gr.089391.108
1087 71. Gubala AM, Schmitz JF, Kearns MJ, Vinh TT, Bornberg-Bauer E, Wolfner MF, et
1088 al. The Goddard and Saturn Genes Are Essential for Drosophila Male Fertility and
1089 May Have Arisen de Novo. Mol Biol Evol. 2017. doi:10.1093/molbev/msx057

1090 72. Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, et al. On the origin of new
1091 genes in Drosophila. Genome Res. 2008;18: 1446–1455.
1092 doi:10.1101/gr.076588.108
1093 73. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. Genome
1094 Research. Cold Spring Harbor Laboratory Press; 2010. pp. 1313–1326.
1095 doi:10.1101/gr.101386.109
1096 74. Langley CH, Stevens K, Cardeno C, Lee YCG, Schrider DR, Pool JE, et al.
1097 Genomic variation in natural populations of Drosophila melanogaster. Genetics.
1098 2012. doi:10.1534/genetics.112.142018
1099 75. O'Brien LE, Soliman SS, Li X, Bilder D. Altered modes of stem cell division drive
1100 adaptive intestinal growth. Cell. 2011. doi:10.1016/j.cell.2011.08.048
1101
1102

**Fig 1. *Bam* shows lineage-specific signals of positive selection within the melanogaster species subgroup and no evidence of positive selection outside of the melanogaster species subgroup.** Cladogram of *Drosophila* species we sampled in this study or previously published. Red lineages represent lineages where we have evidence of positive selection at *bam* with the McDonald Kreitman test. Blue lineages are lineages of which we could not reject neutrality with the McDonald Kreitman test. The 2X2 table from the McDonald Kreitman test is shown on the right (N = nonsynonymous, S = synonymous). Alpha is the proportion of nonsynonymous fixations predicted to be due to positive selection. p-value is from the chi-sq test, n is sample size. [1 *]2X2 table and statistics from Bauer DuMont et al. [3] [2]polymorphism data from Choi et al. [24]

**Fig 2. Species chosen to generate *bam* null alleles using CRISPR/Cas9 to disrupt *bam* with a trackable eye marker cassette.** (A) The five species we chose to generate *bam* null alleles span approximately 15 million years of divergence and show different patterns of evolution at *bam*. The lineages in red show evidence of positive selection (MKT), we did not observe evidence of positive selection for the lineages in blue. (B) Schematic for generating null alleles with trackable eye markers using CRISPR/Cas9. We generate two null alleles using different color eye markers and cross the heterozygotes together to generate an easily trackable null genotype.

**Fig 3. *Bam* is necessary for female and male fertility in *D. melanogaster, D. simulans, and D. yakuba,* but not for *D. ananassae* males nor both *D. teissieri* males and females.** Swarm plots showing the total progeny per female or per male for each *bam* genotype. Dotted boxes show genotypes where the phenotype did not match that of *D. melanogaster.* For *D. melanogaster, D. simulans, and D. yakuba*, the *bam* null DsRed/YFP females and males were sterile ($P < 0.0001$, permutation test), and the heterozygous genotypes did not have a significant effect on fertility ($P > 0.05$, permutation test). For *D. teissieri,* the *bam* null DsRed/YFP females and males did not have significantly different fertility from wildtype, nor did the heterozygotes. For *D. ananassae*, the *bam* null DsRed/YFP females were sterile ($P < 0.0001$, permutation test), and the heterozygotes did not affect fertility. However, the *bam* null DsRed/YFP males were fertile and there was no significant difference between the null and heterozygotes and wildtype fertility ($P>0.05$, permutation test).

**Fig 4. *Bam* is necessary for female and male germline stem cell differentiation in *D. melanogaster, D. simulans, and D. yakuba,* but not for *D. ananassae* males nor *D. teissieri* males and females.** (A) Schematics of ovaries and testes showing the *bam* wildtype and null cytological phenotypes in *D. melanogaster*. Wildtype *bam* ovaries exhibit developing egg chambers with differentiated nurse cells and testes exhibit large differentiated spermatocytes. Null *bam* ovaries exhibit egg chambers filled with small GSC-like tumors that contain undeveloped fusomes and testes exhibit cysts filled with small over-proliferating spermatogonia. (B) Ovaries and testes from all five species of wildtype ($bam^+/bam^+$), heterozygous (YFP/+ or DsRed/+), and null (DsRed/YFP) *bam* genotypes. All tissue is labeled with antibodies to Vasa (germline) and Hts-1B1 (fusome). Dashed boxes indicate genotypes and tissue that are not consistent with the *D. melanogaster* phenotype. Ovaries and testes from all five species exhibit developing nurse cells or spermatocytes (respectively) for wildtype and heterozygous *bam* genotypes. We did not observe any tumorous cysts (arrowheads). Ovaries and testes from *bam* null *D. melanogaster, D. simulans, and D. yakuba* all exhibited cysts filled with GSC-like tumors (ovaries) or over-proliferating spermatogonia (testes) (arrows). The *bam* null ovaries exhibit undeveloped fusomes. *D. ananassae bam* null ovaries exhibited the tumorous egg chamber phenotype (arrow), however testes exhibited large developing spermatocytes (arrowhead) and

we did not observe any tumors. The *bam* null *D. teissieri* ovaries and testes exhibited nurse cell positive egg chambers and developing spermatocytes (respectively) (arrowheads) and we did not observe any tumors.

**Fig 5. Loss of *bam* function in *D. ananassae* males shows mis-expression of Fas3, a marker expressed in the hub.** All testes tissue is labeled with antibodies for Vasa (germline) and Fas3 (hub, arrowheads). (A) *D. melanogaster bam* wildtype testes with Fas3 expression limited to the hub (inset) and featuring large differentiating spermatogonia. (B) *D. melanogaster bam* null testes feature ectopic Fas3 expression and small over-proliferating spermatogonia. (C) *D. ananassae bam* wildtype testes with Fas3 expression limited to the hub and featuring large differentiating spermatogonia. (D) *D. ananassae bam* null testes with ectopic Fas3 expression and large differentiating spermatogonia.

**Fig S1. Extended plots from Fig 3 including the Cumming estimation plot below the swarm plot.** *bam* is necessary for female and male fertility in *D. melanogaster, D. simulans, and D. yakuba,* but not for *D. ananassae* males nor both *D. teissieri* males and females. The Cumming estimation plot displays the 95% confidence interval effect size. As stated in Fig 3*, D. melanogaster, D. simulans, and D. yakuba*, the *bam* null DsRed/YFP females and males were sterile ($P < 0.0001$, permutation test), and the heterozygous genotypes did not have a significant effect on fertility ($P > 0.05$, permutation test). For *D. teissieri,* the *bam* null DsRed/YFP females and males did not have significantly different fertility from wildtype, nor did the heterozygotes. For *D. ananassae*, the *bam* null DsRed/YFP females were sterile ($P < 0.0001$, permutation test), and the heterozygotes did not affect fertility. However, the *bam* null DsRed/YFP males were fertile and there was no significant difference between the null and heterozygotes and wildtype fertility ($P>0.05$, permutation test)

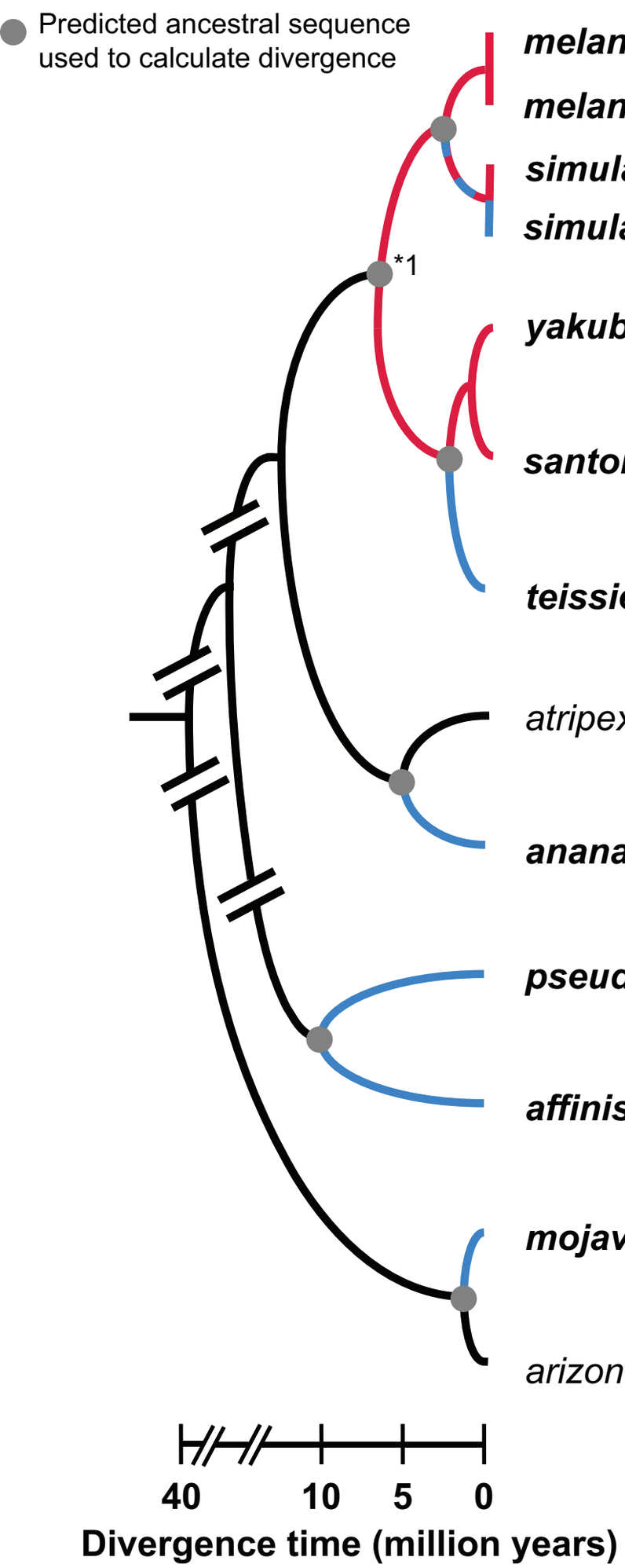
**Fig S2. RT-qPCR experiment to assess *bam* mRNA in *D. ananassae bam* null genotypes reported as the relative quantity (RQ) of *bam* expression to the housekeeping control RP49.** *bam* expression in testes of the null (DsRed/YFP) genotype is significantly reduced (standard error) relative to *bam* expression in testes of the wildtype genotype at approximately 10%, consistent with expectations for NMD. The heterozygous genotypes DsRed/+ and YFP/+ do not show reduced *bam* expression relative to wildtype.

**Fig S3. RT-qPCR experiment to detect patterns consistent with nonsense mediated mRNA decay of the *bam* disruption allele in *D. teissieri* reported as the relative quantity (RQ) of mRNA from the *bam* disruption allele to the *bam* wildtype allele in heterozygotes.** Significance reported as standard error (A) Expression of the *bam* null allele relative to the *bam* wildtype allele in ovaries. Both heterozygous genotypes DsRed/+ and YFP/+ show less than 10% of expression of the *bam* null allele relative to the wildtype allele, consistent with expectations for NMD. The wildtype negative control shows little to no expression of the null allele as expected. (B) Expression of the *bam* null allele relative to the *bam* wildtype allele in testes. Both heterozygous genotypes DsRed/+ and YFP/+ show less than 50% of expression of the *bam* null allele relative to the wildtype allele, consistent with expectations for NMD. The wildtype negative control shows little to no expression of the null allele as expected.

**Fig S4. RT-qPCR experiment to detect evidence of exon skipping of the exon containing the *bam* disruption reported as the relative quantity (RQ) of *bam* exon 1 to exon 2.** Significance reported as standard error (A) Expression of *bam* exon 1 relative to exon 2 in ovaries. The RQ of exon 1 to exon 2 in the *bam* null DsRed/YFP genotype is not significantly different than the RQ of exon 1 to exon 2 in the *bam* wildtype genotype, indicating that

expression of the exon containing the disruption is comparable to the unaffected exon in both the null and wildtype genotype, and therefore exon skipping is not likely occurring in females. (B) Expression of *bam* exon 1 relative to exon 2 in testes. The RQ of exon 1 to exon 2 in the *bam* null DsRed/YFP genotype is not significantly different than the RQ of exon 1 to exon 2 in the *bam* wildtype genotype, indicating that expression of the exon containing the disruption is comparable to the unaffected exon in both the null and wildtype genotype, and therefore exon skipping is not likely occurring in males.

**Fig 1**




| | Divergence | | Polymorphism | | | | |
|---|---|---|---|---|---|---|---|
| | N | S | N | S | alpha | P-value | n |
| ***melanogaster* - ZM** | 20 | 6 | 4 | 13 | 0.907 | ***0.000*** | 187 |
| ***melanogaster** - US, ZI*[1] | 28 | 11 | 8 | 14 | 0.776 | ***0.008*** | 12,7 |
| ***simulans** - US, NC*[1] | 26 | 16 | 9 | 15 | 0.631 | ***0.049*** | 10 |
| ***simulans* - US, CA** | 46 | 25 | 13 | 15 | 0.528 | 0.093 | 182 |
| ***yakuba*- Bioko** | 26 | 18 | 10 | 25 | 0.723 | ***0.006*** | 14 |
| ***santomea*- São Tomé** | 24 | 21 | 5 | 16 | 0.726 | ***0.024*** | 11 |
| ***teissieri*- Bioko** | 12 | 13 | 14 | 15 | -0.011 | 0.983 | 11 |
| *atripex* | | | | | | | |
| ***ananassae* - Thailand**[2] | 28 | 36 | 1 | 7 | 0.816 | 0.089 | 14 |
| ***pseudoobscura* - US, CA** | 40 | 41 | 3 | 6 | 0.487 | 0.360 | 11 |
| ***affinis* - US** | 31 | 31 | 5 | 7 | 0.285 | 0.597 | 14 |
| ***mojavensis* - US** | 2 | 5 | 3 | 8 | 0.250 | 1 | 22 |
| *arizonae* | | | | | | | |

Fig 2

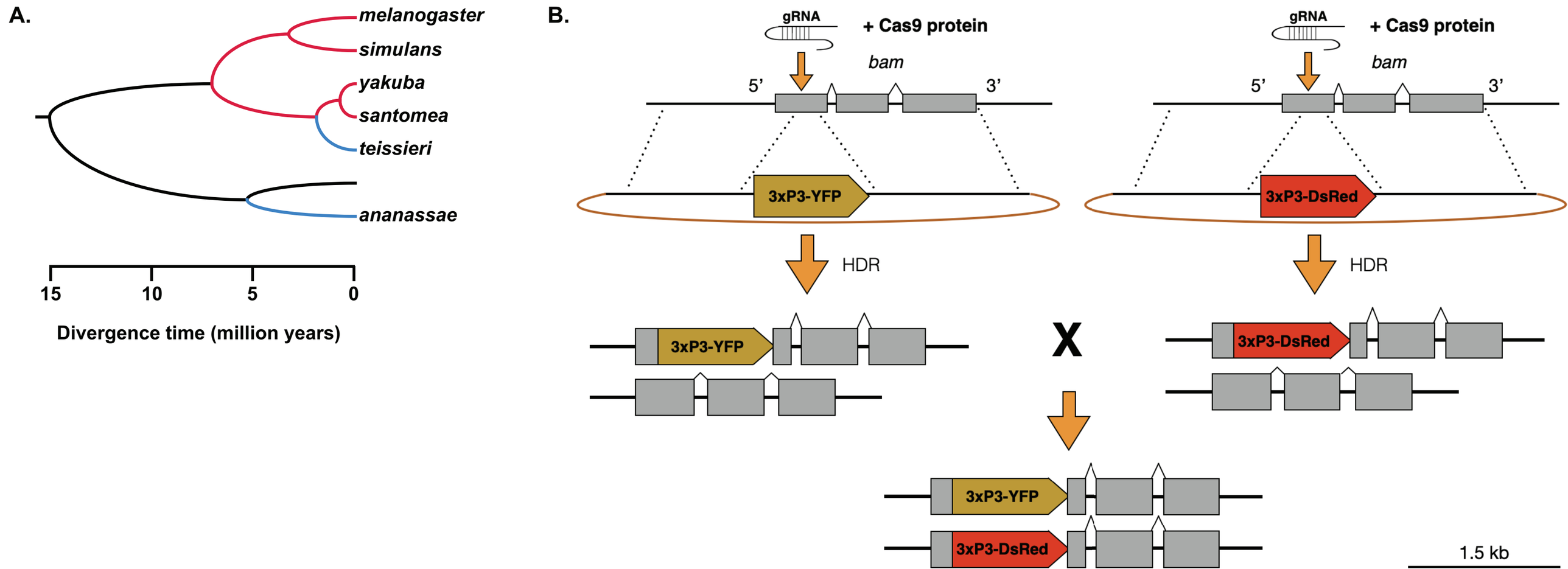

A.
melanogaster
simulans
yakuba
santomea
teissieri
ananassae
15
10
5
0
Divergence time (million years)
B.
gRNA
+ Cas9 protein
bam
5'
3'
3xP3-YFP
HDR
gRNA
+ Cas9 protein
bam
5'
3'
3xP3-DsRed
HDR
3xP3-YFP
X
3xP3-DsRed
3xP3-YFP
3xP3-DsRed
1.5 kb

**Table 1.** Abbreviations for *bam* genotypes for all species and the expected phenotypes as observed in *D. melanogaster.*

| | | ***D. melanogaster* Phenotype Expectations** | |
|---|---|---|---|
| **Genotype** | **Abbreviation** | **Fertility** | **Cytological** |
| $bam^{+}/bam^{+}$ | wildtype | wildtype | wildtype |
| $bam^{3xP3\text{-}DsRed}/bam^{+}$ | DsRed/+ | wildtype | wildtype |
| $bam^{3xP3\text{-}YFP}/bam^{+}$ | YFP/+ | wildtype | wildtype |
| $bam^{3xP3\text{-}DsRed}/bam^{3xP3\text{-}YFP}$ | DsRed/YFP | sterile | tumorous ovaries and testes |

Fig 3

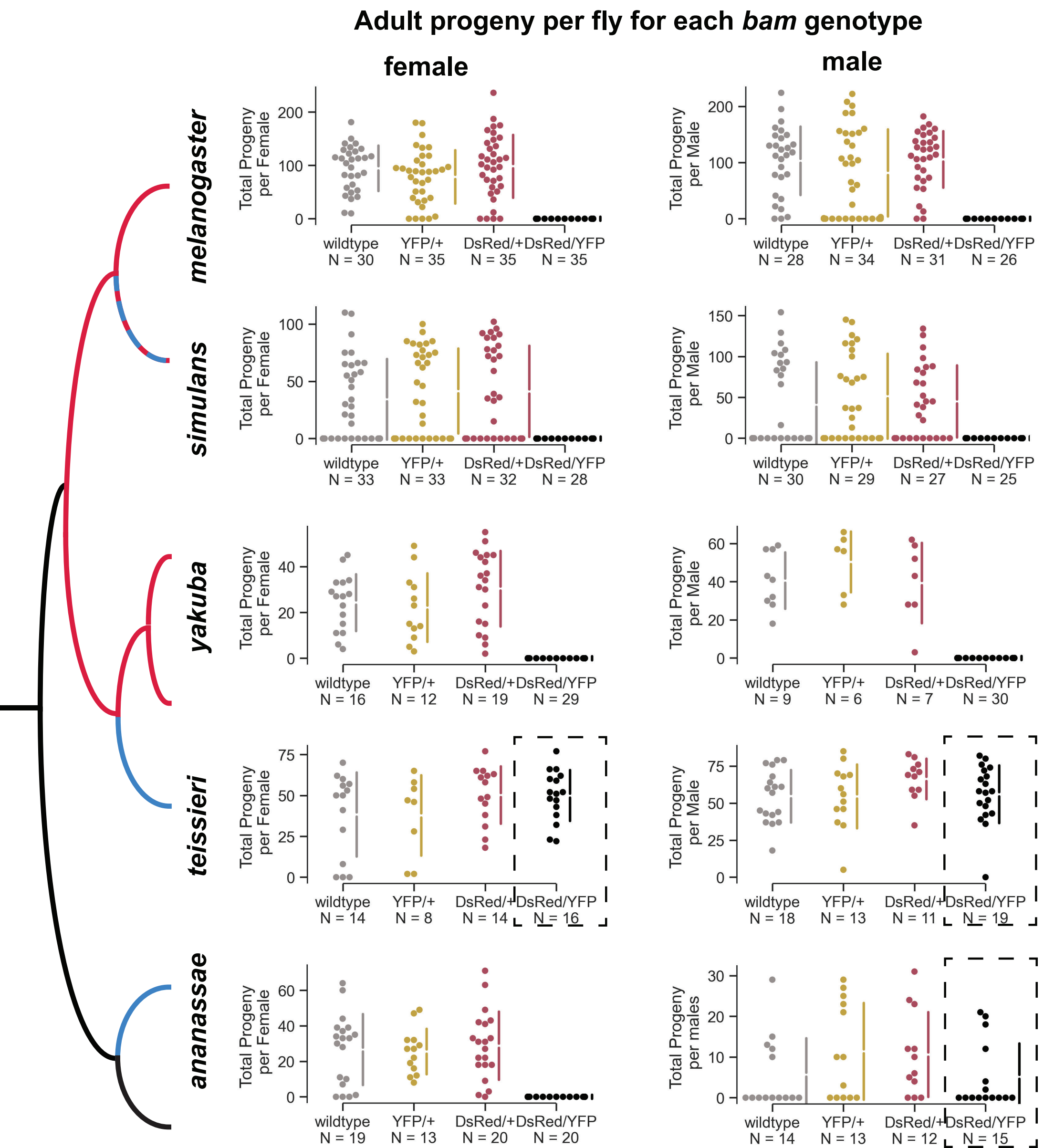
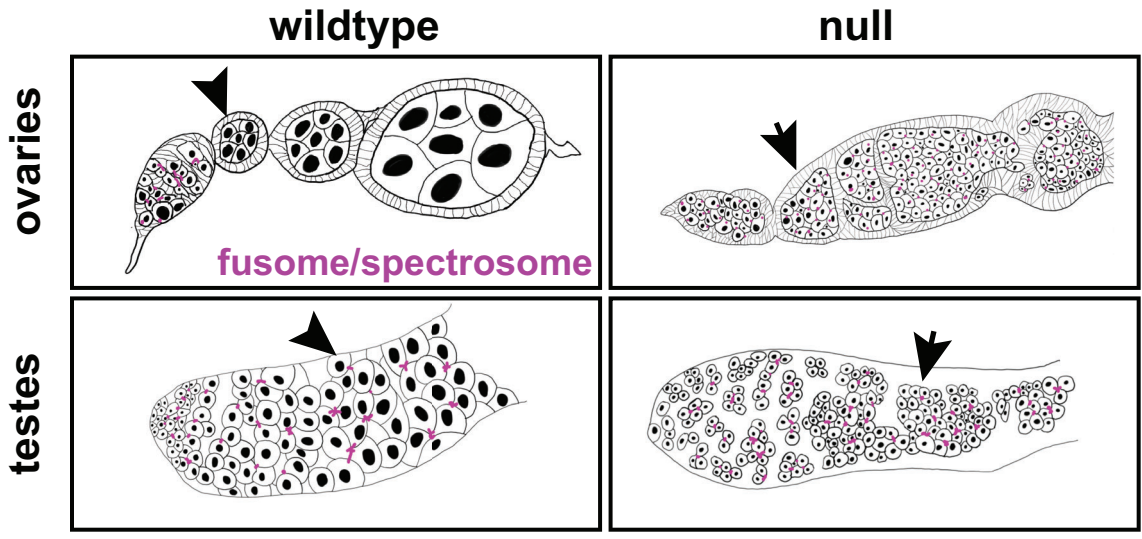
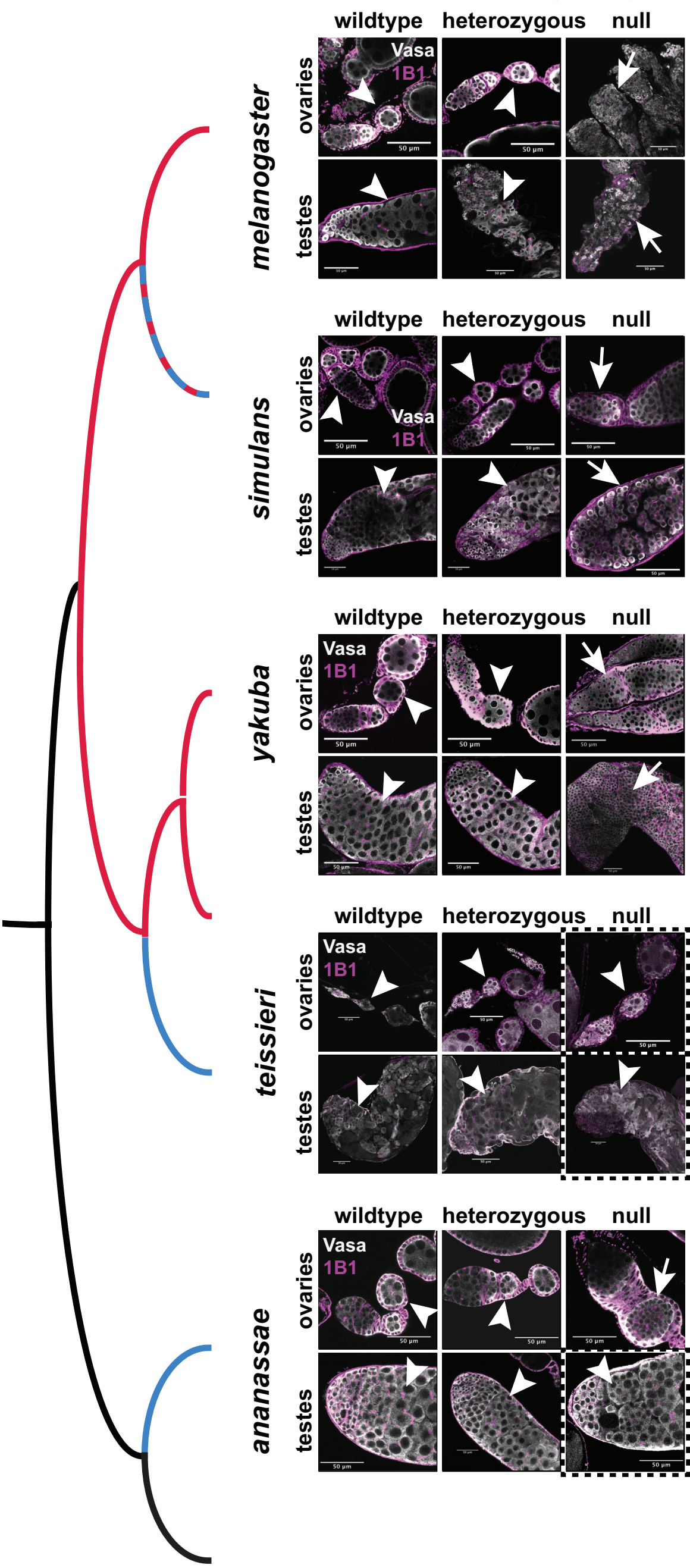

Adult progeny per fly for each bam genotype
female
male
melanogaster
simulans
yakuba
teissieri
ananassae
Total Progeny per Female
Total Progeny per Male
Total Progeny per males
wildtype N = 30
YFP/+ N = 35
DsRed/+ N = 35
DsRed/YFP N = 35
wildtype N = 28
YFP/+ N = 34
DsRed/+ N = 31
DsRed/YFP N = 26
wildtype N = 33
YFP/+ N = 33
DsRed/+ N = 32
DsRed/YFP N = 28
wildtype N = 30
YFP/+ N = 29
DsRed/+ N = 27
DsRed/YFP N = 25
wildtype N = 16
YFP/+ N = 12
DsRed/+ N = 19
DsRed/YFP N = 29
wildtype N = 9
YFP/+ N = 6
DsRed/+ N = 7
DsRed/YFP N = 30
wildtype N = 14
YFP/+ N = 8
DsRed/+ N = 14
DsRed/YFP N = 16
wildtype N = 18
YFP/+ N = 13
DsRed/+ N = 11
DsRed/YFP N = 19
wildtype N = 19
YFP/+ N = 13
DsRed/+ N = 20
DsRed/YFP N = 20
wildtype N = 14
YFP/+ N = 13
DsRed/+ N = 12
DsRed/YFP N = 15

Fig 4

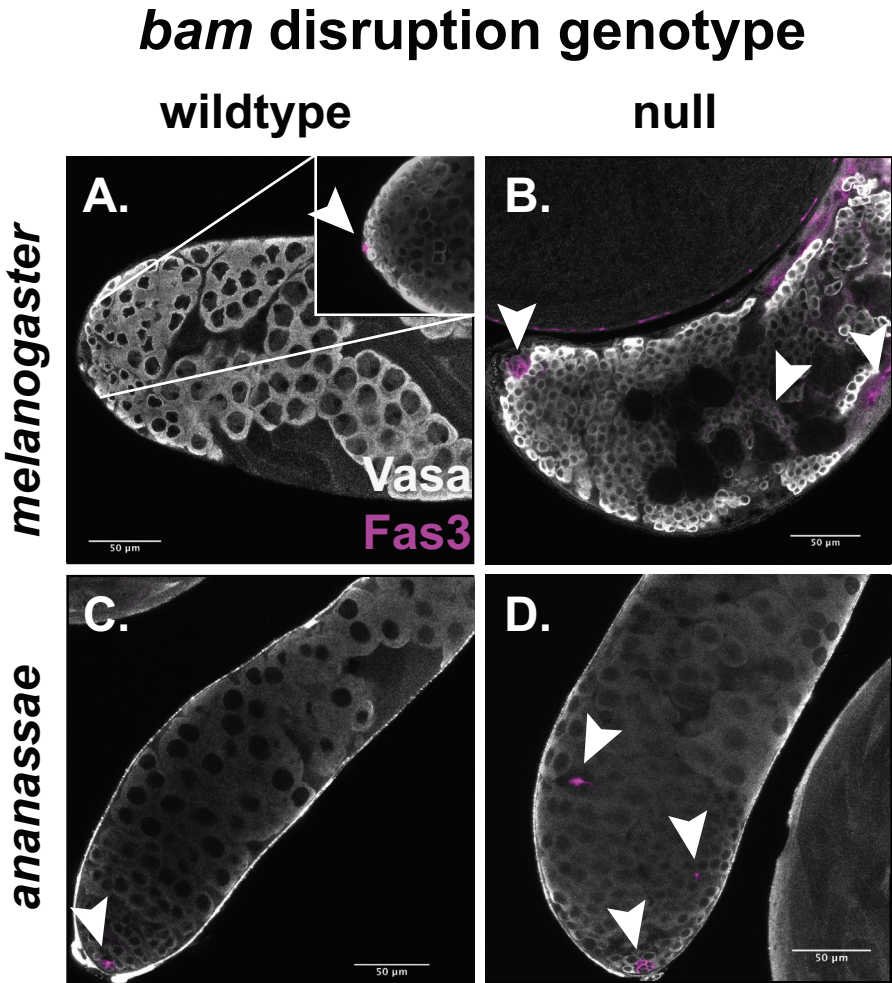
A. *D. melanogaster bam* cytological phenotypes



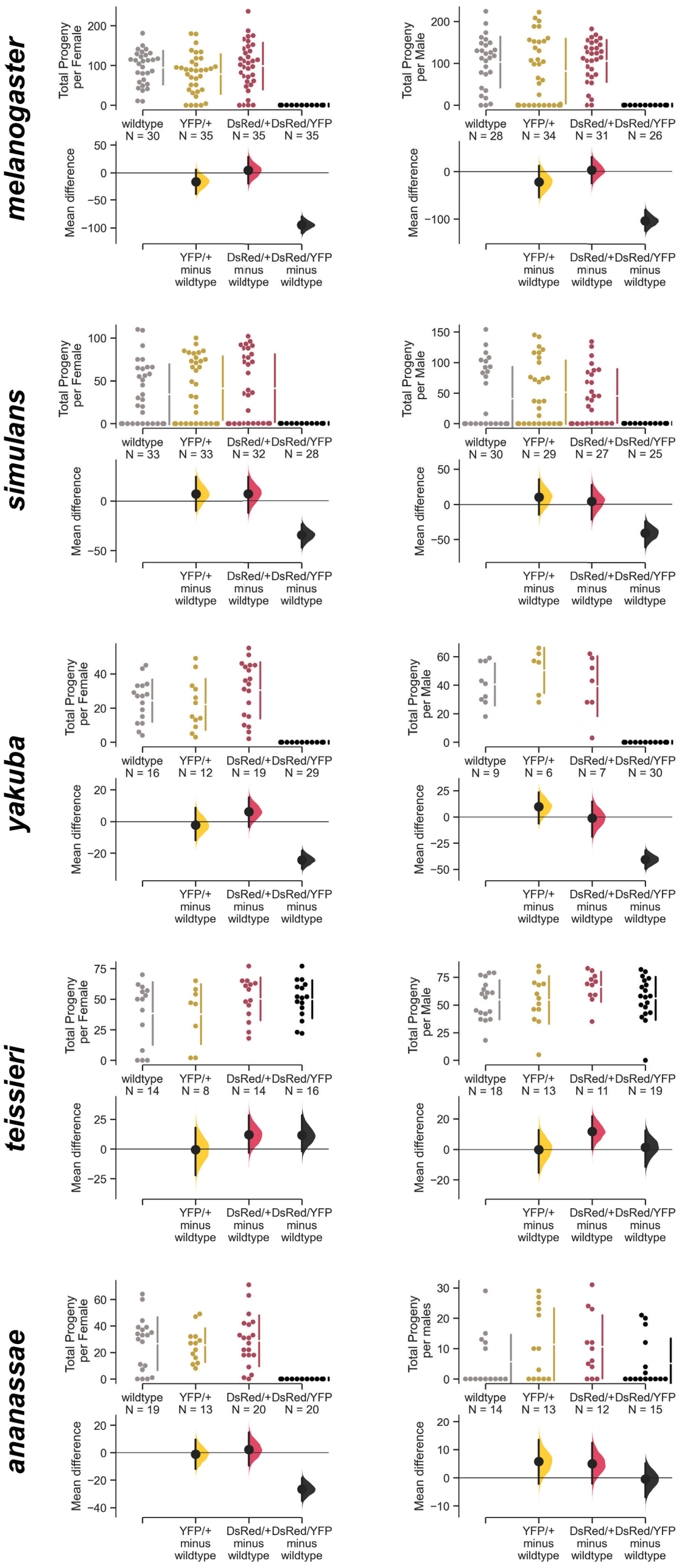

B. *bam* disruption genotype



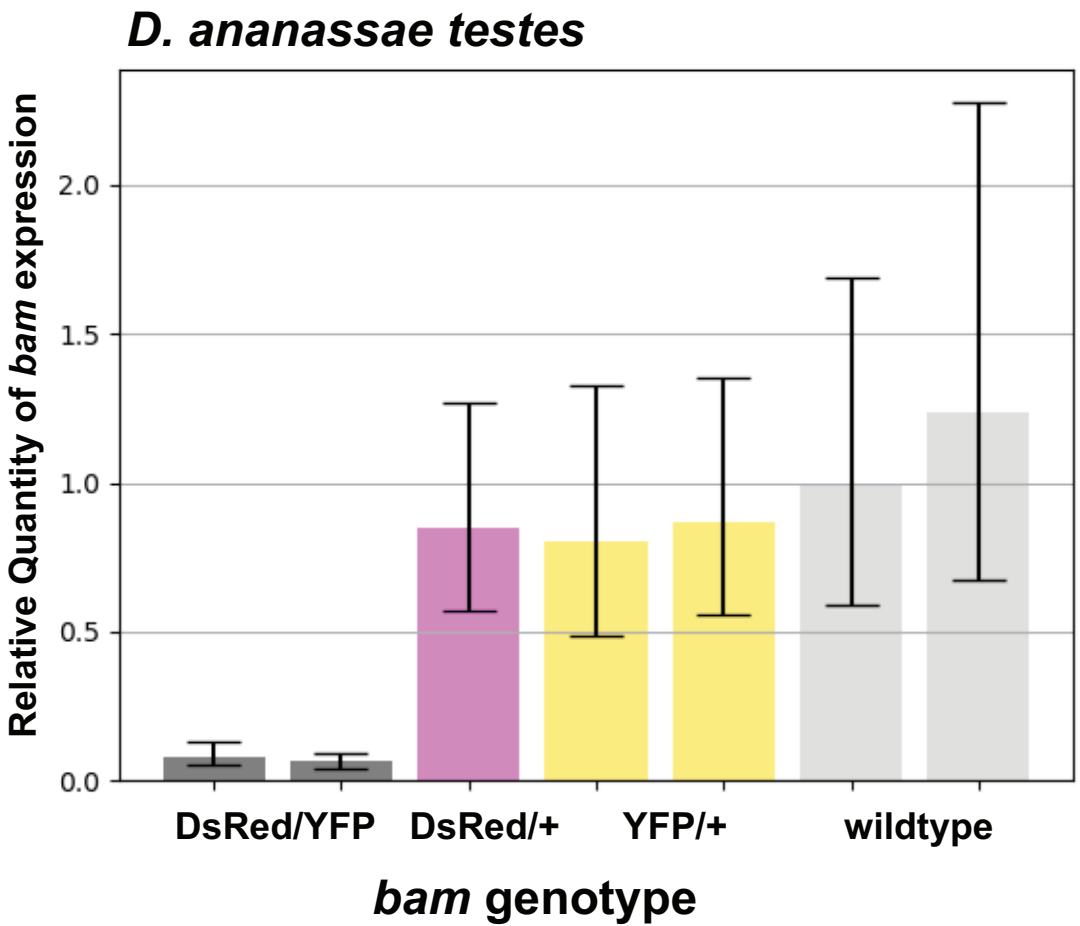
# Fig 5


*bam* disruption genotype
wildtype
null
*melanogaster*
*ananassae*
A.
B.
C.
D.
Vasa
Fas3
50 µm
50 µm
50 µm
50 µm

# Fig S1

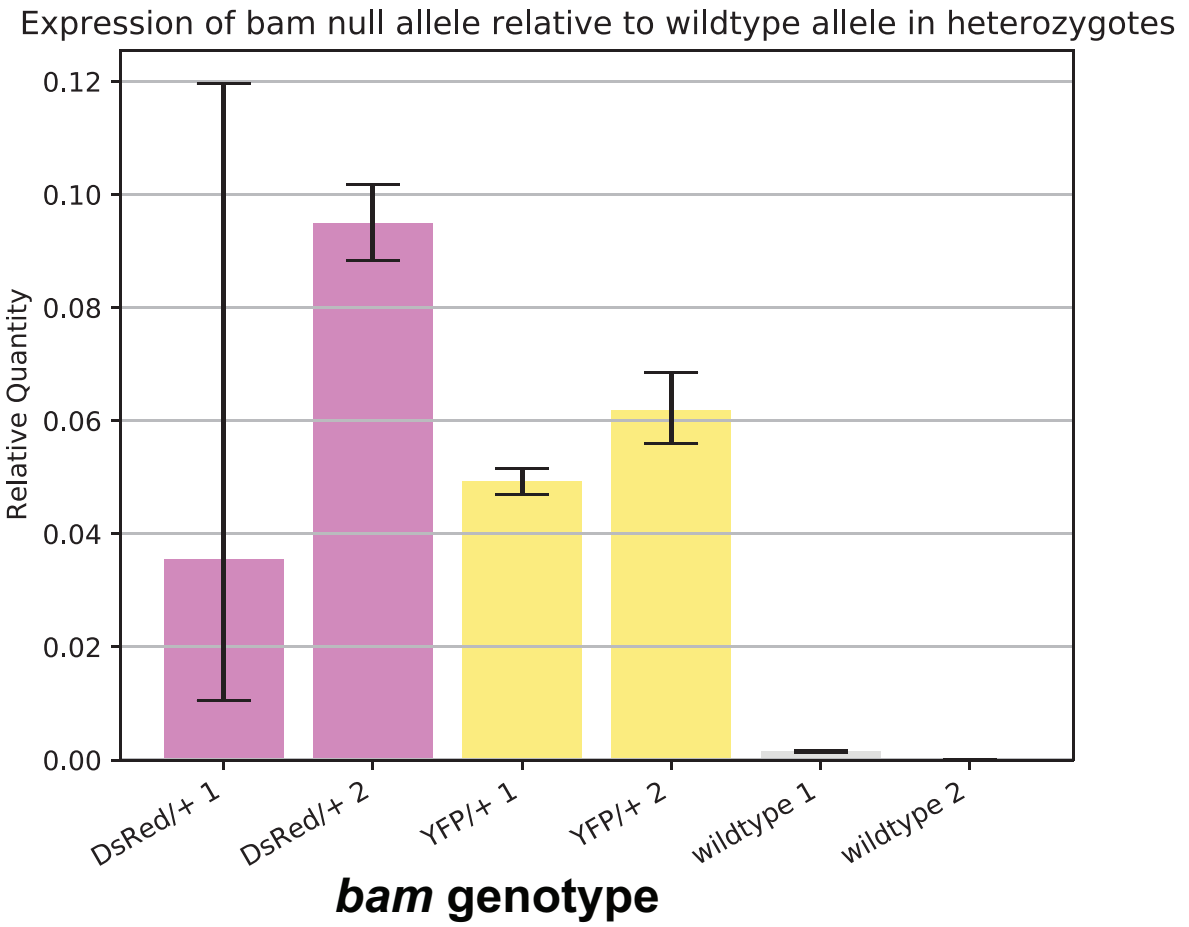

melanogaster
Total Progeny per Female
wildtype N = 30
YFP/+ N = 35
DsRed/+ N = 35
DsRed/YFP N = 35
Mean difference
YFP/+ minus wildtype
DsRed/+ minus wildtype
DsRed/YFP minus wildtype
Total Progeny per Male
wildtype N = 28
YFP/+ N = 34
DsRed/+ N = 31
DsRed/YFP N = 26
simulans
wildtype N = 33
YFP/+ N = 33
DsRed/+ N = 32
DsRed/YFP N = 28
wildtype N = 30
YFP/+ N = 29
DsRed/+ N = 27
DsRed/YFP N = 25
yakuba
wildtype N = 16
YFP/+ N = 12
DsRed/+ N = 19
DsRed/YFP N = 29
wildtype N = 9
YFP/+ N = 6
DsRed/+ N = 7
DsRed/YFP N = 30
teissieri
wildtype N = 14
YFP/+ N = 8
DsRed/+ N = 14
DsRed/YFP N = 16
wildtype N = 18
YFP/+ N = 13
DsRed/+ N = 11
DsRed/YFP N = 19
ananassae
wildtype N = 19
YFP/+ N = 13
DsRed/+ N = 20
DsRed/YFP N = 20
Total Progeny per males
wildtype N = 14
YFP/+ N = 13
DsRed/+ N = 12
DsRed/YFP N = 15

# Fig S2

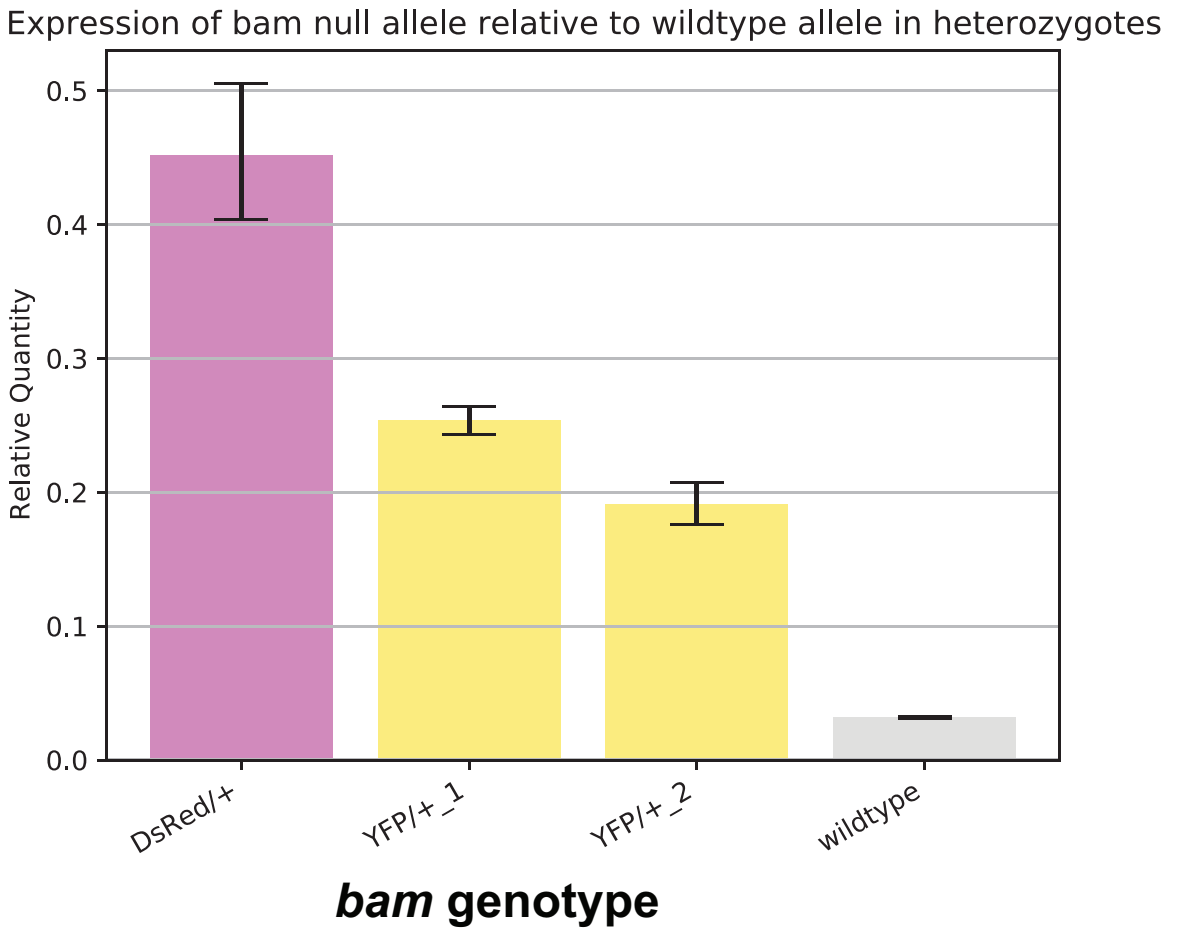

D. ananassae testes
Relative Quantity of bam expression
2.0
1.5
1.0
0.5
0.0
DsRed/YFP
DsRed/+
YFP/+
wildtype
bam genotype

# Fig S3

## A. *D. teissieri* ovaries

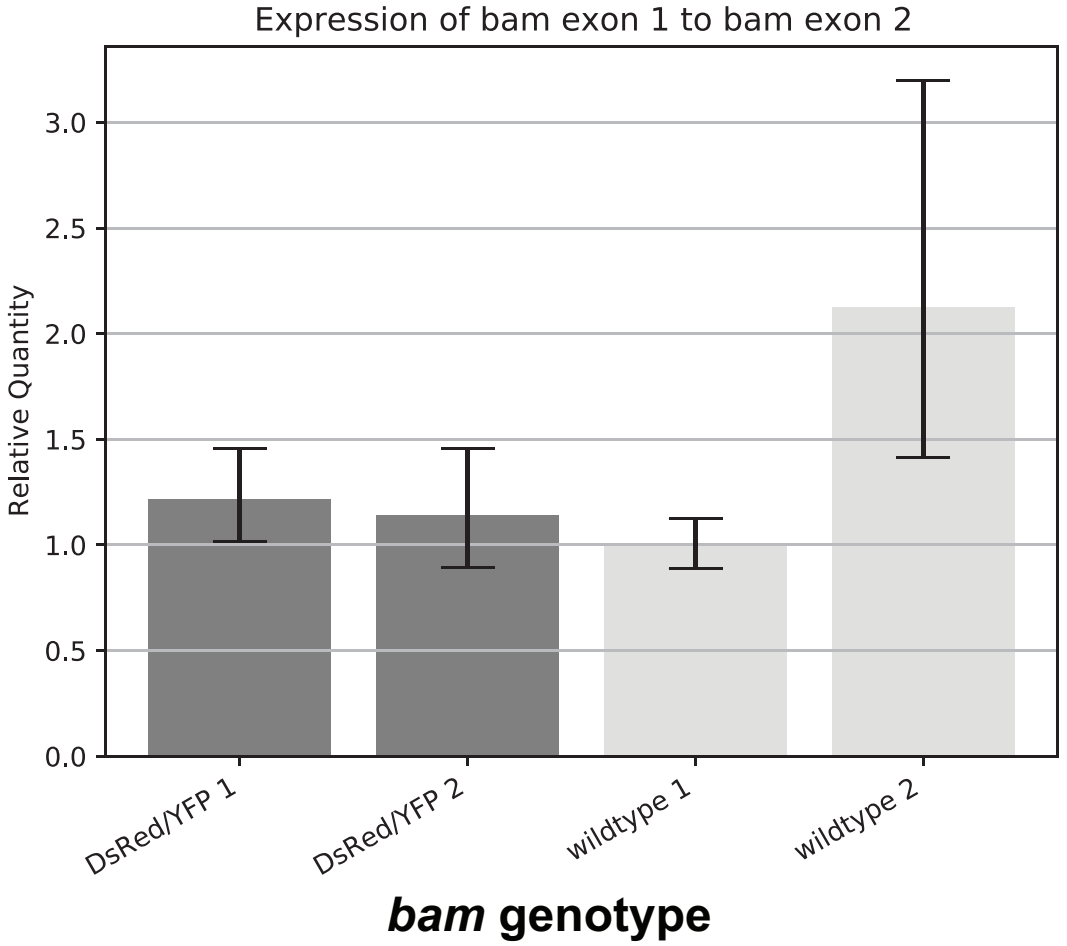


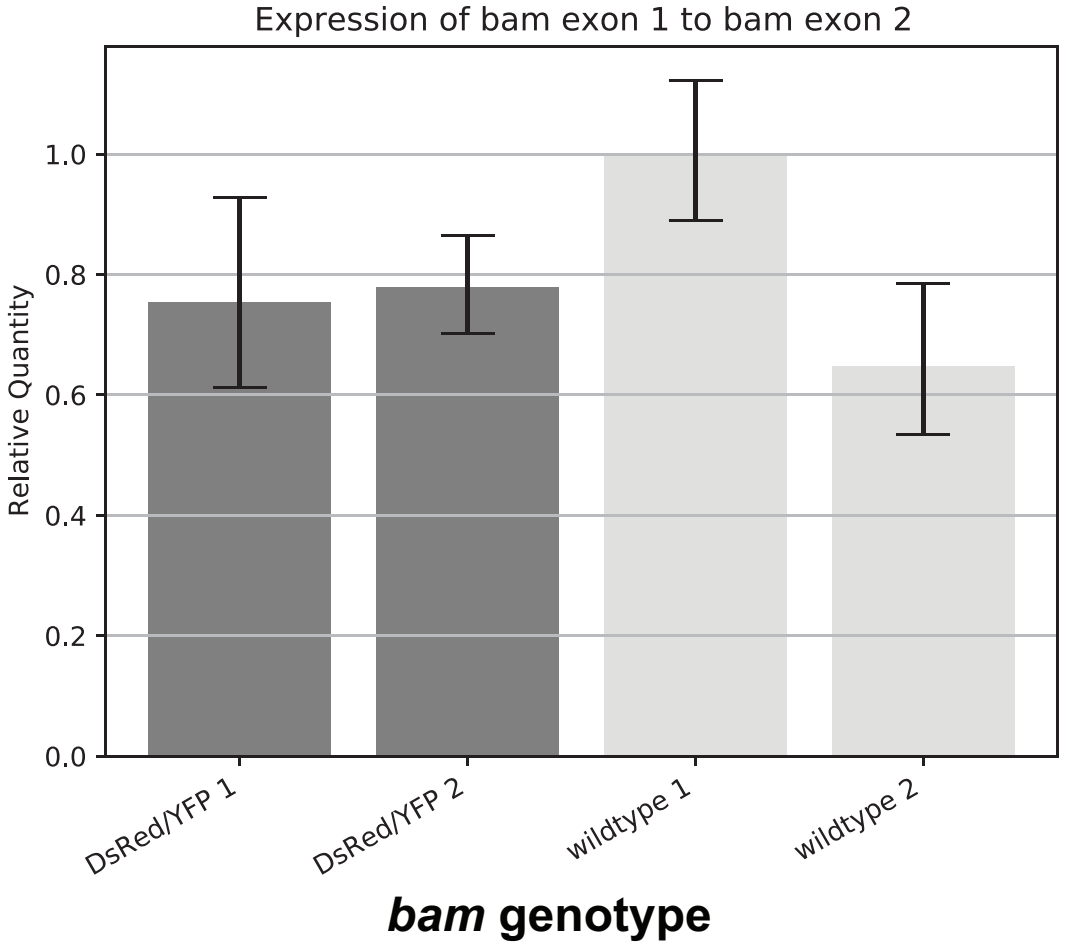

## B. *D. teissieri* testes



# Fig S4

## A. *D. teissieri* ovaries




## B. *D. teissieri* testes