

1 **Fitness effects of CRISPR endonucleases in *Drosophila melanogaster* populations**

2

3 Anna M. Langmüller^{1,2,3+}, Jackson Champer^{3,4+*}, Sandra Lapinska^{3,4}, Lin Xie^{3,4}, Matthew
4 Metzloff^{3,4}, Jingxian Liu^{3,4}, Yineng Xu^{3,4}, Andrew G. Clark^{3,4}, Philipp W. Messer^{3*}

5

6 ¹Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, 1210 Vienna, Austria

7 ²Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210
8 Vienna, Austria

9 ³Department of Computational Biology, Cornell University, Ithaca, NY, USA 14853

10 ⁴Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA 14853

11

12 +Equal contribution

13 *Corresponding authors: JC (jc3248@cornell.edu), PWM (messer@cornell.edu)

14

15 **Abstract**

16

17 CRISPR/Cas9 systems provide a highly efficient and flexible genome editing technology with
18 numerous potential applications in areas ranging from gene therapy to population control.
19 Some proposed applications involve CRISPR/Cas9 endonucleases integrated into an
20 organism's genome, which raises questions about potentially harmful effects to the transgenic
21 individuals. One application where this is particularly relevant are CRISPR-based gene drives,
22 which promise a mechanism for rapid genetic alteration of entire populations. The performance
23 of such drives can strongly depend on fitness costs experienced by drive carriers, yet relatively
24 little is known about the magnitude and causes of these costs. Here, we assess the fitness effects
25 of genomic CRISPR/Cas9 expression in *Drosophila melanogaster* cage populations by
26 tracking allele frequencies of four different transgenic constructs, designed to disentangle
27 direct fitness costs due to the integration, expression, and target-site activity of Cas9 from costs
28 due to potential off-target cleavage. Using a maximum likelihood framework, we find a
29 moderate level of fitness costs due to off-target effects but do not detect significant direct costs.
30 Costs of off-target effects are minimized for a construct with Cas9HF1, a high-fidelity version
31 of Cas9. We further demonstrate that using Cas9HF1 instead of standard Cas9 in a homing
32 drive achieves similar drive conversion efficiency. Our results suggest that gene drives should
33 be designed with high-fidelity endonucleases and may have implications for other applications
34 that involve genomic integration of CRISPR endonucleases.

35

36 **Introduction**

37

38 The ability to make specific edits of genetic material has been a long-standing goal in molecular
39 biology. Until recently, such DNA engineering was cumbersome, expensive, and difficult since
40 it relied on site-specific nucleases or random insertions. CRISPR technology represents a
41 milestone in genome editing because it makes DNA engineering highly efficient, relatively
42 simple to use, and cost-effective through the use of endonucleases that can be flexibly
43 programmed to cut specific sequences dictated by a guide RNA (gRNA) (1, 2).

44

45 The programmability of CRISPR/Cas9 systems allows for numerous potential applications (1),
46 including cancer and disease treatment (3–7), stimuli tracking in living cells (8), and crop
47 improvement (9). While most applications of CRISPR use this technology to engineer specific
48 modifications in a given gene sequence, some proposed applications take the idea one step
49 further by integrating the CRISPR machinery itself into an organism’s genome. In that case,
50 endonuclease activity can continue to produce genetic changes in the cells of the living
51 organism. When present in the germline, these genetic changes might even be passed on to
52 future generations, such as in CRISPR-based gene drives — “selfish” genetic elements that are
53 engineered to rapidly spread a desired genetic trait through a population into which they are
54 released (10–14).

55

56 However, major questions loom large about the technical feasibility of these proposed
57 applications. For example, it remains unclear whether activity of CRISPR endonucleases could
58 entail unintended and potentially harmful consequences in the transgenic organisms, for
59 instance due to the tendency to produce non-specific DNA modifications (so-called “off-target
60 effects”) (15). Such off-target cleavage could be substantially higher when Cas9 is
61 continuously expressed from a genome and inherited by offspring, where further off-target
62 cleavage can occur.

63

64 In this study, we seek to address this question in the context of CRISPR gene drive, a new
65 technology that could potentially be used for applications ranging from the control of vector-
66 borne diseases to the suppression of invasive species (10, 12, 14, 16). One class of CRISPR-
67 based gene drives are so-called “homing drives”. These genetic constructs are programmed to
68 cleave a wild-type sister chromatid and get copied to the target site through homology-directed

69 repair. Since “homing” occurs in the germline, the drive allele will be inherited at a super
70 Mendelian rate and can thereby spread quickly through the population. The effectiveness of
71 such systems has now been demonstrated in various organisms, including yeast (17–20),
72 mosquitoes (21–24), fruit flies (25–32), and mice (33). Another class of CRISPR gene drives
73 operate by the “toxin-antidote” principle (34). Here, the drive allele serves as the “toxin” by
74 carrying a CRISPR endonuclease programmed to target and disrupt an essential wild-type
75 gene. At the same time, the construct also contains a recoded version of that gene (the
76 “antidote”), which is immune to cleavage by the drive. Over time, such a drive will
77 continuously remove wild-type alleles from the population, while the drive allele will increase
78 in frequency (35). Both homing and toxin-antidote drives can be “modification drives”,
79 intended to spread a desired genetic payload through the population (e.g., a gene that prevents
80 mosquitoes from transmitting malaria), or “suppression drives”, where the goal is to diminish
81 or outright eliminate the target population (34, 36).

82

83 A key factor in determining the expected population dynamics of any type of gene drive is the
84 fitness cost imposed by the drive (37). Such fitness costs could come in the form of reduced
85 viability, fecundity, or mating success of the individuals that carry drive alleles. In suppression
86 drives, some fitness costs are typically an intended feature of the drive, necessary to ultimately
87 achieve population suppression. However, these costs are usually recessive to allow the drive
88 to spread to high frequency, and there is generally a limit as to how high other costs can be
89 before the drive will lose its ability to spread effectively (34, 36, 38, 39). For modification
90 drives, fitness costs tend to slow the spread of the drive and can thereby increase the chance
91 that resistance alleles evolve, which could ultimately outcompete the drive (12). For such
92 applications, it is therefore desirable to minimize any fitness costs. In drives with frequency-
93 dependent invasion dynamics, such as most CRISPR toxin-antidote systems (32, 40), fitness
94 costs can further determine the frequency threshold required for the drive to spread through the
95 population (34, 36, 38).

96

97 We believe it is useful to distinguish between two types of fitness costs of a gene drive. The
98 first class of “direct” costs comprise any effects resulting from the genomic integration of the
99 drive construct (e.g., when this disrupts a functionally important region), costs of potential
100 “payload” genes included in the drive construct, costs resulting directly from the expression of
101 the endonuclease or other drive elements such as gRNAs, and costs due to cleavage of the
102 intended target site. The second class comprise any potential fitness costs due to “off-target”

103 activity of the CRISPR endonuclease, referring to cleavage and disruption of any unintended
104 sites in the genome. Despite their critical importance, we still know surprisingly little about the
105 specific types of fitness costs imposed by gene drives. In particular, it remains unclear whether
106 there are certain “baseline” fitness costs that would be difficult to avoid in any gene drive
107 construct, for instance because they are inherent to the expression and activity of the CRISPR
108 endonuclease or result from their tendency to generate off-target effects.

109

110 In this study, we conduct a comprehensive assessment of the fitness effects resulting from
111 genomic expression of CRISPR/Cas9 in experimental *Drosophila melanogaster* populations.
112 We specifically investigate four different transgenic constructs that allow us to disentangle
113 direct fitness costs from those due to off-target effects. We estimate these fitness costs both
114 through a statistical analysis of allele frequency trajectories in cage populations and a direct
115 evaluation of individual fitness components using viability, fecundity, and mate choice assays.

116

117 **Results**

118

119 Construct design

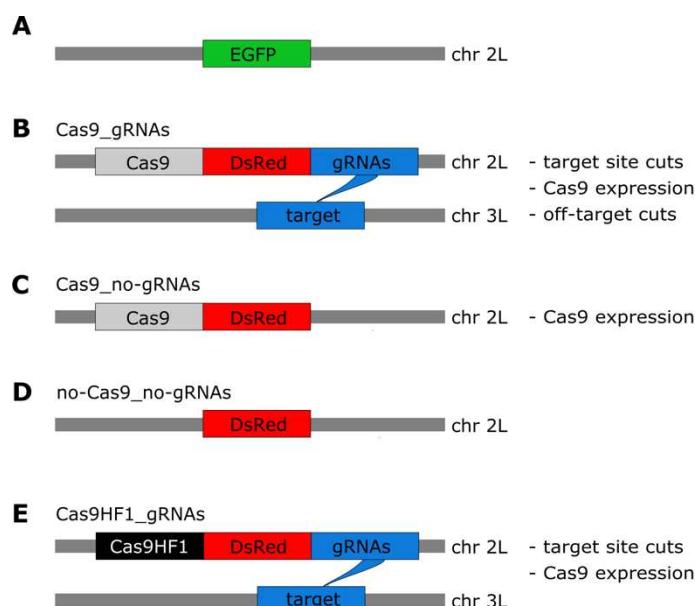
120 We designed four constructs to assess the fitness costs of *in vivo* CRISPR/Cas9 expression in
121 *D. melanogaster*. As a starting point for our transgenic fly lines, we engineered an EGFP
122 fluorescent marker driven by the 3xP3 promoter into a gene-free, non-heterochromatic position
123 on chromosome 2L (region targeted by gRNA: 20,368,542 - 20,368,561; Figure 1A). This
124 EGFP marker was then used as insertion point for the four constructs we tested. Our first
125 construct, “Cas9_gRNAs”, contains Cas9 expressed by the *nanos* promoter, the fluorescence
126 marker DsRed driven by the 3xP3 promoter, and four gRNAs driven by the U6:3 promoter
127 (Figure 1B), which are separated by tRNAs that are removed after transcription (29). The
128 gRNAs of the Cas9_gRNAs construct target a gene-free, non-hetero-chromatic position on a
129 different chromosome (3L, region targeted by gRNAs: 18,297,270 – 18,297,466), preventing
130 any homing activity. In addition to Cas9_gRNAs, three other constructs were designed:
131 “Cas9_no-gRNAs” has a similar architecture as Cas9_gRNAs, but lacks the four gRNAs
132 driven by the U6:3 promoter (Figure 1C); “no-Cas9_no-gRNAs” contains neither Cas9, nor
133 the gRNAs, but only the fluorescence marker DsRed driven by the 3xP3 promoter (Figure 1D);
134 the last construct, “Cas9HF1_gRNAs” (Figure 1E), has the same architecture as Cas9_gRNAs,
135 except that Cas9 is replaced by a high-fidelity version (Cas9HF1), which has been reported to
136 largely eliminate off-target cleavage (41). As expected, all progeny of individuals with the

137 Cas9_gRNAs and Cas9HF1_gRNAs alleles had at least one of their gRNA target sites mutated,
138 together indicating that all four gRNAs were active in both these constructs.

139
140 The specific designs of these four different constructs allow us to identify and disentangle
141 different types of Cas9-related fitness costs. If double strand breaks at the target site impose
142 fitness costs, such costs should be present for the Cas9_gRNAs and Cas9HF1_gRNAs
143 constructs, but not for the Cas9_no-gRNAs and no-Cas9_no-gRNAs constructs, since
144 Cas9_no-gRNAs has no gRNAs expressed to guide Cas9 to the target site, and the no-Cas9_no-
145 gRNAs construct neither expresses Cas9 nor the gRNAs. If the expression of Cas9 imposes a
146 fitness cost, all constructs except for no-Cas9_no-gRNAs should incur such a cost, because
147 only this construct does not express Cas9. If off-target effects of Cas9 impose fitness costs,
148 only the Cas9_gRNAs construct should incur them, because the designs of Cas9_no-gRNAs
149 and no-Cas9_no-gRNAs prevent cutting events, and Cas9HF1_gRNAs reportedly has a much
150 lower rate of off-target cleavage (41). Figure 1 summarizes the designs and different potential
151 fitness costs for our four constructs.

152

153



154
155 Figure 1. Overview of constructs and the potential types of fitness costs in the four constructs. (A) The starting
156 point for our constructs is an EGFP marker inserted into chromosome 2L (~20.4 Mb). The four constructs are
157 then inserted into this EGFP locus (thereby disrupting EGFP). (B) The Cas9_gRNAs construct contains Cas9,
158 DsRed, and gRNAs. The gRNAs target chromosome 3L (~18.3 Mb), instead of the sister chromatid. (C) The
159 Cas9_no-gRNAs construct carries Cas9 and DsRed, but no gRNAs are expressed. (D) The no-Cas9_no-gRNAs
160 construct carries only the fluorescent marker DsRed. (E) The Cas9HF1_gRNAs construct has the same structure
161 as Cas9_gRNAs but carries Cas9HF1 instead of Cas9.

162

163 Population cage experiments

164 To assess the fitness effects of the four constructs, we tracked their population frequencies
165 relative to the baseline EGFP construct over several generations in large cage populations.
166 Overall, we assessed 13 cages: seven with the Cas9_gRNAs construct, and two each with the
167 Cas9_no-gRNAs, no-Cas9_no-gRNAs, and Cas9HF1_gRNAs construct (Figure 2). In each
168 cage population, the construct frequency was tracked for at least eight consecutive, non-
169 overlapping generations. The median population size across all experiments was 3,602 (Figure
170 S1 & Supplementary Results). To avoid potentially confounding maternal fitness effects on the
171 construct frequency dynamics, we excluded the first generation of five cage populations
172 (Cas9_gRNAs construct: replicates 1, 2, 5, 6, and 7) from the analysis, because their founding
173 construct homozygotes and EGFP homozygotes were raised in potentially different
174 environments.

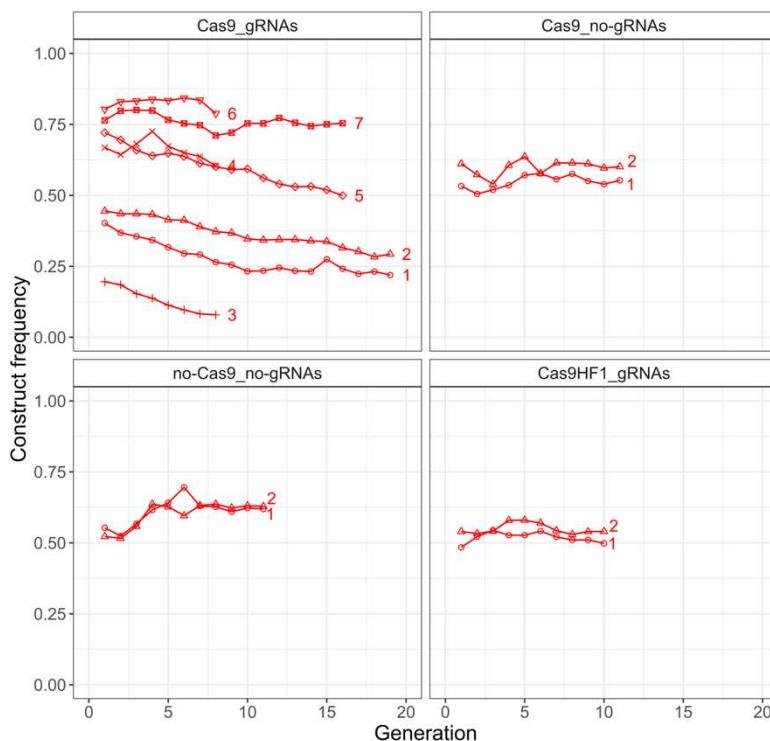
175

176 We found Cas9_gRNAs to be the only construct that systematically decreased in frequency
177 across all replicate cages (Figure 2). Interestingly, the allele frequency change was not
178 consistent with fixed direct fitness costs. Instead, the construct frequency “bottomed out” in
179 most replicates, and this occurred more quickly when the starting frequency was higher (Figure
180 2). We do not expect that the different frequency trajectories were caused by replicate-specific
181 maternal effects, since replicates 3 and 4, which had very different frequency dynamics and
182 starting frequencies, originated from the same pool of founding construct and EGFP
183 homozygotes. In contrast to Cas9_gRNAs, the three other constructs did not decrease in
184 frequency consistently across replicates, suggesting Cas9 off-targets effects are the primary
185 driver of the fitness costs we detected (Figure 1).

186

187 A model in which fitness costs are predominantly caused by a limited set of potential off-target
188 sites (e.g., because they have similar sequence to the actual target site) also suggests a possible
189 mechanism for the observed “bottoming out” of the Cas9_gRNA frequency trajectories. At the
190 beginning of the experiments, all genomes in individuals without the CRISPR construct will
191 harbor uncut alleles at all these sites. In individuals carrying the construct, such sites may be
192 cut and then repaired by end joining, which typically results in a mutated sequence. Some of
193 these mutations could be deleterious (e.g., if they change the sequence of an important gene).
194 A mutated site will also be protected to future cutting, similar to the creation of resistance
195 alleles in a homing drive (28). Early in the experiment, mutated off-target sites will be found

196 primarily in individuals that also carry drive alleles. This will lower the fitness of these
197 individuals and, consequently, impose negative selection against construct alleles. However,
198 as mutated off-target sites accumulate over the course of an experiment, they will increasingly
199 segregate independently from construct alleles, thereby reducing selection against these alleles.
200 By the time all potential off-target sites in the population have been cut, construct alleles would
201 no longer experience any negative selection if such off-target effects were indeed the only
202 mechanism underlying the construct's fitness costs. Due to the higher overall rate of cleavage
203 events in the population, cages where the construct is introduced at a higher frequency will
204 experience this effect faster than cages where it is introduced at lower frequency, consistent
205 with our experiments.



206
207 Figure 2. Construct frequency trajectories in the cage populations. Each line is one cage experiment.
208

209 Maximum likelihood analysis

210 To quantify the fitness costs of the different constructs from the observed frequency trajectories
211 in our experiments, we adopted a previously developed maximum likelihood framework (42)
212 and modified it to model two unlinked autosomal loci, representing the construct and a single
213 idealized off-target site (see Methods). Each of the two loci is biallelic (EGFP/construct;
214 uncut/cut off-target site). In individuals that carry a construct, all uncut off-target alleles are
215 assumed to be cut in the germline, which are then passed on to offspring that could suffer
216 negative fitness consequences. In the early embryo, all uncut off-target alleles are assumed to

217 be cut by maternally deposited Cas9 if the mother carries at least one construct allele, changing
218 the individual's genotype at the off-target site and exposing it to the potential fitness costs
219 associated with this new genotype. Fitness costs due to carrying the construct and/or the
220 presence of cut off-target sites are assumed to be multiplicative across the two loci, as well as
221 for the two alleles at each locus. We studied models where fitness costs affect only viability,
222 and models where they affect only mate choice and fecundity (both equally). Overall, our
223 maximum likelihood model infers three parameters: the effective population size N_e , the “direct
224 fitness estimate” (defined as the relative fitness of construct/EGFP heterozygotes versus
225 EGFP/EGFP homozygotes), and the “off-target fitness estimate” (defined as the relative fitness
226 of cut/uncut heterozygotes versus uncut/uncut homozygotes.). Note that in our idealized model
227 with a single cleavage site, this site could in principle also represent “on-target” cleavage.
228 However, due to the intergenic location of all gRNA target sites in our constructs, we do not
229 expect such fitness costs to be present. Furthermore, if on-target cleavage had a measurable
230 negative fitness effect, this should have been apparent in the frequency trajectories of the
231 Cas9HF1_gRNAs construct. Since this construct had no apparent reduction in fitness, we refer
232 to this fitness parameter as exclusively “off-target”.

233
234 For each construct, five different models were studied: In the “full inference model”, both the
235 construct and cut off-target alleles can impose fitness costs. In the “construct” model, only
236 construct alleles impose a fitness cost. In the “off-target” model, only cut off-target alleles
237 impose a fitness cost. In the “initial off-target model”, we assumed that fitness costs originated
238 before the experiment (e.g., through the injection process or maternal effects in the ancestral
239 generation). For the “initial off-target model”, the construct homozygotes in the ancestral
240 generation all had cut off-target alleles, but no additional off-target cutting occurred during the
241 experiment (i.e., the germline and embryo cut rate were set to 0). Finally, in the “neutral”
242 model, no fitness costs were present at all. Inferences were performed on the combined data of
243 the replicated experimental populations for each construct. The individual models were
244 compared using the corrected Akaike Information Criterion (AICc) (43) — a goodness-of-fit
245 measure that also penalizes for complexity (i.e., number of parameters) in a given model. A
246 lower AICc value indicates a higher quality model.

247
248 Table 1 shows the results for the Cas9_gRNAs construct. Here, we found the full inference
249 model with viability selection to yield the highest quality, with a “direct fitness estimate” of
250 0.98 and an “off-target fitness estimate” of 0.84. Note, however, that the 95% confidence

251 interval of the direct fitness estimate includes 1, and the simpler “off-target” model with
252 viability selection in which the direct fitness estimate is set to 1 in fact has an equal AICc value
253 to the “full” model. Thus, direct fitness costs are likely smaller than 5% in construct/EGFP
254 heterozygotes. Models with fecundity/mate choice selection generally had lower quality than
255 models with viability selection. The “initial off-target” and “neutral” models yielded the lowest
256 AICc values. Taken together, these results suggest that the observed frequency trajectories of
257 the Cas9_gRNAs construct in our cage populations are best explained by a model where direct
258 effects are less than a few percent and off-target effects impose moderate fitness costs of ~30%
259 ($= 1 - 0.84^2$) in cut/cut homozygotes in our idealized single off-target site model.

260

261 To test whether this model can accurately capture the observed frequency-dependent construct
262 dynamics of the Cas9_gRNAs construct, we simulated construct trajectories under the model
263 with the best AICc value (full inference model, viability selection) under its maximum
264 likelihood parameter estimates ($\hat{N}_e = 175$, direct fitness estimate = 0.98, off-target fitness
265 estimate = 0.84). The simulations do not only resemble the observed decrease in construct
266 frequency, but also capture the bottoming out of individual replicates depending on their
267 construct starting frequency (Figure 3). Additionally, we compared simulated trajectories for
268 this model with simulated trajectories from the “construct” model with viability selection
269 (Figure S2). We found that the full inference model captures the observed frequency dependent
270 construct dynamics better than the model that only considers direct fitness costs, with most of
271 the improvement due to better matching trajectories from cages with low starting frequencies,
272 where off-target effects would be expected to have a more drastic impact on the relative fitness
273 of construct-carrying individuals.

274

275 To further support our hypothesis that fitness costs are primarily driven by off-target effects,
276 we applied the maximum likelihood inference framework to the experimental cage data of the
277 three other constructs (Cas9_no-gRNAs, no-Cas9_no-gRNAs, and Cas9HF1_gRNAs).
278 Because none of these three constructs should be capable of producing substantial amounts of
279 off-target cuts by design, we set the germline and embryo cut rate to 0 and inferred viability
280 fitness effects for the construct. Except for the “initial off-target” model, construct
281 homozygotes of the ancestral population were assumed to not carry any cut off-target alleles.
282 For Cas9_no-gRNAs, and no-Cas9_no-gRNAs, the “neutral” model without any fitness costs
283 explains the observed construct frequency trajectories best (Table 2, Figure S3), corroborating
284 the notion that off-target cuts are the main driver of Cas9 fitness costs in our experimental

285 populations (Figure 1). However, the construct frequency dynamics of Cas9HF1_gRNAs are
286 best explained by an “initial off-target” model, where cut off-target alleles are beneficial,
287 closely followed by the neutral model (Table 2). While we cannot rule out that the initial
288 construct homozygotes of Cas9HF1_gRNAs had a fitness advantage due to cut off-target
289 alleles or transgenerational beneficial effects, the 95% confidence interval for the off-target
290 fitness parameter is broad and includes 1. This putative fitness advantage could also potentially
291 be explained by maternal effects that persisted for 2-3 generations. Although we do not
292 anticipate that any other construct than Cas9_gRNAs can produce substantial off-target effects,
293 we repeated the analysis of the three other constructs with cut rates set to 1 and inferred viability
294 selection, which yielded similar results (Table S1).

295

296 Table 1. Model comparison and parameter estimates for Cas9_gRNAs.

model	selection	\hat{N}_e	direct fitness estimate	off-target fitness estimate	$\ln\hat{L}$	P	AICc
full	viability	175 [140 – 215]	0.98 [0.95 – 1.00]	0.84 [0.77 – 0.91]	384.7	3	-763
full	mate choice = fecundity	163 [131 – 200]	0.96 [0.94 – 0.98]	1.00 [0.95 – 1.06]	378.8	3	-751
construct	viability	164 [131 – 201]	0.96 [0.93 – 0.98]	1*	378.9	2	-754
construct	mate choice = fecundity	163 [131 - 200]	0.96 [0.94 – 0.98]	1*	378.8	2	-754
off-target	viability	173 [139 - 212]	1*	0.80 [0.74 – 0.88]	383.6	2	-763
off-target	mate choice = fecundity	157 [126 - 192]	1*	0.95 [0.90 – 1.01]	375.1	2	-746
initial off-target	viability	156 [125 - 191]	1*	0.92 [0.82 – 1.02]	374.8	2	-745
initial off-target	mate choice = fecundity	156 [125 - 191]	1*	0.96 [0.91 – 1.01]	374.8	2	-745
neutral	none	154 [123 – 189]	1*	1*	373.6	1	-745

297 Each row shows the parameter estimates (\hat{N}_e = effective population size), maximum log Likelihood ($\ln\hat{L}$), number
298 of free parameters in the maximum likelihood framework (P), and corrected Akaike Information Criterion value
299 ($AICc = 2p - 2\ln\hat{L} + (2p^2 + 2p)/(n - p - 1)$ where $n = 87$ is the number of generation transitions) for a
300 specific model and selection type. 1* entries indicate that a parameter was fixed at 1 (= no fitness effect is
301 estimated). Values in squared brackets in the parameter estimate columns represent the 95 % confidence intervals
302 estimated from a likelihood ratio test with one degree of freedom.

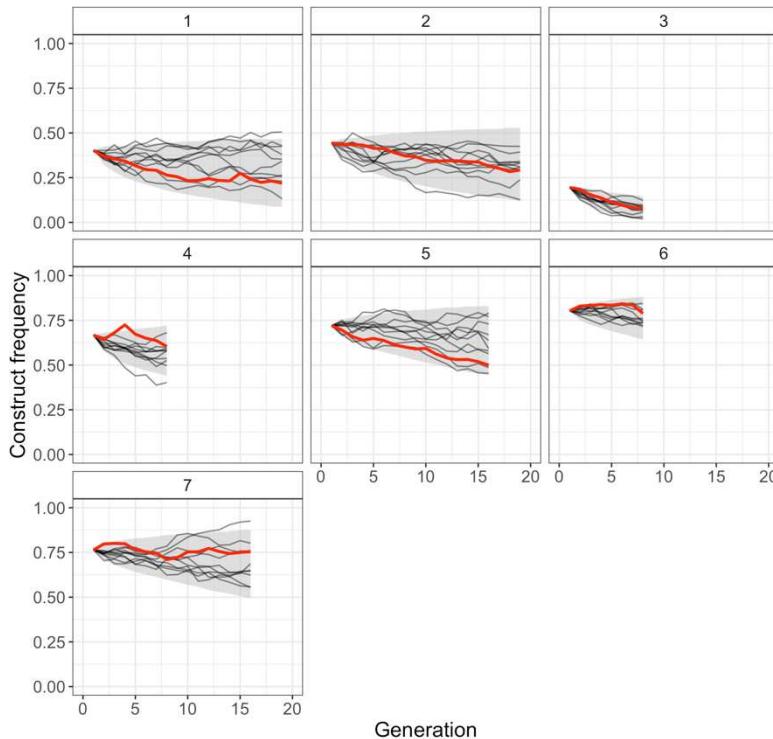
303

304 Table 2. Model comparison and parameter estimates for Cas9_no-gRNAs, no-Cas9_no-gRNAs, and
305 Cas9HF1_gRNAs.

construct	model	selection	\hat{N}_e	direct fitness estimate	off-target fitness estimate	$\ln\hat{L}$	P	AICc
Cas9_no-gRNAs	construct	viability	243 [152 – 366]	1.0 [0.96 – 1.04]	1*	88.6	2	-173
Cas9_no-gRNAs	initial	viability	250	1*	0.84 [0.65 – 1.18]	89.2	2	-174
	off-target		[156 – 377]					
Cas9_no-gRNAs	neutral	none	243 [152 – 366]	1*	1*	88.6	1	-175
no-Cas9_no-gRNAs	construct	viability	162 [101 – 243]	1.0 [0.97 – 1.10]	1*	81.5	2	-158
no-Cas9_no-gRNAs	initial	viability	162 [101 – 243]	1*	1.12 [0.84 – 1.63]	81.5	2	-158
	off-target							
no-Cas9_no-gRNAs	neutral	none	162 [101 – 243]	1*	1*	81.5	1	-161
Cas9HF1_gRNAs	construct	viability	396 [240 – 608]	1.0 [0.97 – 1.04]	1*	88.1	2	-171
Cas9HF1_gRNAs	initial	viability	433 [263 – 655]	1*	1.18 [0.99 – 1.45]	89.7	2	-175
	off-target							
Cas9HF1_gRNAs	neutral	none	396 [240 – 608]	1*	1*	88.1	1	-174

306 Each row shows the parameter estimates (\hat{N}_e = effective population size), maximum log Likelihood ($\ln\hat{L}$), number
307 of free parameters in the maximum likelihood framework (P), and corrected Akaike Information Criterion value
308 ($AICc = 2p - 2\ln\hat{L} + (2p^2 + 2p)/(n - p - 1)$ where n = number of generation transitions; n= 20 for Cas9_no-
309 gRNAs, no-Cas9_no-gRNAs, and n= 18 for Cas9HF1_gRNAs) for a specific construct, model and selection type.
310 1* entries indicate that a parameter was fixed at 1 (= no fitness effect is estimated). Values in squared brackets in
311 the parameter estimate columns represent the 95 % confidence intervals estimated from a likelihood ratio test with
312 one degree of freedom.

313



314

315 Figure 3. Comparison of observed Cas9_gRNAs construct frequencies with simulated trajectories of a full model
316 with viability selection ($\hat{N}_e = 175$, direct fitness estimate = 0.98, off-target fitness estimate = 0.84). Solid red
317 lines present observed construct frequencies, black lines show ten simulated trajectories for each cage, and the
318 shaded area represents the range between the 2.5 and 97.5 percentile of the simulated trajectories (10,000
319 simulations per cage).

320 Phenotypic fitness assays

321 As a complementary validation of our fitness measurements from the cage experiments, we
322 conducted three independent phenotypic assays (mate choice, fecundity, and viability) to
323 estimate the fitness costs of the Cas9_gRNAs construct (see Supplementary Methods &
324 Results). These assays broadly confirmed our previous findings. In particular, we found that
325 Cas9_gRNAs homozygous males were 46.15 % less likely to be picked as mates by EGFP
326 homozygous females (Figure S4A), and Cas9_gRNAs homozygous females laid on average
327 24.5% less eggs than EGFP homozygous females (Figure S4B). However, in contrast to our
328 cage experiments where a model of viability-based fitness effect best matched the data (Table
329 1), we did not observe reduced viability for Cas9_gRNAs carrying flies in the individual assay
330 (Figure S4C). However, this lack of difference in viability between EGFP and Cas9_gRNAs
331 carrying flies could be explained by the assay environment. All phenotypic assays were
332 conducted in vials, an environment where larvae experience much less resource competition
333 than in the densely populated cage populations, which can significantly influence relative
334 viability of different genotypes (44). Indeed, individuals that showed reduced fecundity or

335 mating success in individual assays may have not survived to the adult stage in cage
336 environments, representing a viability cost in that system. In addition, the viability assay
337 examined only Cas9_gRNAs/EGFP heterozygotes, which may not have suffered from off-
338 target effects to the full extent because they received "wild-type" off-target sites from one
339 parent that did not carry the construct.

340

341 Cas9HF1 homing drive

342 Our finding that off-target effects appear to be the primary driver for fitness costs from genomic
343 Cas9 expression raises the question of whether Cas9HF1 would constitute a superior choice
344 for gene drive strategies. As a proof-of-principle that Cas9HF1 is indeed a feasible alternative,
345 we designed a homing drive that is identical to a previous drive (45), except that it uses
346 Cas9HF1 instead of standard Cas9. This drive targets an artificial EGFP target locus with a
347 single gRNA (see Methods).

348

349 We first crossed male flies carrying one of the two drives to females with the same EGFP target
350 site used in our cage experiments. Individuals heterozygous for the homing drive and an EGFP
351 allele were then further crossed to flies homozygous for EGFP, or to *w¹¹¹⁸* females for several
352 of the male drive heterozygotes. The progeny of these crosses was phenotyped for DsRed,
353 indicating presence of a drive allele, and EGFP, indicating the presence of an intact target allele
354 (or more rarely, a resistance allele that preserved the function of EGFP). Disrupted EGFP
355 alleles that did not show green fluorescence indicated the presence of a resistance allele.

356

357 We observed similar performance between the Cas9HF1 drive (Data Set S1) and the standard
358 Cas9 drive (Data Set S2). Drive conversion efficiency for the Cas9HF1 drive was estimated at
359 80±2% for females and 59±3% for males, which was not significantly different than the rates
360 for the standard Cas9 drive (83±2% for females and 61±2% for males) ($P = 0.321$ for female
361 heterozygotes and $P = 0.5513$ for male heterozygotes, Fisher's Exact Test). For both drives, all
362 EGFP alleles in male heterozygotes that had not been converted to drive alleles were converted
363 to resistance alleles, as indicated by the lack of EGFP phenotype in all progeny from crosses
364 with *w¹¹¹⁸* females. Both drives also had similar rates of resistance allele formation in the early
365 embryo due to maternally deposited Cas9 (95±1% for alleles that disrupt EGFP for Cas9HF1,
366 and 96±1% for standard Cas9, $P = 0.3956$, Fisher's Exact Test). Together, these data
367 demonstrate that homing drives with Cas9HF1 are capable of similar performance as drives
368 using standard Cas9.

369

370 **Discussion**

371

372 Negative selection will tend to displace any alleles from the population that are sufficiently
373 deleterious. This effect can be quantified by an allele's fitness, specifying the relative
374 reproductive success between carriers and non-carriers of the allele. In this study, we measured
375 the fitness of transgenic Cas9/gRNA alleles in *D. melanogaster*, which constitute an essential
376 component of proposed applications such as CRISPR gene drives. A quantitative estimate of
377 the fitness costs imposed by such constructs is critical for assessing the expected performance
378 and limitations of these proposed applications.

379

380 Our constructs were designed to mimic a gene drive, yet without homing or any other
381 mechanism that would facilitate super-Mendelian inheritance. This allowed us to estimate the
382 "baseline" fitness costs of such systems. In particular, we inferred fitness by tracking allele
383 frequencies in cage populations, which provides a powerful method for fitness estimation by
384 integrating selective effects over all life stages and affected phenotypes (42). We did not
385 observe detectable fitness costs due to Cas9/gRNA integration, expression, and on-target
386 activity (which we refer to as "direct" costs) in our experiments. However, we did detect a
387 moderate level of fitness cost resulting presumably from off-target effects. Such off-target costs
388 were avoided when we used a high-fidelity version of Cas9 designed to minimize off-target
389 cleavage (41). While the effects of off-target cutting in cells transiently exposed to Cas9 have
390 already been extensively studied (46–49), our results demonstrate that such cleavage may have
391 more substantial negative consequences over multiple generations when the cells are
392 continuously expressing Cas9 in the germline from a genomic source.

393

394 Our study design does have some limitations that reduce the generality of our conclusions. First
395 and foremost, off-target effects can vary substantially depending on the specific target
396 sequence(s), genome composition, and expression patterns of Cas9 and gRNAs (15, 50). While
397 we specifically selected gRNAs with a low number of predicted off-target sites to minimize
398 such effects, this may not be possible for every application. Furthermore, the prediction of off-
399 target sites may not always be accurate, potentially missing important sites. Some applications
400 may also require the use of different promoters with higher somatic expression rates than *nanos*
401 (22, 24, 28, 29), which could increase fitness costs caused by off-target cleavage. On the other
402 hand, Cas9 expression may be lower in other organisms or at other genomic sites, and some

403 applications might require fewer than the four gRNAs we included in our constructs, thereby
404 potentially reducing off-target effects.

405
406 Another limitation is that our maximum likelihood framework for fitness estimation
407 necessarily relied on a simplistic and highly idealized model. Most importantly, we modeled
408 only a single genomic site to represent fitness costs from off-target cutting, and we used
409 codominant fitness costs. In reality, there could be many off-target cut sites, with variable types
410 of alleles after cleavage, different fitness costs, dominance relationships, degrees of genetic
411 linkage, and possibly even epistatic interactions between them. Given the limited number of
412 data points in our cage experiments, together with the large number of conceivable models, it
413 is questionable whether our maximum likelihood framework could robustly infer the “correct”
414 model.

415
416 The same holds true for the inference of different fitness components. While we did compare
417 models where fitness affected viability versus models where it affected fecundity and mate
418 choice, we believe that any conclusions from these comparisons should be taken with a grain
419 of salt, given how many assumptions still went into each model (e.g., multiplicative fitness
420 costs, same costs in males and females, and equal costs for fecundity and mate choice). Indeed,
421 while our maximum likelihood analysis of the cage experiments ranked the viability model
422 higher than the fecundity/mate choice model, we did observe reduced fecundity and mating
423 success of genotypes carrying the Cas9_gRNAs construct in our phenotypic assays, while not
424 finding a substantial effect on viability. However, as explained above, this could be due to a
425 lower power of the phenotypic assays where only heterozygotes were studied, and/or the fact
426 that many of those individuals with reduced fecundity and mating success in the phenotypic
427 assays would not actually have survived into adult stage in the cage populations due to
428 increased larval competition at higher densities.

429
430 Finally, we note that we did not attempt to identify the specific off-target cut sites that
431 presumably caused the observed fitness effects and then track their allele frequencies. Such an
432 analysis would have required time-resolved whole-genome sequencing on the population level.
433 While potentially interesting, off-target sites would be construct-specific, and such an elaborate
434 analysis may thus be more suitable when developing strategies to address off-target effects in
435 a particular construct intended for release. Indeed, many studies have analyzed such individual
436 off-target mutations in a large number of settings, including a study in mosquito gene drive

437 (51). The focus of our study, however, was to elucidate the combined effects of such off-target
438 cleavage on reproductive success, which has not been previously studied in the context of
439 genomic integration of CRISPR elements. By directly measuring the “fitness” of a given
440 construct on the population level, our approach is complementary to molecular studies that
441 seek to identify all off-target mutations and then score their potential phenotypic effects. It also
442 allows us to avoid the complexity of characterizing large numbers of potentially rare mutations,
443 determining whether they are actually caused by off-target cleavage, and predicting what
444 potential effect they may have on an organism’s fitness.

445

446 Though considerable uncertainty remains regarding the precise nature of the fitness effects of
447 our constructs, the overall finding that genomic Cas9/gRNA-expression imposes a moderate
448 fitness cost is robust. The fact that we only detected such costs for the construct that expressed
449 both Cas9 and gRNAs, but not those lacking gRNAs, suggests that these costs are primarily
450 due to cleavage activity, rather than just the expression of Cas9 or its genomic integration.
451 Further, the fact that we did not observe any fitness costs when Cas9 was replaced with
452 Cas9HF1 suggests that off-target effects are likely the driving factor for these costs. All of this
453 is also consistent with the “bottoming out” phenomenon observed for construct frequencies in
454 our cage experiments with Cas9_gRNAs, which can be explained by the accumulation of cut
455 alleles at the off-target sites over the course of the experiment. It would be more difficult to
456 reconcile with models where direct fitness costs are the driving factor or where fitness costs
457 would be due to the specific genetic background of the construct flies or health-related effects
458 of the initially released flies.

459

460 Our results have important implications for the modeling of gene drive approaches. Thus far,
461 only direct fitness costs have been modeled in such studies, arising from the CRISPR nuclease
462 itself, a payload gene, or cleavage of the intended target site. It is well known that such direct
463 fitness costs can reduce the power of a suppression drive (38, 39, 52) and reduce the persistence
464 of a modification drive in the face of resistance alleles (12). If such fitness costs are in fact
465 lower than 5%, as suggested by our study in *Drosophila* at least, then they would not be
466 expected to substantially impede the spread of suppression drives, though such drives could
467 still suffer from other forms of direct fitness costs such as haploinsufficiency of the target gene
468 and somatic Cas9 expression and cleavage. The direct fitness of modification drives would be
469 largely determined by their cargo gene(s) and possibly their rescue efficiency if they involve
470 use of a recoded gene (32, 40, 53, 54).

471
472 On the other hand, if off-target effects are in fact the primary driver of fitness costs of a drive
473 with otherwise low direct fitness costs, as suggested by our study for *Drosophila* with the *nanos*
474 promoter, this should result in different population dynamics. In a modification drive, such off-
475 target effects would likely only slow the drive initially. After the drive has spread through most
476 of the population, and cleaved off-target alleles had time to accumulate, resistance alleles
477 would not be as selectively advantageous. Thus, it would take them much longer to outcompete
478 drive alleles in the long run as compared to a scenario where direct fitness costs are the primary
479 driver. A suppression drive may still suffer from cuts at off-target sites in a manner more
480 closely resembling direct fitness costs, because these effects come into play during the early
481 spread of a drive, often the most critical period in determining the fate of a suppression drive
482 (38, 39, 52). However, if the rate at which off-target mutations form is sufficiently low, then
483 mutated off-target sequences may not have a large effect on population dynamics. This has
484 been shown in a recent study on a suppression drive with a germline-restricted promoter and a
485 single gRNA that successfully eliminated a mosquito cage population before substantial
486 amounts of off-target cleavage could occur (51).

487
488 We demonstrated that Cas9HF1, which largely eliminates off-target cleavage (41), does not
489 induce substantial negative fitness effects when used as a replacement for standard Cas9 in our
490 cage populations. Furthermore, we showed that homing drives with either form of Cas9
491 perform similarly. We therefore recommend that gene drives, as well as other applications that
492 require the genomic integration of CRISPR endonucleases, should move from standard
493 *Streptococcus pyogenes* Cas9 to higher fidelity versions that can effectively minimize off-
494 target effects. This would have the added advantage of reducing the generation of unanticipated
495 genetic changes in natural populations from off-target cleavage and repair. One potential
496 drawback of these nucleases is that they tend to have a lower cleavage rate that can depend on
497 the specific gRNA sequence employed (55–60). In practice, this may reduce the number of
498 available gRNA target sites and increase the need for initial evaluation of gRNA targets.
499 However, newer improved forms of Cas9 (55, 56, 58–60), including ones with an expanded
500 range of target sites (57), promise to ameliorate this issue.

501
502 In conclusion, we demonstrated that genomic CRISPR/Cas9 expression in *D. melanogaster*
503 can impose a moderate level of fitness costs, most likely via off-target effects. Our results
504 further indicate that fitness costs can be effectively minimized by using a high-fidelity

505 endonuclease with reduced off-target cleavage. Future studies should investigate whether these
506 conclusions hold in other experiments involving different constructs, target sites, and other
507 organisms.

508

509 **Methods**

510

511 Plasmid construction

512 The starting plasmid pDsRed (Addgene plasmid #51019) was provided by Melissa Harrison,
513 Kate O'Connor-Giles, and Jill Wildonger, pnos-Cas9-nos (61) (Addgene plasmid #62208) was
514 provided by Simon Bullock, and VP12 (41) (Addgene plasmid #72247) was provided by Simon
515 Bullock. Starting plasmids ATSacG, TTTgRNAtRNAi, TTTgRNAt, BHDgN1c, and
516 BHDgN1cv3 were constructed in a previous study (45). Restriction enzymes for plasmid
517 digestion, Q5 Hot Start DNA Polymerase for PCR, and Assembly Master Mix for Gibson
518 assembly were acquired from New England Biolabs. Oligonucleotides and gBlocks were
519 obtained from Integrated DNA Technologies. JM109 competent cells and ZymoPure Midiprep
520 kit from Zymo Research were used to transform and purify plasmids. Cas9 gRNA target
521 sequences were identified by the use of CRISPR Optimal Target Finder (62). A list of DNA
522 fragments, plasmids, primers, and restriction enzymes used for cloning of each construct can
523 be found in the Supplemental Information, together with annotated sequences of the final drive
524 insertion plasmids (ApE format, <http://biologylabs.utah.edu/jorgensen/wayned/ape>).
525

526

526 Generation of transgenic lines

527 Injections were conducted by Rainbow Transgenic Flies. The donor plasmid (Cas9_gRNAs,
528 Cas9_no-gRNAs, no-Cas9_no-gRNAs, Cas9HF1_gRNAs, or BHDgNf1v2) (~500 ng/µL) was
529 injected along with plasmid BHDgg1c (or TTTgU1 for BHDgNf1v2) (45) (~100 ng/µL), which
530 provided additional gRNAs for transformation, and pBS-Hsp70-Cas9 (~500 ng/µL, from
531 Melissa Harrison & Kate O'Connor-Giles & Jill Wildonger, Addgene plasmid #45945)
532 providing Cas9. A 10 mM Tris-HCl, 100 µM EDTA solution at pH 8.5 was used for the
533 injection. Most constructs were injected into *w¹¹¹⁸* flies, but BHDgNf1v2 was injected into flies
534 with ATSacG (45). Transformants were identified by the presence of DsRed fluorescent
535 protein in the eyes, which usually indicated successful construct insertion.
536

537 Maintenance of transgenic flies with active Cas9HF1 gene drive

538 To minimize risk of accidental release, all flies with an active homing gene drive system were
539 kept at the Sarkaria Arthropod Research Laboratory at Cornell University under Arthropod
540 Containment Level 2 protocols in accordance with USDA APHIS standards. In addition, the
541 synthetic target site drive system (30) prevents drive conversion in wild-type flies, which lack
542 the EGFP target site. All safety standards were approved by the Cornell University Institutional
543 Biosafety Committee.

544

545 Experimental fly populations

546 The experimental fly populations were maintained on Bloomington Standard medium in
547 30x30x30 cm fly cages (Bugdorm). Flies were kept at constant temperature (25°C, 14 hours
548 light, 10 hours dark), with non-overlapping generations. 0 – 2 day-old flies of one generation
549 were allowed to lay eggs on fresh medium (8 food bottles per cage) for 24 hours. After that,
550 the adults were frozen at -20°C for later phenotyping, and the new generation was allowed to
551 develop for 11-12 days, before fresh medium was provided and a new generation cycle starts.
552 The ancestral generation of each cage was generated by allowing homozygous EGFP flies and
553 flies homozygous for the construct to deposit eggs for 24 hours separately from each other in
554 four food bottles each. These eight egg-containing bottles were put in the fly cages to start one
555 experimental fly population. Seven replicates of Cas9_gRNAs, and two replicates each for
556 Cas9_no-gRNAs, no-Cas9_no-gRNAs, and Cas9HF1_gRNAs were maintained.

557

558 Phenotyping experimental fly populations

559 The dominant fluorescent markers, EGFP and DsRed, allow a direct readout of the genotype
560 by screening the fluorescent phenotype of an individual fly. Flies that are only red fluorescent
561 are construct homozygotes, flies that are only green fluorescent do not carry any construct, and
562 flies that are fluorescent for both colors carry one construct copy.

563

564 For each experimental population and generation, all individuals were screened for their
565 genotypes using either a stereo dissecting microscope in combination with the NIGHTSEA
566 system, or an automated image-based screening pipeline we specifically developed for this
567 purpose. Quantifying phenotypic traits (e.g. pupae size, the amount of laid eggs) in an
568 automated way has been done successfully before in *Drosophila* (63, 64). In our image-based
569 screening pipeline, three pictures were taken for each batch of flies: a white light picture to
570 determine the number and the position of the flies, one fluorescent picture filtered to screen for
571 DsRed, and one fluorescent picture filtered to screen for EGFP expression.

572
573 We used a Canon EOS Rebel T6 with a 18-55 mm lens for image acquisition. The camera was
574 held in a fixed position by a bracket 25 cm above the frozen flies spread on a black poster
575 board. NIGHTSEA light heads (Green and Royalblue) were used as light sources. The light
576 sources both for white and fluorescent light were covered with a paper tissue for diffusion. For
577 the fluorescent pictures, barrier filters (Tiffen 58 mm Dark Red #29; Tiffen 58 mm Green #58)
578 were used, attached with a magnetic XUME Lens/Filter system to the camera. Except for the
579 filter change, the camera was fully controlled through a PC interface (EOS Utility 2 software).
580 Focus was set automatically under white light and was kept constant for the fluorescent
581 pictures. First, a white light picture (F 5.6, ISO 100, exposure time 1'') was taken to determine
582 the number and positions of the flies. Second, a picture under NIGHTSEA Green light with the
583 Tiffen Dark red #29 filter (F 5.6, ISO 400, exposure time 30'') was taken to determine, whether
584 flies express DsRed. Third, a picture under NIGHTSEA Royal Blue with the Tiffen Dark Green
585 #58 barrier filter (F 5.6, ISO 400, exposure time 25'') was taken to screen flies for EGFP
586 expression.

587
588 We used the ImageJ distribution Fiji (v 2.0.0-rc-69/1.52p) (65, 66) to process and analyze the
589 picture sets with an in-house ImageJ macro: The three multi-channel images were split into the
590 respective red, green, and blue image components. Further analysis included the red and the
591 green image component of the white light picture, the red image component of the red
592 fluorescent picture, and the green image component of the green fluorescent picture. The four
593 remaining images were merged into a stack, and we performed slice alignment (matching
594 method: normalized correlation coefficient) based on a selected landmark using the plugin
595 Template_Matching.jar (67). We used a rectangular piece of white tape on the black poster
596 board as landmark. To obtain the contours of the flies, we calculated the difference between
597 the red and the green image component of the white light picture and applied a median and a
598 Gaussian filter (radius = 3 pixels). After that, the picture was binarized using global
599 thresholding (option: Max Entropy) (68). The binary image was post-processed (functions: Fill
600 Holes, Open) before the position and the size of individual particles (=flies) were determined
601 with the Analyze Particles method of ImageJ (minimum size = 750 pixels²). To account for
602 translocations that have not been corrected for by the slice alignment (e.g., when the position
603 of the fly changed slightly), the convex hull for each particle was calculated and enlarged by
604 20 pixels. A median filter (radius = 2 pixels) was applied to both fluorescent pictures before
605 each particle (= fly) was scanned by a human investigator for the eye fluorescent pattern in

606 both fluorescent pictures. We compared the image-based screening pipeline to the screening
607 method using a stereo dissecting microscope and found that the estimated genotype frequencies
608 deviate not more than 1% from each other (n = 646 flies, 4 picture sets).

609

610 Phenotype data analysis, Cas9HF1 homing gene drive

611 When calculating drive parameters, we pooled offspring from the same type of cross together
612 and calculated rates from the combined counts. A potential issue of this pooling approach is
613 that batch effects could distort rate and error estimates (offspring were raised in separate vials
614 with different parents). To account for such effects, we performed an alternate analysis as in
615 previous studies (32, 45) by fitting a generalized linear mixed-effects model with a binomial
616 distribution using the function glmer and a binomial link function (fit by maximum likelihood,
617 Adaptive Gauss-Hermite Quadrature, nAGQ = 25). This allows for variance between batches,
618 usually resulting in different rate estimates and increased error estimates. Offspring from a
619 single vial were considered a distinct batch. This analysis was performed using the R statistical
620 computing environment (v3.6.1) (69) with packages lme4 (1.1-21, <https://cran.r-project.org/web/packages/lme4/index.html>) and emmeans (1.4.2, <https://cran.r-project.org/web/packages/emmeans/index.html>). The R script we used for this analysis is
621 available on Github (<https://github.com/MesserLab/Binomial-Analysis>). The results were
622 similar to the pooled analysis and are provided in Supplementary Data Sets S1- S2.

623

624 Genotyping

625 Flies were frozen, and DNA was extracted by grinding in 30 μ L of 10 mM Tris-HCl pH 8,
626 1mM EDTA, 25 mM NaCl, and 200 μ g/mL recombinant proteinase K (ThermoScientific),
627 followed by incubation at 37°C for 30 minutes and then 95°C for 5 minutes. The DNA was
628 used as a template for PCR using Q5 Hot Start DNA Polymerase from New England Biolabs.
629 The region of interest containing gRNA target sites was amplified using DNA oligo primers
630 AutoDLeft_S2_F and AutoDRight_S2_R. PCR products were purified after gel
631 electrophoresis using a gel extraction kit (Zymo Research). Purified products were Sanger
632 sequenced and analyzed with ApE (<http://biologylabs.utah.edu/jorgensen/wayned/ape>).

633

634 Fitness cost estimation framework

635 To estimate the fitness costs of the different transgenic constructs in our *D. melanogaster* cage
636 experiments, we modified a previously developed maximum likelihood inference framework
637 (42). Specifically, we extended the original model to a two-locus model, where the first locus

640 represents the construct insertion site and the second locus represents an idealized cut site. In
641 this model, cleavage at the cut site could represent in principle the effects of non-specific DNA
642 modifications (“off-target” effects) as well as the effects of cleavage at the desired gRNA target
643 site (i.e., target site activity). However, the latter is not expected to impose any fitness costs for
644 our constructs due to the intergenic location of the target site. Thus, we refer to the idealized
645 cut site as “off-target” site. At the construct locus, the two possible allele states are
646 EGFP/construct (observed by fluorescence); at the off-target site, the two possible states are
647 uncut/cut (not directly observed). The two loci are assumed to be autosomal and unlinked.
648 Thus, there are nine possible genotype combinations an individual could have in our model.
649 Unless stated otherwise, we assumed that the construct homozygotes used for the ancestral
650 generation of a cage are cut/cut homozygotes at the idealized off-target site. Since the construct
651 is not homing, the allelic state of a single individual cannot change at the construct locus. By
652 contrast, the allelic state at the off-target locus can be altered by cutting events in the germline
653 or in the early embryo phase. Germline cutting will only impact the genotype of offspring in
654 the next generation, while embryo cutting will directly change the individual’s genotype and
655 hence expose it to any potential fitness effects of this new genotype. Both the germline and
656 embryo off-target cut rates were set to 1 in our model. This means that any uncut allele at the
657 off-target locus will be cut in the germline if the individual carries at least one construct allele
658 (germline cute rate = 1). Furthermore, individuals will become cut/cut homozygotes if their
659 mother carried a least one construct allele (embryo cute rate = 1; we assume that maternally
660 deposited Cas9/gRNA is present in all such embryos).

661
662 A full inference model for the potential fitness costs of construct alleles and cut off-target
663 alleles that includes all three previously implemented types of selection (mate choice,
664 fecundity, viability) would feature a vast number of parameters that would be difficult to
665 disentangle (42). For simplicity and to avoid overfitting, we therefore reduced model
666 complexity with a series of assumptions: First, potential fitness costs were assumed to be equal
667 for both sexes. Second, we either included only viability selection in the model, or included
668 only mate choice (i.e., relative mating success for males with a particular genotype, reference
669 value = 1) and fecundity selection (i.e., relative fecundity for females with a particular
670 genotype, reference value = 1), both of equal magnitude. We further considered all fitness
671 effects to be multiplicative across the two loci and for the two alleles at each locus (e.g., a
672 construct homozygote would have a fitness equal to the square of a construct/EGFP
673 heterozygote, given the same genotype at the off-target site). This results in two much more

674 tractable inference models (viability and fecundity/mate choice) with only three parameters
675 overall: the effective population size (N_e), the relative fitness of construct/EGFP heterozygotes
676 versus EGFP homozygotes (the “direct fitness parameter”), and the relative fitness of cut/uncut
677 heterozygotes versus uncut homozygotes (the “off-target fitness parameter”).

678

679 Availability of data and materials

680 The annotated sequences of the final construct insertions are available in ApE format
681 (Supplemental file 1; constructs.zip). The raw counts of each experimental population
682 (different constructs and the Cas9HF1 homing drive) can be found in Supplemental file 2
683 (Supplemental_Data_Sets.xlsx). The macro of the image-based screening pipeline is available
684 on GitHub (<https://github.com/MesserLab/CRISPR-Cas9-fitness-effects>), a picture sample set
685 for the image-based screening pipeline can be found in the Supplemental file 3 (example-
686 pictures.zip). The raw data of the phenotypic assays can be found in Supplemental file 4
687 (phenotype_assays.zip). The maximum likelihood inference framework was implemented in R
688 (v 3.6.0) (69), and is available together with all necessary scripts to reproduce the results on
689 GitHub (<https://github.com/MesserLab/CRISPR-Cas9-fitness-effects>).

690

691 **Acknowledgements**

692

693 This study was supported by the National Institutes of Health awards R21AI130635 to JC,
694 AGC, and PWM, award F32AI138476 to JC, and award R01GM127418 to PWM. AML was
695 supported by Vetmeduni Vienna Funds and an Austrian Science Funds grant (FWF; DK
696 W1225-B20) awarded to Christian Schlotterer and an Austrian Marshall Plan Foundation
697 fellowship. We thank Marlies Dolezal for helpful advice on the statistical analysis of the
698 phenotypic assays. Special thanks to Charles Mazel from NIGHTSEA, and Guy Reeves for
699 support developing the image-based screening pipeline.

700

701 **References**

702

- 703 1. S. Bin Moon, D. Y. Kim, J. H. Ko, Y. S. Kim, Recent advances in the CRISPR
704 genome editing tool set. *Exp. Mol. Med.* **51** (2019).
- 705 2. E. K. and C. G. Mali Prashant, Cas9 as a versatile tool for engineering biology. *Nat
706 Methods* **10**, 957–963 (2013).
- 707 3. M. Chen, *et al.*, CRISPR-Cas9 for cancer therapy: Opportunities and challenges.

708 *Cancer Lett.* **447**, 48–55 (2019).

709 4. C. A. Hodges, R. A. Conlon, Delivering on the promise of gene editing for cystic
710 fibrosis. *Genes Dis.* **6**, 97–108 (2019).

711 5. D. H. Jo, *et al.*, CRISPR-Cas9-mediated therapeutic editing of Rpe65 ameliorates the
712 disease phenotypes in a mouse model of Leber congenital amaurosis. *Sci. Adv.* **5**, 1–10
713 (2019).

714 6. Y. Zhang, *et al.*, Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy
715 in mice by a self-complementary AAV delivery system. *Sci. Adv.* **6**, 1–11 (2020).

716 7. S. Yang, *et al.*, CRISPR / Cas9-mediated gene editing ameliorates neurotoxicity in
717 mouse model of Huntington ' s disease. *J. Clin. Invest.* **127**, 2719–2724 (2017).

718 8. W. Tang, D. R. Liu, Rewritable multi-event analog recording in bacterial and
719 mammalian cells. *Science (80-)* **360** (2018).

720 9. Y. Zhang, M. Pribil, M. Palmgren, C. Gao, A CRISPR way for accelerating
721 improvement of food crops. *Nat. Food* **1**, 200–205 (2020).

722 10. K. M. Esvelt, A. L. Smidler, F. Catteruccia, G. M. Church, Concerning RNA-guided
723 gene drives for the alteration of wild populations. *Elife* **3**, 1–21 (2014).

724 11. J. Champer, A. Buchman, O. S. Akbari, Cheating evolution: Engineering gene drives
725 to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**, 146–159 (2016).

726 12. R. L. Unckless, A. G. Clark, P. W. Messer, Evolution of resistance against
727 CRISPR/Cas9 gene drive. *Genetics* **205**, 827–841 (2017).

728 13. C. Noble, J. Olejarz, K. M. Esvelt, G. M. Church, M. A. Nowak, Evolutionary
729 dynamics of CRISPR gene drives. *Sci. Adv.* **3**, 3–10 (2017).

730 14. A. Burt, Heritable strategies for controlling insect vectors of disease. *Philos. Trans. R.
731 Soc. B Biol. Sci.* **369** (2014).

732 15. X. H. Zhang, L. Y. Tee, X. G. Wang, Q. S. Huang, S. H. Yang, Off-target effects in
733 CRISPR/Cas9-mediated genome engineering. *Mol. Ther. - Nucleic Acids* **4**, e264
734 (2015).

735 16. L. Alphey, Genetic Control of Mosquitoes. *Annu. Rev. Entomol.* **59**, 205–224 (2014).

736 17. E. Roggenkamp, *et al.*, Tuning CRISPR-Cas9 gene drives in *Saccharomyces
737 cerevisiae*. *G3 Genes, Genomes, Genet.* **8**, 999–1018 (2018).

738 18. R. Shapiro, *et al.*, A CRISPR–Cas9-based gene drive platform for genetic interaction
739 analysis in *Candida albicans*. *Nat. Microbiol.* **3** (2018).

740 19. J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, G. M. Church, Safeguarding
741 CRISPR-Cas9 gene drives in yeast. *Nat. Biotechnol.* **33**, 1250–1255 (2015).

742 20. E. M. Basgall, *et al.*, Gene drive inhibition by the anti-CRISPR proteins AcrIIA2 and
743 AcrIIA4 in *Saccharomyces cerevisiae*. *Microbiol. (United Kingdom)* **164**, 464–474
744 (2018).

745 21. A. M. Hammond, *et al.*, The creation and selection of mutations resistant to a gene
746 drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**, 1–16 (2017).

747 22. V. M. Gantz, *et al.*, Highly efficient Cas9-mediated gene drive for population
748 modification of the malaria vector mosquito *Anopheles stephensi*. *Proc. Natl. Acad.
749 Sci. U. S. A.* **112**, E6736–E6743 (2015).

750 23. K. Kyrou, *et al.*, A CRISPR–Cas9 gene drive targeting doublesex causes complete
751 population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* **36**,
752 1062–1066 (2018).

753 24. A. Hammond, *et al.*, A CRISPR-Cas9 gene drive system targeting female reproduction
754 in the malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* **34**, 78–83 (2016).

755 25. G. Oberhofer, T. Ivy, B. A. Hay, Behavior of homing endonuclease gene drives
756 targeting genes required for viability or female fertility with multiplexed guide RNAs.
757 *Proc. Natl. Acad. Sci. U. S. A.* **115**, E9343–E9352 (2018).

758 26. M. KaramiNejadRanjbar, *et al.*, Consequences of resistance evolution in a Cas9-based
759 sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad.
760 Sci. U. S. A.* **115**, 6189–6194 (2018).

761 27. V. M. Gantz, E. Bier, The mutagenic chain reaction: A method for converting
762 heterozygous to homozygous mutations. *Science (80-.).* **348**, 442–444 (2015).

763 28. J. Champer, *et al.*, Novel CRISPR/Cas9 gene drive constructs reveal insights into
764 mechanisms of resistance allele formation and drive efficiency in genetically diverse
765 populations. *PLoS Genet.* **13**, 1–18 (2017).

766 29. J. Champer, *et al.*, Reducing resistance allele formation in CRISPR gene drive. *Proc.
767 Natl. Acad. Sci. U. S. A.* **115**, 5522–5527 (2018).

768 30. J. Champer, *et al.*, Molecular safeguarding of CRISPR gene drive experiments. *Elife* **8**,
769 1–10 (2019).

770 31. J. Champer, *et al.*, CRISPR gene drive efficiency and resistance rate is highly heritable
771 with no common genetic loci of large effect. *Genetics* **212**, 333–341 (2019).

772 32. J. Champer, *et al.*, A toxin-antidote CRISPR gene drive system for regional population
773 modification. *Nat. Commun.* **11**, 1–10 (2020).

774 33. H. A. Grunwald, *et al.*, Super-Mendelian inheritance mediated by CRISPR–Cas9 in
775 the female mouse germline. *Nature* **566**, 105–109 (2019).

776 34. J. Champer, I. K. Kim, S. E. Champer, A. G. Clark, P. W. Messer, Performance
777 analysis of novel toxin-antidote CRISPR gene drive systems. *BMC Biol.* **18**, 27 (2020).

778 35. A. Burt, A. Crisanti, Gene Drive: Evolved and Synthetic. *ACS Chem. Biol.* **13**, 343–
779 346 (2018).

780 36. J. Champer, S. E. Champer, I. K. Kim, A. G. Clark, P. W. Messer, Design and analysis
781 of CRISPR-based underdominance toxin-antidote gene drives. *Evol. Appl.* **14**, 1052–
782 1069 (2021).

783 37. N. Wedell, T. A. R. Price, A. K. Lindholm, Gene drive: progress and prospects. *Proc.
784 R. Soc. B Biol. Sci.* **286**, 20192709 (2019).

785 38. J. Champer, I. K. Kim, S. E. Champer, A. G. Clark, P. W. Messer, Suppression gene
786 drive in continuous space can result in unstable persistence of both drive and wild-type
787 alleles. *Mol. Ecol.* **30**, 1086–1101 (2021).

788 39. A. Deredec, A. Burt, H. C. J. Godfray, The population genetics of using homing
789 endonuclease genes in vector and pest management. *Genetics* **179**, 2013–2026 (2008).

790 40. G. Oberhofer, T. Ivy, B. A. Hay, Cleave and Rescue, a novel selfish genetic element
791 and general strategy for gene drive. *Proc. Natl. Acad. Sci.* **116**, 6250 LP – 6259
792 (2019).

793 41. B. P. Kleinstiver, *et al.*, High-fidelity CRISPR-Cas9 nucleases with no detectable
794 genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

795 42. Y. L. Lee, *et al.*, Maximum Likelihood Estimation of Fitness Components in
796 Experimental Evolution. *Genetics* **211**, 1005–1017 (2019).

797 43. H. Akaike, “Information Theory and an Extension of the Maximum Likelihood
798 Principle” in *Selected Papers of Hirotugu Akaike*, (Springer New York, 1973), pp.
799 199–213.

800 44. A. Moya, F. González-Candelas, J. L. Ménsua, Larval competition in *Drosophila*
801 *melanogaster*: frequency-dependence of viability. *Theor. Appl. Genet.* **75**, 366–377
802 (1988).

803 45. S. E. Champer, *et al.*, Computational and experimental performance of CRISPR
804 homing gene drive strategies with multiplexed gRNAs. *Sci. Adv.* **6** (2020).

805 46. M. Kimberland, *et al.*, Strategies for controlling off-target effects and biological
806 variations in mammalian CRISPR/Cas9 genome editing experiments. *J. Biotechnol.*
807 **284** (2018).

808 47. G. Palermo, *et al.*, Molecular Mechanism of Off-Target Effects in CRISPR-Cas9.
809 *Biophys. J.* **116**, 319a (2019).

810 48. M. Klein, B. Eslami-Mossallam, D. G. Arroyo, M. Depken, Hybridization Kinetics
811 Explains CRISPR-Cas Off-Targeting Rules. *Cell Rep.* **22**, 1413–1423 (2018).

812 49. D. Carroll, Collateral damage: Benchmarking off-target effects in genome editing.
813 *Genome Biol.* **20**, 19–21 (2019).

814 50. S. Gisler, *et al.*, Multiplexed Cas9 targeting reveals genomic location effects and
815 gRNA-based staggered breaks influencing mutation efficiency. *Nat. Commun.* **10**,
816 1598 (2019).

817 51. W. T. Garrood, *et al.*, Analysis of off-target effects in CRISPR-based gene drives in
818 the human malaria mosquito. *Proc. Natl. Acad. Sci.*, 202004838 (2021).

819 52. A. R. North, A. Burt, H. C. J. Godfray, Modelling the suppression of a malaria vector
820 using a CRISPR-Cas9 gene drive to reduce female fertility. *BMC Biol.* **18**, 98 (2020).

821 53. J. Champer, *et al.*, A CRISPR homing gene drive targeting a haplolethal gene removes
822 resistance alleles and successfully spreads through a cage population. *Proc. Natl.*
823 *Acad. Sci.* **117**, 24377 LP – 24383 (2020).

824 54. A. Adolfi, *et al.*, Efficient population modification gene-drive rescue system in the
825 malaria mosquito *Anopheles stephensi*. *Nat. Commun.* **11**, 5553 (2020).

826 55. A. Casini, *et al.*, A highly specific SpCas9 variant is identified by in vivo screening in
827 yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).

828 56. Y. Tan, *et al.*, Rationally engineered *Staphylococcus aureus* Cas9 nucleases with high
829 genome-wide specificity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20969–20976 (2019).

830 57. P. Chatterjee, *et al.*, An engineered ScCas9 with broad PAM range and high specificity
831 and activity. *Nat. Biotechnol.* (2020) <https://doi.org/10.1038/s41587-020-0517-0>.

832 58. H. Xie, *et al.*, High-fidelity SaCas9 identified by directional screening in human cells.
833 *PLoS Biol.* **18**, e3000747 (2020).

834 59. I. M. Slaymaker, *et al.*, Rationally engineered Cas9 nucleases with improved
835 specificity. *Science (80-)* **351**, 84–88 (2016).

836 60. J. Lee, M. H. Jung, E. Jeong, J. K. Lee, Using Sniper-Cas9 to Minimize Off-target
837 Effects of CRISPR-Cas9 Without the Loss of On-target Activity Via Directed
838 Evolution. *J. Vis. Exp.*, 1–8 (2019).

839 61. F. Port, H.-M. Chen, T. Lee, S. L. Bullock, Optimized CRISPR/Cas tools for efficient
840 germline and somatic genome engineering in *Drosophila*. *Proc. Natl. Acad. Sci.* **111**,
841 E2967 LP-E2976 (2014).

842 62. S. J. Gratz, *et al.*, Highly specific and efficient CRISPR/Cas9-catalyzed homology-
843 directed repair in *Drosophila*. *Genetics* **196**, 961–971 (2014).

844 63. R. G. Reeves, D. Tautz, Automated Phenotyping Indicates Pupal Size in *Drosophila* Is
845 a Highly Heritable Trait with an Apparent Polygenic Basis. *G3*
846 *Genes|Genomes|Genetics* **7**, 1277–1286 (2017).

847 64. P. Nouhaud, F. Mallard, R. Poupardin, N. Barghi, C. Schlötterer, High-throughput
848 fecundity measurements in *Drosophila*. *Sci. Rep.* **8**, 4469 (2018).

849 65. J. Schindelin, C. T. Rueden, M. C. Hiner, K. W. Eliceiri, The ImageJ ecosystem: An
850 open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82**, 518–529 (2015).

851 66. J. Schindelin, *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat.*
852 *Methods* **9**, 676–82 (2012).

853 67. Q. Tseng, Template Matching and Slice Alignment--- ImageJ Plugins - ImageJ plugins
854 by Qingzong TSENG. *Template Matching and Slice Alignment*.

855 68. J. N. Kapur, P. K. Sahoo, A. K. C. Wong, A new method for gray-level picture
856 thresholding using the entropy of the histogram. *Comput. Vision, Graph. Image*
857 *Process.* **29**, 273–285 (1985).

858 69. R Core Team, R: A Language and Environment for Statistical Computing (2019).

859 70. J. Fox, G. Monette, Generalized Collinearity Diagnostics. *J. Am. Stat. Assoc.* **87**, 178–
860 183 (1992).

861 71. J. Fox, S. Weisberg, *An R Companion to Applied Regression*, Third (Sage, 2019).

862 72. R. Lenth, emmeans: Estimated Marginal Means, aka Least-Squares Means (2020).

863 73. J. Gross, U. Ligges, nortest: Tests for Normality (2015).

864

865

866 **Supplemental Information**

Plasmid construction

Construct with just DsRed (“no-Cas9_no-gRNAs”):

FACacR	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	pDsRed	FACacR_F	FACacR_R
<i>Plasmid Digest</i>	ATSacG	NcoI	HincII

Intermediate for the gRNAs:

TTTacU4	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	TTTgRNA _{At} RNAi	Acg4_41_F	Acg4_41_R
<i>PCR Product</i>	TTTgRNA _{At}	Acg4_12_F	Acg4_12_R
<i>PCR Product</i>	TTTgRNA _{At}	Acg4_23_F	Acg4_23_R
<i>PCR Product</i>	TTTgRNA _{At}	Acg4_34_F	Acg4_34_R

867

Construct with Cas9 and no gRNAs (“Cas9_no-gRNAs”):

FACacN	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	none	acN_F	acN_R
<i>Plasmid Digest</i>	BHDgN1c	StuI	XbaI

Construct with Cas9 and four gRNAs (“Cas9_gRNAs”):

FACacN4	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	none	U6_3_gRNA1_v4_F	gRNA_f_R
<i>Plasmid Digest</i>	BHDgN1c	StuI	XbaI

Intermediate for Cas9HF1:

Nos-Cas9HF1	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	VP12	Cas9HF1_F	Cas9HF1_R
<i>Plasmid Digest</i>	nos-Cas9-nos	Bsu36I	FspI

Construct with Cas9HF1 and four gRNAs (“Cas9HF1_gRNAs”):

FACacNf4	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	Nos-Cas9HF1	HF1_F	HF1_R
<i>Plasmid Digest</i>	FACacN4	Bsu36I	AscI

Homing drive with Cas9HF1:

BHDgNf1v2	<i>Template</i>	<i>Oligo/Enzyme 1</i>	<i>Oligo/Enzyme 2</i>
<i>PCR Product</i>	Nos-Cas9HF1	HF1_F	HF1_R
<i>Plasmid Digest</i>	BHDgN1cv3	Bsu36I	AscI

868

Construction oligonucleotides

Acg4_12_F: GGCAATATAGGAATGCACGTTTAGAGCTAGAAATAGCAAGTTAA
Acg4_12_R: AACACTCGGTATAATTGGTTATGCACCAGCCGGGAATCG
Acg4_23_F: GCATAAACCAATTATACCGAGTGTAGAGCTAGAAATAGCAAGTTAA
Acg4_23_R: AACTCCCCGCAAGTTCTGCCCCGTGACCAGCCGGGAATCG
Acg4_34_F: GCAAGGGACAGAACTTGCAGGGAGTTAGAGCTAGAAATAGCAAGTTAA
Acg4_34_R: GGTGGTCTCGTTTCCACTTGCACCAGCCGGGAATCG
Acg4_41_F: GTGCAAGTGGAAAACGGAGACCACCGTTAGAGCTAGAAATAGCAAGTTAA
Acg4_41_R: AAAACGTGCATTCTATATATTGCCTGCATCGGCCGGGAATCG
acN_F: CAAACTCATCAATGTATCTAACCGGTAGGAGCAAGCTGCCGTGCCCTGCCACCTC
acN_R: GAGGGTGGGCCAGGGCACGGGCAGCTGCTCCTACCGGTTAAGATACTTGATGAGTTG
Cas9HF1_F: CACCTGGCGAAGTCACGCTATCCTCAGGAGGCAGGAGGATTTTATCCGT
Cas9HF1_R: ACCACTGCATTCAAGGTAGGCATCATGCGCATGGTGGTAGTTATTATCTCCCTAACCTT
FACacR_F: CTAAACAATCGGCTCGAACG
FACacR_R: GTAACCATTATAAGCTGCAATAACAA
gRNA_f_R: GAGGGTGGGCCAGGGCACGGCAGCTGCTCTAGAATGCATACGCATTAAGCGAAC
HF1_F: GGTGGTGTGCAAGTACTTGAAG
HF1_R: AGATTCACCTGGCGAAGT
U6_3_gRNA1_v4_F:
GTCCAAACTCATCAATGTATCTAACCGTAGGCCTTTTGCTCACGTGATTGCTC

Sequencing oligonucleotides

869 AutoDLeft_S2_F: CTTACGCTGAAGCCATTCAA
870 AutoDRight_S2_R: ATCTGGTTCTCACTCCATTAAAT
871 Cas9_HF1_S1_R: GGACTTTCTTGTCACTCCATGCG
872 Cas9HF1_S_R: CTACCCCGGAGATCTGACAG
873 Cas9mid_S_F: CGACCCAGTACGCAGACCTTT
874 DsRed_S_F: CTGAAGGGCGAGATCCACAAG
875 EGFP_S_R: AGTTGTACTCCAGCTTGTGCC
876 pCFD5_S_R: ACGTCAACGGAAAACCATTGTCTA
877
878

879 **Supplementary methods**

880

881 Phenotypic assays

882

883 We measured three fitness proxies for flies carrying Cas9_gRNAs constructs: mate choice,
884 fecundity, and viability. All phenotypic assays were conducted on Bloomington Standard
885 medium and under the same temperature (25°C) and light conditions (14 hours light/10 hours
886 dark) as the caged populations. The statistical analysis of the phenotypic assays was conducted
887 in R (v3.6.0). (69)

888

889 *Mate choice*

890 We conducted a mate choice assay to test for mating preferences of EGFP homozygous
891 females. Individual 2-day-old virgin EGFP homozygous females were set up with one EGFP
892 homozygous male and one Cas9_gRNAs homozygous male of the same age in a vial. After 24
893 hours, the adult flies were removed, and the genotypes of the eclosed offspring were assessed
894 after 11 to 12 days. If the EGFP homozygous female has mated only with the male of the same
895 genotype, only homozygous offspring is expected. We tested for deviations from an expected
896 equal frequency of offspring genotypes under the null hypothesis of no mate preference via a
897 binomial test.

898

899 *Fecundity*

900 We assessed the fecundity of EGFP homozygous, heterozygous, and Cas9_gRNAs
901 homozygous females in individual crosses with EGFP homozygous males. Each individual
902 single 2-day-old virgin female of a distinct genotype was crossed with one EGFP homozygous
903 male of the same age. Crosses were flipped on fresh medium every 24 hours, and eggs were
904 counted manually using a stereo dissecting microscope. Fecundity was defined as the total
905 number of laid eggs per female over three consecutive days. To assess the impact of female
906 genotype we fitted a linear model using function lm() with fecundity as response. The female
907 genotype was the only fixed effect in the model. The residuals were both normally distributed
908 and showed variance homogeneity, meeting all assumptions of a linear model. None of the
909 used model diagnostics (Cook's distance, DFbetas, leverage (70); calculated with the R
910 package car (v3.0-3) (71)) indicated strongly influential cases or outliers. We used the R
911 package emmeans (v1.4.7) (72) to conduct pairwise comparisons of the three assessed female
912 genotypes.

913

914 *Viability*

915 We measured relative viability of heterozygous offspring of single crosses between EGFP
916 homozygous males and heterozygous females. Single 2-day-old heterozygous virgin females
917 were crossed each with one EGFP homozygous male of the same age. After 24 hours, the adult
918 flies were removed, and the genotypes of the eclosed offspring was assessed after 11 to 12
919 days. Relative heterozygote viability was defined as the fraction of heterozygous offspring out
920 of the total number of offspring, ranging between 0 and 1. If the genotype does not influence
921 viability, we expect a relative heterozygote viability of 0.5. Relative heterozygote viability was
922 tested for normality with an Anderson-Darling test (function ad.test() in the R package nortest
923 (v1.0-4) (73)). We then used a one-sample t-test against a population mean of 0.5 for
924 heterozygotes viability.

925

926 **Supplementary results**

927

928 Population cage experiments setups

929

930 All experiments started by crossing construct and EGFP homozygotes, except for replicates 1,
931 and 2 of Cas9_gRNAs. These two experimental populations were set up with all three
932 genotypes that originated from the same batch that included heterozygotes and both
933 homozygotes. While construct homozygotes of Cas9_no-gRNAs, no-Cas9_no-gRNAs, and
934 Cas9HF1_gRNAs were of the same age as the EGFP homozygotes they were mixed with to
935 start the experiments, the age differed between EGFP and construct homozygotes for
936 Cas9_gRNAs replicates 1, 2, 5, 6, and 7. To avoid confounding maternal effects on the
937 construct frequency dynamics, we excluded for each of these replicates the first generation
938 from the analysis. The full data set including the removed time points can be found in
939 Supplemental File 2.

940

941 Population sizes were controlled via the limited egg-lay time period, which led to fluctuations
942 in the number of flies per generation (Figure S1). Some experiments experienced bottlenecks
943 due to high variation in food moisture content (resulting in either high or low larvae density).

944

945

946

947 Phenotypic assays

948

949 *Mate choice*

950 We assessed the mate choice of 40 independent EGFP homozygous females that were each set
951 up with one EGFP homozygous and one Cas9_gRNAs homozygous male in one single vial.
952 38 samples had exclusively EGFP homozygous or heterozygous offspring, whereas 2 samples
953 displayed offspring of both genotypes. As this suggests multiple matings of the female, we
954 excluded these two data points from the analysis. The estimated frequency of 0.684 of EGFP
955 homozygous females choosing EGFP homozygous males ($n=26$) over Cas9_gRNAs
956 homozygous males ($n=12$) as mates was significantly different from 0.5 (exact binomial test
957 $P=0.033$; Figure S4A).

958

959 *Fecundity*

960 In total, we measured the fecundity of 128 independent females (27 EGFP homozygotes, 55
961 heterozygotes, and 46 Cas9_gRNAs homozygotes). Overall, the female genotype significantly
962 influenced fecundity, which was defined as the total number of laid eggs per female over the
963 course of three consecutive days (full-null model comparison $F_{2,125}=5.885$, $P=0.004$).
964 Cas9_gRNAs homozygous females are significantly less fecund than EGFP homozygous
965 females. However, no significant difference was detected between EGFP homozygotes and
966 heterozygotes, or heterozygotes and Cas9_gRNAs homozygotes respectively (Figure S4B).

967

968 *Viability*

969 We determined the relative heterozygote viability (=fraction of heterozygous offspring out of
970 the total number of offspring, ranging between 0 and 1) of 56 independent fly crosses. We
971 observed that the relative heterozygote viability is normally distributed (mean = 0.486, standard
972 deviation = 0.098; $A=0.405$, $P=0.343$) and average heterozygote viability does not significantly
973 differ from 0.5 ($t_{55}=-1.057$, $P=0.295$; Figure S4C).

974 **Supplementary tables**

975

976 Table S1 Model comparison of Cas9_no-gRNAs, no-Cas9_no-gRNAs, Cas9HF1_gRNAs – all cut parameters
977 set to 1

construct	model	selection	\hat{N}_e	direct fitness estimate	off-target fitness estimate	$\ln\hat{L}$	P	AICc
Cas9_no-gRNAs	full	viability	252	1	0.76	89.3	3	-171
			[157 - 379]	[0.97 – 1.05]	[0.57 – 1.29]			
Cas9_no-gRNAs	construct	viability	243	1	1*	88.6	2	-173
			[152 - 366]	[0.96 – 1.04]				
Cas9_no-gRNAs	off-target	viability	252	1*	0.76	89.3	2	-174
			[157 - 379]		[0.57 – 1.29]			
Cas9_no-gRNAs	neutral	none	243	1*	1*	88.6	1	-175
			[152 – 366]					
no-Cas9_no-gRNAs	full	viability	162	1	1.06	81.5	3	-155
			[101 - 244]	[0.97 - 1.10]	[0. 74 – 2.06]			
no-Cas9_no-gRNAs	construct	viability	162	1	1*	81.5	2	-158
			[101 - 244]	[0.97 - 1.10]				
no-Cas9_no-gRNAs	off-target	viability	162	1*	1.06	81.5	2	-158
			[101 - 244]		[0. 74 – 2.06]			
no-Cas9_no-gRNAs	neutral	none	162	1*	1*	81.5	1	-161
			[101 - 244]					
Cas9HF1_gRNAs	full	viability	444	0.99	1.35	90.2	3	-173
			[240 – 682]	[0.96 – 1.02]	[1.04 – 1.97]			
Cas9HF1_gRNAs	construct	viability	396	1	1*	88.1	2	-171
			[240 - 608]	[0.97 – 1.04]				
Cas9HF1_gRNAs	off-target	viability	440	1*	1.30	90.0	2	-175
			[267 – 675]		[1.00 – 1.88]			
Cas9HF1_gRNAs	neutral	none	396	1*	1*	88.1	1	-174
			[240 - 608]					

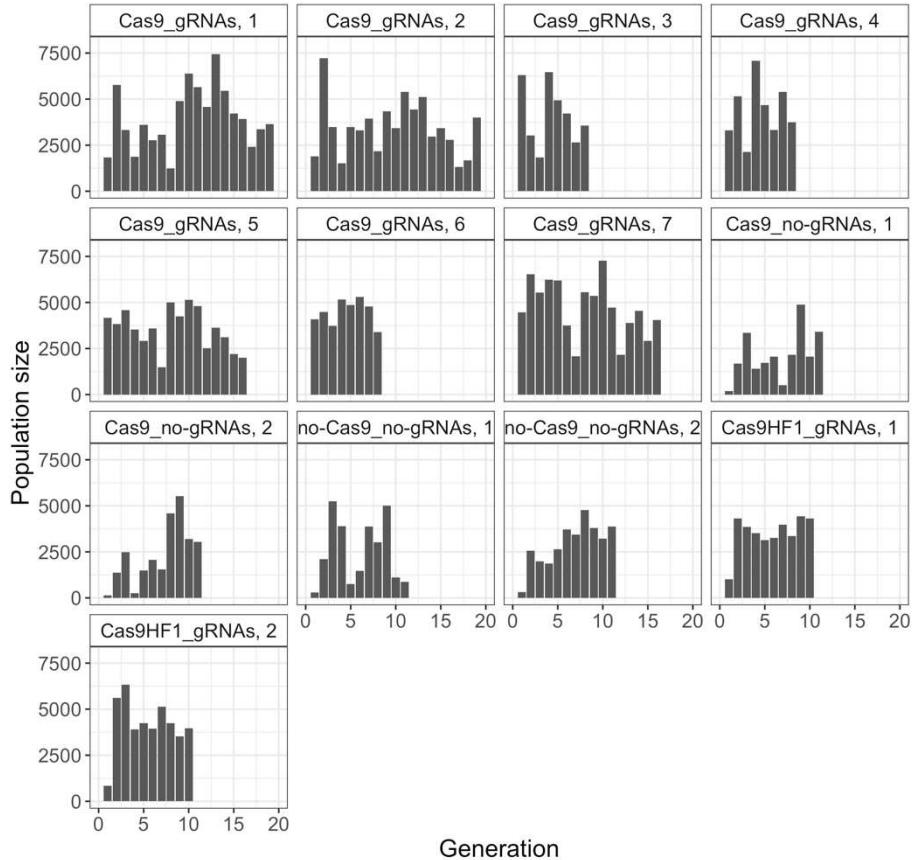
978

979 Each row shows the parameter estimates (\hat{N}_e = effective population size), maximum log Likelihood ($\ln\hat{L}$), number
980 of free parameters in the maximum likelihood framework (P), and corrected Akaike Information Criterion value
981 ($AICc = 2p - 2\ln\hat{L} + (2p^2 + 2p)/(n - p - 1)$ where n = number of generation transitions; n= 20 for Cas9_no-
982 gRNAs, no-Cas9_no-gRNAs, and n= 18 for Cas9HF1_gRNAs) for a specific construct, model and selection type.
983 1* entries indicate that a parameter was fixed at 1 (= no fitness effect is estimated). Values in squared brackets in
984 the parameter estimate columns represent the 95 % confidence intervals estimated from a likelihood ratio test with
985 one degree of freedom.

986

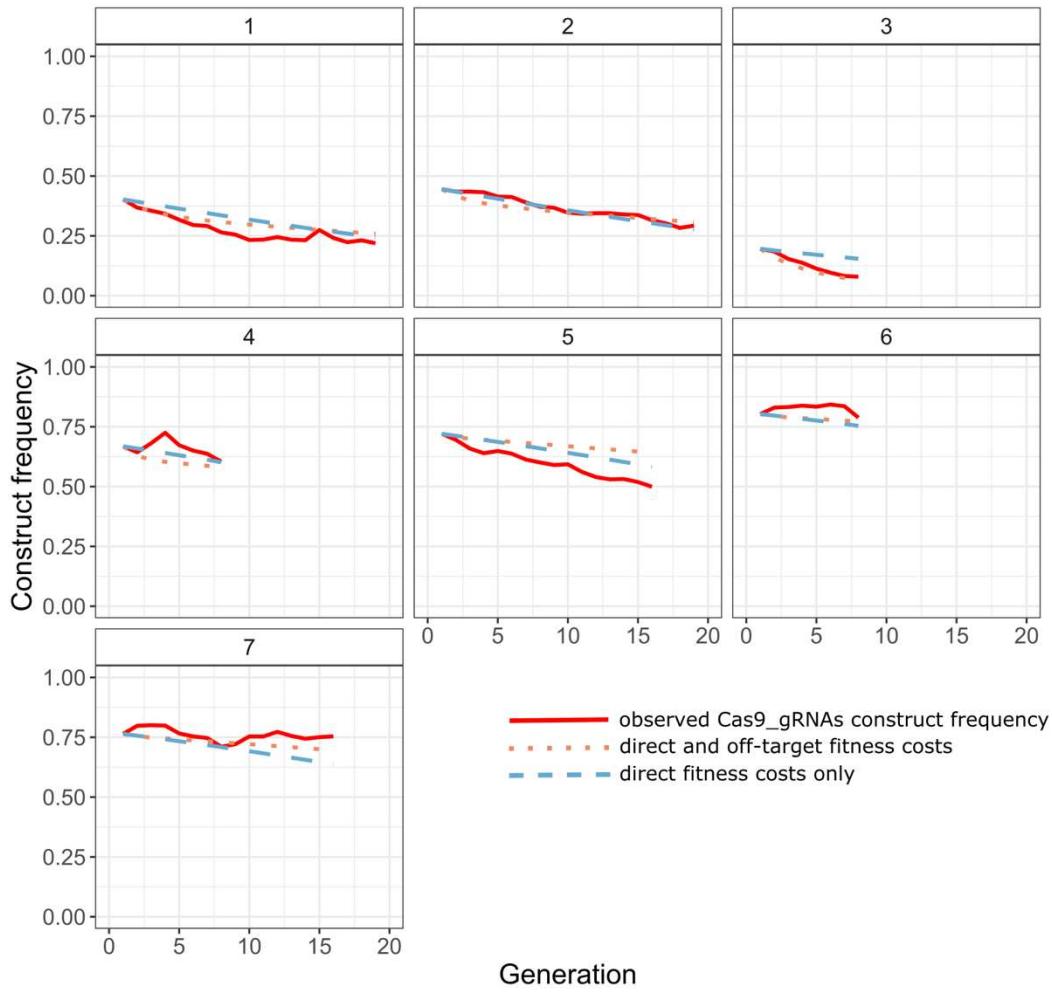
987
988
989

Supplementary figures



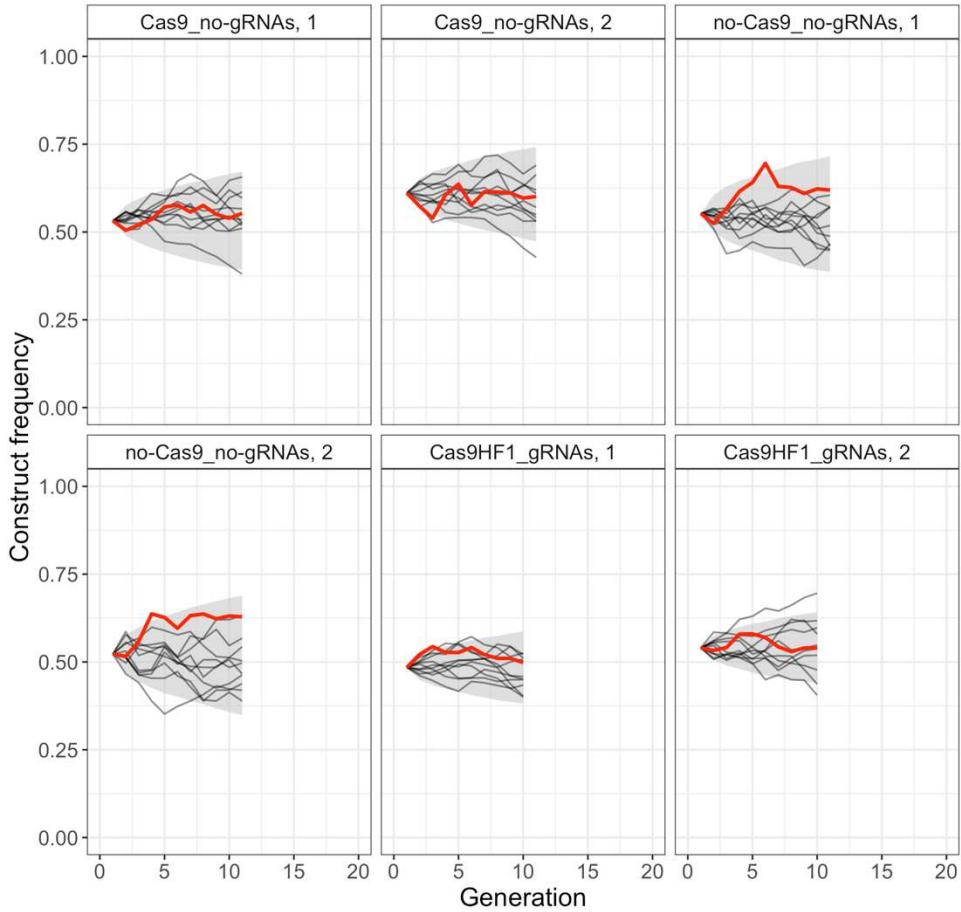
990
991

Figure S1. Population sizes of all *D. melanogaster* cage populations.



992

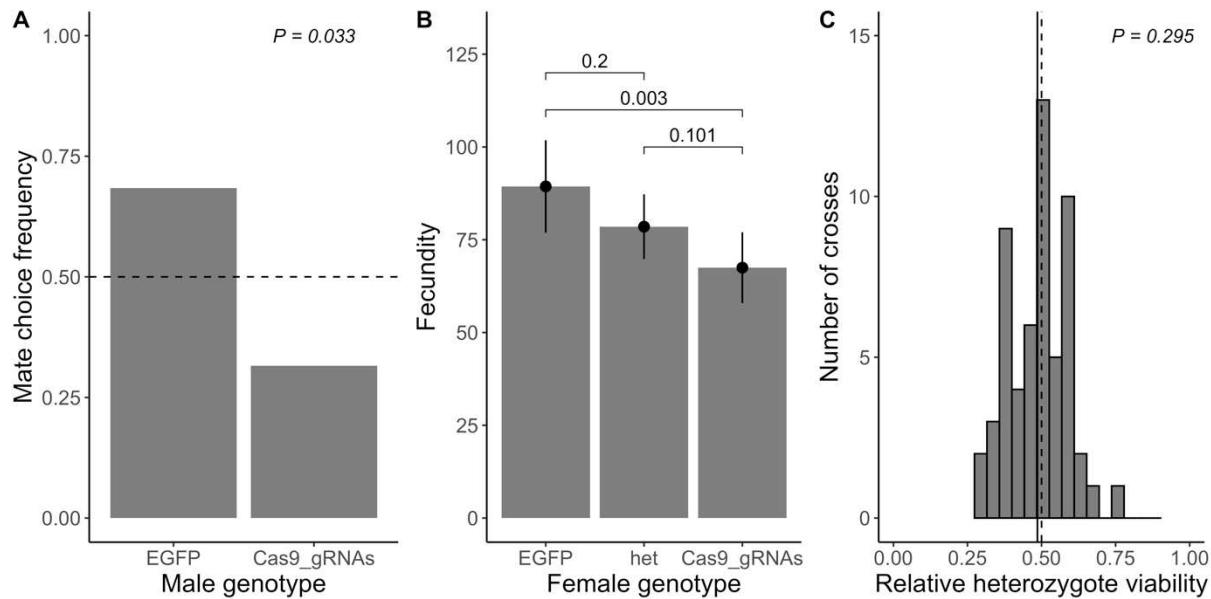
993 Figure S2. Comparison of observed construct frequencies (solid red line) in our experimental Cas9_gRNAs cages
994 with the predicted trajectories of the full inference model with viability selection (dotted, orange line; off-target
995 fitness = 0.84, direct fitness = 0.98), and the construct model with viability selection (dashed, blue line; direct
996 fitness = 0.96), using the inferred maximum likelihood parameter estimates (Table 1). Genetic drift was not
997 simulated.



998

999 Figure S3. Comparison of observed construct frequencies with simulated trajectories of a neutral model (Cas9_no-
1000 gRNAs: $\hat{N}_e = 243$; no-Cas9_no-gRNAs: $\hat{N}_e = 162$; Cas9HF1_gRNAs: $\hat{N}_e = 396$). Solid red lines present
1001 observed construct frequencies, black lines show ten simulated trajectories for each cage, and the shaded area
1002 represents the range between the 2.5 and 97.5 percentile of the simulated trajectories (10,000 simulations per
1003 cage).

1004



1005

1006 Figure S4. Direct measurement of fitness parameters. (A) Observed mate choice frequency (y-axis) of EGFP
1007 homozygous females choosing between EGFP and Cas9_gRNAs homozygous males (x-axis; as only two
1008 genotypes were tested, the frequencies sum up to 1). In case of no mate choice preference, the expected mate
1009 choice frequency is 0.5 (horizontal dashed line). The observed mate choice frequency of EGFP homozygous males
1010 was significantly different from 0.5 (exact binomial test; $P = 0.033$). (B) Average fecundity (y-axis) for each
1011 female genotype (x-axis). The observed average fecundity (= total number of eggs per female laid over the course
1012 of three consecutive days) is plotted for each female genotype separately (EGFP = EGFP homozygous females,
1013 het = heterozygous females, Cas9_gRNAs = Cas9_gRNAs homozygous females). All females were mated in
1014 individual crosses to EGFP homozygous males of the same age. The fitted model is shown as black dots with
1015 error bars displaying the 95 % confidence interval. P-values of pairwise genotype comparisons adjusted with the
1016 Tukey method are displayed above the bars. (C) Relative heterozygote viability (= fraction of heterozygous
1017 offspring of crosses between heterozygous females and EGFP homozygous males). If the genotype does not
1018 influence viability, we expect a relative heterozygote viability of 0.5 (vertical dashed line). The observed average
1019 heterozygote viability (vertical solid line) does not differ from 0.5 (one sample t-test; $P = 0.295$).