Causes of variability in estimates of mutational variance from mutation accumulation experiments

Cara Conradsen[1*], Mark W. Blows[1] and Katrina McGuigan[1]

[1]School of Biological Sciences; The University of Queensland; St. Lucia, Queensland, Australia 4072

Running Title: Variability of mutational variance

Keywords: meta-analysis, Drosophila, life history, morphology, heritability, coefficient of variance, wing, sampling error

[*]Corresponding Author: Cara Conradsen

School of Biological Sciences

The University of Queensland;

St. Lucia, QLD, 4072 AUSTRALIA

T: +61(0)450645318

E-mail: cara.conradsen@uqconnect.edu.au

ABSTRACT

Characteristics of the new phenotypic variation introduced via mutation have broad implications in evolutionary and medical genetics. Standardised estimates of this mutational variance, V_M, span two orders of magnitude, but the causes of this remain poorly resolved. We investigated estimate heterogeneity using two approaches. First, meta-analyses of ~150 estimates from 37 mutation accumulation (MA) studies did not support a difference among taxa (which differ in mutation rate) in standardised V_M, but provided equivocal support for standardised V_M to vary with trait type (life history versus morphology, predicted to differ in mutation rate). Notably, several experimental factors were confounded with taxon and trait, and further empirical data are required to resolve their influences. Second, we analysed morphological data from an experiment in *Drosophila serrata* to determine the potential for unintentional heterogeneity among environments in which phenotypes were measured (i.e., among laboratories or time points) or transient segregation of mutations within MA lines to affect standardised V_M. Approximating the size of an average MA experiment, variability among repeated estimates of (accumulated) mutational variance was comparable to variation among published estimates of standardised V_M for morphological traits. This heterogeneity was (partially) attributable to unintended environmental variation or within line segregation of mutations only for wing size, not wing shape traits. We conclude that sampling error contributed substantial variation within this experiment, and infer that it will also contribute substantially to differences among published estimates. We suggest a logistically permissive approach to improve the precision of estimates, and consequently our understanding of the dynamics of mutational variance of quantitative traits.

INTRODUCTION

The magnitude of per-generational increase in genetic variance due to spontaneous mutations (V_M) is important for a wide range of genetic and evolutionary phenomena, including the maintenance of quantitative genetic variance (LYNCH 1988; BARTON AND TURELLI 1989; JOHNSON AND BARTON 2005). Much of our understanding of V_M comes from mutation accumulation (MA) experiments, where populations diverge phenotypically due solely to the neutral fixation of new mutations (MUKAI 1964; HALLIGAN AND KEIGHTLEY 2009). Reviews of MA experiments in a range of traits and taxa have reported that mutation increases phenotypic variance in quantitative traits by 10^{-4} to 10^{-2} times the environmental variance of the trait, or 0.02 to 5.1 percent of the trait mean per generation (HOULE *et al.* 1996; LYNCH *et al.* 1999; HALLIGAN AND KEIGHTLEY 2009). Differences in V_M may cause differences in the magnitude of standing quantitative genetic variation and, ultimately, in rates of phenotypic evolution (HOULE 1998; LYNCH *et al.* 1999; HOULE *et al.* 2017; WALSH AND LYNCH 2018). However, the causes of variation among estimates of V_M, and thus the evolutionary interpretation of this variability, are not well resolved.

Mutation rate is known to vary widely among species (reviewed in KATJU AND BERGTHORSSON 2019), with further opportunity for differences in per-generation mutation number arising through differences in ploidy, genome size, and/or effective population size (LYNCH *et al.* 1999; LYNCH 2010; SUNG *et al.* 2012). Marked variation in mutation rate has also been observed within species, both among replicated MA experiments (i.e., different founder genotypes: NESS *et al.* 2012; SUNG *et al.* 2012; SCHRIDER *et al.* 2013; HO *et al.* 2020) and among lines within a single MA panel (HUANG *et al.* 2016; HO *et al.* 2020). Resulting differences in mutation number may explain variation in V_M estimates, such as, for example, the fourfold range of h^2_M of body size estimated for different MA in *Caenorhabditis elegans* (AZEVEDO *et al.* 2002; ESTES *et al.* 2005; OSTROW *et al.* 2007).

65 Traits have also been hypothesised to differ in magnitude of V_M due to differences in mutation
66 rate, arising due to differences in the number of contributing loci. Specifically, life history traits are
67 hypothesised to be affected by more loci than morphological traits (HOULE 1991; HOULE 1992;
68 HOULE *et al.* 1996; HOULE 1998; MERILÄ AND SHELDON 1999). The magnitude of V_M depends not
69 only on the rate of mutation, but also on their effects, and the relationship between rate and effect
70 size is not well characterised. BESNARD *et al.* (2020) demonstrated that the high mutational variance
71 (and relatively rapid evolution) of a vulval phenotype in nematodes was due to a broad mutational
72 target size, rather than large-effect mutation. Whether trait types differ systematically in mutational
73 target size is difficult to assess, as a full catalogue of causal loci is unknown for most quantitative
74 traits (BARTON AND KEIGHTLEY 2002; MACKAY *et al.* 2009; YANG *et al.* 2010; ROCKMAN 2012).
75 Indeed, emerging evidence that diverse traits, including morphology, are all highly polygenic (YANG
76 *et al.* 2010; BOYLE *et al.* 2017) suggests that differences in the distribution of mutational effect sizes
77 (SIMONS *et al.* 2018), rather than simply in number of contributing loci, might cause heterogeneity in
78 estimates of V_M among traits.

79 Comparison among trait types is complicated by differences among them in variability and
80 measurement scale, which may influence standardised values. Low mutational heritability ($h^2_M = V_M/$
81 V_E, where V_E is the environmental variance) of life history traits relative to morphological traits has
82 been attributed to greater environmental variance (larger V_E) for life history traits, rather than lower
83 V_M (HOULE *et al.* 1996). Thus, comparison on the coefficient of mutational variance scale (CV_M;
84 $100 \times \sqrt{V_M}/\bar{X}$, where $\bar{X}$ is the trait mean) reveals a different picture, one of greater mutational
85 variance in life history than morphological traits, consistent with the prediction of greater mutational
86 target size (HOULE *et al.* 1996).

87 Other contributions to variation in magnitude of V_M might be revealed by consideration of the
88 MA experimental design itself. The timeframe over which mutations accumulate might influence

89 estimates of V_M. When MA lines are established from a homozygous (heterozygous) base population,
90 estimates of V_M will be downwardly (upwardly) biased before $6N_e$ generations (LYNCH AND HILL
91 1986). However, V_M is typically estimated after > $6N_e$ generations, suggesting limited contribution of
92 ancestral variation to variability of V_M. Conversely, long-running MA experiments might under-
93 estimate V_M when the cumulative effect of low fitness mutation causes line extinction, or within-line
94 selection against further accumulation (LYNCH *et al.* 1999; ESTES *et al.* 2004; MCGUIGAN AND
95 BLOWS 2013). A decline in V_M over time has been observed in some studies (MACKAY *et al.* 1995),
96 but not in others (GARCÍA-DORADO *et al.* 2000; HALL *et al.* 2008).

97 Stochastic sampling from the distribution of mutational effects could also introduce temporal
98 heterogeneity among estimates of V_M. For example, among-line variance estimated before versus
99 after a line(s) fixed a large effect mutation(s) could result in inference of a much larger per-
100 generation increase in variance at the second time-point relative to the first. Transient within-line
101 segregation of mutations might generate variability in estimates, for example causing temporary
102 inflation of within-line variance (V_E), impacting power to detect among-line variance, and potentially
103 biasing estimates of h^2_M (downward) and CV_M (upward; see HOFFMANN *et al.* 2016). Notably, several
104 studies in nematode have suggested that within-line variance increased over the duration of the MA
105 experiment (BAER 2008; BAER *et al.* 2010; BRAENDLE *et al.* 2010), which may contribute to a pattern
106 of lower estimated V_M in longer-running MA experiments.

107 Environmental context within which MA lines are assayed could also contribute to variation
108 among V_M estimates. Several studies have considered the effect of replicable, experimenter-imposed,
109 changes in the environment, including in temperature (WAYNE AND MACKAY 1998), light
110 (KAVANAUGH AND SHAW 2005) and density (FRY *et al.* 1996). Although the magnitude of V_M often
111 varies under such environmental manipulations, there is only weak evidence for predicable patterns,
112 such as novel or stressful environments increasing the magnitude of V_M (KONDRASHOV AND HOULE

113 1994; MARTIN AND LENORMAND 2006). Even in carefully controlled laboratory experiments, factors
114 such as food quality or quantity, light, humidity and diurnal timing of collection will vary among
115 individuals or lines within a phenotyping assay, and among assays conducted in different laboratories
116 or at different times within the same laboratory. Such variation may impact estimates of standardized
117 V_M through inflation of within-line variance, similar to the effect of transient, within-line segregation
118 of new mutations. Furthermore, if MA lines differ in their response to this unintended environmental
119 heterogeneity, then genotype by environment (GxE) variance could contribute variation among MA
120 lines, and variability among estimates of V_M, a potential source of variation that has received little
121 attention (but see GARCÍA-DORADO *et al.* 2000).

122 Here, we combined two approaches to investigate causes of variability in estimates of
123 mutational variance. Given that it has been over twenty years since this variability has been broadly
124 documented and investigated (HOULE *et al.* 1996; HOULE 1998; LYNCH *et al.* 1999), we first
125 conducted a meta-analysis to update tests of the previously implicated causal factors of taxon (LYNCH
126 AND WALSH 1998; LYNCH *et al.* 1999; HALLIGAN AND KEIGHTLEY 2009) and trait type (HOULE *et al.*
127 1996; HOULE 1998). We had intended to examine how the number of generations affected V_M
128 (LYNCH AND HILL 1986; MACKAY *et al.* 1995), but MA duration was strongly confounded with taxon
129 (detailed in the Results). Second, we conducted a new empirical experiment in *Drosophila serrata*, in
130 which we repeatedly estimate the among-line (mutational) variance to investigate whether unintended
131 environmental heterogeneity, or transient within-line segregation of mutations can contribute
132 variation among estimates. After accounting for these effects within the data, we finally quantify the
133 magnitude of variation among estimates from a set of 10 wing shape traits to characterise the
134 magnitude of variation among estimates within a trait category.

135 **METHODS**

136 **Meta-analysis of empirical estimates of mutational variance**

137 *Literature search*

138 We extracted all papers in seven reviews of mutational variance: LYNCH (1988); KEIGHTLEY *et al.*
139 (1993); HOULE *et al.* (1996); LYNCH AND WALSH (1998); LYNCH *et al.* (1999); HALLIGAN AND
140 KEIGHTLEY (2009); WALSH AND LYNCH (2018). We then searched the Web of Science database on
141 the 11/12/2019 at 4:38 PM AEST for *journal article* document types meeting the topic criteria of
142 "mutation accumulation" and (varia* or "mutat* coefficient*") and published between 1998 and
143 2019. These years overlapped HALLIGAN AND KEIGHTLEY (2009) (fitness traits only) and WALSH
144 AND LYNCH (2018) (brief update on LYNCH AND WALSH 1998), allowing us to capture papers that
145 may have been excluded from those reviews, as well as those published subsequently.

146 Further details on the papers identified and preliminary handling steps can be found in Figure
147 S1. For 473 unique papers identified, we screened titles and abstracts, then the full text, for relevance,
148 applying four strict criteria, retaining only studies where the estimates of mutational variance were:
149 1) quantitative; 2) from spontaneous MA; 3) from MA environmental conditions and; 4) not re-
150 reporting of previously published estimates. We excluded six studies of transcriptomic data as the
151 number of traits was much larger than for other trait categories.

152 *Meta-analysis data collection*

153 For each of the 65 papers retained after applying the above criteria, mutational parameter estimates
154 were extracted (as described in Table S1), associated with taxon and trait identifiers, and details of
155 the experimental design. Twenty papers not reporting error for the mutational parameters were
156 excluded (Table S1.C). Following initial qualitative assessments of data, we excluded five studies (15
157 traits) due to low representation of taxon type (one vertebrate, *Mus musculus*; one alga,
158 *Chlamydomonas reinhardtii*, and; two non-*Drosophila* insects: *Daktulosphaira vitifoliae* and
159 *Nasonia vitripennis*), and one study due to low representation of trait type: mitotic cell division traits
160 (Table S1.B).

161 Where possible, we extracted (or calculated from provided information) both the coefficient
162 of variance (CV_M; $100 \times \sqrt{V_M}/\bar{X}$, where $\bar{X}$ is the trait mean) and mutational heritability (h^2_M; V_M/V_E,
163 where V_E is the environmental variance) for each trait. As detailed below, estimates were weighted
164 by the inverse of their standard error (SE) in the meta-analysis. Where these were not reported for h^2_M
165 or CV_M, but were for V_M and V_E or $\bar{X}$, we used a sampling approach to obtain estimates. We sampled
166 from $N \sim (\hat{\theta}, V)$ 10,000 times, using the *rnorm* function in R [v. 3.6.1], where $\hat{\theta}$ and V were
167 respectively the reported parameter value and its standard error. We then calculated CV_M or h^2_M for
168 each of these simulated samples, and obtained the standard error of this sample of estimates. Samples
169 with negative values of V_M are undefined for CV_M; to ensure unbiased estimates of the magnitude of
170 error we calculated CV_M as: $100 \text{ x } (\text{sign of } V_M) \text{ x } \sqrt{|V_M|}/\bar{X}$. This sampling approach was used to
171 estimate the error for 28% of the h^2_M estimates and 61% of the CV_M estimates analysed (Table S1.A).
172 Two studies (17 estimates) were excluded due to nonsensically large standard error estimates, while a
173 further three estimates (from three studies) were excluded due nonsensical scaled parameter estimates
174 (detailed in Table S1.B). Two extreme values (> 3 SD) of h^2_M and two of CV_M were excluded from
175 the analyses (Table S1.B). There were 11 cases with extremely small SE (> 5 IQR below the
176 median); notably, for six of these came from studies where confidence intervals were constrained to
177 be positive, suggesting that this boundary condition had reduced the SE estimate, inflating meta-
178 analysis weights for traits where the mutational variance was not supported. These outliers were
179 excluded from analyses, although results and conclusions were qualitatively consistent when they
180 were included.

181 *Predictor variables for the meta-analysis*

182 Estimates came from 11 species, and based on the distribution of estimates, we defined five taxon
183 categories (Figure 1A): Daphnia (*Daphnia pulex* only); Drosophila (*Drosophila melanogaster* [*n* =
184 68] and *D. serrata* [*n* = 5]); Plant (*Arabidopsis thaliana* [*n* = 12], *Amsinckia douglasiana* [*n*=2] and

185 *Amsinckia gloriosa* [n = 1]) and; Nematode (*Caenorhabditis elegans* [n=62], *C. brenneri* [n=4], *C.*
186 *briggsae* [n=8], *C. remanei* [n=5] and *Oscheius myriophila* [n=5]). We differ from a previous study
187 (HOULE *et al.* 1996; 1998) in considering size of juveniles as morphological (not growth) traits.
188 Reflecting more recent publications, we defined a physiology category (33% of estimates; Figure
189 1B), which included locomotive, enzymatic and metabolic activity traits (Table S2), which may
190 differ from life-history or morphological traits in mutational target size or environmental sensitivity.
191 We assigned the relatively well-represented life history traits (52% of estimates; Figure 1B) into
192 more narrowly defined sub-categories: total fitness, survival, productivity, and a miscellaneous
193 category (capturing traits such as development time, phenology, longevity and mating success, which
194 were individually less well represented) (Figure 1B; Table S2).

195 *Meta-analyses of mutational variance estimates*

196 We implemented a mixed model analyses via PROC MIXED in SAS v.9.4 (SAS Institute Inc., Cary,
197 N.C.), using restricted-maximum likelihood (REML) and applying the Satterthwaite approximation
198 to correct the denominator degrees of freedom, to fit the model:

199 $$y_{ijkl} = \mu + Taxon_i + Trait_j + Study_k + \varepsilon_{ijkl} \quad (1)$$

200 where y is the vector of published estimates (either h^2_{M} or CV_{M}), and μ was the grand mean; the
201 categorical predicators of taxon and trait (defined above) were fit as fixed effects. Estimates were
202 weighted by the inverse of the standard error (SE) of h^2_{M} or CV_{M}, obtained as detailed above. Study
203 was fit as a random effect, accounting for non-independence among estimates within a paper (1–29
204 estimates per study; median = 3). Studies reporting multiple estimates varied widely in whether these
205 were estimates from different trait types, species (strains), sexes, or time points. Likely reflecting
206 this, most variation not accounted for by the fixed effects was observed at the residual, not study,
207 level (99% for h^2_{M}; 85% for CV_{M}). Similarly, while some studies shared the same MA lines (Table

208 S3) fitting a further random effect to account for this non-independence did not explain any variation,
209 a likely consequence of both the unbalanced design (only some studies share lines), and the relative
210 variation of estimates. We investigated different options for fitting heterogenous residuals (e.g.,
211 allowing separate estimation of residuals for studies grouped depending on the number of estimates
212 reported), but interpretation of the fixed effects (taxon and trait) were consistent across all
213 investigated models, and we report results only from model (1) above.

214 **Variation in estimates of among-line variance within the same taxon and trait type: An**
215 **experiment in *Drosophila serrata***

216 To what extent do differences in magnitude among estimates of V_M reflect differences in mutation
217 number and/or effect sizes (correlated with the above-investigated proxies of taxon and trait), versus
218 factors such as mutations segregating within MA lines or environmental dependency of mutational
219 effects? We conducted a further experiment to address this question. *Drosophila serrata* is a member
220 of the *montium* species group, endemic to Australia and Papua New Guinea, which has been
221 extensively used in quantitative genetic research, including study of mutational variance (e.g.,
222 McGuigan and Blows 2013; Hine *et al.* 2018; Dugand *et al.* 2021). A panel of 200 MA lines were
223 founded from one of the *D. serrata* Reference Genome Panel (DsRGP) lines described in Reddiex *et*
224 *al.* (2018). These MA lines were each maintained by brother-sister inbreeding for 20 generations,
225 following protocols established by McGuigan *et al.* (2011) to minimise selection. Genome-wide
226 heterozygosity was very low (0.3%) in the DsRGP line that founded the MA lines, and among-line
227 variance for wing traits (defined below) was not statistically supported in the first generation of the
228 MA (S. Chenoweth, pers. comm.).

229 As detailed below, we applied an experimental design to this MA panel that allowed us to generate
230 repeated estimates of the magnitude of among-line variance over six sequential generations, and
231 characterise the relative contribution to differences among these sequential estimates of i) mutations

232 segregating within the MA lines or ii) unintentional variation in environment. We randomly chose 42
233 of the MA lines for this investigation based on the median number of MA lines in the reviewed
234 published studies (see Results). The number of MA lines is the relevant degrees of freedom for the
235 among-line variance, and this value (42) allows us to consider the other two effects against a relevant
236 level of sampling error. Quantitative genetic parameters are associated with large sampling errors
237 (KLEIN *et al.* 1973; KLEIN 1974; LYNCH AND WALSH 1998), and the relatively low signal (i.e., few
238 genetic differences) among MA lines will make mutational variance particularly vulnerable to
239 "noisy" estimation, and as such, it is important to document the potential for statistical sampling error
240 to contribute to the observed variation among published estimates.

241 There were three key aspects of the experimental design that allowed us to test whether segregating
242 variation or unintended environmental variation could explain differences among repeated estimates
243 of among-line variance. First, we increased the population size within each MA line to a minimum of
244 12 males and 12 females (Figure 2A). Empirical evidence suggests that population sizes as low as ten
245 may be sufficient to prevent fixation of mutations (ESTES *et al.* 2004; KATJU *et al.* 2015; LUIJCKX *et*
246 *al.* 2018). Therefore, we expect no ongoing fixation of mutations among lines during this experiment,
247 and for the repeated estimates of among-line variance to be true replicate sampling of the same
248 mutations (but also test this assumption, as detailed below). We note that these changes in census
249 population size complicate calculation of a per-generation rate of increase in phenotypic variance
250 (LYNCH AND HILL 1986; LYNCH AND WALSH 1998); here, we instead focus on the among-line
251 variance, V_L, and do not interpret a per-generation rate of change.

252 The second key aspect of the experimental design was to manipulate the mutation-selection-
253 drift dynamics within an MA line; this was achieved by imposing two, substantially different,
254 population sizes on sublines of each of the 42 MA lines (N = 24 versus 288 flies, referred to hereafter
255 as small, S, and large, L, population size treatments: Figure 2A). Segregating variants within MA
256 lines (i.e., mutations that have not yet been fixed or lost) could cause transient inflation of among

257 and/or within line variance (V_E), impacting on both the estimation and scaling of V_L, and this
258 manipulation allowed us to determine the magnitude of this effect. The treatments contrast
259 deterministic evolution of mutations with relatively strong (s > ~0.038: N_e ~ 13), versus weak (s >
260 0.003: N_e ~ 158) fitness effects, based on $s = 1/2N_e$ (WRIGHT 1931; KIMURA 1983) and genomic
261 estimates of N_e in MA lines of *D. melanogaster* maintained similarly to our small population size
262 treatment (10 males and 10 females: HUANG *et al.* 2016). The S and L treatments therefore had
263 different opportunities for new mutations to increase in frequency within a line, and thus for the
264 magnitude of within-line variance.

265 The final key aspect of the experimental design was the repeated measures themselves,
266 allowing us to observe the effect of environmental variation on among-line variance. If the
267 phenotypic effects of a mutation are context-dependent (i.e., exhibit genotype by environment [GxE]
268 variance), then unintended differences in assay conditions could contribute heterogeneity among
269 estimates when phenotypic data is collected at different timepoints (or in different laboratories). We
270 randomly sampled the average environmental conditions present within our laboratory by repeatedly
271 sampling the lines (genotypes) over six consecutive generations. Thus, our experiment consisted of
272 applying two population size treatments (S, L) to each of 42 lines (derived from a classical MA
273 experiment, with low among-line variation), where these 84 lines were maintained under the same
274 conditions (12 flies per sex per vial founding each generation, with S and L differing in the number
275 of vials) for six generations (Figure 2A). As detailed below, we consider 11 wing shape and size
276 traits. This allows us to understand the general influences of segregating variation, environment and
277 sampling error for a set of related morphological traits. After accounting for those three factors that
278 are the main focus of the investigation, we also determine whether the magnitude of V_L varies among
279 these traits, allowing insight into potential magnitude of differences in mutational variance among
280 traits within the same category (morphology).

281 *Data collection*

282 Each generation, 12 males (six from each of two rearing vials) from each of the 84 sublines were
283 randomly sampled for wing phenotypes (Figure 2). Wings were mounted on microscope slides and
284 photographed using a Leica MZ6 microscope camera and the software LAS EZ v2.0.0 (Leica
285 Microsystems Ltd, Switzerland). A total of 5,135 wings were landmarked for nine positions, defined
286 by wing vein and margin intersections (Figure 2B), using the software tpsDIG2 (ROHLF 2013). The
287 number of wings were evenly distributed across treatments (2,583 in S and 2,552 in L) and
288 generations (~425 per generation, per treatment). Landmarks were aligned using a General Procrustes
289 fit in tpsRelw (ROHLF 2007). Centroid size (CS), the square root of the sum of squared deviations of
290 the coordinates from the centroid (ROHLF 1999), was recorded as a metric of wing size. The aligned
291 X-Y coordinates for each landmark were then used to calculate ten inter-landmark distances (ILDs)
292 (Figure 2B). ILDs scores were re-scaled prior to analysis (multiplied by 100) to aid model
293 convergence. Outliers >3.0 SD from the mean were removed for each of the 11 traits (10 ILD and
294 size) (329 measures across the 56,485 total measures).

295 *Analyses of variation in among-line variance estimates*

296 Our experimental design allows us to repeatedly estimate variance among mutation accumulation
297 lines under conditions where we expect the number of mutations fixed among the lines, and their
298 phenotypic effects, to be constant, and thus to investigate other potential causes of variability in
299 estimates. We first treat the data from each generation and population size treatment as independent
300 experiments of similar size (number of lines and individuals sampled per line) to typical MA
301 experiments. To estimate among line variance from these 12 "experiments" for each of the 11 traits
302 we fit the following model using REML in PROC MIXED in SAS v9.4 (SAS Institute Inc., Cary,
303 NC.):

$$y_{klm} = \mu + Line_k + Vial_{l(k)} + \varepsilon_{klm} \quad (2)$$

where y_{klm} was the trait value for the *m*th wing (individual), from the *l*th vial, within the *k*th line, μ was the trait mean; Line and replicate rearing Vial (nested within line) were fit as random effects, along with the among-individual variation (residual error, ε). We used REML-MVN sampling (MEYER AND HOULE 2013; HOULE AND MEYER 2015; SZTEPANACZ AND BLOWS 2017) to estimate confidence intervals, sampling 10,000 times from $N \sim (\hat{\theta}, \mathbf{V})$ using the rnorm in R [v. 3.6.1], where $\hat{\theta}$ was the vector of REML random effect parameter estimates, and **V** was their inverse Fisher information matrix, $\boldsymbol{I}(\hat{\theta})^{-1}$. We similarly estimated the confidence intervals for the trait mean, sampling based on least-squares mean and standard error estimates output from model (2). The samples of random effect variances were not constrained to the parameter space (i.e., could be negative), allowing inference of statistical support when the lower 5% CI did not encompass zero (a one-tailed test); this approach is equivalent to a Log Likelihood Ratio test (LRT) (DUGAND *et al.* 2021). Here we are interested in general patterns of variability among these 12 "experiments", and thus do not correct for multiple testing.

As detailed in the Results, substantial heterogeneity in magnitude was observed among the 12 replicate estimates of V_L per trait. We considered the potential contribution to this heterogeneity from unintentional heterogeneity in the culture conditions among sampling time points (generations) or between replicate measures of the same MA line within a generation (the S and L treatments). First, to determine if simple effects of variability in culture conditions on trait scale could account for the variability of among-line (mutational) variation, we placed estimates on a heritability (V_L / V_E where V_E was the sum of among and within vial variances) or coefficient of variance ($100 \times \sqrt{|V_L|}/\bar{X}$) scales, and calculated confidences intervals by applying these equations to each of the 10,000 samples described above (and applying the sign correction for coefficients of variance as detailed in

327 the meta-analysis methods). We further explored the relationship between V_L and the scaling
328 parameters by regressing the 12 estimates of V_L on the corresponding estimates of V_E or trait mean.

329 Second, we determined whether mutational effects changed in response to the unintended
330 changes in culture conditions, with such genotype by environment causing differences among
331 sequential estimates of V_L. Therefore, we extended this investigation, following GARCÍA-DORADO *et*
332 *al.* (2000) in treating different generations as different environments to formally test the null
333 hypothesis that there was no GxE variance within the S or L treatments, using PROC MIXED and
334 REML to fit:

335 $$y_{jklm} = \mu + G_j + Line_k + (G \times Line)_{jk} + Vial_{l(jk)} + \varepsilon_{jklm} \quad (3)$$

336 where the fixed effect of generation (G) accounted for differences in trait mean among generations
337 and the random effect of G(eneration) × Line estimated the variation in genetic effects among
338 generations (where generations represent different local environments). For the component of V_E
339 (i.e., Vial and residual), generation-specific effects were modelled (using the GROUP statement) to
340 account for among-generation heterogeneity in the magnitude of V_E. This model was applied to each
341 trait within each population size treatment, and statistical support for G × Line (and for Line) was
342 determined using log-likelihood ratio tests (LRT, 0.5 d.f.: SELF AND LIANG 1987; LITTELL *et al.*
343 2006) to compare model (3) to reduced models that did not fit G x Line (or did not fit Line). We
344 applied the Benjamini-Hochberg method (BENJAMINI AND HOCHBERG 1995) to correct for multiple
345 hypothesis testing (within each random effect), using a conservative 5% false discovery rate (FDR).
346 Sampling based on the REML variance estimate and the Fisher information matrix, as detailed above,
347 was used to estimate confidence intervals for plotting.

348 While non-zero generation by line variance could reveal the presence of environment-specific
349 mutational effects, it could alternatively be explained by changes in the frequency at which mutations

350 were segregating within or among lines. In contrast to environmental heterogeneity, we expect these
351 evolutionary processes to systematically differ between the two population size treatments due to the
352 different efficacy of selection in the S versus L sublines, and the independent sampling of mutations
353 in the sublines after they were established. Differences between S and L are predicted to increase
354 with increasing time since divergence. For each of the 11 traits, we analysed all data (from both L
355 and S) collected within a single generation, using PROC MIXED and REML to fit:

356 $$y_{jklm} = \mu + Treat_i + Line_k + Vial_{l(ik)} + \varepsilon_{iklm} \quad (4)$$

357 where Treatment was fit as a fixed effect to account for differences in trait mean between L and S
358 panels of sublines within that generation. Vial and residual are as described for model (2). At the
359 among-line level, we took advantage of the paired subline design to model the between treatment
360 variance-covariance matrix. We employed LRT to test two hypotheses. First, we determined whether,
361 for these analyses within a generation, there was support for differences between treatments in the
362 magnitude of V_E. Mutations that are segregating (*i.e.,* occur at frequencies other than 0 or 1) within
363 an MA line will contribute to variation both among-vials and the residual. We compared a model in
364 which one (common to both Treatments) among Vial variance and one residual variance were
365 estimated to a model in which Treatment-specific variances were modelled at both levels (fit using
366 the GROUP statement). Second, we tested whether the two copies of the MA lines had diverged from
367 one another by testing the hypothesis that the correlation between the paired sublines was <1.00
368 (implemented using the PARMS statement). To correct for multiple hypothesis testing (within each
369 hypothesis), we employed a FDR correction as described above.

370 Finally, as there was little support for varying mutational effects (no GxE) or number (no
371 divergence between S and L) contributing to the apparent heterogeneity among repeated estimates
372 per trait (detailed in Results), we use our data to revisit the question of whether traits inherently differ

373 from one another in the magnitude of mutational variance. We obtained a single estimate of V_L per
374 trait by using PROC MIXED in SAS to fit:

375 $$y_{ijklm} = \mu + T_i + G_j + TG_{ij} + Line_k + (G \times Line)_{jk} + Vial_{l(ijk)} + \varepsilon_{ijklm} \quad (5)$$

376 where all effects are as described above, including the fixed effects of population size Treatment (T),
377 generation (G) and their interaction (TG), as well as the random effects of Line, Generation by
378 among-Line, Vial and residual. We obtained REML-MVN confidence intervals for each parameter,
379 as described above. To test whether observed differences in V_L among traits were due to differences
380 among them in scale, we took the among-line (V_L) estimates from model (5) and regressed them on
381 the corresponding estimates of environmental variance or on the squared trait mean. These
382 regressions were applied to the REML parameter estimates, and to each of 10,000 samples of these
383 parameters to determine statistical significance (95% CI of slope did not include zero).

384 **RESULTS**

385 **Meta-analysis of published mutational variance estimates**

386 Our final meta-analysis data set consisted of 154 estimates of h^2_{M} and 148 estimates of CV_{M}. These
387 estimates (excluding extreme outliers; Table S1.B) of h^2_{M} ranged from 2.50 x 10^{-5} to 1.02 x 10^{-2},
388 while CV_{M} ranged from 0.13 to 7.32. We predicted that differences in genome size and/or genomic
389 mutation rate would cause differences in the magnitude of mutational variance among taxa. However,
390 there was no statistical support for a difference in mutational variance among the taxon categories
391 (h^2_{M}: $F_{3,24.2} = 2.28$, $P = 0.1044$; CV_{M}: $F_{3,17.5} = 1.24$, $P = 0.3261$), although h^2_{M} estimates from Plants
392 were markedly lower than estimates from Daphnia and Drosophila (Figure 3A). We predicted that
393 differences among traits in the number of contributing loci would cause differences in the magnitude
394 of mutational variance but that due to differences among trait categories in environmental sensitivity,
395 the rank of trait types would differ between h^2_{M} and CV_M. These predictions for trait type were

396 partially met as the pattern did differ between the two standardisations, but trait categories only
397 differed in the magnitude of CV_M ($F_{5,37.1} = 3.86$, $P = 0.0064$), not h^2_M ($F_{5,85.9} = 0.40$, $P = 0.8497$)
398 (Figure 4). Overall, these factors (taxon and trait category) accounted for 1.64% of the variation in
399 estimates of h^2_M and 9.88% of variation among CV_M estimates.

400 Although not statistically supported, it is notable that the among-trait trend did not follow
401 predictions for h^2_M: fitness traits had the largest average h^2_M, not the lowest as expected (Figure 4A).
402 Following HOULE *et al.* (1996), we also analysed V_E (fit model (1) to $CV_E = \sqrt{V_E}/\bar{X}$). There was no
403 statistical support for a difference among traits in the magnitude of CV_E ($F_{5,25.2} = 1.32$, $P = 0.2887$);
404 morphology (average $CV_E = 7.4$) and physiology (71.6) differed the most, with life history traits
405 having intermediate values (e.g., fitness = 42.6) (Figure S2A). For CV_M, the statistically supported
406 differences did follow the predicted pattern, with morphological traits having the smallest CV_M and
407 fitness the largest (Figure 4B). Survival notably had lower CV_M than productivity and fitness (Figure
408 4B), although surviving to reproduce was a component of fitness. Physiological traits had similar
409 magnitude of CV_M to morphological traits, and lower than any life history trait category (Figure 4B).

410 While the lack of observed difference in scaled estimates of V_M among taxa may reflect a true
411 commonality among species in this important evolutionary parameter, aspects of the MA design also
412 differed markedly among taxa. Estimates from Plants were derived from MA experiments that were
413 of short duration (maximum 25 generations) relative to other taxa (median 44, 75 and 214 for
414 Drosophila, Daphnia and Nematode, respectively) (Figure S2B). As mutations arise independently in
415 each MA line, the number of MA lines maintained may also influence the number of mutations that
416 similar duration MA could sample, and the potential for sampling of rarer mutational effects (i.e.,
417 from the tails of the distribution of effect sizes) to influence estimates; while the median number of
418 MA lines was similar in Nematode (43) Plant (50) and Drosophila (52), it was substantially lower in
419 Daphnia (8) (Figure S2C). While mutational variance was estimated for multiple types of traits in

420 every taxon, most data from Plants was for fitness, while most data from Daphnia was for
421 morphological traits, and Drosophila and Nematode were the only two taxon categories that
422 contributed estimates for physiological traits (Figure S2D).

423 **Variation in estimates of among-line variance within the same taxon and trait type**

424 Within the *D. serrata* experiment, we first determined the heterogeneity in among-line variance (V_L)
425 under the assumption that mutation number and effects were constant, treating the 12 V_L estimates
426 per trait as independent mutation accumulation experiments. There was substantial variation among
427 the 12 estimates per trait, with some differences of over an order of magnitude (Figure 5; Table S4).
428 Notably, there was a fourfold difference between the smallest and largest estimates of V_L for size
429 (CS), comparable to the fourfold differences among reported estimates of h^2_{M} for body size in
430 *Caenorhabditis elegans* (AZEVEDO *et al.* 2002; ESTES *et al.* 2005; OSTROW *et al.* 2007). Predictably
431 given this heterogeneity in magnitude of effect (i.e., in V_L), there was also inconsistent statistical
432 support for the presence of V_L for most traits, despite consistent sample sizes in each of the 12
433 "experiments" (Figure 5; Table S4). Thus, we might draw very different conclusions about the
434 magnitude of V_L for a trait, depending on which "experiment" we had conducted (Figure 5).

435 Due to the changes in N_e within this experiment, we do not place these V_L estimates on a per-
436 generation scale (i.e., do not calculate V_M). However, there is no trend for V_L to increase through time
437 (i.e., no signal of ongoing divergence through fixation of mutations), or to diverge between the
438 different population size (N_e) treatments (Figure 5) (addressed further below). Therefore, calculating
439 V_M is not expected to eliminate the heterogeneity in estimates.

440 Reporting mutational variance estimates as h^2_{M}or CV_M facilitates comparison among estimates
441 by accounting for inherent differences in scale. Although here the 12 estimates come from the same
442 trait, scale differences may still arise through typical effects of any unintended variation in culture
443 conditions (occurring among generations or between the replicate S and L sublines) on non-genetic

trait variance (V_E) or mean. Both V_E and the trait mean varied substantially among the 12 repeated estimates for all traits (Figures S3, S4; Table S4). However, this variation in V_E and trait mean was independent of the observed variation in V_L; regressing the 12 estimates of V_L on their corresponding estimate of V_E or trait mean supported only one slope (ILD3.7, V_L on mean) as statistically different from zero (although this did not remain significant following FDR correction) (Figure 6; Table S5). Consistent with this pervasive independence of V_L from the scaling factors for these repeated measures of the same trait, when the 12 estimates were placed on either a heritability (Figure S5; Table S4) or coefficient of variance scale (Figure S6; Table S4), the variation among them was of a similar magnitude to that observed for V_L itself (i.e., the variation plotted in Figure 5). Plotting the scaled estimates (h^2 or CV) against their respective numerator (V_L or $\sqrt{V_L}$) and denominator (V_E or mean) illustrates the predominant contribution from variation in V_L to variation in the scaled estimates (Figure S7). Overall, the 12 estimates of V_L are more variable than the corresponding estimates of V_E or trait mean, with variability of the scaled estimates (h^2 or CV) more similar to V_L than to their respective scaling factor (Figure S8).

To compare the variability among published estimates of h^2_{M} and CV_M in similar morphological traits (excluding bristle traits: 19 estimates, eight each from Daphnia and Nematode, three Drosophila and one Plant) to the variability among the V_L estimates for *D. serrata* wing trait traits, we calculated the coefficient of variance (cv = standard deviation / mean) for each set of estimates. The cv of all 19 published h^2_{M} (0.80) and CV_M (0.82) estimates was above the median cv of the 11 *D. serrata* traits on the observed V_L scale (0.55), V_E-scale (h^2: 0.54) or mean-scale (CV: 0.39), but nonetheless within the same range: cv of V_L (V_E-scale; mean-scaled) ranged from 0.30 (0.37; 0.15) up to 0.92 (0.85; 0.62) across the 11 traits (Figure S8). Within taxa, cv of published h^2_{M} (CV_M) estimates ranged from 0.29 (0.08) for the three Drosophila estimates up to 0.97 (0.38) in Daphnia, with a median of the three within-taxon (Drosophila, Daphnia and Nematode) cv's of 0.64 (0.24). Thus, overall, the heterogeneity

468 among repeated V_L estimates in *D. serrata* is of a similar magnitude to the variation among published
469 estimates of the same trait type.

470 Having established that variation in the magnitude of V_L is not a simple consequence of
471 varying scale (V_E or mean), we investigated other putative causes. In addition to the general effects
472 on scale, unintended differences in culture conditions among generations could also affect V_L if
473 mutations had context-dependent effects on the trait (i.e., GxE), as characterised by generation by
474 among-line variance. GxE was statistically supported in only four cases, with only two remaining
475 significant at a 5% FDR (LRT for CS, in L: χ_1^2 = 13.9, P = 0.0001; LRT for ILD2.5 in L: χ_1^2 = 11.1,
476 P = 0.0005) (Figure 7A). Thus, for these two traits, the analysis suggests that mutational effects, and
477 the magnitude of among-line variance, may depend on the specific conditions under which the traits
478 were assayed.

479 Ongoing mutation-drift-selection processes could contribute to variation among the 12
480 estimates, where the S and L treatments are expected differ in the potential effects of these processes
481 on both within and among-line variance. Segregating variation within a line will contribute to the
482 estimate of V_E, and we determined whether the S and L treatments differed in the magnitude of V_E,
483 analysing each of the 11 traits within each of the six generations separately. Eight of the 66 estimates
484 of V_E differed significantly between S and L at $P < 0.05$, but only one remained significant at 5%
485 FDR (CS in generation 5; Table S6). There was no statistical support for the S and L sub-lines
486 founded by each of the 42 original MA to have diverged from one another in the mutations they
487 carried, either through initial sampling when lines were founded, or through (near) fixation of
488 mutations arising after establishment of the sub-lines. Specifically, the among-line correlation
489 between S and L sub-lines was not statistically distinguishable from 1.0 for any trait in any
490 generation (Table S6).

491 Finally, we obtained a single estimate of among-line variance for each trait to determine
492 whether the magnitude of V_L was consistent among the 10 wing shape traits, which are expected to
493 share a genetic basis, and developmental pathways (e.g., MEZEY *et al.* 2005; NETO-SILVA *et al.*
494 2009). The among-trait heterogeneity (i.e., non-overlapping confidence intervals: Figure 8A; cv =
495 0.77) was larger than the median variability among the repeated estimates per trait (0.55; see above),
496 and comparable to the variability among published estimates of mutational variance in morphological
497 traits (0.80, detailed above). Variation among shape traits (excluding size) in V_E accounted for ~18%
498 (95% CI: 0.049 - 0.383) of this variation (β = 0.064 [95% CI: 0.050 - 0.106]), but variation in trait
499 mean did not account for any ($\beta = 5.3 \times 10^{-7}$ [$-0.74 \times 10^{-8} - 1.86 \times 10^{-6}$]; R^2 = 0.003 [95% CI: <0.001
500 - 0.027]) (Figure 8B). Establishing whether these differences are informative of the inherent genetic
501 architecture, or are a manifestation of the stochastic nature of mutation, requires repeating the
502 estimation using either the same or a different genetic background to determine if consistent
503 differences among traits persist. When wing size was also considered, overall scale influenced V_L,
504 with much of the variation among estimates accounted from by the scaling factors (V_E: R^2 = 0.995
505 [0.978 - 0.998]; Mean^2: R^2 = 0.921 [0.903 - 0.925]), suggesting that there was little difference in the
506 magnitude of underlying mutational variance between wing size and the shape traits (Figure 8B).

507 **DISCUSSION**

508 Although numerous estimates of mutational variance have been published, it remains unclear what
509 contributes to the ~two orders of magnitude difference among these estimates. Our meta-analytic
510 investigation provided some support for a difference among trait types in the magnitude of
511 mutational variance, but also revealed substantial confoundment between potential causal factors.
512 Analyses of data from a manipulative experiment in *D. serrata* suggests that, for the morphological
513 traits under consideration, factors such as unintended heterogeneity in environmental conditions or
514 transient segregation of mutations within MA lines may contribute little to the variation among

515 estimates. Given this experimental design, and the evidence that mutation number and effect did not
516 typically cause differences among repeated estimates, we conclude that substantial variability among
517 repeated estimates of the among-line variance must reflect sampling error. Below we discuss the
518 specific outcomes and limitations of both approaches, and the implications our analyses have for
519 future work characterising mutational input to quantitative genetic variation.

520 *Effects of taxon and trait type on the magnitude of mutational variance*

521 Given the ~four-fold higher per site mutation rate (KATJU AND BERGTHORSSON 2019), and slightly
522 larger genome of *A. thalania* relative to *C. elegans* we predict (assuming the same mutational effect
523 sizes) ~five times more mutational variance in the Plant than Nematode taxon categories. However,
524 the meta-analysis did not support a difference among taxa in the magnitude of mutational variance,
525 and the observed (strong but non-significant) pattern in h^2_{M} contradicted this rank prediction, with
526 Plants (to which *Arabidopsis* contributed most estimates) having substantially smaller h^2_{M} than other
527 taxa (Figure 3A). Taxon categories differed substantially in the number of generations (Figure S2B),
528 and the number of genomes (MA lines) (Figure S2C) sampled. However, scaling predicted genomic
529 mutation by this opportunity for mutation also fails to predict the trend, with Daphnia MA
530 experiments predicted to sample the fewest mutations but observed to have the largest h^2_{M} (Figure
531 3B). We suggest that further MA experiments, decoupling the confounded effects of MA duration
532 and trait type from taxon, are warranted to determine whether V_{M} does vary among taxa. Advances in
533 accessibility of genome data provides substantial scope for such experiments to explicitly estimate
534 relevant genomic parameters (e.g., frequency spectra for different types of mutations across
535 putatively causal genes) alongside the phenotypic variation generated by those mutations (KATJU
536 AND BERGTHORSSON 2019). Furthermore, given evidence that epigenetic mutations arise more
537 frequently than genetic mutations (e.g., VAN DER GRAAF *et al.* 2015; BELTRAN *et al.* 2020), we

538 suggest that the potential contribution of epimutations to patterns of heterogeneity of V_M should be
539 explicitly assessed in future studies.

540 HOULE *et al.* (1996); (see also LYNCH AND WALSH 1998; LYNCH *et al.* 1999) concluded that
541 life history traits had lower h^2_M and higher CV_M than morphological traits, a pattern that is also
542 observed in standing genetic variation (e.g., HOULE 1992; HANSEN *et al.* 2011; but see HOFFMANN *et*
543 *al.* 2016). However, this conclusion was not supported by our analysis, where the trend was for
544 fitness and productivity to have the highest h^2_M (non-significant) as well as highest CV_M (Figure 4).
545 As expected given this shared pattern, h^2_M and CV_M estimates were positively correlated (Spearman's
546 correlation coefficient: 0.309, N =117, P = 0.0007). Although with reverse rank (life history traits
547 having lowest values), standing genetic variation estimates have also been reported to be positively
548 correlated between the two scales (h^2 and CV) when biologically uninformative CV estimates were
549 excluded (HOFFMANN *et al.* 2016). GARCIA-GONZALEZ *et al.* (2012) highlight the potential for
550 skewed data distributions to inflate (deflate) CV, an issue that may be particularly relevant to
551 estimates of CV_M. While strong bias toward mutations that decrease mean fitness has been reported
552 (HALLIGAN AND KEIGHTLEY 2009), bias in other traits is less well-established. If trait types differ in
553 the magnitude of directional bias of mutational effects, this may also result in differences in skew,
554 and exaggerate differences between trait types on the CV_M scale. Again, resolution of the key
555 question of whether differences among traits in h^2_M and CV_M reflect differences in mutation number
556 and/or effect size may depend on further genomic data.

557 *The contributions of unintended environmental variation, mutation-drift-selection processes, and*
558 *sampling error to variation in the magnitude of mutational variance*

559 We observed substantial variation among repeated estimates of V_L each of 11 wing traits measured in
560 *D. serrata* (Figure 5), resulting in variation among scaled (h^2 or CV) estimates that was of
561 comparable magnitude to the differences observed among published estimates. Although both V_E and

562 trait mean also varied among repeated measures (Figures S3, S4), this heterogeneity was substantially
563 less than that observed for V_L (or h^2 or CV) (Figure S8). Given the evidence that mutational effects
564 can vary among environments (KONDRASHOV AND HOULE 1994; MARTIN AND LENORMAND 2006),
565 we tested the effects on the magnitude of V_L resulting from unintentional and undocumented minor
566 changes in culture conditions (e.g., density, humidity or temperature), such as may occur among
567 phenotype assays conducted at different times or in different laboratories. Variation in mutational
568 effects among phenotypic assays (generations) was supported in only two cases (Figure 7A).
569 GARCÍA-DORADO *et al.* (2000) also found evidence of GxE among consecutive generations for one
570 (sternopleural bristle count) of four traits investigated. Notably, in *D. serrata*, wing size, which might
571 be particularly sensitive to variation in energy availability (or competing energetic demands)
572 (CAVICCHI *et al.* 1985; BITNER-MATHÉ AND KLACZKO 1999), exhibited the strongest GxE (Figure
573 7A). Our results, and those of GARCÍA-DORADO *et al.* (2000) suggest that changes in mutational
574 effects with environment may contribute to heterogeneity among published estimates of some traits,
575 which may reflect differences in trait environmental sensitivity, or potentially in the covariation of
576 environmental sensitivity and mutational effect size (LYNCH *et al.* 1999; GARCÍA-DORADO *et al.*
577 2000).

578 The mutation-drift process itself may also contribute to variability among published estimates
579 due to effects on both the within-line variance (transient inflation leading to increased magnitude of
580 V_E but not V_L) and among-line variance (transient contribution to V_L of additive or dominant
581 mutations that are subsequently lost via random sampling). We introduced an ~order of magnitude
582 difference in census size in paired sets of MA sub-lines to manipulate the mutation-drift-selection
583 processes. However, analyses did not support an effect of population size on either within- or among-
584 line variation, with wing size again an exception, where there was some evidence that relaxed
585 selection allowed the S treatment to accumulate greater within-line variance (Table S6). The effect of
586 segregating variation can be expected to be greater at smaller population sizes (e.g., when $N = 2$,

587 mutations can reach within-line frequency of 75% before being lost by drift) than considered here,
588 and so may play a greater role in explaining variation among estimates from classical MA breeding
589 designs. But, nonetheless, this factor did not account for the substantial heterogeneity that was
590 observed among the 12 estimates per trait within the current study.

591 Rejecting general contributions from environmental variation and transient segregation of
592 mutations as explanations of the heterogeneity among the 12 repeated estimates of V_L for the wing
593 shape traits, we conclude that the observed variability is largely the consequence of sampling error.
594 LYNCH *et al.* (1999) suggested that a substantial part of the order of magnitude range of h^2_{M} reported
595 for *D. melanogaster* may be due to sampling error. Here, we observed the magnitude of
596 heterogeneity among the 12 repeated estimates of V_L to be similar to the heterogeneity among
597 published, scaled estimates of mutational variance in morphological traits, consistent with their
598 prediction. We observed that V_L estimates varied markedly more than the other estimated parameters
599 (Figure S8), as expected given that quantitative genetic parameters are associated with relatively
600 large sampling errors. Notably, V_E was more variable among the 12 estimates than trait mean was,
601 which may lead to greater variability among estimates of h^2_{M} than CV_M (Figure S8). We designed this
602 experiment to mimic an average MA sample size, considered a trait expected to have relatively low
603 experimental noise (residual variation), but expected small effect size relative to many MA (due to
604 the relatively few generations; see e.g., VASSILIEVA *et al.* 2000). While traits and MA panels will
605 vary in their vulnerability to sampling error, we nonetheless suggest that greater consideration must
606 be given to the consequences of this error when designing experiments. The heterogeneity among
607 repeated estimates resulted in the total confidence range for each trait spanning a far greater region
608 than suggested by the error estimated for each repeated estimation of V_L (Figure 5), indicating that
609 within-study estimates of error do not fully capture the uncertainty in estimates.

610 The sequential repeated-measures experimental design provided greater statistical control
611 over the experimental noise, allowing us to consistently detect statistically significant mutational
612 variance in all traits (Figure 8A), including in traits for which very few of the 12 estimates were
613 distinguishable from zero (e.g., ILD2.8; Figure 5). While increasing sample sizes within a generation
614 is likely to have similarly improved estimate precision, this can be logistically prohibitive in some
615 systems. Given these limits, our analysis highlights the potential benefits of short-term repeated
616 measures (sequential generations) to improve estimate precision, and power to detect small effects.
617 Repeated measures of lines at relatively large generation intervals have also been utilised to estimate
618 V_M as the slope of the regression of among-line variance on generation (VASSILIEVA *et al.* 2000;
619 HOULE AND NUZHDIN 2004; MCGUIGAN *et al.* 2011), which may also improve estimation.

620 Understanding the contribution that mutations make to evolutionary and genetic phenomena
621 relies on accurate estimates of the phenotypic variance generated by new mutation. Our meta-
622 analysis of empirical estimates of mutational variance was unsuccessful in clearly resolving causes of
623 variation due to confounding of predictors, and inconsistent patterns. Our manipulative experiment
624 suggested that sampling error may contribute substantially to estimate variability, and demonstrated
625 that repeated measures over few (e.g., sequential) generations provides a simple but effective
626 approach to address this and improve inference. Overall, further empirical studies are needed to fully
627 assess how both general and study specific factors influence V_M estimates, where improved precision
628 and replicability in estimates will consequently advance broader evolutionary questions such as those
629 addressing the maintenance of quantitative genetic variance (BARTON AND TURELLI 1989; JOHNSON
630 AND BARTON 2005; WALSH AND LYNCH 2018).

631

632 **Data Availability**

633 Both analysed datasets are available at doi 10.6084/m9.figshare.14913051
634

635 **Acknowledgments**

636 We thank Stephen F. Chenoweth for providing us with the *Drosophila serrata* MA lines, and Adam
637 Reddiex, Nicholas Appleton, Jack Price and Derek Sun for their help with maintaining the flies and
638 collecting the data. We also thank Jan Engelstädter for suggesting the combinational approach, and
639 Emma Hine for contributions to preparing figures. This manuscript was improved by the input of
640 three anonymous reviewers and the AE.

641 # References

642 Azevedo, R. B., P. D. Keightley, C. Laurén-Määttä, L. L. Vassilieva, M. Lynch *et al.*, 2002 Spontaneous mutational variation
643 for body size in *Caenorhabditis elegans*. Genetics 162**:** 755-765.
644 Baer, C. F., 2008 Quantifying the decanalizing effects of spontaneous mutations in rhabditid nematodes. American
645 Naturalist 172**:** 272-281.
646 Baer, C. F., J. Joyner-Matos, D. Ostrow, V. Grigaltchik, M. P. Salomon *et al.*, 2010 Rapid decline in fitness of mutation
647 accumulation lines of gonochoristic (outcrossing) *Caenorhabditis nematodes*. Evolution 64**:** 3242-3253.
648 Barton, N. H., and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nature Reviews Genetics 3**:** 11-21.
649 Barton, N. H., and M. Turelli, 1989 Evolutionary quantitative genetics: how little do we know? Annual Review of
650 Genetics 23**:** 337-370.
651 Beltran, T., V. Shahrezaei, V. Katju and P. Sarkies, 2020 Epimutations driven by small RNAs arise frequently but most
652 have limited duration in *Caenorhabditis elegans*. Nature Ecology & Evolution 4**:** 1539-1548.
653 Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple
654 testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 57**:** 289-300.
655 Besnard, F., J. Picao-Osorio, C. Dubois and M. A. Felix, 2020 A broad mutational target explains a fast rate of phenotypic
656 evolution. Elife 9**:** e54928.
657 Bitner-Mathé, B. C., and L. B. Klaczko, 1999 Plasticity of *Drosophila melanogaster* wing morphology: effects of sex,
658 temperature and density. Genetica 105**:** 203.
659 Boyle, E. A., Y. I. Li and J. K. Pritchard, 2017 An expanded view of complex traits: From polygenic to omnigenic. Cell 169**:**
660 1177-1186.
661 Braendle, C., C. F. Baer and M.-A. Félix, 2010 Bias and evolution of the mutationally accessible phenotypic space in a
662 developmental system. PLoS Genetics 6**:** e1000877.
663 Cavicchi, S., D. Guerra, G. Giorgi and C. Pezzoli, 1985 Temperature-related divergence in experimental populations of
664 *Drosophila melanogaster*. I. Genetic and developmental basis of wing size and shape variation. Genetics 109**:**
665 665-689.
666 Dugand, R. J., J. D. Aguirre, E. Hine, M. W. Blows and K. McGuigan, 2021 The contribution of mutation and selection to
667 multivariate quantitative genetic variance in an outbred population of *Drosophila serrata*. Proceedings of the
668 National Academy of Sciences, USA 118**:** e2026217118.
669 Estes, S., B. C. Ajie, M. Lynch and P. C. Phillips, 2005 Spontaneous mutational correlations for life-history, morphological
670 and behavioral characters in *Caenorhabditis elegans*. Genetics 170**:** 645-653.
671 Estes, S., P. C. Phillips, D. R. Denver, W. K. Thomas and M. Lynch, 2004 Mutation accumulation in populations of varying
672 size: The distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics 166**:** 1269-
673 1279.
674 Fry, J. D., S. L. Heinsohn and T. F. Mackay, 1996 The contribution of new mutations to genotype-environment interaction
675 for fitness in *Drosophila melanogaster*. Evolution 50**:** 2316-2327.
676 García-Dorado, A., J. Fernández and C. López-Fanjul, 2000 Temporal uniformity of the spontaneous mutational variance
677 of quantitative traits in *Drosophila melanogaster*. Genetics Research 75**:** 47-51.
678 Garcia-Gonzalez, F., L. W. Simmons, J. L. Tomkins, J. S. Kotiaho and J. P. Evans, 2012 Comparing evolvabilities: Common
679 errors surrounding the calculation and use of coefficients of additive genetic variation. Evolution 66**:** 2341-
680 2349.
681 Hall, D. W., R. Mahmoudizad, A. W. Hurd and S. B. J. G. r. Joseph, 2008 Spontaneous mutations in diploid *Saccharomyces*
682 *cerevisiae*: another thousand cell generations. Genetics Research 90**:** 229-241.
683 Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annual
684 Review of Ecology, Evolution and Systematics 40**:** 151-172.
685 Hansen, T. F., C. Pelabon and D. Houle, 2011 Heritability is not Evolvability. Evolutionary Biology 38**:** 258-277.
686 Hine, E., D. E. Runcie, K. McGuigan and M. W. Blows, 2018 Uneven distribution of mutational variance across the
687 transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. Genetics 209**:**
688 1319-1328.
689 Ho, E. K. H., F. Macrae, L. C. Latta, P. McIlroy, D. Ebert *et al.*, 2020 High and highly variable spontaneous mutation rates
690 in Daphnia. Molecular Biology and Evolution 37**:** 3258-3266.
691 Hoffmann, A. A., J. Merila and T. N. Kristensen, 2016 Heritability and evolvability of fitness and nonfitness traits: Lessons
692 from livestock. Evolution 70**:** 1770–1779.
693 Houle, D., 1991 Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters.
694 Evolution 45**:** 630-648.

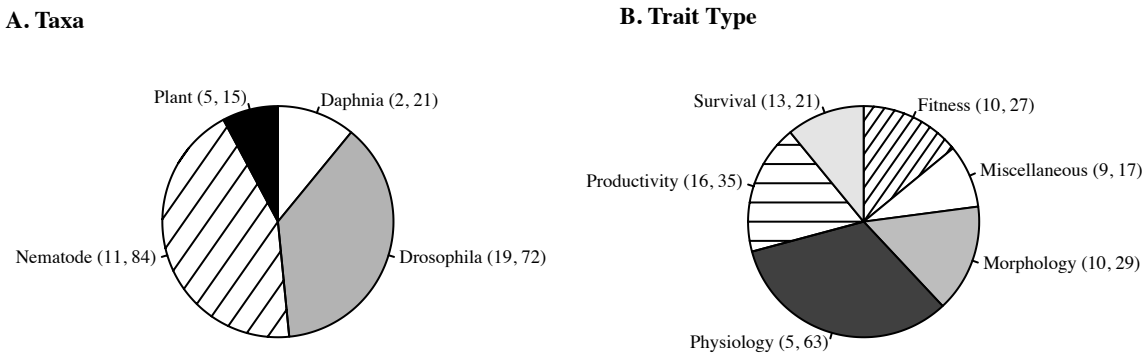
Houle, D., 1992 Comparing evolvability and variability of quantitative traits. Genetics 130: 195-204.
Houle, D., 1998 How should we explain variation in the genetic variance of traits? Genetica 102: 241-253.
Houle, D., G. H. Bolstad, K. van der Linde and T. F. Hansen, 2017 Mutation predicts 40 million years of fly wing evolution. Nature 548: 447-450.
Houle, D., and K. Meyer, 2015 Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. Journal of Evolutionary Biology 28: 1542-1549.
Houle, D., B. Morikawa and M. Lynch, 1996 Comparing mutational variabilities. Genetics 143: 1467-1483.
Houle, D., and S. V. Nuzhdin, 2004 Mutation accumulation and the effect of *copia* insertions in *Drosophila melanogaster*. Genetical Research 83: 7-18.
Huang, W., R. F. Lyman, R. A. Lyman, M. A. Carbone, S. T. Harbison *et al.*, 2016 Spontaneous mutations and the origin and maintenance of quantitative genetic variation. Elife 5: e14625.
Johnson, T., and N. Barton, 2005 Theoretical models of selection and mutation on quantitative traits. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 360: 1411-1425.
Katju, V., and U. Bergthorsson, 2019 Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. Genome Biology and Evolution 11: 136-165.
Katju, V., L. B. Packard, L. Bu, P. D. Keightley and U. Bergthorsson, 2015 Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective population sizes. Evolution 69: 104-116.
Kavanaugh, C. M., and R. G. Shaw, 2005 The contribution of spontaneous mutation to variation in environmental responses of *Arabidopsis thaliana*: Responses to light. Evolution 59: 266-275.
Keightley, P. D., T. F. Mackay and A. Caballero, 1993 Accounting for bias in estimates of the rate of polygenic mutation. Proceedings of the Royal Society of London, Series B: Biological Sciences 253: 291-296.
Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.
Klein, T. W., 1974 Heritability and genetic correlation: Statistical power, population comparisons and sample size. Behavior Genetics 4: 171-189.
Klein, T. W., J. C. De Vries and C. T. Finkbeiner, 1973 Heritability and genetic correlation: Standard errors of estimates and sample size. Behavior Genetics 3: 355-364.
Kondrashov, A. S., and D. Houle, 1994 Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. Proceedings of the Royal Society of London, Series B: Biological Sciences 258: 221-227.
Littell, R., G. Milliken, W. Stroup, R. Wolfinger and O. Schabenberger, 2006 *SAS for Mixed Models*. SAS Institute, Cary, NC.
Luijckx, P., E. K. H. Ho, A. Stanic and A. F. Agrawal, 2018 Mutation accumulation in populations of varying size: Large effect mutations cause most mutational decline in the rotifer *Brachionus calyciflorus* under UV-C radiation. Journal of Evolutionary Biology 31: 924-932.
Lynch, M., 1988 The rate of polygenic mutation. Genetical Research 51: 137-148.
Lynch, M., 2010 Evolution of the mutation rate. Trends in Genetics 26: 345-352.
Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz *et al.*, 1999 Perspective: spontaneous deleterious mutation. Evolution 53: 645-663.
Lynch, M., and W. G. Hill, 1986 Phenotypic evolution by neutral mutation. Evolution 40: 915-935.
Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.
Mackay, T., R. F. Lyman and W. G. Hill, 1995 Polygenic mutation in *Drosophila melanogaster*: non-linear divergence among unselected strains. Genetics 139: 849-859.
Mackay, T. F. C., E. A. Stone and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10: 565-577.
Martin, G., and T. Lenormand, 2006 The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60: 2413-2427.
McGuigan, K., and M. W. Blows, 2013 Joint allelic effects on fitness and metric traits. Evolution 67: 1131-1142.
McGuigan, K., D. Petfield and M. W. Blows, 2011 Reducing mutation load through sexual selection on males. Evolution 65: 2816-2829.
Merilä, J., and B. C. Sheldon, 1999 Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. Heredity 83: 103-109.
Meyer, K., and D. Houle, 2013 Sampling based approximation of confidence intervals for functions of genetic covariance matrices. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 523-526.
Mezey, J. G., D. Houle and S. V. Nuzhdin, 2005 Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics 169: 2101-2113.

751 Mukai, T., 1964 Genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of
752 polygenes controlling viability. Genetics 50**:** 1-19.
753 Ness, R. W., A. D. Morgan, N. Colegrave and P. D. Keightley, 2012 Estimate of the spontaneous mutation rate in
754 *Chlamydomonas reinhardtii*. Genetics 192**:** 1447-1454.
755 Neto-Silva, R. M., B. S. Wells and L. A. Johnston, 2009 Mechanisms of growth and homeostasis in the *Drosophila* wing.
756 Annual Review of Cell and Developmental Biology 25**:** 197-220.
757 Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs *et al.*, 2007 Mutational bias for body size in Rhabditid nematodes.
758 Genetics 176**:** 1653-1661.
759 Reddiex, A. J., S. L. Allen and S. F. Chenoweth, 2018 A genomic reference panel for *Drosophila serrata*. G3 8**:** 1335-1346.
760 Rockman, M. V., 2012 The QTN program and the alleles that matter for evolution: All that's gold does not glitter.
761 Evolution 66**:** 1-17.
762 Rohlf, F., 2007 tpsRelw version 1.45. Department of Ecology and Evolution, State University of New York, Stony Brook.
763 Rohlf, F. J., 1999 Shape statistics: Procrustes superimpositions and tangent spaces. Journal of Classifation 16**:** 197-223.
764 Rohlf, F. J., 2013 tpsDig, digitize landmarks and outlines, version 2.17. Department of Ecology and Evolution, State
765 University of New York at Stony Brook.
766 Schrider, D. R., D. Houle, M. Lynch and M. W. J. G. Hahn, 2013 Rates and genomic consequences of spontaneous
767 mutational events in *Drosophila melanogaster*. 194**:** 937-954.
768 Self, S. G., and K.-Y. Liang, 1987 Asymptotic properties of maximum likelihood estimators and likelihood ratio tests
769 under nonstandard conditions. Journal of the American Statistical Association 82**:** 605-610.
770 Simons, Y. B., K. Bullaughey, R. R. Hudson and G. Sella, 2018 A population genetic interpretation of GWAS findings for
771 human quantitative traits. Plos Biology 16**:** e2002985.
772 Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate
773 evolution. Proceedings of the National Academy of Sciences, USA 109**:** 18488-18492.
774 Sztepanacz, J. L., and M. W. Blows, 2017 Accounting for sampling error in genetic eigenvalues using random matrix
775 theory. Genetics 206**:** 1271-1284.
776 van der Graaf, A., R. Wardenaar, D. A. Neumann, A. Taudt, R. G. Shaw *et al.*, 2015 Rate, spectrum, and evolutionary
777 dynamics of spontaneous epimutations. Proceedings of the National Academy of Sciences, USA 112**:** 6676-
778 6681.
779 Vassilieva, L. L., A. M. Hook and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*.
780 Evolution 54**:** 1234-1246.
781 Walsh, B., and M. Lynch, 2018 *Evolution and Selection of Quantitative Traits*. Oxford University Press.
782 Wayne, M. L., and T. F. Mackay, 1998 Quantitative genetics of ovariole number in *Drosophila melanogaster*. II.
783 Mutational variation and genotype-environment interaction. Genetics 148**:** 201-210.
784 Wright, S., 1931 Evolution in Mendelian populations. Genetics 16**:** 97-159.
785 Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of
786 the heritability for human height. Nature Genetics 42**:** 565-569.

787

788

789 **FIGURES AND TABLES**



790

791 **Figure 1** The distribution of published mutational variance estimates by taxon (A) and trait (B)

792 categories. The number of studies (first value in brackets) and estimates (second value) per category

793 are shown. See Methods (and Table S2) for details of the categories and Table S1 for the studies.

794

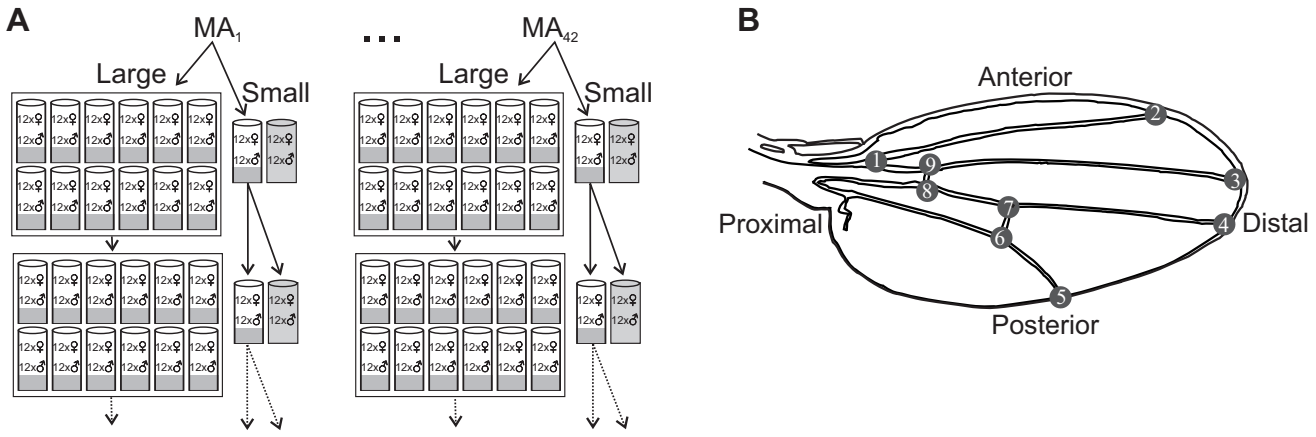


Figure 2 Schematic of design (A) and phenotypes (B) from a manipulative experiment in *Drosophila serrata*. (A) 42 MA lines (evolved through 20 generations of brother-sister mating) each founded two sublines: Small (S; 12 virgin males and 12 virgin females) and Large (L; 144 virgin males and 144 virgin females, distributed evenly among 12 vials). These 84 lines (S and L subline per 42 MA lines) were maintained at these census population sizes for six generations (only two shown here). Each generation, all emergent flies from the 12 vials per L subline were pooled prior to virgin collection. For S sublines, two vials were established each generation; the focal vial contributed offspring to the next generation, while the replicate vial (grey shaded) did not. Each generation, one wing was sampled from each of six randomly chosen males from each of two vials per line (focal and replicate vials for S; randomly chosen two for L). (B) Wing size and shape were characterised from landmarks recorded on an image of each wing: proximal (1) and distal intersections of the radial vein (2); distal intersections of medial (3), cubital vein (4), and distal veins (5) and; the posterior (6, 7) and anterior (8, 9) cross-veins. Inter-landmark distance traits were described by their end-point landmarks (e.g., ILD1.2 was the distance between landmark1 and landmark 2).

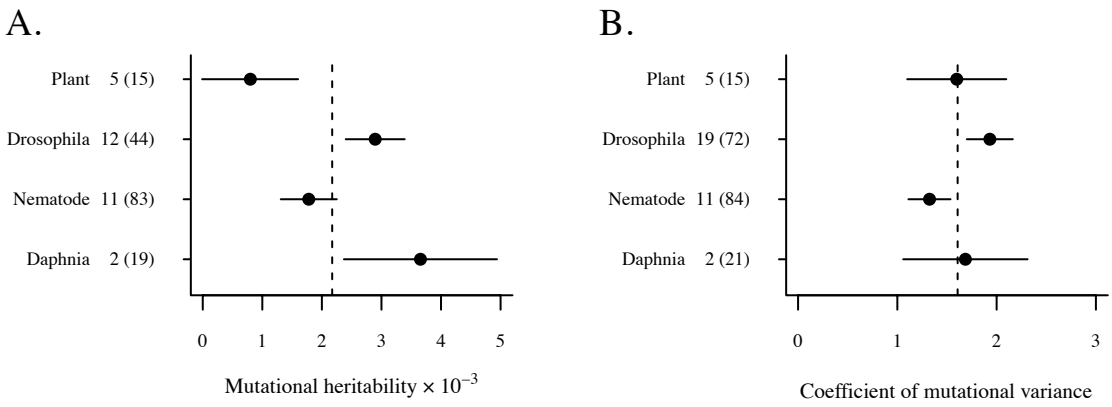


Figure 3 Variation in estimates of (A) mutational heritability and (B) coefficient of mutational variance across taxon categories. Plotted are the least-squares mean estimate (± SE) from the analysis of model (1). The number of studies (and estimates) analysed for each category are shown. The dashed line indicates the global mean value.

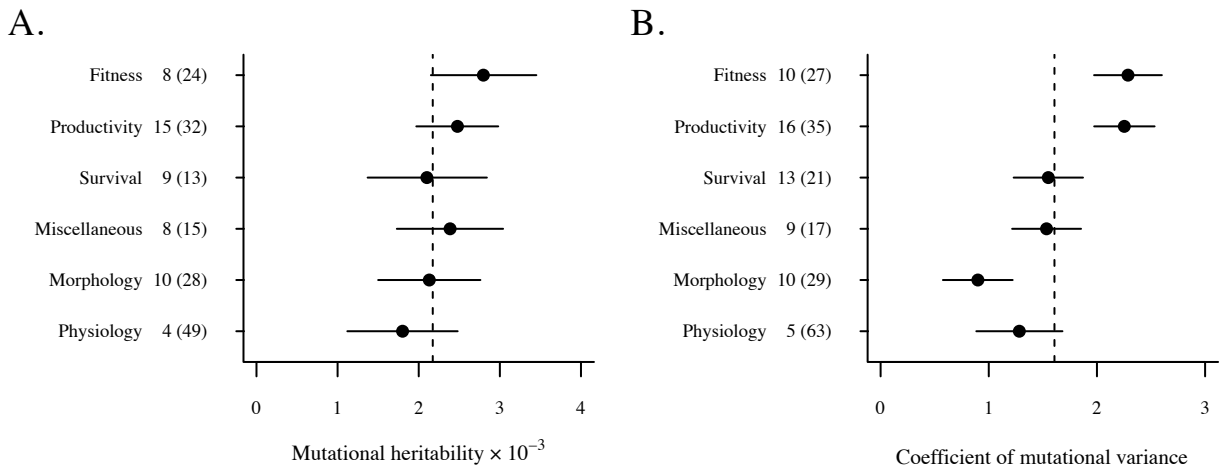
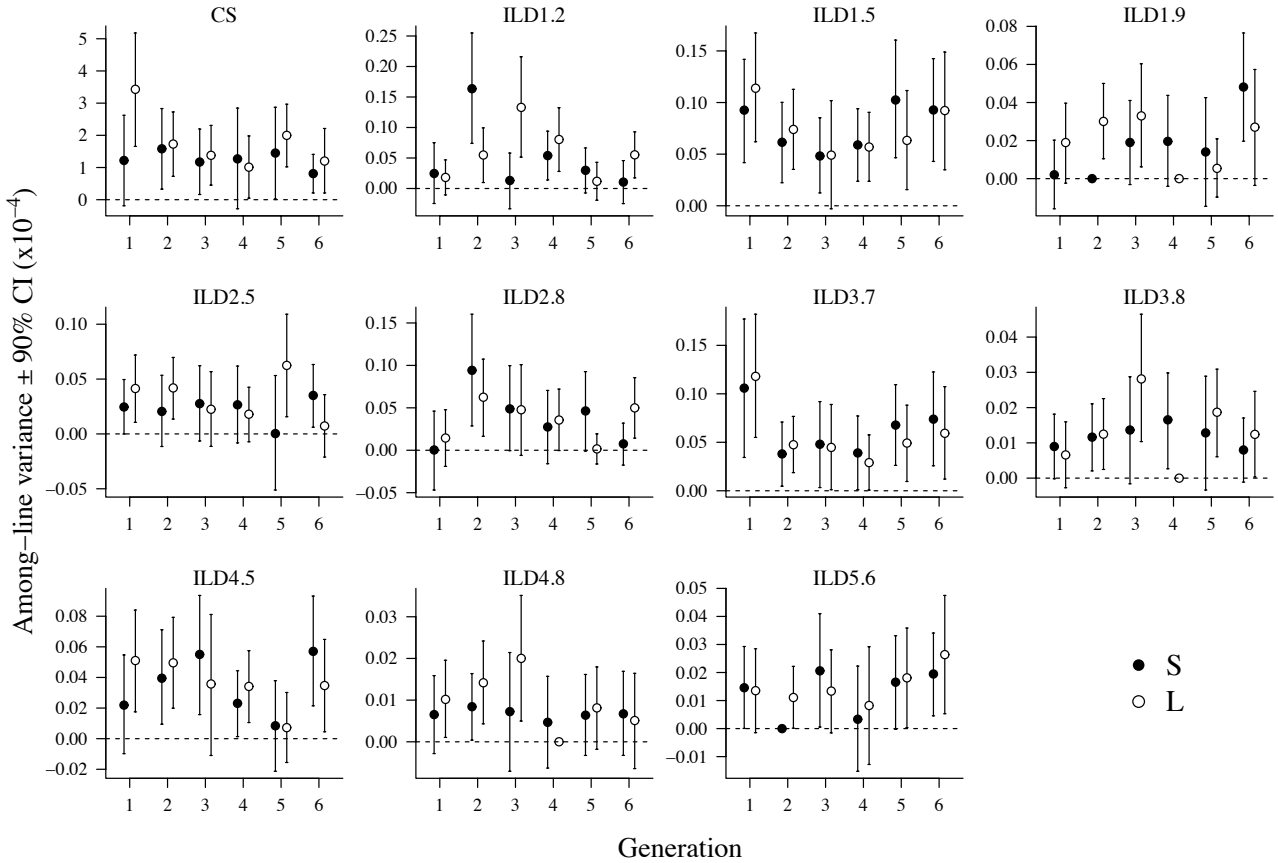


Figure 4 Variation in estimates of (A) mutational heritability and (B) coefficient of mutational variance across trait categories. Plotted are the least-squares mean estimate (± SE) from the analysis of model (1). The number of studies (and estimates) analysed for each category are shown. The dashed line indicates the global mean value

825



826

827 **Figure 5** Among-line variance estimates across six generations in an experiment in *Drosophila*

828 *serrata.* Variances were estimated independently for each trait (panel; see Figure 2B for trait

829 definitions) in each generation (x-axis) for each of the two population size treatments (Small: solid

830 circles; Large: open circles). Plotted are the REML point estimate, and the REML-MVN 90%

831 confidence intervals (CI). The dashed horizontal line indicates zero; estimates for which the lower CI

832 did not overlap zero were interpreted as statistically supported. Where REML estimates of among-

833 line variance were zero, no CI are plotted.

834

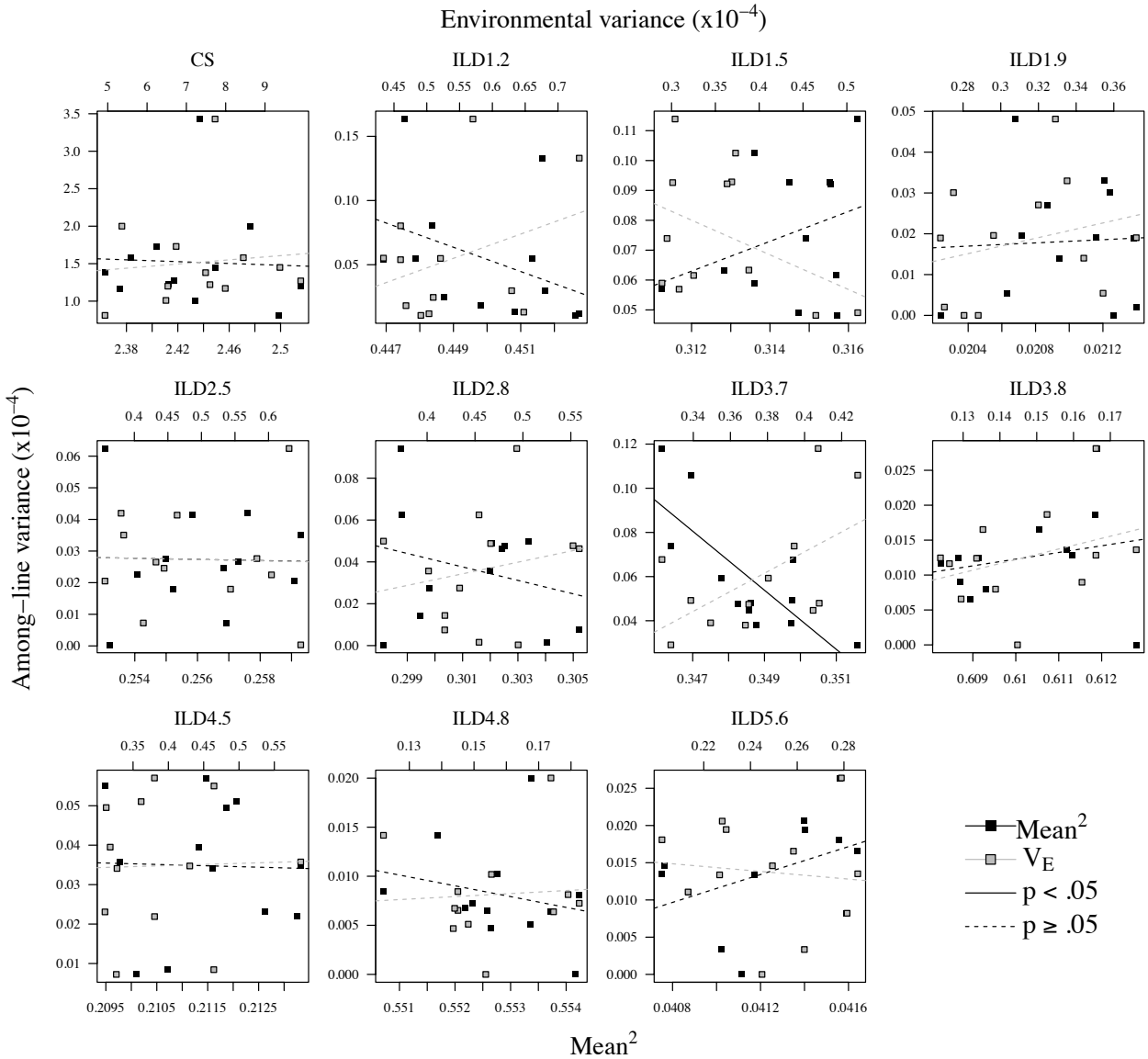
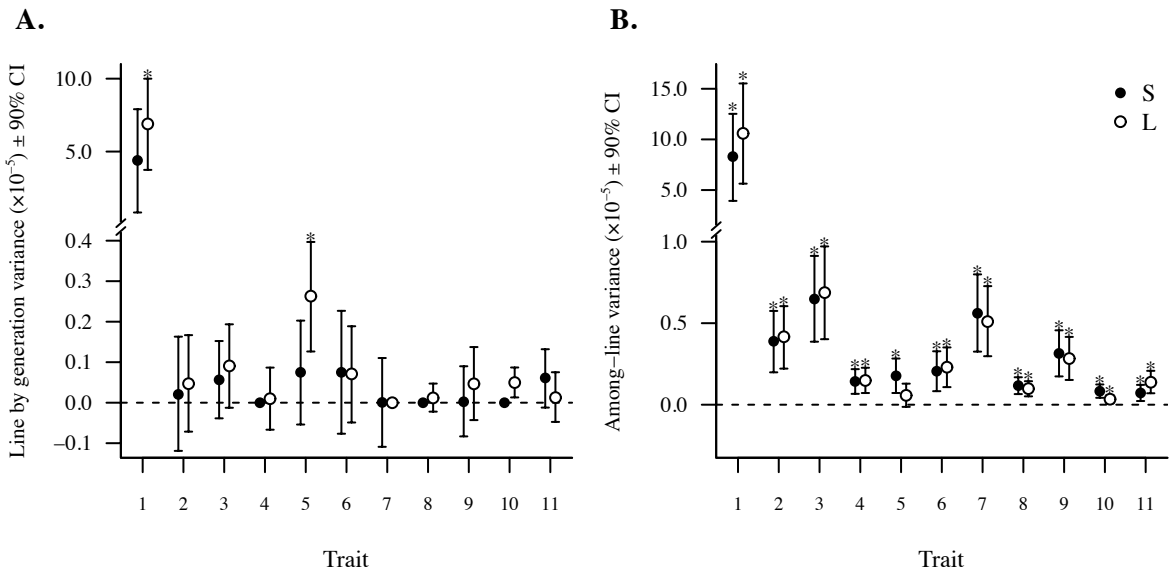


Figure 6 Among-line variance estimates for *Drosophila serrata* wing traits plotted as a function of trait mean or variance. The 12 estimates of among-line variances for each of the eleven wing traits (panels; see Figure 2B for trait definitions) are plotted against the corresponding (i.e., same generation and treatment) squared trait mean (bottom x-axis, black symbols) or environmental variance (summed among and within vial variances; top x-axis, grey symbols). All regression statistics are reported in Table S4; only the effect of mean^2 on V_L of ILD3.7 was significant at $P <$ 0.05, although it does not remain significant after applying a 5% FDR correction.

843

844



845

846 **Figure 7** Estimates of variance from an experiment in *Drosophila serrata*. (A) Among-line by

847 generation (GxE) variance and (B) among-line variance estimated for 11 *D. serrata* wing traits (x-

848 axis), in two different population size treatments (Small: solid circles; Large: open circles). Plotted

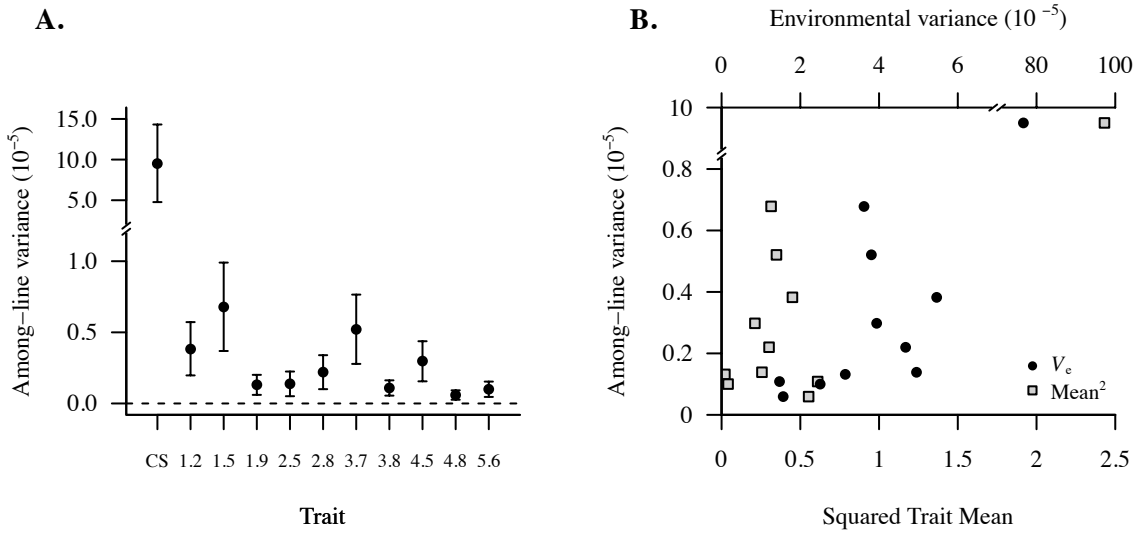
849 are the REML point estimates (from model 3) and the REML-MVN 90% confidence intervals (CI)

850 around this. The dashed horizontal line indicates zero; statistical significance was inferred where the

851 lower 5% CI did not overlap zero. After applying a conservative 5% FDR correction, two estimates

852 in (A) and 21 in (B) remained significant (asterisk above CI).

853



854

855 **Figure 8** Among-line variances for eleven wing traits in *Drosophila serrata*. (A) Among-line

856 variance, V_L REML estimates (and 90% CI) (model 5; see Figure 2B for trait definitions) are plotted.

857 Dashed line indicates zero. (B) REML estimates of among line variance (points in panel A) were

858 plotted against the corresponding estimate of environmental variance (black circles, top x-axis) or

859 mean squared (grey squares, bottom x-axis) of the trait. Regression results are reported in text.

SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table S1 **Information on mutational variance estimates obtained from the literature.**

Supplementary Table S2 **Broad and narrow trait categories.**

Supplementary Table S3 **Mutation accumulation (MA) lines shared across studies.**

Supplementary Table S4 **Estimates of mean and variances within each trait, generation and treatment in *D. serrata***

Supplementary Table S5 **Results of regression of the among-line variance estimates on either the trait mean or on the environmental variance in *D. serrata*.**

Supplementary Table S6 **Statistical support for the effect of population size treatment on (A) among-line and (B) among-line by generation variance estimates.**

Supplementary Figure S1 **Diagram of systematic literature search for the meta-analysis of mutational variance estimates from mutation accumulation (MA) experiments.**

Supplementary Figure S2. **Distribution of the (log-transformed) number of generations (A) and the (log-transformed) number of lines (B) of mutation accumulation experiments in different taxon categories.**

Supplementary Figure S3 **Variation in wing trait mean across six generations of an experiment in *Drosophila serrata.***

Supplementary Figure S4 **Variation in V_E across six generations of an experiment in *Drosophila serrata.***

Supplementary Figure S5 **Contribution of variability in phenotypic variance to estimates of among-line variance in *D. serrata*.**

Supplementary Figure S6 **Contribution of variability in trait mean to variability of among-line variance estimates in *D. serrata*.**

Supplementary Figure S7 **Relative contributions of heterogeneity in the scaling parameter and among-line variance to the variation in the scaled estimates from *D. serrata*.**

Supplementary Figure S8 **Variation in estimates of each genetic variation parameter in *D. serrata.***

Table S1. Information on mutational variance estimates obtained from the literature. We detail estimates that were included in the meta-analysis (A) and those that were disqualified (B, C). For each paper, we detail the source of the estimates of scaled mutational variance (mutational heritability, h_M^2 and, coefficient of mutational variance, CV_M) and associated error estimates†. Studies (or individual estimates within studies) with scaled estimate (and error) required for the meta-analysis, but which were excluded on based on some additional data criteria are listed (and exclusion justified) in (B). In C, we list papers (estimates) that were excluded due to insufficient information to obtain standard error estimates for the scaled mutational parameters. NB: Some papers appear in (A) and (B) or (C). Only estimates identified in (A) were included in the meta-analysis. Estimates are available at: https://figshare.com/articles/dataset/MA_dserrata_wings_csv/14913051.

Reference	Description of where estimates were obtained from
(A) ESTIMATES USED IN THE META-ANALYSIS	
Ávila, V. and A. García-Dorado, 2002 The effects of spontaneous mutation on competitive fitness in *Drosophila melanogaster*. J. Evol. Biol. **15**: 561–566	CV_M from Halligan & Keightley[1] Table 1.
Azevedo, R. L., P. D. Keightley, C. Laurén-Määttä, L. L. Vassilieva, M. Lynch, *et al.*, 2002 Spontaneous mutational variation for body size in *Caenorhabditis elegans*. Genetics **162**: 755–765	Table 1 of paper with scale of mean corrected based on Figure 1.
Baer, C. F., F. Shaw, C. Steding, M. Baumgartner, A. Hawkins, *et al.*, 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. Proc. Natl. Acad. Sci. **102**: 5785–5790	h^2_M, V_M and CV_M estimates calculated from information in paper's Table 1 and Supplementary Tables 3, 4 and 5. Error estimated by sampling†.
Baer, C. F., J. Joyner-Matos, D. Ostrow, V. Grigaltchik, M. P. Salomon, *et al.*, 2010 Rapid decline in fitness of mutation accumulation lines of gonochoristic (outcrossing) *Caenorhabditis nematodes*. Evolution **64**: 3242–3253	Table 2 of paper; V_M reported as $I = {V_M}/{\bar{X}^2}$ and therefore CV_M calculated as: $100*\sqrt{I}$. Confidence intervals from bootstrapping where parameter constrained to be positive.
Caballero, A., E. Cusi, C. Garcia, and A. García-Dorado, 2002 Accumulation of deleterious mutations: Additional *Drosophila melanogaster* estimates and a simulation of the effects of selection. Evolution **56**: 1150–1159	CV_M calculated from in-text information, taking V_B from $\Delta V = V_B / (\text{mean}_c \times F_c)$; where F_c=203 and mean_c = 0.625. Error estimated by sampling†.
Charlesworth, L., H. Borthwick, C. Bartolomé, and P. Pignatelli, 2004 Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. Genetics **167**: 815–826	CV_M (and error) estimates taken from Halligan and Keightley[1] Table 1.
Clark, A. G., L. Wang, and T. Hulleberg, 1995 Spontaneous mutation rate of modifiers of metabolism in *Drosophila*. Genetics **139**: 767–779	Trait means were taken from Table 1. h^2_m and V_M estimates taken from Table 4; V_M was multiplied by 2.5 (scale for whole genome, as stated by authors) before calculating CV_M. Error in CV_M estimated by sampling†.
Davies, S. K., A. Leroi, A. Burt, J. G. Bundy, and C. F. Baer, 2016 The mutational structure of metabolism in *Caenorhabditis elegans*. Evolution **70**: 2239–2246	Table S1 of paper. Note: No estimate of CV_M for citric acid as mean mis-reported in Table S1 and correct value unknown. Confidence intervals from bootstrapping where parameter constrained to be positive.
Estes, S., L. C. Ajie, M. Lynch, and P. C. Phillips, 2005 Spontaneous mutational correlations for life-history, morphological and behavioral characters in *Caenorhabditis elegans*. Genetics **170**: 645–653	We calculated CV_M using V_M (Table 1) and the mean from Table 1 of the companion paper by Ajie *et al.*[3]. h^2_m was from Table 1 of Ajie *et al.*[3]
Etienne,V.,E.C.Andersen,J.M.Ponciano,D.Blanton,A.Cadavid,*etal.*,2015 The red death meets the abdominal bristle: polygenic mutation for susceptibility to a bacterial pathogen in *Caenorhabditis elegans*. Evolution **69**: 508–519	Table 1 of paper. Error in h^2_m estimated by sampling†. Confidence intervals from bootstrapping where parameter constrained to be positive.
Fernández, J. and C. López-Fanjul, 1996 Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. Genetics **143**: 829–837	V_M and h^2_m taken from Table 3 & trait mean from Table1 of paper; CV_M calculated from those estimates (but no error estimate available).

Fry, J. D. and S. L. Heinsohn, 2002 Environment dependence of mutational parameters for viability in *Drosophila melanogaster*. Genetics **161**: 1155–1167	Trait means were from Table 1 and mutational variances were from Table 3; CV_M was calculated from those estimates. Error estimated by sampling†.
Fry, J. D., P. D. Keightley, S. L. Heinsohn, and S. V. Nuzhdin, 1999 New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. **96**: 574–579	Trait means and mutational variances were from Fry and Heinsohn (2002)'s Table 1 and Table 3, respectively, where CV_M was calculated from those estimates.
García-Dorado, A., J. Fernández, and C. López-Fanjul, 2000 Temporal uniformity of the spontaneous mutational variance of quantitative traits in *Drosophila melanogaster*. Genet. Res. **75**: 47–51	Table 2 of paper.
Harada, K., 1995 A quantitative analysis of modifier mutations which occur in mutation accumulation lines in *Drosophila melanogaster*. Heredity **75**: 589–598	CV_M was calculated from information in Table 3 and 4 of paper; V_M was calculated using $4V_a$(reported) and divided by 2× the number of generations. Error estimated by sampling†.
Houle, D., K. A. Hughes, D. K. Hoffmaster, J. Ihara, S. Assimacopoulos, *et al.*, 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. Genetics **138**: 773–785	V_M from Table 4 of paper (x2.5 to scale to genomewide as indicated by the authors); data was standardised to mean=1 prior to estimation, and $CV_M = 100*\sqrt{V_M}$. Error estimated by sampling†.
Houle, D., and S. V. Nuzdhin, 2004 Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. Genetical Research **83**: 7-18	Table 3 of paper (for IVe-39 sublines only, constrained intercept), where $CV_M = 100*\sqrt{\Delta V}$ as described in text of paper on page 9. Error estimated by sampling†.
Joyner-Matos, J., A. Upadhyay, M. P. Salomon, V. Grigaltchik, and C. F. Baer, 2009 Genetic (co)variation for life span in rhabditid nematodes: Role of mutation, selection, and history. J. Gerontol. A Biol. Sci. Med. Sci. **64**: 1134–1145	Table 1 and 2 of paper.
Keightley, P. D. and A. Caballero, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. **94**: 3823–3827	h^2_m estimate (and SE) provided in text on pg 3825. Estimates for productivity determined from among-line variances and trait mean taken from Figure 2 using the R package *metaDigitise*. Error in CV_M estimated by sampling†.
Latimer, C. A. L., K. McGuigan, R. S. Wilson, M. W. Blows, and S. F. Chenoweth, 2014 The contribution of spontaneous mutations to thermal sensitivity curve variation in *Drosophila serrata*. Evolution **68**: 1824–1837	Table 1 of paper
López, M. A. and C. López-Fanjul, 1993 Spontaneous mutation for a quantitative trait in *Drosophila melanogaster*. II. Distribution of mutant effects on the trait and fitness. Genet. Res. **61**: 117–126	We used estimates from B (full-sib mating) lines only, where trait means were from figure 1 (males=30, females = 36.8). Among-line variance and h^2_m for bristle number from Table 6, CV_M calculated from the provided information, but no error estimate available.
Lynch, M., 1985 Spontaneous mutations for life-history characters in an obligate parthenogen. Evolution **39**: 804–818	h^2_m from Table 1 and CV_M calculated from information in Table 1 (but no SE provided for mean and hence unavailable for CV_M). Error estimated by sampling†.
Lynch, M., L. Latta, J. Hicks, and M. Giorgianni, 1998 Mutation, selection, and the maintenance of life-history variation in a natural population. Evolution **52**: 727–733	Estimates of h^2_m, V_M, V_E and trait mean (MA mean calculated for the final generation as: z_0+R_m*30) from Table 1 of paper and CV_M calculated from this information. Error in CV_M estimated by sampling†: z_0 and R_m were sampled to allow the mean to be determined per sample.
Mackay, T. F., R. F. Lyman, M. S. Jackson, C. Terzian and W. G. Hill, 1992 Polygenic mutation in *Drosophila melanogaster*: Estimates from divergence among inbred strains. Evolution **46**: 300-316	Table 3 of paper, assuming $N = 0.7N_e$. Estimates available for per generation (Table 1), but SE not reported.
Martorell, C., M. Toro, and C. Gallego, 1998 Spontaneous mutation for life-history traits in *Drosophila melanogaster*. Genetica **102/103**: 315–324	Table 2 of paper.
McGuigan, K., D. Petfield, and M. W. Blows, 2011 Reducing mutation load through sexual selection on males. Evolution **65**: 2816–2829	h^2_m was taken from text. CV_M was calculated by determining V_M as (mean (V_E) × h^2_M), where V_E and MA means were taken from Tables S1 and S2 of paper for the Control MA population. Error estimated by sampling†
Mukai, T., S. I. Chigusa, J. F. Crow, and L. E. Mettler, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72**: 335–355	Among-line and mean estimates from Tables 1, 2 and 3 of paper for Generation 40 only were used to calculate V_M and CV_M.
Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs, *et al.*, 2007 Mutational bias for body size in Rhabditid nematodes. Genetics **176**: 1653–1661	Trait mean from Table 1 and CV_M (MA mean), V_M and h^2_m from Table 2 of paper. While log transformation had little impact on scaled estimates, V_M and V_E were extreme on the raw scale, while mean of the log-transformed data was negative.
Pletcher, S. D., D. Houle, and J. W. Curtsinger, 1998 Age-specific properties of spontaneous mutations affecting mortality in *Drosophila melanogaster*. Genetics **148**: 287–303	h^2_m, V_M and mean from Table 5 of paper and CV_M estimated from that information. Error estimated by sampling†

Roles, A. J., M. T. Rutter, I. Dworkin, C. L. Fenster, and J. K. Conner, 2016 Field measurements of genotype by environment interaction for fitness caused by spontaneous mutations in *Arabidopsis thaliana*. Evolution **70**: 1039–1050	Table 2 of paper. We included CV_M for fruit number and total fitness however authors have calculated CV_M as $100(e^{\sqrt{VM}}-1)$ as per Poisson values. Parameter confidence intervals from MCMCglmm, and constrained to be positive.
Rutter, M. T., F. H. Shaw, and C. L. Fenster, 2010 Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. Evolution **64**: 1825–35	Table 2 of paper.
Santiago, E., J. Albornoz, A. Domínguez, M. Toro, and C. López-Fanjul, 1992 The distribution of spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. Genetics **132**: 771–781	h^2_m and V_E from Table 4 for full-sib mating lines only (B lines; all lines). Trait means were taken from Figure 3 (sternopleural bristles =21.7, wing length = 1210 and wing width = 465). CV_M estimated from that information but error estimates not available.
Schoen, D. J., 2005 Deleterious mutation in related species of the plant genus *Amsinckia* with contrasting mating systems. Evolution **59**: 2370–2377	Trait means and mutational variances were taken from Table 1 and 3 for flower number and weight respectively. V_M, h^2_m, and CV_M were estimated from that information, where V_L for MA was corrected for initial (control variance) if that was non-zero; error estimates were similar between control and mutant and MA SE were used to approximate error for meta-analysis.
Schultz, S. T., M. Lynch, and J. H. Willis, 1999 Spontaneous deleterious mutation in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. 96: 11393-11398	From Table 3; CV_M not available as calculated from among-line variance, not the per-generation V_M.
Shaw, R. G., D. L. Byers, and E. Darmo, 2000 Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. Genetics **155**: 369–78	Estimates of mean and among-line and environmental variances taken from Table 1 of paper (generation 17), and estimates of V_M, h^2_m, and CV_M from Table 2. To estimate error for h^2_m, and CV_M, the among-line and mean estimates for generation 17 were used, along with V_M from the joint analysis. Error estimated by sampling†
Vassilieva, L. L., A. M. Hook, and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution **54**: 1234–1246	h^2_m was taken from Table 2; trait mean (intercept) from Table 1 and V_M (slope/2) from Table 2 were used to calculate CV_M. Error in CV_M estimated by sampling† following the approach detailed for Lynch *et al.* (1998) above. For one trait (generation time) the reported standard error of the V_M was 0, precluding calculation of errors.
Vassilieva, L. L. and M. Lynch, 1999 The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics **151**: 119–129	h^2_M and V_M estimates were taken from Table 1 of paper. Error in CV_M estimated by sampling† following the approach detailed for Lynch *et al.* (1998) above. For one trait (generation time) the estimated mean was ~0, resulting in nonsensical CV_M and hence this was excluded.

(B) ESTIMATES REMOVED FROM META-ANALYSIS DUE TO ISSUES WITH MUTATIONAL PARAMETER ESTIMATES

Baer, C. F., J. Joyner-Matos, D. Ostrow, V. Grigaltchik, M. P. Salomon, *et al.*, 2010 Rapid decline in fitness of mutation accumulation lines of gonochoristic (outcrossing) *Caenorhabditis nematodes*. Evolution **64**: 3242–3253	As in (A); Excluded *Caenorhabditis brenneri* strain PB2801 h^2_m and strain PB2802 CV_M generation 100 estimates as they were outliers (>3.5 standard deviations from the mean and >5 times the inter-quartile range from the median) and .
Bailey, D. W., 1959 Rates of subline divergence in highly inbred strains of mice. J. Hered **50**: 26–30.	From Table 1, and in-text information on number of generations; h^2_m calculated as per Lynch 1988[4]. Excluded due to low representation of vertebrate taxa.
Casellas, J. and J. F. Medrano, 2008 Within-generation mutation variance for litter size in inbred mice. Genetics **179**: 2147–55	MA litter mean was taken for generation 46 in Table 1. The V_M and h^2_m estimates were taken from text and Table 3. CV_M was calculated from reported information. Excluded due to low representation of vertebrate taxa.
Clark, A. G., L. Wang, and T. Hulleberg, 1995 Spontaneous mutation rate of modifiers of metabolism in *Drosophila*. Genetics **139**: 767–779	As in (A). ADH h^2_m excluded as it was an extreme outlier (>4 standard deviations from the mean and >5 times the inter-quartile range from the median).
Downie, D., 2003 Effects of short-term spontaneous mutation accumulation for life history traits in grape phylloxera, *Daktulosphaira vitifoliae*. Genetica **119**: 237–251	V_M and h^2_m were taken from Table 1. Line means were estimated from figure 3 and used to calculate CV_M (but not error). Excluded due to low representation of non-Drosophila insect taxa.
Farhadifar, R., J. M. Ponciano, E. C. Andersen, D. J. Needleman, and C. F. Baer, 2016 Mutation is a sufficient and robust predictor of genetic variation for mitotic spindle traits in *Caenorhabditis elegans*. Genetics **203**: 1859–1870	Trait means from Table 1 of paper, and variances from Table S4. One estimate (for posterior centrosome oscillation frequency) was removed as the estimate for V_M was negative. Excluded due to low representation of trait type (mitotic cell division).
Harada, K., 1995 A quantitative analysis of modifier mutations which occur in mutation accumulation lines in *Drosophila melanogaster*. Heredity **75**: 589–598	As in (A); Could not calculate CV_M (negative mean) for one metabolite (Isocitrate-dehydrogenase activity) which was therefore excluded. Alpha-amalyse was excluded as the estimated SE of CV_M was considerably larger than the estimate itself.

Houle, D., K. A. Hughes, D. K. Hoffmaster, J. Ihara, S. Assimacopoulos, *et al.*, 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. Genetics **138**: 773–785	As in (A). Early male mating success excluded as CV_M estimate was negative; late mating CV_M excluded due extreme value (> 4 standard deviations from the mean and >5 times the inter-quartile range from the median)
Joyner-Matos, J., L. C. Bean, H. L. Richardson, T. Sammeli, and C. F. Baer, 2011 No evidence of elevated germline mutation accumulation under oxidative stress in *Caenorhabditis elegans*. Genetics **189**: 1439–1447	Information in Table 1 of paper used to calculate $CV_M = 100*\sqrt{V_M}$, where V_M (and V_L) were reported on intensity of selection scale, as described on pg 1441 of paper. We sampled from the distribution of parameters used to calculate V_M, but the standard error of the sample was orders of magnitude larger than the estimate. Therefore, this study was excluded.
Latta, L. C., M. Peacock, D. J. Civitello, J. L. Dudycha, J. M. Meik, *et al.*, 2015 The phenotypic effects of spontaneous mutations in different environments. Am. Nat. **185**: 243–252	Among-line variance (designated by authors as ΔV) provided in Table S1 and used to calculated using $V_M = \Delta V/(2*96)$, where 96 was the number of generations of MA reported by the authors on pg 245 of paper. V_E and $\bar{X}$ were provided in Table S1 and, along with the calculated V_M, used to calculate h^2_m and CV_M. However, the reported confidence intervals for ΔV are extremely large (infinite for several traits), and the standard errors calculated are always larger than the parameter estimate, and all estimates from this study were excluded. Notably, this study included only five MA lines, and poor quality estimates could be expected.
López, M. A. and C. López-Fanjul, 1993 Spontaneous mutation for a quantitative trait in *Drosophila melanogaster*. II. Distribution of mutant effects on the trait and fitness. Genet. Res. **61**: 117–126	
Lynch, M., L. Latta, J. Hicks, and M. Giorgianni, 1998 Mutation, selection, and the maintenance of life-history variation in a natural population. Evolution **52**: 727–733.	As in (A). Second clutch size excluded as it was an extreme outlier (>7 standard deviations from the mean and >5 times the inter-quartile range from the median). Size at birth was excluded because calculation of h^2_m from the data in Table 1 returned an estimate several orders of magnitude larger than reported, suggesting some incorrect parameter values reported.
Morgan, A. D., R. W. Ness, P. D. Keightley, and N. Colegrave, 2014 Spontaneous mutation accumulation in multiple strains of the green alga, *Chlamydomonas reinhardtii*. Evolution **68**: 2589–2602	CV_M estimates were taken from Table 3 (second assay only, as one strain had missing data for the first assay), but no SE reported. To estimate SE it was necessary to use V_L estimates for MA and ancestor from Table 3 (and reported CI), the generation number estimates from Table 2, and estimates of trait mean (CI) from Table 4. Excluded due to low representation of micro-organisms.
Pannebakker, L. A., D. L. Halligan, K. T. Reynolds, G. A. Ballantyne, D. M. Shuker, *et al.*, 2008 Effects of spontaneous mutation accumulation on sex ratio traits in a parasitoid wasp. Evolution **62**: 1921–1935	MA mean and line numbers taken from Table 2 of paper. Variance estimates were taken from Table 3, where CV_M has been converted to a percentage scale. Excluded due to low representation of non-Drosophila insect taxa.

(C) ESTIMATES REMOVED FROM META-ANALYSIS DUE TO LACK OF INFORMATION ON ESTIMATION ERROR OF A MUTATIONAL PARAMETER

Burch, C. L., S. Guyader, D. Samarov, and H. Shen, 2007 Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus φ6. Genetics **176**: 467–476	V_M was calculated from the log fitness variances reported on pg. 473, and MA mean was calculated as 1 - E(log W_1) to calculate CV_M.
Chang, S. and R. G. Shaw,2003 The contribution of spontaneous mutation to variation in environmental response in *Arabidopsis thaliana*: Responses to nutrients. Evolution **57**: 984–994	Table 3 of paper.
Clarke, D. K., E. A. Duarte, A. Moya, S. F. Elena, E. Domingo, *et al.*, 1993 Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. J. Virol. **67**: 222–228	Halligan & Keightley[1] Table 1; no further detail in original paper to estimate SE.
Festing, M., 1973 A multivariate analysis of subline divergence in the shape of the mandible in c57bl/gr mice. Genet. Res. **21**: 121–132	Houle *et al.*[2] Table 1; insufficient information in the original paper to determine error for scaled variance estimates.
Gong, Y., R. Woodruff, and J. Thompson Jr, 2005 Deleterious genomic mutation rate for viability in *Drosophila melanogaster* using concomitant sibling controls. Biol. Lett. **1**: 492–495	CV_M estimates were taken from Halligan & Keightley[1] Table 1, but were noted by. Halligan & Keightley[1] as extreme outliers. Insufficient information to calculate errors for scaled estimates.
Hall, D. W., R. Mahmoudizad, A. W. Hurd, and S. L. Joseph, 2008 Spontaneous mutations in diploid *Saccharomyces cerevisiae*: Another thousand cell generations. Genet. Res. **90**: 229–241	Table 2 of paper.
Huang, W., R. F. Lyman, R. A. Lyman, M. A. Carbone, S. T. Harbison, *et al.*, 2016 Spontaneous and the origin and maintenance of quantitative genetic variation. Elife **5**: e14625	h^2_m estimates were taken from supplementary information 1D but no estimate of error provided.
Joseph, S. L. and D. W. Hall, 2004 Spontaneous mutations in diploid *Saccharomyces cerevisiae* more beneficial than expected. Genetics **168**: 1817–1825	From in-text information on page 1819; NB: heritability estimate was incorrectly reported and corrected in Hall et al. 2008 (citation details above).
Kavanaugh, C. M. and R. G. Shaw, 2005 The contribution of spontaneous mutation to variation in environmental responses of *Arabidopsis thaliana*: Responses to light. Evolution **59**: 266–275	Estimates from information in Table 5 of paper, but no errors reported in that table. Notably several estimates of mutational variance were negative (Table 5), and mutational variance was not statistically supported for any trait in this study.
Keightley, P. D. and A. Caballero, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. **94**: 3823–3827	Lifespan estimates were taken from text on p. 3826 of paper but no error estimate reported.
Kibota, T. T. and M. Lynch, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature **381**: 694–696	Halligan & Keightley[1] Table 1; no further detail in original paper to estimate SE.
Mackay, T. F. and R. F. Lyman, 1998 Polygenic mutation in *Drosophila melanogaster*: Genotype × environment interaction for spontaneous mutations affecting bristle number. Genetica **102**: 199-215	Table 9 of paper: Only the MA maintenance temperature was recorded. No error estimates.
Mukai, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. spontaneous mutation rate of polygenes controlling viability. Genetics **50**: 1–19	Houle *et al.* (1996)[2] Table 1; V_E complexity detailed in Houle *et al.* (1996)[2] and lack of information on mean error mean that no SE of scaled V_M were obtained.
Mukai, T. and T. Yamazaki, 1968 The genetic structure of natural populations of *Drosophila melanogaster*. V. coupling-repulsion effect of spontaneous mutant polygenes controlling viability. Genetics **59**: 513–535	Table 1 of paper; no further detail in original paper to estimate SE of scaled V_M.
Mukai,T., K. Harada,and H. Yoshimaru, 1984 Spontaneous mutations modifying the activity of alcohol dehydrogenase (adh) in *Drosophila melanogaster*. Genetics **106**: 73–84	Table 1 and in text. Insufficient information to obtain error estimate for scaled estimate of V_M..
Ohnishi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster* .II. homozygous effect of polygenic mutations. Genetics **87**: 529–545	Table 1 and in text (for zero level of mutagen only), but insufficient information to calculate SE for scaled estimates; Houle *et al.* (1996)[2] Table 1.

Paxman, G., 1957 A study of spontaneous mutation in *Drosophila melanogaster*. Genetica **29**: 39–57	Houle *et al.* (1996)[2] Table 1; no further detail in original paper to estimate SE.
Pletcher, S. D., D. Houle, and J. W. Curtsinger, 1999 The evolution of age-specific mortality rates in *Drosophila melanogaster*: Genetic divergence among unselected lines. Genetics **153**: 813–823	Estimates of V_M from Table 4 of paper, but no estimates of error of trait mean or V_E reported and no scaled estimate with associated error derived.
Russell, W., G. Sprague, and L. Penny, 1963 Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. **3**: 175–178	Houle *et al.* (1996)[2] Table 1; no further detail in original paper to estimate SE. The authors note (pp 177) that their genetic variation are "rather poorly estimated".
Wayne, M. L. and T. F. Mackay, 1998 Quantitative genetics of ovariole number in *Drosophila melanogaster*. II. Mutational variation and genotype-environment interaction. Genetics **148**: 201– 210	Table 3 of paper.
Xu, J., 2004 Genotype-environment interactions of spontaneous mutations for vegetative fitness in the human pathogenic fungus *Cryptococcus neoformans*. Genetics **168**: 1177–1188	h^2_m estimates from Table 2 of paper.

† Where 95% confidence intervals were reported, we converted to approximate standard errors as: (upper limit – lower limit) / 3.92. Where standard errors were not reported for h^2_M or CV_M, but were reported for the parameters used to calculate these scaled mutational variance estimates (e.g., V_M, V_E, mean), as described in the Methods, we used the `rnorn` function in R to generate 10,000 random samples from the distribution defined by the estimate (mean) and standard error (variance), and calculated the scaled metric (for h^2_M or CV_M) in each sample, then estimated the standard error.

[1] Halligan, D. L. and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. **40**: 151–172

[2] Houle, D., L. Morikawa, and M. Lynch, 1996 Comparing mutational variabilities. Genetics **143**: 1467–1483

[3] Ajie, L. C., S. Estes, M. Lynch, and P. C. Phillips, 2005 Behavioural degradation under mutation accumulation in *Caenorhabditis elegans*. Genetics **170**: 655–660

[4] Lynch, M., 1988 The rate of polygenic mutation. Genetical Research **51:** 137-148.

Table S2. Trait types included in the broad categories and within the more narrowly defined life-history trait categories. The trait description is a list of traits included within each category.

Broad Category	Narrow Category	Trait Description
Life-history	Fitness	Fitness; Intrinsic rate of increase; Competitive mating success; Rate of convergence to the stable age distribution.
	Miscellaneous	Longevity, Lifespan, Male mating success, Age at maturity, Age at reproduction, Growth rate, Instar duration, The reciprocal of the mean age at reproduction, Time to maturity.
	Productivity	Clutch size, Fecundity, Litter size, Number of flowers, Seed count, Productivity, Reproductive mass
	Survival	Survival, Age-specific survivorship or viability, Germination, Pathogen susceptibility.
Morphology	-	Abdominal and sternopleural bristles; Body length, width or size; Limb/Organ/Spine length or width, Weight.
Physiology	-	Enzyme activity, Metabolites, Locomotor activity, Velocity.

Table S3. Mutation accumulation (MA) lines shared across studies. The first published description of a set of MA lines is indicated in bold; the following non-bolded entries represent subsequent investigations of the same set of MA lines. See Table S1 for citation details.

Species name	Reference	Generation	Trait Description
Arabidopsis thaliana	**Shaw *et al.* (2000)**	17	Fecundity, Reproductive mass
	Rutter *et al.* (2010)	25	Fecundity
	Roles *et al.* (2016)	25	Fecundity, Germination, Age-specific survivorship, Fitness
Caenorhabditis briggsae	**Baer *et al.* (2005)**	200	Productivity
	Ostrow *et al.* (2007)	200	Body size
	Joyner-Matos *et al.* (2009)	250	Life span
Caenorhabditis elegans	**Baer *et al.* (2005)**	200	Productivity
	Ostrow *et al.* (2007)	200	Body size
	Joyner-Matos *et al.* (2009)	250	Life span
	Etienne *et al.* (2015)	250	Pathogen susceptibility
	Davies *et al.* (2016)	250	Metabolites
	Keightley & Caballero (1997)	60	Lifespan, Productivity, Population growth
	Azevedo *et al.* (2002)	60	Body size
	Vassilieva *et al.* (1999)	49.4	Longevity, Productivity, Survivorship, Population growth
	Vassilieva *et al.* (2000)	214	Longevity, Productivity, Survivorship, Population growth
	Azevedo *et al.* (2002)	152	Body size
	Estes *et al.* and Ajie *et al.* (2005)	370	Population growth parameter, Body width Locomotor activity, Productivity, Velocity
Drosophila melanogaster	**Houle *et al.* (1994)**	44	Fecundity, Fitness, Longevity, Male mating, Productivity
	Pletcher *et al.* (1998)	19	Female longevity, Male longevity
	Santiago *et al.* (1992)	46	Sternopleural bristles, Wing dimensions
	Fernandez & Lopez-Fanjul (1996)	106	Viability, Fecundity
	Martorell *et al.* (1998)	82	Productivity, Longevity
	Garcia-Dorado *et al.* (2000)	170	Wing length
		204	Abdominal bristles, Sternopleural bristles
		209	Viability
	Caballero *et al.* (2002)	210	Viability
	Avila & Garca-Dorado (2002)	288	Fitness
Drosophila serrata	**McGuigan *et al.* (2011a)**	23	Mating Success
		26	Productivity
	Latimer *et al.* (2014)	32	Locomotor activity
Oscheius myriophila	**Baer *et al.* (2005)**	200	Productivity
	Ostrow *et al.* (2007)	200	Body size

Table S4. Parameter estimates from analyses of data for each of the 11 traits in each generation for each treatment in the Drosophila serrata experiment. Model (2) was fit to data for each trait in each generation for each population size treatment. The least squares trait mean was estimated, and the phenotypic variance was partitioned to among-line (due to mutation), among the two replicate rearing vials nested within each line, and to the residual (among individuals within each vial) variances. We estimated the environmental variance as the sum of among-vial and among-individual (residual) variances. We placed the estimates of among-line variance on both a heritability (among-line variance / environmental variance) and coefficient of variance (100* √Among-line / Mean) scales but note that these should not be interpreted as mutational heritability or coefficient of variance because they are calculated from the total accumulated mutational variance (among-line variance) not the per generation rate at which that accrued. For each parameter, confidence intervals were estimated by drawing 10,000 random samples from a normal distribution with a mean equal to the REML model estimate and variance equal to either the inverse of the Fisher information matrix (for variance parameters) or the standard deviation (for the trait mean), as described in the Methods. From the 10,000 samples, 95% (Mean) or 90% (all other parameters, which are constrained to be positive) confidence intervals (CI) are reported. The environmental variance, heritability and coefficient of variance) were calculated for each of the 10,000 samples to estimate the CI of these calculated parameters. Statistical support for non-zero estimates was inferred where the lower CI did not include zero. Where REML parameter estimates were zero, no CI were estimated.

			Mean		Among-Line Variance (x10^4)		Among-Vial Variance (x10^4)		Among-Individual Variance (x10^4)		Environmental Variance (x10^4)		Heritability		Coefficient of Variance	
Trait	Generation	Treatment	Estimate	CI	Estimate	CI	Estimate	CI	Estimate	CI	Estimate	CI	Estimate	CI	Estimate	CI
Size	1	Small	1.5533	(1.5479, 1.5590)	1.22	(-0.1893, 2.6234)	2.86	(1.3896, 4.3164)	4.74	(4.1377, 5.3282)	7.60	(6.0591, 9.1344)	0.1605	(-0.0225, 0.3952)	0.7111	(-0.2797, 1.0420)
Size	1	Large	1.5610	(1.5544, 1.5676)	3.43	(1.6556, 5.1803)	1.16	(0.2797, 2.0447)	6.57	(5.7532, 7.3629)	7.73	(6.6219, 8.8173)	0.4437	(0.2100, 0.7128)	1.1864	(0.8249, 1.4583)
Size	2	Small	1.5438	(1.5384, 1.5492)	1.58	(0.3280, 2.8324)	1.71	(0.6214, 2.8125)	6.74	(5.9335, 7.5567)	8.45	(7.1647, 9.7280)	0.1870	(0.0363, 0.3644)	0.8142	(0.3710, 1.0909)
Size	2	Large	1.5504	(1.5455, 1.5553)	1.73	(0.7260, 2.7259)	0.69	(0.0253, 1.3597)	6.05	(5.3302, 6.7605)	6.74	(5.8657, 7.6052)	0.2567	(0.1037, 0.4260)	0.8484	(0.5501, 1.0650)
Size	3	Small	1.5410	(1.5362, 1.5459)	1.17	(0.1614, 2.1962)	0.91	(-0.0213, 1.8340)	7.08	(6.1208, 8.0087)	7.99	(6.8023, 9.1452)	0.1464	(0.0191, 0.2963)	0.7019	(0.2605, 0.9613)
Size	3	Large	1.5373	(1.5326, 1.5422)	1.38	(0.4517, 2.3084)	0.23	(-0.4827, 0.9642)	7.26	(6.2514, 8.2785)	7.49	(6.4357, 8.5600)	0.1842	(0.0574, 0.3275)	0.7642	(0.4369, 0.9881)
Size	4	Small	1.5547	(1.5489, 1.5606)	1.27	(-0.2832, 2.8471)	3.49	(1.8026, 5.1612)	6.42	(5.6468, 7.1644)	9.91	(8.1122, 11.6892)	0.1282	(-0.0258, 0.3233)	0.7249	(-0.3418, 1.0854)
Size	4	Large	1.5599	(1.5554, 1.5644)	1.01	(0.0468, 1.9807)	1.63	(0.7131, 2.5526)	4.85	(4.2651, 5.4373)	6.48	(5.4665, 7.5360)	0.1559	(0.0070, 0.3326)	0.6443	(0.1385, 0.9018)
Size	5	Small	1.5651	(1.5593, 1.5708)	1.45	(0.0225, 2.8730)	2.07	(0.6662, 3.4709)	7.31	(6.2938, 8.2947)	9.38	(7.7903, 10.9640)	0.1546	(0.0020, 0.3378)	0.7694	(0.0959, 1.0826)
Size	5	Large	1.5736	(1.5687, 1.5786)	2.00	(1.0200, 2.9678)	0.15	(-0.3053, 0.6057)	5.21	(4.5577, 5.8661)	5.36	(4.6609, 6.0466)	0.3731	(0.1851, 0.5770)	0.8987	(0.6420, 1.0950)
Size	6	Small	1.5807	(1.5770, 1.5845)	0.81	(0.2089, 1.4088)	0.61	(0.0976, 1.1138)	4.32	(3.8044, 4.8274)	4.93	(4.2726, 5.5820)	0.1643	(0.0400, 0.3024)	0.5694	(0.2890, 0.7510)
Size	6	Large	1.5861	(1.5813, 1.5909)	1.20	(0.2083, 2.2130)	1.30	(0.3949, 2.2133)	5.23	(4.5325, 5.9235)	6.53	(5.4997, 7.5842)	0.1838	(0.0299, 0.3707)	0.6907	(0.2875, 0.9374)
ILD1.2	1	Small	0.6699	(0.6689, 0.6709)	0.0246	(-0.0246, 0.0751)	0.0769	(0.0163, 0.1372)	0.4329	(0.3798, 0.4859)	0.5098	(0.4362, 0.5829)	0.0482	(-0.0455, 0.1598)	0.2340	(-0.2344, 0.4089)
ILD1.2	1	Large	0.6707	(0.6699, 0.6715)	0.0181	(-0.0106, 0.0469)	0.0098	(-0.0248, 0.0448)	0.4585	(0.4030, 0.5143)	0.4683	(0.4116, 0.5254)	0.0387	(-0.0219, 0.1068)	0.2007	(-0.1534, 0.3228)
ILD1.2	2	Small	0.6690	(0.6675, 0.6705)	0.1635	(0.0739, 0.2550)	0.0420	(-0.0097, 0.0926)	0.5284	(0.4659, 0.5916)	0.5704	(0.4981, 0.6415)	0.2867	(0.1249, 0.4670)	0.6044	(0.4063, 0.7554)
ILD1.2	2	Large	0.6692	(0.6682, 0.6702)	0.0549	(0.0098, 0.0993)	0.0239	(-0.0161, 0.0651)	0.4970	(0.4399, 0.5549)	0.5209	(0.4598, 0.5842)	0.1054	(0.0181, 0.2003)	0.3501	(0.1475, 0.4711)
ILD1.2	3	Small	0.6714	(0.6705, 0.6724)	0.0131	(-0.0333, 0.0581)	0.0267	(-0.0340, 0.0879)	0.6212	(0.5388, 0.7050)	0.6479	(0.5596, 0.7379)	0.0202	(-0.0490, 0.0965)	0.1704	(-0.2718, 0.3591)
ILD1.2	3	Large	0.6720	(0.6705, 0.6735)	0.1330	(0.0514, 0.2160)	0	-	0.7324	(0.6349, 0.8286)	0.7324	(0.6349, 0.8286)	0.1816	(0.0699, 0.3066)	0.5427	(0.3371, 0.6919)
ILD1.2	4	Small	0.6685	(0.6676, 0.6695)	0.0539	(0.0139, 0.0940)	0.0084	(-0.0250, 0.0416)	0.4516	(0.3983, 0.5050)	0.4600	(0.4058, 0.5147)	0.1172	(0.0291, 0.2132)	0.3474	(0.1765, 0.4585)
ILD1.2	4	Large	0.6696	(0.6685, 0.6707)	0.0803	(0.0279, 0.1325)	0.0263	(-0.0126, 0.0662)	0.4338	(0.3813, 0.4868)	0.4601	(0.4030, 0.5175)	0.1745	(0.0592, 0.3021)	0.4232	(0.2498, 0.5437)
ILD1.2	5	Small	0.6721	(0.6711, 0.6731)	0.0298	(-0.0073, 0.0665)	0	-	0.6294	(0.5479, 0.7107)	0.6294	(0.5479, 0.7107)	0.0473	(-0.0114, 0.1101)	0.2568	(-0.1274, 0.3837)
ILD1.2	5	Large	0.6729	(0.6721, 0.6736)	0.0118	(-0.0193, 0.0430)	0.0066	(-0.0331, 0.0461)	0.4968	(0.4337, 0.5596)	0.5034	(0.4395, 0.5683)	0.0235	(-0.0367, 0.0910)	0.1618	(-0.2063, 0.3080)
ILD1.2	6	Small	0.6728	(0.6720, 0.6736)	0.0105	(-0.0248, 0.0457)	0.0467	(0.0004, 0.0932)	0.4443	(0.3918, 0.4972)	0.4910	(0.4277, 0.5522)	0.0215	(-0.0480, 0.0997)	0.1526	(-0.2341, 0.3179)
ILD1.2	6	Large	0.6718	(0.6709, 0.6728)	0.0551	(0.0174, 0.0928)	0	-	0.4340	(0.3797, 0.4873)	0.4340	(0.3797, 0.4873)	0.1271	(0.0388, 0.2221)	0.3495	(0.1960, 0.4535)
ILD1.5	1	Small	0.5617	(0.5606, 0.5628)	0.0926	(0.0418, 0.1419)	0.015	(-0.0109, 0.0413)	0.2863	(0.2502, 0.3221)	0.3013	(0.2628, 0.3407)	0.3074	(0.1347, 0.4953)	0.5418	(0.3642, 0.6708)
ILD1.5	1	Large	0.5624	(0.5612, 0.5635)	0.1139	(0.0620, 0.1674)	0	-	0.3040	(0.2692, 0.3391)	0.3040	(0.2692, 0.3391)	0.3747	(0.2012, 0.5622)	0.6001	(0.4427, 0.7273)
ILD1.5	2	Small	0.5619	(0.5609, 0.5628)	0.0615	(0.0223, 0.1002)	0.0184	(-0.0094, 0.0460)	0.3069	(0.2706, 0.3428)	0.3253	(0.2855, 0.3652)	0.1892	(0.0660, 0.3221)	0.4415	(0.2660, 0.5632)
ILD1.5	2	Large	0.5612	(0.5602, 0.5621)	0.0739	(0.0352, 0.1128)	0.0069	(-0.0147, 0.0283)	0.2879	(0.2542, 0.3216)	0.2948	(0.2601, 0.3292)	0.2507	(0.1167, 0.3948)	0.4845	(0.3345, 0.5987)
ILD1.5	3	Small	0.5619	(0.5609, 0.5628)	0.0482	(0.0124, 0.0852)	0	-	0.4646	(0.4060, 0.5247)	0.4646	(0.4060, 0.5247)	0.1037	(0.0260, 0.1917)	0.3906	(0.1985, 0.5196)
ILD1.5	3	Large	0.5610	(0.5599, 0.5621)	0.0491	(-0.0030, 0.1018)	0.0323	(-0.0211, 0.0852)	0.4801	(0.4146, 0.5467)	0.5124	(0.4390, 0.5872)	0.0958	(-0.0055, 0.2114)	0.3949	(-0.0969, 0.5691)
ILD1.5	4	Small	0.5600	(0.5591, 0.5609)	0.0590	(0.0238, 0.0940)	0.0154	(-0.0079, 0.0388)	0.2737	(0.2418, 0.3060)	0.2891	(0.2540, 0.3240)	0.2040	(0.0788, 0.3399)	0.4336	(0.2752, 0.5477)
ILD1.5	4	Large	0.5579	(0.5570, 0.5588)	0.0569	(0.0237, 0.0905)	0.0032	(-0.0192, 0.0252)	0.3054	(0.2691, 0.3423)	0.3086	(0.2720, 0.3448)	0.1845	(0.0739, 0.3057)	0.4277	(0.2760, 0.5394)
ILD1.5	5	Small	0.5600	(0.5588, 0.5612)	0.1025	(0.0465, 0.1605)	0.0056	(-0.0273, 0.0375)	0.3675	(0.3180, 0.4168)	0.3731	(0.3230, 0.4233)	0.2747	(0.1201, 0.4509)	0.5717	(0.3851, 0.7156)
ILD1.5	5	Large	0.5593	(0.5583, 0.5604)	0.0633	(0.0156, 0.1116)	0.0379	(-0.0029, 0.0781)	0.3503	(0.3062, 0.3945)	0.3882	(0.3347, 0.4419)	0.1632	(0.0387, 0.3076)	0.4500	(0.2236, 0.5975)
ILD1.5	6	Small	0.5608	(0.5597, 0.5619)	0.0929	(0.0429, 0.1425)	0.0099	(-0.0182, 0.0376)	0.3588	(0.3159, 0.4006)	0.3687	(0.3248, 0.4119)	0.2518	(0.1135, 0.4031)	0.5434	(0.3691, 0.6731)
ILD1.5	6	Large	0.5617	(0.5606, 0.5629)	0.0922	(0.0348, 0.1489)	0.0349	(-0.0048, 0.0749)	0.3285	(0.2850, 0.3720)	0.3634	(0.3111, 0.4151)	0.2537	(0.0910, 0.4380)	0.5406	(0.3321, 0.6872)
ILD1.9	1	Small	0.1463	(0.1457, 0.1469)	0.0020	(-0.0158, 0.0202)	0.0185	(-0.0066, 0.0434)	0.2516	(0.2205, 0.2835)	0.2701	(0.2350, 0.3064)	0.0075	(-0.0554, 0.0804)	0.3076	(-0.8604, 0.9727)
ILD1.9	1	Large	0.1462	(0.1456, 0.1469)	0.0190	(-0.0023, 0.0397)	0.0134	(-0.0088, 0.0360)	0.2546	(0.2237, 0.2855)	0.2680	(0.2341, 0.3017)	0.0707	(-0.0083, 0.1568)	0.9416	(-0.3307, 1.3607)
ILD1.9	2	Small	0.1458	(0.1453, 0.1464)	0	-	0.0036	(-0.0115, 0.0184)	0.2844	(0.2501, 0.3184)	0.2880	(0.2566, 0.3190)	0	-	0	-
ILD1.9	2	Large	0.1457	(0.1450, 0.1465)	0.0301	(0.0104, 0.0500)	0	-	0.2749	(0.2442, 0.3050)	0.2749	(0.2442, 0.3050)	0.1095	(0.0374, 0.1879)	1.1902	(0.6994, 1.5340)
ILD1.9	3	Small	0.1455	(0.1447, 0.1462)	0.0190	(-0.0031, 0.0411)	0	-	0.3719	(0.3260, 0.4193)	0.3719	(0.3260, 0.4193)	0.0512	(-0.0086, 0.1150)	0.9484	(-0.3839, 1.3922)
ILD1.9	3	Large	0.1456	(0.1448, 0.1465)	0.0330	(0.0062, 0.0603)	0	-	0.3352	(0.2903, 0.3790)	0.3352	(0.2903, 0.3790)	0.0984	(0.0182, 0.1869)	1.2468	(0.5405, 1.6876)
ILD1.9	4	Small	0.1439	(0.1432, 0.1447)	0.0196	(-0.0040, 0.0438)	0.0228	(-0.0034, 0.0489)	0.2734	(0.2414, 0.3060)	0.2962	(0.2594, 0.3334)	0.0661	(-0.0133, 0.1562)	0.9723	(-0.4425, 1.4530)
ILD1.9	4	Large	0.1423	(0.1417, 0.1428)	0	-	0.0148	(-0.0020, 0.0315)	0.2657	(0.2341, 0.2973)	0.2805	(0.2504, 0.3108)	0	-	0	-
ILD1.9	5	Small	0.1447	(0.1439, 0.1455)	0.0140	(-0.0145, 0.0426)	0.0232	(-0.0125, 0.0579)	0.3209	(0.2778, 0.3652)	0.3441	(0.2957, 0.3940)	0.0407	(-0.0397, 0.1338)	0.8182	(-0.8322, 1.4253)
ILD1.9	5	Large	0.1436	(0.1430, 0.1442)	0.0054	(-0.0097, 0.0209)	0	-	0.3543	(0.3120, 0.3969)	0.3543	(0.3120, 0.3969)	0.0153	(-0.0261, 0.0617)	0.5132	(-0.6856, 1.0065)
ILD1.9	6	Small	0.1438	(0.1430, 0.1446)	0.0481	(0.0197, 0.0765)	0	-	0.3290	(0.2920, 0.3655)	0.3290	(0.2920, 0.3655)	0.1463	(0.0597, 0.2407)	1.5255	(0.9770, 1.9243)
ILD1.9	6	Large	0.1445	(0.1436, 0.1453)	0.0271	(-0.0036, 0.0574)	0.0220	(-0.0107, 0.0548)	0.2980	(0.2581, 0.3372)	0.3200	(0.2759, 0.3646)	0.0846	(-0.0103, 0.1914)	1.1389	(-0.4124, 1.6596)
ILD2.5	1	Small	0.5068	(0.5060, 0.5076)	0.0246	(-0.0000, 0.0496)	0	-	0.4444	(0.3918, 0.4962)	0.4444	(0.3918, 0.4962)	0.0553	(-0.0001, 0.1161)	0.3093	(-0.0120, 0.4394)

ILD2.5	1	Large	0.5058	(0.5049, 0.5067)	0.0414	(0.0106, 0.0719)	0	-	0.4638	(0.4107, 0.5180)	0.4638	(0.4107, 0.5180)	0.0892	(0.0222, 0.1604)	0.4021	(0.2037, 0.5301)
ILD2.5	2	Small	0.5090	(0.5082, 0.5098)	0.0205	(-0.0115, 0.0534)	0.0439	(0.0058, 0.0815)	0.3124	(0.2742, 0.3501)	0.3563	(0.3076, 0.4052)	0.0576	(-0.0309, 0.1627)	0.2814	(-0.2104, 0.4545)
ILD2.5	2	Large	0.5075	(0.5067, 0.5084)	0.0420	(0.0135, 0.0697)	0	-	0.3803	(0.3373, 0.4227)	0.3803	(0.3373, 0.4227)	0.1103	(0.0348, 0.1894)	0.4036	(0.2292, 0.5201)
ILD2.5	3	Small	0.5050	(0.5040, 0.5059)	0.0276	(-0.0065, 0.0622)	0	-	0.5827	(0.5092, 0.6565)	0.5827	(0.5092, 0.6565)	0.0474	(-0.0110, 0.1111)	0.3291	(-0.1591, 0.4939)
ILD2.5	3	Large	0.5041	(0.5031, 0.5050)	0.0225	(-0.0114, 0.0568)	0	-	0.6044	(0.5259, 0.6830)	0.6044	(0.5259, 0.6830)	0.0372	(-0.0183, 0.0984)	0.2973	(-0.2115, 0.4732)
ILD2.5	4	Small	0.5072	(0.5064, 0.5081)	0.0265	(-0.0082, 0.0619)	0.0339	(-0.0047, 0.0730)	0.3983	(0.3507, 0.4453)	0.4322	(0.3771, 0.4863)	0.0613	(-0.0182, 0.1526)	0.3210	(-0.1788, 0.4905)
ILD2.5	4	Large	0.5052	(0.5044, 0.5060)	0.0179	(-0.0072, 0.0426)	0	-	0.5434	(0.4804, 0.6054)	0.5434	(0.4804, 0.6054)	0.0330	(-0.0130, 0.0819)	0.2651	(-0.1677, 0.4084)
ILD2.5	5	Small	0.5032	(0.5022, 0.5042)	0.0003	(-0.0512, 0.0532)	0.0698	(-0.0049, 0.1443)	0.5782	(0.4987, 0.6562)	0.6480	(0.5519, 0.7433)	0.0005	(-0.0738, 0.0905)	0.0357	(-0.4494, 0.4586)
ILD2.5	5	Large	0.5030	(0.5020, 0.5041)	0.0624	(0.0157, 0.1091)	0	-	0.6306	(0.5564, 0.7066)	0.6306	(0.5564, 0.7066)	0.0990	(0.0243, 0.1786)	0.4967	(0.2487, 0.6570)
ILD2.5	6	Small	0.5092	(0.5084, 0.5100)	0.0350	(0.0061, 0.0633)	0.0004	(-0.0255, 0.0267)	0.3838	(0.3388, 0.4295)	0.3842	(0.3403, 0.4286)	0.0912	(0.0155, 0.1714)	0.3677	(0.1534, 0.4941)
ILD2.5	6	Large	0.5069	(0.5062, 0.5076)	0.0072	(-0.0211, 0.0357)	0.0165	(-0.0213, 0.0553)	0.3969	(0.3445, 0.4512)	0.4134	(0.3579, 0.4705)	0.0175	(-0.0488, 0.0939)	0.1679	(-0.2867, 0.3732)
ILD2.8	1	Small	0.5460	(0.5451, 0.5469)	0.0003	(-0.0469, 0.0461)	0.0939	(0.0299, 0.1592)	0.4011	(0.3518, 0.4509)	0.4950	(0.4210, 0.5730)	0.0007	(-0.0877, 0.1037)	0.0339	(-0.3964, 0.3935)
ILD2.8	1	Large	0.5472	(0.5464, 0.5481)	0.0144	(-0.0190, 0.0477)	0.0436	(0.0027, 0.0854)	0.3750	(0.3298, 0.4194)	0.4186	(0.3631, 0.4732)	0.0344	(-0.0429, 0.1223)	0.2192	(-0.2517, 0.3995)
ILD2.8	2	Small	0.5466	(0.5454, 0.5479)	0.0942	(0.0287, 0.1604)	0.0613	(0.0084, 0.1139)	0.4323	(0.3799, 0.4843)	0.4936	(0.4246, 0.5597)	0.1908	(0.0560, 0.3477)	0.5614	(0.3097, 0.7325)
ILD2.8	2	Large	0.5466	(0.5456, 0.5476)	0.0625	(0.0165, 0.1074)	0.0282	(-0.0093, 0.0661)	0.4260	(0.3763, 0.4759)	0.4542	(0.3988, 0.5098)	0.1375	(0.0356, 0.2479)	0.4572	(0.2350, 0.5997)
ILD2.8	3	Small	0.5496	(0.5485, 0.5506)	0.0488	(-0.0005, 0.0997)	0.0448	(-0.0070, 0.0962)	0.4215	(0.3645, 0.4771)	0.4663	(0.3979, 0.5320)	0.1046	(-0.0011, 0.2288)	0.4019	(-0.0426, 0.5747)
ILD2.8	3	Large	0.5500	(0.5490, 0.5512)	0.0477	(-0.0061, 0.1009)	0.0339	(-0.0223, 0.0913)	0.5184	(0.4451, 0.5913)	0.5523	(0.4737, 0.6324)	0.0864	(-0.0105, 0.1962)	0.3972	(-0.1417, 0.5774)
ILD2.8	4	Small	0.5475	(0.5466, 0.5485)	0.0275	(-0.0159, 0.0705)	0.0726	(0.0219, 0.1220)	0.3614	(0.3196, 0.4049)	0.4340	(0.3734, 0.4952)	0.0633	(-0.0350, 0.1757)	0.3028	(-0.2303, 0.4851)
ILD2.8	4	Large	0.5495	(0.5487, 0.5505)	0.0357	(-0.0004, 0.0721)	0.0339	(-0.0034, 0.0713)	0.3676	(0.3241, 0.4112)	0.4015	(0.3502, 0.4520)	0.0888	(-0.0010, 0.1899)	0.3436	(-0.0349, 0.4886)
ILD2.8	5	Small	0.5499	(0.5489, 0.5510)	0.0463	(-0.0011, 0.0926)	0.0010	(-0.0442, 0.0476)	0.5578	(0.4818, 0.6319)	0.5588	(0.4834, 0.6342)	0.0829	(-0.0019, 0.1761)	0.3913	(-0.0592, 0.5535)
ILD2.8	5	Large	0.5514	(0.5508, 0.5521)	0.0016	(-0.0162, 0.0193)	0	-	0.4542	(0.4004, 0.5089)	0.4542	(0.4004, 0.5089)	0.0035	(-0.0344, 0.0446)	0.0725	(-0.2308, 0.2520)
ILD2.8	6	Small	0.5525	(0.5518, 0.5531)	0.0075	(-0.0175, 0.0320)	0.0152	(-0.0168, 0.0476)	0.4032	(0.3563, 0.4514)	0.4184	(0.3691, 0.4684)	0.0179	(-0.0398, 0.0803)	0.1567	(-0.2394, 0.3236)
ILD2.8	6	Large	0.5508	(0.5499, 0.5517)	0.0499	(0.0143, 0.0855)	0.0004	(-0.0293, 0.0298)	0.3541	(0.3067, 0.4016)	0.3545	(0.3079, 0.4011)	0.1408	(0.0392, 0.2541)	0.4056	(0.2172, 0.5308)
ILD3.7	1	Small	0.5890	(0.5878, 0.5904)	0.1059	(0.0343, 0.1772)	0.0707	(0.0185, 0.1230)	0.3580	(0.3149, 0.4021)	0.4287	(0.3653, 0.4929)	0.2470	(0.0768, 0.4461)	0.5525	(0.3144, 0.7144)
ILD3.7	1	Large	0.5883	(0.5871, 0.5896)	0.1180	(0.0550, 0.1822)	0.0267	(-0.0092, 0.0631)	0.3806	(0.3345, 0.4268)	0.4073	(0.3555, 0.4581)	0.2897	(0.1308, 0.4671)	0.5839	(0.3988, 0.7253)
ILD3.7	2	Small	0.5906	(0.5897, 0.5914)	0.0380	(0.0047, 0.0709)	0.0204	(-0.0105, 0.0516)	0.3477	(0.3056, 0.3896)	0.3681	(0.3211, 0.4152)	0.1033	(0.0119, 0.2039)	0.3301	(0.1158, 0.4507)
ILD3.7	2	Large	0.5901	(0.5893, 0.5910)	0.0475	(0.0187, 0.0767)	0	-	0.3699	(0.3285, 0.4105)	0.3699	(0.3285, 0.4105)	0.1283	(0.0487, 0.2133)	0.3692	(0.2313, 0.4693)
ILD3.7	3	Small	0.5904	(0.5894, 0.5914)	0.0479	(0.0033, 0.0920)	0.0340	(-0.0103, 0.0779)	0.3739	(0.3239, 0.4243)	0.4079	(0.3507, 0.4652)	0.1176	(0.0077, 0.2429)	0.3709	(0.0969, 0.5137)
ILD3.7	3	Large	0.5904	(0.5894, 0.5914)	0.0447	(0.0010, 0.0890)	0.0284	(-0.0148, 0.0710)	0.3762	(0.3242, 0.4294)	0.4046	(0.3463, 0.4627)	0.1105	(0.0023, 0.2357)	0.3582	(0.0547, 0.5051)
ILD3.7	4	Small	0.5914	(0.5905, 0.5923)	0.0390	(0.0011, 0.0771)	0.0503	(0.0126, 0.0887)	0.2990	(0.2637, 0.3341)	0.3493	(0.3017, 0.3969)	0.1118	(0.0030, 0.2360)	0.3341	(0.0556, 0.4697)
ILD3.7	4	Large	0.5930	(0.5922, 0.5937)	0.0291	(0.0007, 0.0577)	0.0203	(-0.0081, 0.0491)	0.3076	(0.2711, 0.3447)	0.3279	(0.2878, 0.3682)	0.0887	(0.0023, 0.1865)	0.2875	(0.0461, 0.4052)
ILD3.7	5	Small	0.5914	(0.5904, 0.5924)	0.0677	(0.0262, 0.1094)	0.0038	(-0.0241, 0.0312)	0.3193	(0.2758, 0.3626)	0.3231	(0.2793, 0.3673)	0.2096	(0.0775, 0.3547)	0.4400	(0.2737, 0.5592)
ILD3.7	5	Large	0.5914	(0.5905, 0.5924)	0.0492	(0.0095, 0.0883)	0.0293	(-0.0053, 0.0642)	0.3094	(0.2699, 0.3480)	0.3387	(0.2926, 0.3842)	0.1453	(0.0270, 0.2800)	0.3751	(0.1654, 0.5024)
ILD3.7	6	Small	0.5885	(0.5875, 0.5896)	0.0738	(0.0257, 0.1226)	0.0335	(-0.0029, 0.0699)	0.3609	(0.3191, 0.4042)	0.3944	(0.3438, 0.4452)	0.1872	(0.0629, 0.3276)	0.4616	(0.2726, 0.5951)
ILD3.7	6	Large	0.5897	(0.5887, 0.5908)	0.0593	(0.0121, 0.1074)	0.0397	(-0.0030, 0.0819)	0.3407	(0.2946, 0.3867)	0.3804	(0.3257, 0.4346)	0.1558	(0.0297, 0.3028)	0.4128	(0.1865, 0.5556)
ILD3.8	1	Small	0.7802	(0.7797, 0.7807)	0.0090	(-0.0002, 0.0181)	0	-	0.1623	(0.1434, 0.1815)	0.1623	(0.1434, 0.1815)	0.0553	(-0.0012, 0.1151)	0.1214	(-0.0183, 0.1727)
ILD3.8	1	Large	0.7803	(0.7799, 0.7808)	0.0066	(-0.0027, 0.0160)	0.0063	(-0.0047, 0.0171)	0.1232	(0.1084, 0.1381)	0.1295	(0.1123, 0.1462)	0.0507	(-0.0205, 0.1307)	0.1038	(-0.0670, 0.1619)
ILD3.8	2	Small	0.7799	(0.7794, 0.7804)	0.0116	(0.0020, 0.0211)	0.0004	(-0.0085, 0.0093)	0.1259	(0.1107, 0.1411)	0.1263	(0.1103, 0.1421)	0.0922	(0.0160, 0.1747)	0.1383	(0.0580, 0.1861)
ILD3.8	2	Large	0.7805	(0.7800, 0.7809)	0.0125	(0.0025, 0.0225)	0.0040	(-0.0051, 0.0131)	0.1198	(0.1058, 0.1340)	0.1238	(0.1088, 0.1390)	0.1006	(0.0188, 0.1916)	0.1430	(0.0635, 0.1923)
ILD3.8	3	Small	0.7818	(0.7812, 0.7824)	0.0136	(-0.0016, 0.0287)	0.0061	(-0.0100, 0.0220)	0.1709	(0.1473, 0.1944)	0.1770	(0.1529, 0.2014)	0.0770	(-0.0087, 0.1735)	0.1493	(-0.0513, 0.2169)
ILD3.8	3	Large	0.7822	(0.7815, 0.7829)	0.0281	(0.0104, 0.0465)	0	-	0.1661	(0.1438, 0.1875)	0.1661	(0.1438, 0.1875)	0.1693	(0.0607, 0.2928)	0.2144	(0.1301, 0.2756)
ILD3.8	4	Small	0.7814	(0.7808, 0.7819)	0.0165	(0.0026, 0.0298)	0.0118	(-0.0005, 0.0246)	0.1236	(0.1087, 0.1383)	0.1354	(0.1177, 0.1528)	0.1220	(0.0188, 0.2347)	0.1645	(0.0657, 0.2210)
ILD3.8	4	Large	0.7828	(0.7824, 0.7832)	0	-	0.0063	(-0.0024, 0.0149)	0.1385	(0.1219, 0.1551)	0.1448	(0.1291, 0.1604)	0	-	0	-
ILD3.8	5	Small	0.7819	(0.7813, 0.7825)	0.0128	(-0.0034, 0.0289)	0.0123	(-0.0053, 0.0298)	0.1538	(0.1332, 0.1745)	0.1661	(0.1421, 0.1904)	0.0773	(-0.0193, 0.1870)	0.1449	(-0.0743, 0.2174)
ILD3.8	5	Large	0.7822	(0.7816, 0.7828)	0.0187	(0.0061, 0.0309)	0	-	0.1528	(0.1347, 0.1711)	0.1528	(0.1347, 0.1711)	0.1222	(0.0393, 0.2099)	0.1747	(0.0995, 0.2248)
ILD3.8	6	Small	0.7806	(0.7801, 0.7810)	0.0080	(-0.0011, 0.0170)	0.0020	(-0.0079, 0.0122)	0.1368	(0.1208, 0.1530)	0.1388	(0.1224, 0.1548)	0.0575	(-0.0077, 0.1289)	0.1144	(-0.0431, 0.1673)
ILD3.8	6	Large	0.7802	(0.7796, 0.7807)	0.0124	(0.0003, 0.0246)	0.0049	(-0.0077, 0.0172)	0.1288	(0.1116, 0.1458)	0.1337	(0.1157, 0.1518)	0.0928	(0.0023, 0.1967)	0.1427	(0.0232, 0.2010)
ILD4.5	1	Small	0.4618	(0.4610, 0.4626)	0.0219	(-0.0098, 0.0547)	0.0286	(-0.0084, 0.0645)	0.3519	(0.3091, 0.3954)	0.3805	(0.3291, 0.4305)	0.0576	(-0.0247, 0.1541)	0.3205	(-0.2146, 0.5064)
ILD4.5	1	Large	0.4605	(0.4596, 0.4614)	0.0510	(0.0175, 0.0840)	0.0007	(-0.0247, 0.0260)	0.3608	(0.3181, 0.4052)	0.3615	(0.3190, 0.4049)	0.1412	(0.0476, 0.2430)	0.4907	(0.2872, 0.6294)
ILD4.5	2	Small	0.4597	(0.4588, 0.4605)	0.0395	(0.0094, 0.0712)	0.0158	(-0.0110, 0.0426)	0.3019	(0.2653, 0.3374)	0.3177	(0.2774, 0.3566)	0.1243	(0.0281, 0.2362)	0.4323	(0.2113, 0.5801)
ILD4.5	2	Large	0.4603	(0.4595, 0.4611)	0.0495	(0.0199, 0.0793)	0.0001	(-0.0203, 0.0207)	0.3119	(0.2757, 0.3481)	0.3120	(0.2760, 0.3484)	0.1588	(0.0621, 0.2633)	0.4835	(0.3065, 0.6118)
ILD4.5	3	Small	0.4577	(0.4567, 0.4587)	0.0550	(0.0157, 0.0936)	0	-	0.4646	(0.4057, 0.5236)	0.4646	(0.4057, 0.5236)	0.1184	(0.0333, 0.2099)	0.5124	(0.2740, 0.6687)
ILD4.5	3	Large	0.4580	(0.4570, 0.4590)	0.0357	(-0.0110, 0.0812)	0.0035	(-0.0479, 0.0546)	0.5839	(0.5018, 0.6646)	0.5874	(0.5071, 0.6679)	0.0609	(-0.0178, 0.1463)	0.4128	(-0.2291, 0.6221)
ILD4.5	4	Small	0.4611	(0.4604, 0.4618)	0.0231	(0.0013, 0.0444)	0.0031	(-0.0185, 0.0252)	0.3072	(0.2706, 0.3438)	0.3103	(0.2746, 0.3471)	0.0744	(0.0042, 0.1507)	0.3295	(0.0789, 0.4569)
ILD4.5	4	Large	0.4600	(0.4592, 0.4608)	0.0341	(0.0105, 0.0575)	0	-	0.3273	(0.2900, 0.3643)	0.3273	(0.2900, 0.3643)	0.1042	(0.0312, 0.1818)	0.4014	(0.2230, 0.5210)
ILD4.5	5	Small	0.4590	(0.4583, 0.4598)	0.0084	(-0.0213, 0.0379)	0.0048	(-0.0346, 0.0446)	0.4595	(0.3968, 0.5216)	0.4643	(0.4006, 0.5279)	0.0182	(-0.0439, 0.0867)	0.2001	(-0.3178, 0.4239)
ILD4.5	5	Large	0.4584	(0.4577, 0.4590)	0.0072	(-0.0156, 0.0302)	0.0167	(-0.0131, 0.0467)	0.3096	(0.2699, 0.3484)	0.3263	(0.2824, 0.3691)	0.0222	(-0.0451, 0.0995)	0.1856	(-0.2723, 0.3791)
ILD4.5	6	Small	0.4599	(0.4589, 0.4608)	0.0570	(0.0214, 0.0932)	0.0016	(-0.0242, 0.0286)	0.3787	(0.3327, 0.4228)	0.3803	(0.3353, 0.4247)	0.1499	(0.0550, 0.2550)	0.5193	(0.3182, 0.6638)
ILD4.5	6	Large	0.4619	(0.4610, 0.4627)	0.0347	(0.0045, 0.0648)	0	-	0.4303	(0.3768, 0.4848)	0.4303	(0.3768, 0.4848)	0.0807	(0.0101, 0.1573)	0.4035	(0.1448, 0.5514)
ILD4.8	1	Small	0.7434	(0.7429, 0.7438)	0.0065	(-0.0028, 0.0159)	0.0009	(-0.0096, 0.0119)	0.1441	(0.1266, 0.1623)	0.1450	(0.1274, 0.1633)	0.0450	(-0.0186, 0.1146)	0.1087	(-0.0715, 0.1695)

ILD4.8	1	Large	0.7435	(0.7430, 0.7440)	0.0102	(0.0011, 0.0196)	0	-	0.1554	(0.1375, 0.1735)	0.1554	(0.1375, 0.1735)	0.0656	(0.0065, 0.1309)	0.1358	(0.0438, 0.1882)
ILD4.8	2	Small	0.7421	(0.7416, 0.7425)	0.0084	(0.0004, 0.0164)	0	-	0.1450	(0.1284, 0.1614)	0.1450	(0.1284, 0.1614)	0.0582	(0.0027, 0.1165)	0.1238	(0.0269, 0.1725)
ILD4.8	2	Large	0.7428	(0.7423, 0.7432)	0.0142	(0.0043, 0.0242)	0.0009	(-0.0075, 0.0094)	0.1211	(0.1070, 0.1352)	0.1220	(0.1074, 0.1368)	0.1161	(0.0342, 0.2050)	0.1603	(0.0884, 0.2094)
ILD4.8	3	Small	0.7432	(0.7427, 0.7437)	0.0072	(-0.0071, 0.0214)	0.0089	(-0.0082, 0.0268)	0.1736	(0.1500, 0.1973)	0.1825	(0.1569, 0.2082)	0.0397	(-0.0367, 0.1250)	0.1145	(-0.1131, 0.1969)
ILD4.8	3	Large	0.7439	(0.7433, 0.7445)	0.0200	(0.0050, 0.0351)	0	-	0.1738	(0.1512, 0.1965)	0.1738	(0.1512, 0.1965)	0.1151	(0.0281, 0.2103)	0.1902	(0.0949, 0.2520)
ILD4.8	4	Small	0.7434	(0.7429, 0.7439)	0.0047	(-0.0063, 0.0157)	0.0148	(0.0009, 0.0285)	0.1288	(0.1136, 0.1438)	0.1436	(0.1241, 0.1630)	0.0325	(-0.0423, 0.1163)	0.0919	(-0.1068, 0.1687)
ILD4.8	4	Large	0.7444	(0.7441, 0.7448)	0	-	0.0015	(-0.0065, 0.0096)	0.1521	(0.1343, 0.1704)	0.1536	(0.1370, 0.1706)	0	-	0	-
ILD4.8	5	Small	0.7441	(0.7436, 0.7446)	0.0064	(-0.0033, 0.0162)	0	-	0.1746	(0.1524, 0.1968)	0.1746	(0.1524, 0.1968)	0.0365	(-0.0184, 0.0970)	0.1073	(-0.0768, 0.1709)
ILD4.8	5	Large	0.7445	(0.7440, 0.7450)	0.0081	(-0.0018, 0.0180)	0	-	0.1791	(0.1573, 0.2009)	0.1791	(0.1573, 0.2009)	0.0454	(-0.0096, 0.1045)	0.1211	(-0.0569, 0.1800)
ILD4.8	6	Small	0.7431	(0.7426, 0.7435)	0.0067	(-0.0033, 0.0169)	0.0075	(-0.0046, 0.0193)	0.1365	(0.1205, 0.1527)	0.1440	(0.1268, 0.1612)	0.0467	(-0.0223, 0.1229)	0.1104	(-0.0768, 0.1750)
ILD4.8	6	Large	0.7439	(0.7434, 0.7443)	0.0051	(-0.0064, 0.0164)	0.0071	(-0.0073, 0.0214)	0.1411	(0.1218, 0.1602)	0.1482	(0.1281, 0.1679)	0.0345	(-0.0409, 0.1197)	0.0961	(-0.1078, 0.1723)
ILD5.6	1	Small	0.2019	(0.2013, 0.2025)	0.0146	(0.0000, 0.0293)	0	-	0.2494	(0.2206, 0.2792)	0.2494	(0.2206, 0.2792)	0.0585	(0.0001, 0.1218)	0.5983	(0.0319, 0.8474)
ILD5.6	1	Large	0.2019	(0.2013, 0.2025)	0.0135	(-0.0015, 0.0285)	0	-	0.2859	(0.2533, 0.3190)	0.2859	(0.2533, 0.3190)	0.0473	(-0.0050, 0.1040)	0.5758	(-0.1902, 0.8358)
ILD5.6	2	Small	0.2028	(0.2021, 0.2034)	0	-	0.0454	(0.0240, 0.0670)	0.1995	(0.1752, 0.2234)	0.2449	(0.2158, 0.2737)	0	-	0	-
ILD5.6	2	Large	0.2022	(0.2016, 0.2027)	0.0111	(0.0001, 0.0222)	0	-	0.2133	(0.1898, 0.2370)	0.2133	(0.1898, 0.2370)	0.0519	(0.0007, 0.1082)	0.5204	(0.0605, 0.7370)
ILD5.6	3	Small	0.2035	(0.2028, 0.2041)	0.0206	(0.0005, 0.0410)	0.0053	(-0.0152, 0.0260)	0.2226	(0.1931, 0.2524)	0.2279	(0.1980, 0.2583)	0.0903	(0.0023, 0.1916)	0.7052	(0.1147, 0.9951)
ILD5.6	3	Large	0.2029	(0.2023, 0.2035)	0.0134	(-0.0015, 0.0281)	0	-	0.2268	(0.1969, 0.2574)	0.2268	(0.1969, 0.2574)	0.0590	(-0.0069, 0.1278)	0.5699	(-0.1934, 0.8258)
ILD5.6	4	Small	0.2025	(0.2019, 0.2031)	0.0033	(-0.0152, 0.0223)	0.0286	(0.0023, 0.0542)	0.2345	(0.2071, 0.2618)	0.2631	(0.2285, 0.2965)	0.0127	(-0.0547, 0.0916)	0.2855	(-0.6090, 0.7370)
ILD5.6	4	Large	0.2039	(0.2033, 0.2046)	0.0082	(-0.0128, 0.0292)	0.0263	(-0.0007, 0.0525)	0.2550	(0.2244, 0.2850)	0.2813	(0.2452, 0.3164)	0.0292	(-0.0434, 0.1119)	0.4443	(-0.5544, 0.8382)
ILD5.6	5	Small	0.2041	(0.2034, 0.2047)	0.0165	(-0.0002, 0.0331)	0	-	0.2584	(0.2256, 0.2916)	0.2584	(0.2256, 0.2916)	0.0640	(-0.0009, 0.1336)	0.6304	(-0.0771, 0.8915)
ILD5.6	5	Large	0.2039	(0.2032, 0.2045)	0.0181	(0.0003, 0.0359)	0.0089	(-0.0092, 0.0265)	0.1932	(0.1693, 0.2177)	0.2021	(0.1756, 0.2288)	0.0894	(0.0015, 0.1884)	0.6596	(0.0880, 0.9288)
ILD5.6	6	Small	0.2035	(0.2029, 0.2041)	0.0195	(0.0045, 0.0341)	0	-	0.2296	(0.2035, 0.2554)	0.2296	(0.2035, 0.2554)	0.0848	(0.0195, 0.1533)	0.6856	(0.3304, 0.9071)
ILD5.6	6	Large	0.2039	(0.2031, 0.2046)	0.0264	(0.0053, 0.0475)	0	-	0.2789	(0.2433, 0.3143)	0.2789	(0.2433, 0.3143)	0.0946	(0.0185, 0.1792)	0.7967	(0.3563, 1.0690)

Table S5. Results of regressing the *D. serrata* among-line variance estimates on either the squared trait mean or on the environmental variance. For each trait (defined in Figure 2B) the slope, R^2, *P*-value and 5% FDR adjusted *P*-value are reported. Only one relationship (ILD3.7 on mean, in bold) was significant at $P < 0.05$, but not following FDR multiple test correction. The analysis was conducted on the squared trait mean, rather than the square root of V_L (i.e., matching evolvability *I*, not *CV*) to allow visualization of both scalar relationships on the same graph (Figure 6).

Trait	Relationship	Slope (Beta)	R^2	*P*-value	FDR *P*
Size	Mean^2	-0.0001	0.2%	0.8952	0.9440
	V_E	0.0428	0.9%	0.7735	0.9440
ILD1.2	Mean^2	-0.0009	15.5%	0.2052	0.8333
	V_E	0.1842	11.5%	0.2810	0.8333
ILD1.5	Mean^2	0.0005	10.5%	0.3030	0.8333
	V_E	-0.1296	16.8%	0.1857	0.8333
ILD1.9	Mean^2	0.0002	0.2%	0.8836	0.9440
	V_E	0.1050	6.4%	0.4278	0.8766
ILD2.5	Mean^2	0.0000	0.1%	0.9440	0.9440
	V_E	-0.0043	0.1%	0.9318	0.9440
ILD2.8	Mean^2	-0.0003	7.2%	0.3980	0.8766
	V_E	0.0962	4.3%	0.5180	0.8766
ILD3.7	Mean^2	-0.0013	59.5%	**0.0033**	0.0726
	V_E	0.4493	32.4%	0.0533	0.5863
ILD3.8	Mean^2	0.0001	4.7%	0.5006	0.8766
	V_E	0.1302	11.5%	0.2812	0.8333
ILD4.5	Mean^2	<0.0000	0.1%	0.9357	0.9440
	V_E	0.0054	0.1%	0.9342	0.9440
ILD4.8	Mean^2	-0.0001	5.1%	0.4808	0.8766
	V_E	0.0179	0.4%	0.8412	0.9440
ILD5.6	Mean^2	0.0009	17.6%	0.1741	0.8333
	V_E	-0.0273	1.0%	0.7533	0.9440

Table S6. Statistical support for the effect of *Drosophila serrata* population size treatment on the mutations segregating within and among sub-lines. Log-likelihood ratio tests were used to test two hypotheses: whether the within-line variance differed in magnitude between the two population size treatments (Heterogeneous within-line) and; whether the mutational correlation between the treatments differed from 1 (Correlation). The test statistic (X^2) follows a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters between these models (Heterogeneous within-line: d.f. = 2; Correlation < 0.999: d.f. = 1). The *P*-value is shown, and for the Heterogeneous within-line test, the 5% FDR corrected *P*-value. Following FDR correction, there was only one instance of statistically supported difference between the treatments, highlighted in bold. No FDR correction was applied to Correlation < 0.999 as all $P > 0.05$. The among-line (mutational) correlation between the Small and Large treatments ($r_{S,L}$) is shown and should be interpreted within the context of the corresponding variances (Table S4; Figure 5). The per-treatment estimates of V_E from the model analysed here (5) were indistinguishable from the estimates from model (2) (Table S4) and are not re-reported here.

Trait	Generation	Heterogeneous within-line				Correlation	
		X^2	*P*-value	FDR *P*	$r_{S,L}$	X^2	*P*-value
CS	1	11.40	0.0033	0.0552	0.70	0.70	0.7047
	2	3.50	0.1738	0.7168	0.78	1.10	0.5769
	3	1.00	0.6065	0.9403	0.82	0.40	0.8187
	4	12.00	0.0025	0.0545	1	0.00	1.0000
	5	20.10	<0.0001	**0.0029**	0.44	2.40	0.3012
	6	5.30	0.0707	0.4663	0.63	1.50	0.4724
ILD1.2	1	2.90	0.2346	0.7741	0.93	0.00	1.0000
	2	0.70	0.7047	0.9403	0.76	1.00	0.6065
	3	2.00	0.3679	0.9403	1	0.00	1.0000
	4	0.30	0.8607	0.9659	0.86	0.40	0.8187
	5	4.60	0.1003	0.5914	0.80	0.00	1.0000
	6	2.10	0.3499	0.9403	0.68	0.00	1.0000
ILD1.5	1	0.90	0.6376	0.9403	0.96	0.20	0.9048
	2	0.80	0.6703	0.9403	1.00	0.00	1.0000
	3	1.20	0.5488	0.9403	1	0.00	1.0000
	4	1.20	0.5488	0.9403	0.87	0.60	0.7408
	5	0.80	0.6703	0.9403	0.65	2.10	0.3499

	6	0.90	0.6376	0.9403	0.95	0.20	0.9048
ILD1.9	1	0.00	1.0000	1.0000	1	0.00	1.0000
	2	0.10	0.9512	0.9659	1	0.00	1.0000
	3	0.90	0.6376	0.9403	1	0.00	1.0000
	4	0.40	0.8187	0.9659	1	0.00	1.0000
	5	1.60	0.4493	0.9403	0.96	0.00	1.0000
	6	1.90	0.3867	0.9403	0.65	0.80	0.6703
ILD2.5	1	0.10	0.9512	0.9659	-0.04	2.00	0.3679
	2	5.40	0.0672	0.4663	1	0.00	1.0000
	3	0.10	0.9512	0.9659	0.88	0.00	1.0000
	4	10.00	0.0067	0.0741	0.47	0.70	0.7047
	5	3.20	0.2019	0.7373	1	0.00	1.0000
	6	0.60	0.7408	0.9403	1	0.00	1.0000
ILD2.8	1	1.90	0.3867	0.9403	1	0.00	1.0000
	2	0.80	0.6703	0.9403	0.45	3.90	0.1423
	3	3.10	0.2122	0.7373	0.39	1.50	0.4724
	4	1.10	0.5769	0.9403	1	0.00	1.0000
	5	4.30	0.1165	0.5914	1	0.00	1.0000
	6	2.30	0.3166	0.9403	0.99	0.00	1.0000
ILD3.7	1	1.70	0.4274	0.9403	0.84	0.80	0.6703
	2	1.50	0.4724	0.9403	0.97	0.00	1.0000
	3	0.10	0.9512	0.9659	1	0.00	1.0000
	4	1.20	0.5488	0.9403	0.59	1.10	0.5769
	5	1.00	0.6065	0.9403	0.97	0.00	1.0000
	6	0.30	0.8607	0.9659	1	0.00	1.0000
ILD3.8	1	7.20	0.0273	0.2491	1	0.00	1.0000
	2	0.30	0.8607	0.9659	1	0.00	1.0000
	3	0.60	0.7408	0.9403	0.88	0.10	0.9512
	4	1.50	0.4724	0.9403	1	0.00	1.0000
	5	2.10	0.3499	0.9403	1	0.00	1.0000
	6	0.30	0.8607	0.9659	0.65	0.50	0.7788
ILD4.5	1	1.10	0.5769	0.9403	0.54	0.90	0.6376
	2	0.60	0.7408	0.9403	0.60	2.50	0.2865
	3	4.40	0.1108	0.5914	0.54	1.00	0.6065
	4	0.40	0.8187	0.9659	1	0.00	1.0000
	5	12.50	0.0019	0.0545	1	0.00	1.0000
	6	1.50	0.4724	0.9403	0.43	3.10	0.2122
ILD4.8	1	0.60	0.7408	0.9403	0.38	1.00	0.6065
	2	3.20	0.2019	0.7373	1	0.00	1.0000
	3	0.60	0.7408	0.9403	1	0.00	1.0000
	4	3.70	0.1572	0.7168	1	0.00	1.0000
	5	0.10	0.9512	0.9659	1	0.00	1.0000
	6	0.10	0.9512	0.9659	0.61	0.10	0.9512
ILD5.6	1	1.90	0.3867	0.9403	0.78	0.30	0.8607

2	10.70	0.0047	0.0627	1	0.00	1.0000
3	0.20	0.9048	0.9659	0.86	0.10	0.9512
4	0.70	0.7047	0.9403	1	0.00	1.0000
5	7.00	0.0302	0.2491	-0.34	3.90	0.1423
6	3.60	0.1653	0.7168	0.54	2.50	0.2865

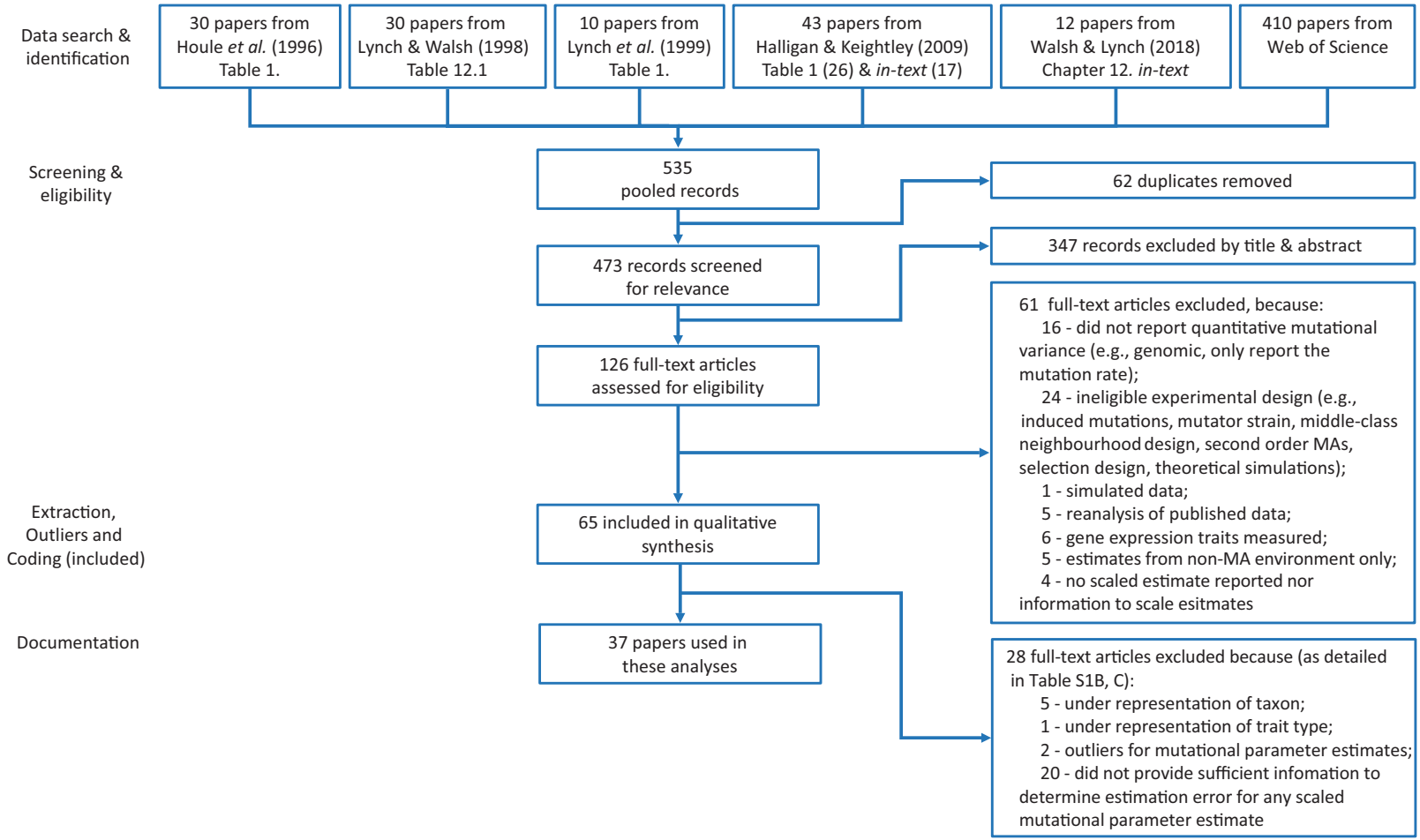


Figure S1. Diagram of systematic literature search for the meta-analysis of mutational variance estimates from mutation accumulation (MA) experiments. The schematic follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis)[1]

[1] Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman and P. Group, 2009 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Medicine 6: e1000097.

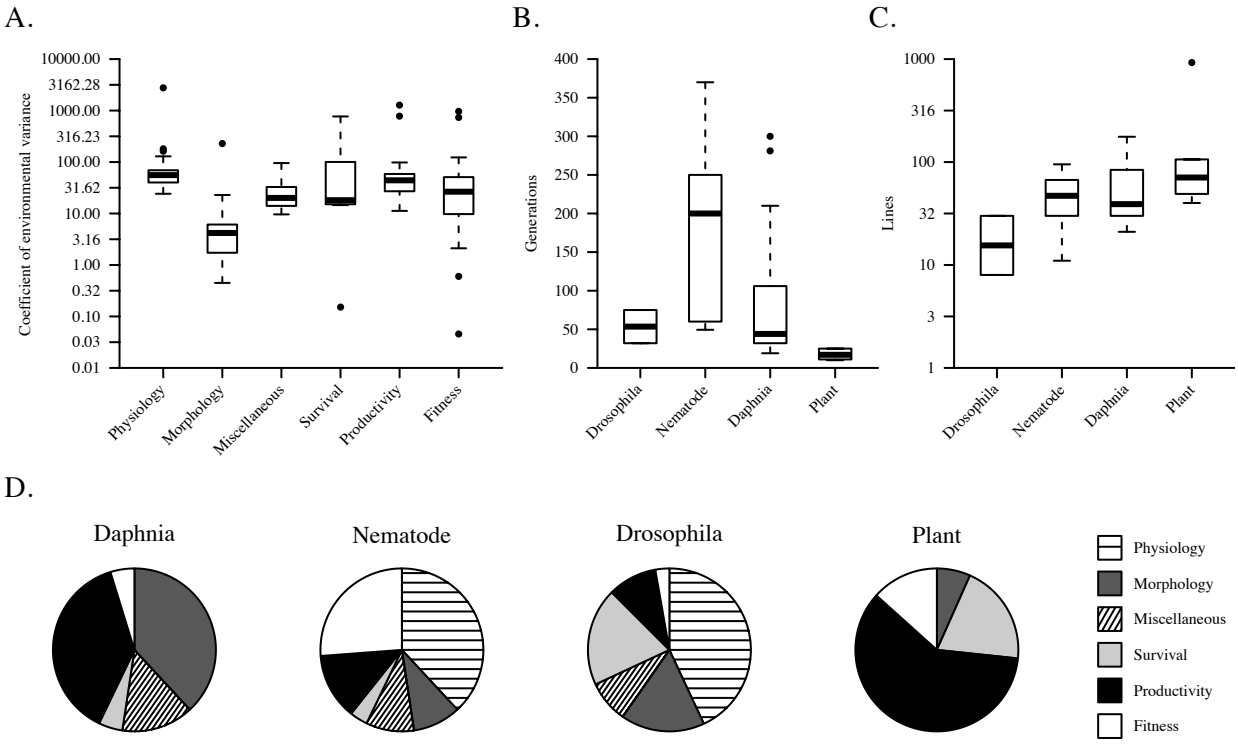


Figure S2. Variation of factors that may influence the magnitude of mutational variance estimates. (A) CV_E was calculated as[1]: $CV_E = \sqrt{V_E}/\bar{X}$ and log-transformed to improve resolution for plotting (NB: the values are shown in the y-axis were back-transformed to the original CV_E scale). CV_E was plotted per trait category, which are defined in Methods and Table S2. (B) the number of generations and (C) the (log-transformed; but as in (A) the y-axis is shown on the original scale) number of lines of mutation accumulation experiments (studies) in different taxon categories. Taxon categories are defined in Methods. Box plots represent the median (bar), interquartile range (IQR, box) and 3 IQR (whiskers) values of the published estimates, with more extreme values indicated by solid circles. (D) The distribution of estimate numbers across the different trait categories within each taxon group, where traits definitions as per Table S2.

[1] Houle, D., B. Morikawa and M. Lynch, 1996 Comparing mutational variabilities. Genetics 143: 1467-1483

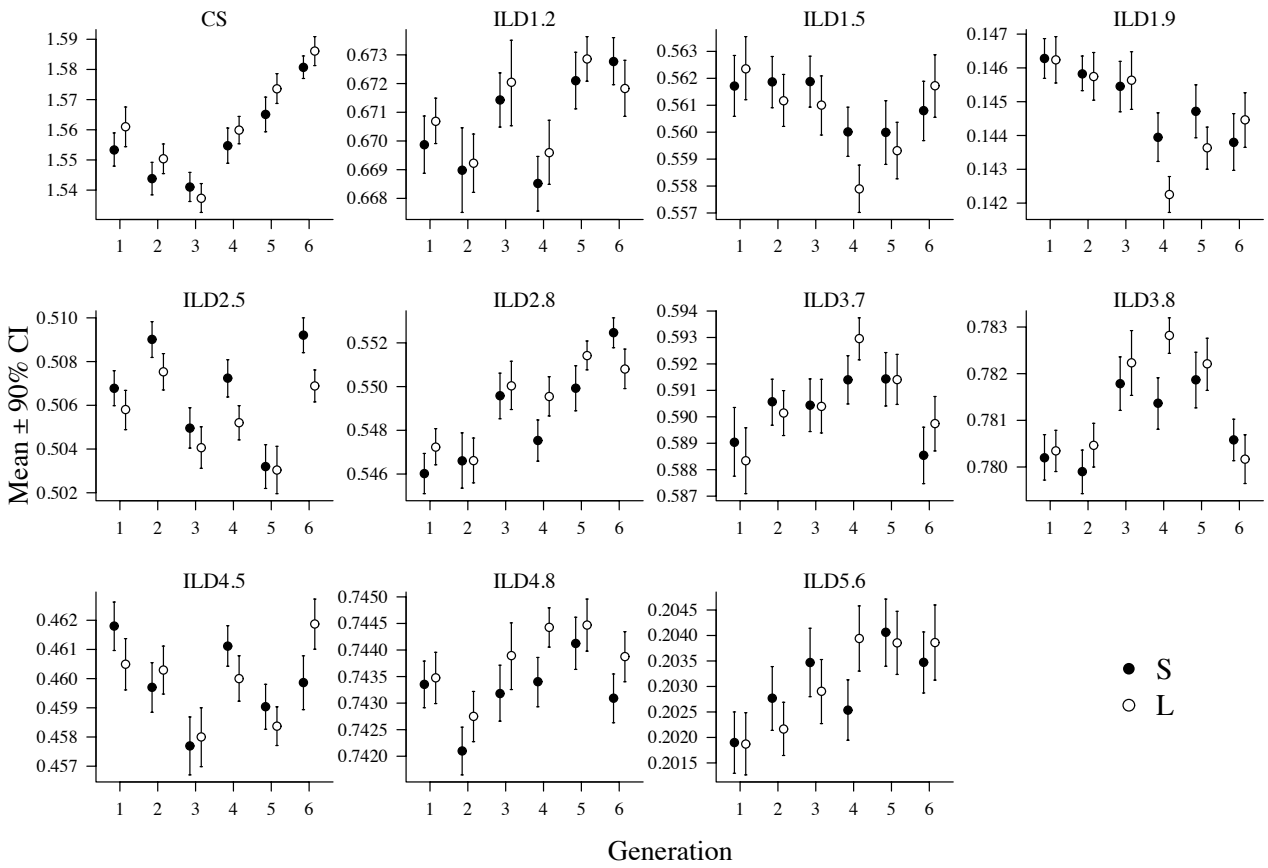


Figure S3. Variation in wing trait mean across six generations of an experiment in *Drosophila serrata.* For each of the 11 wing traits (panels), the least-squares mean (± 95% CI) from model (2) is plotted for each generation for the Small (solid circle) and Large (open circles) population size treatments. Centroid size (CS) is measured in millimetres; the ten inter-landmark distances are in units of centroid size (multiplied by 100; see Figure 2B for trait definitions). Plotted estimates are reported in Table S4.

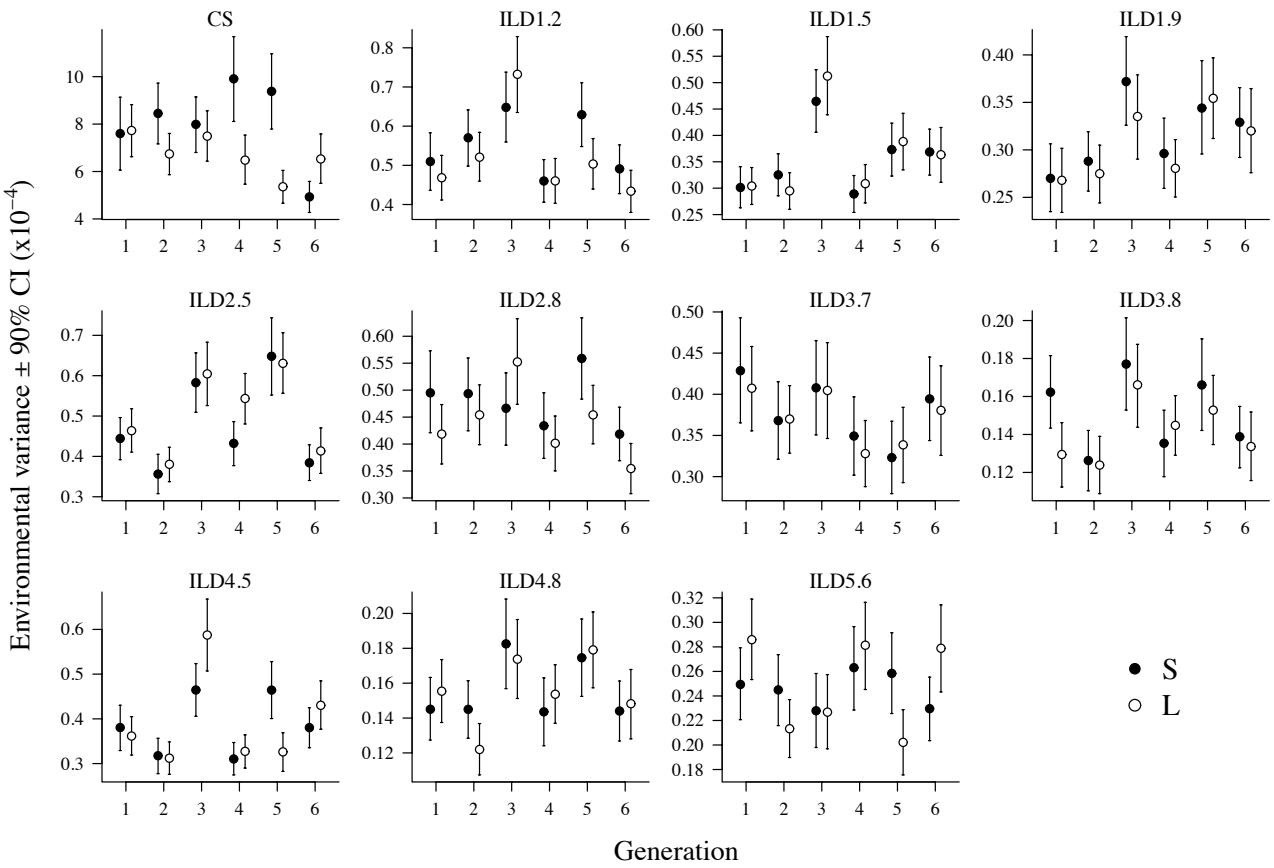


Figure S4. Variation in environmental variance across six generations of an experiment in *Drosophila serrata.* For each of the 11 wing traits (panels), V_E (± 95% CI), estimated as the sum of among vial and residual variances from model (2), is plotted for each generation for the Small (solid circle) and Large (open circles) population size treatments. Plotted estimates are reported in Table S4.

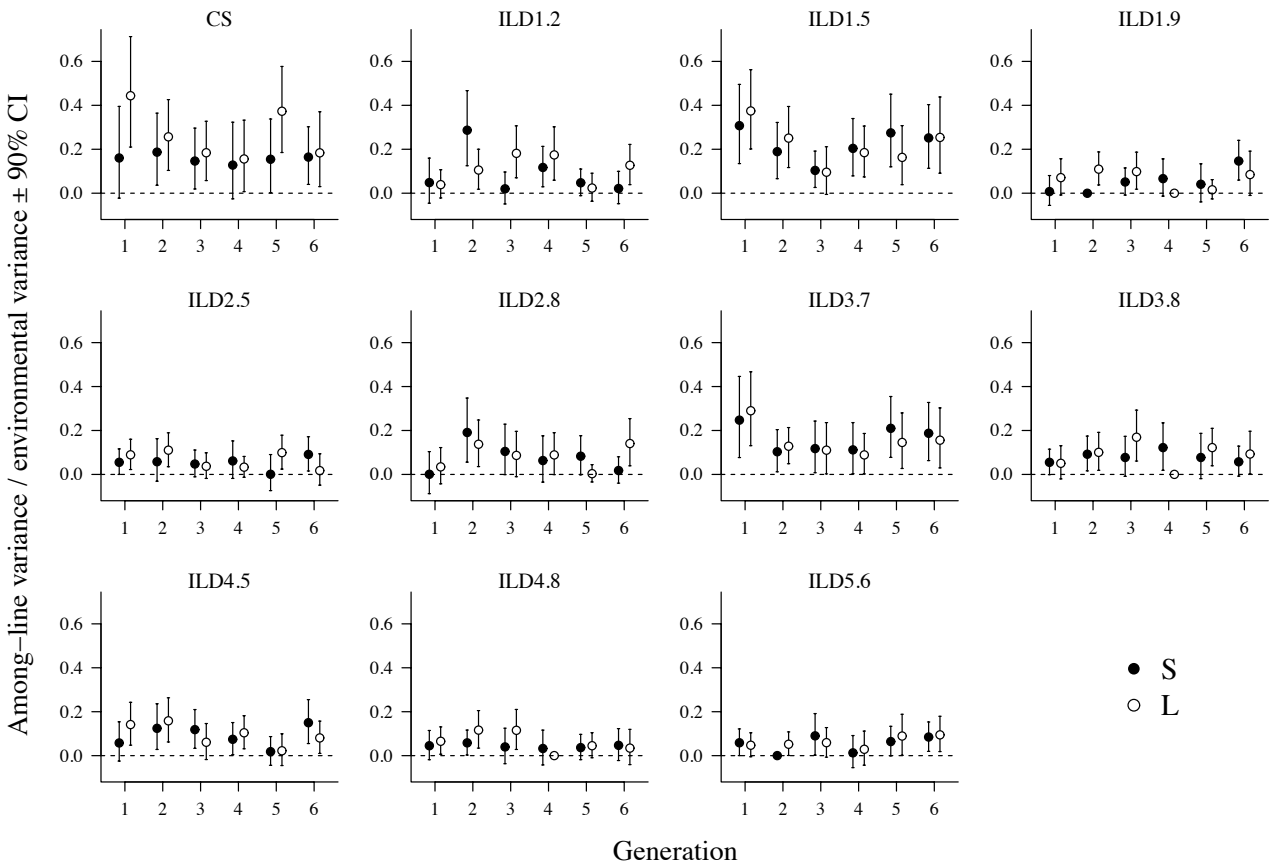


Figure S5. Variance scaled estimates of among-line variance estimates in *D. serrata*. Among-line variance estimates (V_L, Figure 5) were placed on a heritability scale to account for the influence of trait variance on the magnitude of estimates, dividing V_L by the corresponding estimate of environmental variance (summed among-vial and residual variance reported in Table S4). Plotted are these variance-scaled REML estimates (90% CI) for each trait (panel) (trait descriptions in Figure 2B) in each generation (x-axis) for each of the two population size treatments (Small: solid circles; Large: open circles). The dashed horizontal line indicates zero.

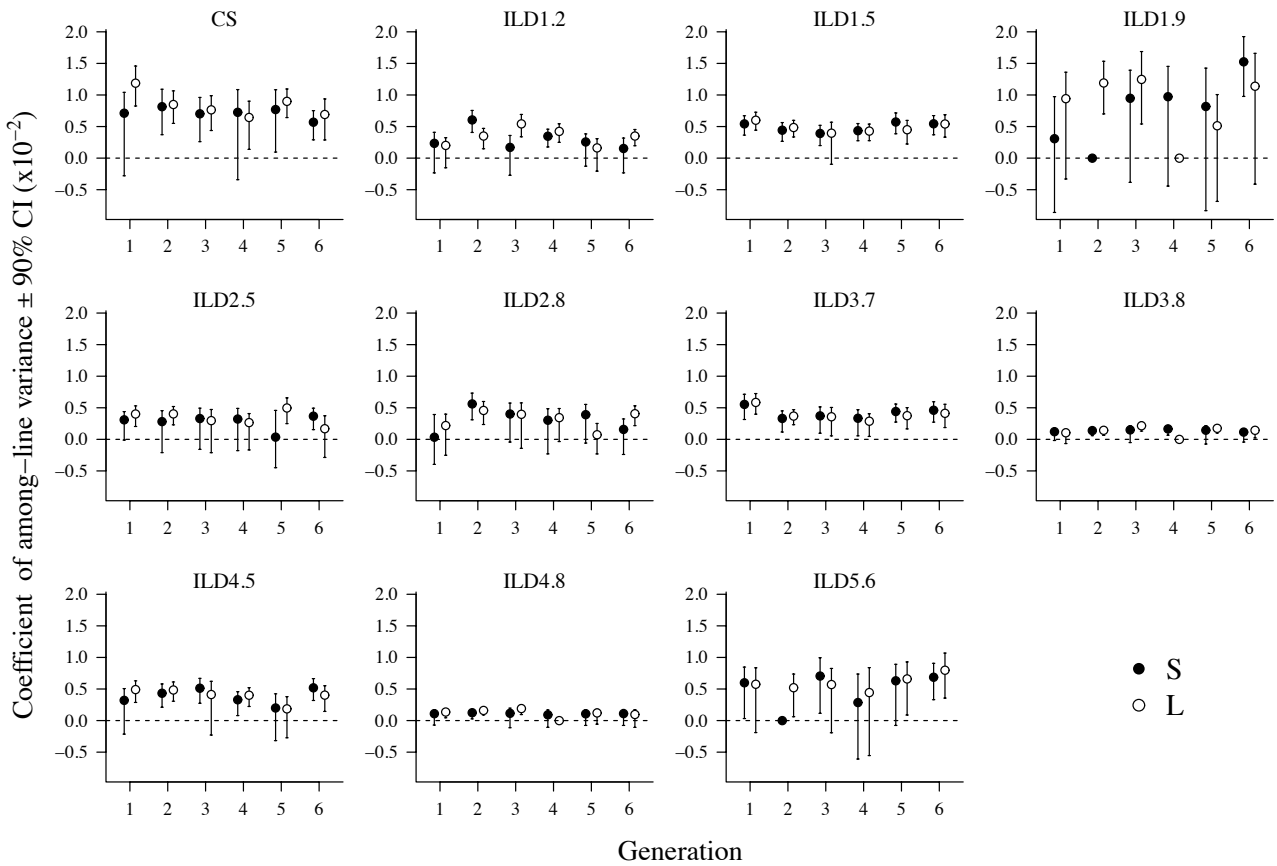


Figure S6. Mean-scaled estimates of among-line variance estimates in *D. serrata*. Among-line variance estimates (V_L, Figure 5) were placed on a coefficient of variance scale to account for the influence of trait mean on magnitude of estimates, using the following formula: $100*\frac{\sqrt{V_L}}{\bar{X}}$ where $\bar{X}$ is the trait mean (Table S4). Plotted are the REML point estimates (and 90% CI) for each trait (panel) (see Figure 2B for trait descriptions) in each generation (x-axis) for each of the two population size treatments (Small: solid circles; Large: open circles). The dashed horizontal line indicates zero.

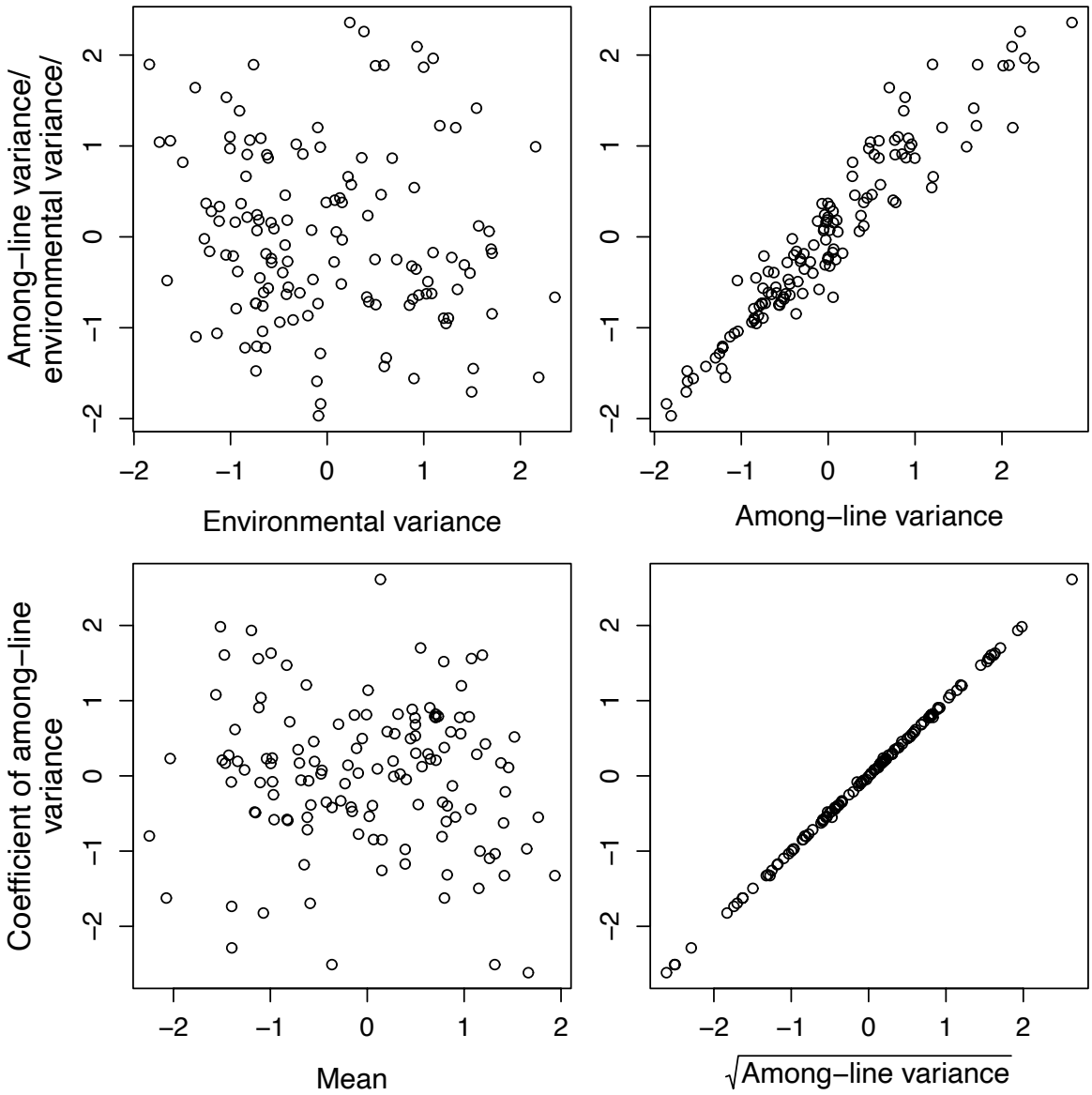


Figure S7. The relative contributions of heterogeneity in the scaling parameter (V_E or trait mean; left panels) and the among-line variance (V_L; right panels) to the variation observed in the variance-scaled (top panels) or mean-scaled (bottom panel) estimates of *Drosophila serrata* among-line variance. All estimates are presented in Table S4, V_L is also presented in Figure 5; V_E and mean in Figures S3 and S4, respectively; variance or mean scaled estimates of V_L in Figures S5 and S6, respectively. The 12 estimates for each of the 11 traits for each parameter (V_L, V_E, mean, variance-scaled and mean-scaled) were first placed on a standardised scale by centring on the mean of the 12 estimates, and dividing by the standard deviation of the 12 estimates (i.e., z-scores were calculated), and all 132 estimates (12 estimates per 11 traits) are plotted.

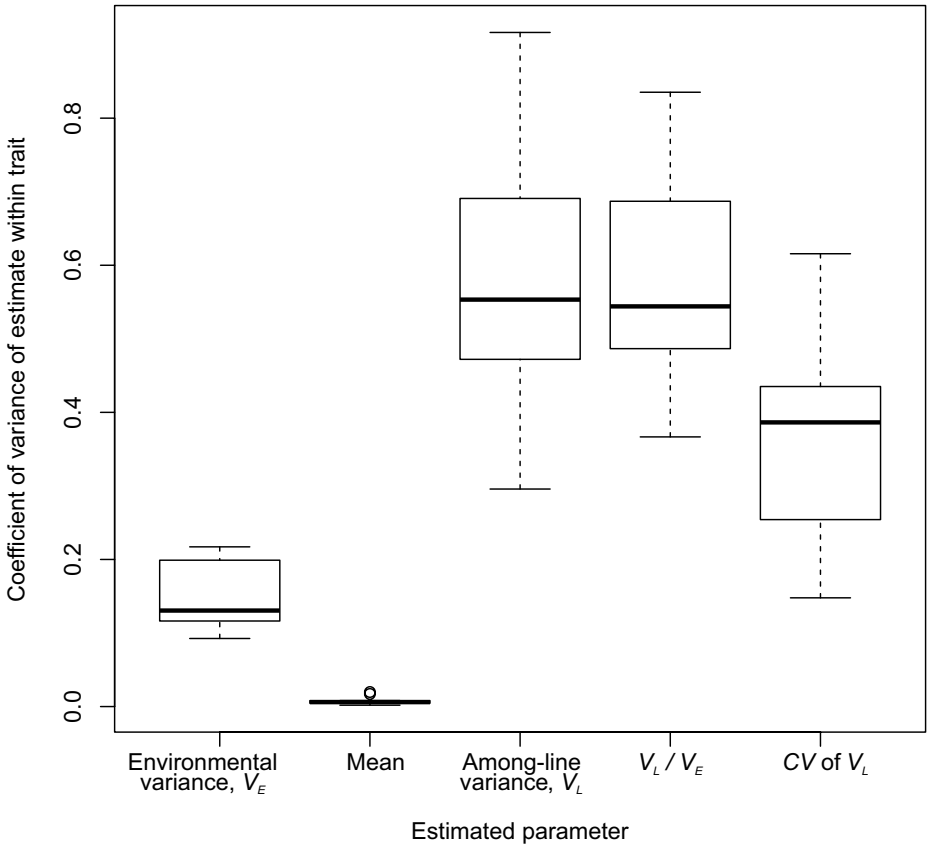


Figure S8. Variation in estimates of each genetic variation parameter in *D. serrata*. For each of these parameters, the 12 estimates per trait (Table S4) were analysed to calculate the coefficient of variance across the 12 observations as: standard deviation / mean. Box plots represent the median (bar), interquartile range (IQR, box) and 3 IQR (whiskers) values for the 11 traits.