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, micro-environment, measurement 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

Data will be made publicly available upon acceptance of the manuscript.

ABSTRACT

Characteristics of the new phenotypic variation introduced via mutation have broad implications in evolutionary and medical genetics. While many estimates of this mutational variance have been published, factors contributing to their range, spanning two orders of magnitude, remain poorly resolved. In this study, we first apply a meta-analytic approach to assess how previously implicated factors affect variability in estimates. While estimates of mutational variance are available from a range of taxa, and quantitative trait type over a range of timescales, these factors are confounded with one another, precluding a clear resolution of causes of observed variability. We call for further directly comparable empirical data to address this question. We then applied a modified mutation accumulation experimental design to generate independent repeated estimates of mutational variance for a single taxon (the vinegar fly, *Drosophila serrata*), and for a single trait type (wing morphology) under the same experimental conditions. Analyses revealed that, in this tightly controlled experiment, variability in among-line (mutational) variance was largely the consequence of sampling error. Micro-environmental variation in mutational effects was supported as a cause of low levels of variability in mutational variance for two of 11 traits analysed. Further, there was no evidence that variation in segregating mutations, as the realisation of the mutation-drift process, impacted estimates. This investigation demonstrates the utility of short-term repeated measures to be broadly applied to improve estimates of mutational variance, consequently expanding our understanding of the dynamics of mutations in natural populations.

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 1988b; BARTON AND TURELLI 1989; JOHNSON AND BARTON 2005). Mutational variance is determined by both the mutation rate and the average effects of new mutations across all contributing loci (KIMURA 1965; LATTER 1970; LANDE 1975). Much of our understanding of V_M comes from mutation accumulation (MA) experiments, where populations diverge phenotypically due solely to the fixation of new mutations (MUKAI 1964; HALLIGAN AND KEIGHTLEY 2009). Estimates from a wide range of traits and taxa indicate that mutation increases the variance in quantitative traits by 10^{-4} to 10^{-2} times the environmental variance of the trait per generation (HOULE *et al.* 1996; LYNCH AND WALSH 1998; LYNCH *et al.* 1999; HALLIGAN AND KEIGHTLEY 2009). This range of values (two orders of magnitude) suggests variation in either the mutation rate or the average magnitude of mutational effects, leading to differences in the magnitude of standing quantitative genetic variation and, ultimately, rates of phenotypic evolution (HOULE 1998; LYNCH *et al.* 1999; HOULE *et al.* 2017; WALSH AND LYNCH 2018). However, the causes of variation in estimates of V_M, and thus the evolutionary interpretation of this variability, are not well resolved.

A relatively simple potential cause of variation in V_M estimates is scale. HOULE *et al.* (1996) showed that life history traits have lower mutational heritability ($h^2_M = V_M / V_E$, where V_E is the environmental variance) than morphological traits, but the reverse pattern holds when comparing coefficients of mutational variance (CV_M; $100 \times \sqrt{V_M}/\bar{X}$, where $\bar{X}$ is the trait mean). While the pattern is reversed on the different scales, the variation in the magnitude of mutational variance between trait types nonetheless persists. Life-history traits are thought to be affected by more loci than morphological traits, with this larger target of mutation resulting in a higher mutation rate (HOULE 1991; HOULE 1992; HOULE *et al.* 1996; MERILÄ AND SHELDON 1999). Reviewing 57

67 published estimates, Houle *et al.* (1996) found a strong correlation between the magnitude of
68 mutational variance and a metric of mutational target size. However, the exact role of mutational
69 target size plays in generating variability of mutational variance is difficult to assess, as a full
70 catalogue of causal loci is unknown for most quantitative traits (MACKAY 2001; BARTON AND
71 KEIGHTLEY 2002; YANG *et al.* 2010; ROCKMAN 2012).

72 Mutation rate varies widely among taxa (reviewed in KATJU AND BERGTHORSSON 2019),
73 which may reflect differences in ploidy levels, genome size, generation time, and/or effective
74 population size (LYNCH *et al.* 1999; LYNCH 2010; SUNG *et al.* 2012). Marked variation in mutation
75 rate has also been observed within species, both among replicated MA experiments (i.e., different
76 founder genotypes: KEIGHTLEY *et al.* 2009; NESS *et al.* 2012; SUNG *et al.* 2012; SCHRIDER *et al.*
77 2013) and among lines within a single MA panel (HUANG *et al.* 2016). V_{M} estimates vary within
78 taxon and trait. For example, h^2_{M} of body size in *Caenorhabditis elegans* ranged from 0.0035–0.0076
79 across two different strains (genetic isolates) (N2 and PB306; OSTROW *et al.* 2007), and from
80 0.0011–0.0044 among four sets of MA lines established from the same strain (AZEVEDO *et al.* 2002;
81 ESTES *et al.* 2005; OSTROW *et al.* 2007). The role of variability in mutation rate to this variability is
82 unclear. Notably, different combinations of rate and effect size apparently cause highly similar
83 h^2_{M} values (Table 1 in HALLIGAN AND KEIGHTLEY 2009), suggesting the potential for complex
84 relationships between allelic effect sizes and genomic mutation rates.

85 Ultimately, the random nature of mutation-drift processes themselves might contribute to
86 variability of estimates, where it is the distribution of estimates obtained from multiple experiments
87 that is most informative of the expected input of mutational variance. The extent to which
88 experimental error or biologically relevant sampling variance contribute to variability in estimates
89 might be revealed by further consideration of the MA experimental design. Firstly, sampling the
90 mutation-drift process using a finite number of lines will yield variation in estimates, representing
91 variation in the realisation of the mutation-drift process. LYNCH (1988a) recommended that the

92 number of replicate lines (L) should be greater than 400 for selfing and 300 for full-sib lines (the two
93 most common modes in bottlenecking in MA designs) to ensure less than 10% error in the estimate
94 of among-line variance. However, maintaining a large number of lines simultaneously is often
95 logistically prohibitive, and this recommendation is rarely met, with experiments typically
96 considering far fewer (< 50) lines (see Methods).

97 Secondly, the timeframe over which mutations accumulate might influence estimates of
98 mutational variance. If the base population was homozygous (heterozygous) when MA lines were
99 established, estimates of V_M will be downwardly (upwardly) biased before $6N_e$ generations (LYNCH
100 AND HILL 1986). However, V_M is typically estimated after more than $6N_e$ generations (see Methods),
101 suggesting little potential for bias due to variation in ancestral segregating variation. Long-running
102 MA experiments might also under-estimate mutational variance, as the cumulative effect of low
103 fitness alleles can cause line extinction, or within-line selection against further accumulation (LYNCH
104 *et al.* 1999; ESTES *et al.* 2004; MCGUIGAN AND BLOWS 2013). While a decline in the magnitude of
105 V_M over time has been observed (MACKAY *et al.* 1995), other studies have reported consistent
106 estimates through time (GARCÍA-DORADO *et al.* 2000). Apart from consistent temporal trends,
107 transient within-line segregation of mutations might generate variability in V_M estimates from
108 generation to generation as a consequence of the mutation-drift process (LYNCH AND HILL 1986).

109 Environmental context within which MA lines are assayed could also contribute variation to
110 estimates of V_M. Several studies have considered the effect of macro-environmental variation (i.e.,
111 known, replicable, experimenter-imposed, variation), including temperature (WAYNE AND MACKAY
112 1998), light (KAVANAUGH AND SHAW 2005), density (FRY *et al.* 1996) or food availability (LATTA *et*
113 *al.* 2015). Although the magnitude of mutational variance often varies among environments, there is
114 only weak evidence for predicable patterns, such as novel or stressful conditions increasing the
115 magnitude of variance (KONDRASHOV AND HOULE 1994; MARTIN AND LENORMAND 2006). Even in
116 carefully controlled laboratory experiments, micro-environmental conditions (due to uncontrolled,

117 un-replicated, and typically unknown sources) vary among replicates and through time, affecting
118 phenotypic variation (LYNCH 1988a; FALCONER AND MACKAY 1996). If all MA lines respond
119 similarly, micro-environmental variation will affect trait mean, but not among-line variance; if lines
120 differ in their response, then this genotype by micro-environment variance could contribute
121 undetected variability to estimates of V_M (LYNCH 1988a; GARCÍA-DORADO *et al.* 2000). The
122 magnitude of inter-generation variation of among-line estimates has rarely been assessed (GARCÍA-
123 DORADO *et al.* 2000) and the relative contributions to any such variation from micro-environment,
124 measurement error and segregation remain unknown.

125 Here, we combined two approaches to investigate causes of variability in estimates of
126 mutational variance. Given that it has been twenty years since (1996); 1998) addressed variability in
127 mutational variance, we first conducted a meta-analysis to update tests of previously implicated
128 causal factors, specifically taxon (LYNCH AND WALSH 1998; LYNCH *et al.* 1999; HALLIGAN AND
129 KEIGHTLEY 2009), trait type (HOULE *et al.* 1996; HOULE 1998) and experimental duration (generation
130 number) (LYNCH AND HILL 1986). Multiple estimates of V_M were available for different values of
131 these factors, allowing us to test the hypothesised effect. In contrast, there were insufficient studies of
132 the same taxon, trait and macro-environment for us to utilise meta-analysis to investigate the effects
133 of segregation, micro-environment and sampling error. Therefore, we conducted a new empirical
134 experiment, assaying 42 MA lines from *Drosophila serrata* after 20 generations of brother-sister
135 mating. This sample size (42) was below that required to minimise sampling error (LYNCH 1988a)
136 but is typical of published MA experiments, allowing us to consider sampling variance comparable
137 with the literature. We sampled our phenotype of interest, the well-studied *Drosophila* wing, every
138 generation for six generations, allowing us to quantify variability in V_M estimates. We used a
139 population size manipulation to allow us to infer the relative influence on V_M of micro-environment
140 and measurement error versus idiosyncratic variation in evolutionary processes.
141

142 **METHODS**

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

144 *Literature search*

145 We extracted all papers from seven articles that have assembled and reviewed empirical estimates of
146 mutational variance published between 1942 and 2014: LYNCH (1988b); KEIGHTLEY *et al.* (1993);
147 HOULE *et al.* (1996); LYNCH AND WALSH (1998); LYNCH *et al.* (1999); HALLIGAN AND KEIGHTLEY
148 (2009); WALSH AND LYNCH (2018). We then searched the Web of Science database on the
149 11/12/2019 at 4:38 PM AEST for *journal article* document types meeting the topic criteria of
150 "mutation accumulation" and (varia* or "mutat* coefficient*") and published between 1998 and
151 2019. This time range overlapped HALLIGAN AND KEIGHTLEY (2009) (which considered fitness traits
152 only) and WALSH AND LYNCH (2018) (which only gives a brief in-text update on their 1998 review),
153 allowing us to capture papers potentially missed in previous reviews.

154 Further details on the papers identified and preliminary handling steps can be found in Figure
155 S1. From the 473 unique papers identified, we screened titles and abstracts, then the full text, for
156 relevance, using four strict criteria: 1) quantitative mutational variance was reported; 2) estimates
157 were from spontaneous mutation accumulation experiments; 3) estimates were made under the same
158 macro-environmental conditions as mutations were accumulated and; 4) estimates were not re-
159 reporting of previously published estimates. We excluded six studies that analysed transcriptomic
160 data due to the extremely large number of traits relative to other trait categories. We also excluded
161 mitotic cell division traits, examined in only one study (FARHADIFAR *et al.* 2016), cases where V_M
162 estimates were not reported on either a heritability or coefficient of variance scale nor the information
163 to calculate these (i.e., the environmental variance or trait mean) provided and, finally, studies where
164 the number of mutation accumulation generations was not reported. After applying these criteria, 69
165 papers were retained.

166 *Meta-analysis data collection*

167 For each of the 69 papers, mutational parameter estimates were extracted, associated with the identity
168 of the taxon and trait, number of generations of mutation accumulation, and the number of lines
169 sampled. Where studies reported estimates from multiple environments, only estimates from the MA
170 maintenance environment were extracted. Papers reported estimates of mutational variance for a
171 median of three traits (range: 1–29). Where possible, we extracted (or calculated from provided
172 information) both the coefficient of variance (CV_M; $100 \times \sqrt{V_M}/\bar{X}$, where $\bar{X}$ is the trait mean) and
173 mutational heritability (h^2_M; V_M/V_E) for each trait. As detailed in Table S1, after further inspection we
174 excluded extreme estimates, and those reported with very low precision (Table S1.B).

175 Before proceeding with the meta-analyses, we transformed estimates to improve normality,
176 applying a $\log_{10}$ transformation to h^2_M estimates and a Box-Cox transformation (BOX AND COX 1964)
177 to CV_M estimates. For the Box-Cox transformation, using the MASS (VENABLES AND RIPLEY 2002)
178 package in R and maximum-likelihood, $\lambda = 0.2$ was determined as the best-fit power term for this
179 data. Both the number of generations and the number of lines sampled were log-transformed to a
180 normal distribution, and an extreme outlier (> 8 SD) excluded (Table S1B).

181 *Predictor variables for the meta-analysis*

182 Estimates came from 22 species, and based on the distribution of estimates, we defined six taxon
183 categories: Drosophila (*Drosophila melanogaster* [n = 89] and *D. serrata* [n = 15]), plant
184 (*Arabidopsis thaliana* [five studies, n = 15]; cumulatively three studies and n = 15 from *Amsinckia*
185 *douglasiana*, *A. gloriosa*, *Hordeum vulgare* and *Zea mays*), microbe (*Cryptococcus neoformans*,
186 *Escherichia coli*, *Chlamydomonas reinhardtii*, *Saccharomyces cerevisiae*, *Cystovirus pseudomonas*
187 *virus* φ6 and *Vesicular stomatitis virus*), nematode (*Caenorhabditis briggsae*, *C. brenneri*, *C.*
188 *elegans*, *C. remanei* and *Oscheius myriophila*), and Daphnia (*Daphnia pulex*) (Figure 1A). We
189 excluded from analyses estimates from two other insects (*Daktulosphaira vitifoliae* and *Nasonia*

190 *vitripennis*) for clarity of inference (insect versus *Drosophila*), and from *Mus musculus*, the only
191 vertebrate, due to low sample size (four estimates: Figure 1A).

192 HOULE *et al.* (1996; 1998) contrasted life-history, morphology and growth traits, which were
193 expected to differ in mutational target size and sensitivity to micro-environmental variation. We
194 differ from HOULE *et al.* (1996; 1998) in considering weight and height of juveniles as morphological
195 (not growth) traits. Within the morphology category (20% of all estimates), 76% of estimates were
196 for size-related traits. Reflecting more recent publications, we defined a physiology category (28% of
197 all estimates), which included locomotive, enzymatic and metabolic activity traits, which may differ
198 from life-history or morphological traits in mutational target size or micro-environmental sensitivity.
199 Life history traits accounted for 53% of estimates, prompting us to subdivide life history traits into
200 relatively evenly represented narrower sub-categories: total fitness (either measured directly, or
201 reported based on combined data from individual components), survival, productivity, development
202 time and a miscellaneous category (capturing traits such as phenology, longevity and mating success,
203 which were individually less well represented) (Figure 1B; Table S2).

204 *Meta-analyses of mutational variance estimates*

205 Our final meta-analysis data set consisted of 201 estimates of h^2_{M} and 211 estimates of CV_{M} from 62
206 studies. To investigate causes of variability among these published estimates, we implemented the
207 following model (separately for h^2_{M} and CV_{M} data) via PROC MIXED in SAS v.9.4 (SAS Institute
208 Inc., Cary, NC.), using maximum likelihood:

209 $$y_{ijklmn} = \mu + tx_i + trt_j + g_k + l_l + s_m + \varepsilon_{ijklmn} \quad (1)$$

210 where y is the vector of published estimates, and μ was the overall mean mutational variance. We fit
211 the categorical predictors taxon (*tx, i* =1, 2, …, 5) and trait (*trt, j*=1, 2, …, 7), and the continuous
212 predictors generation number (*g*) and number of lines sampled (*l*) as fixed effects. The number of
213 MA lines and generations spanned different (non-overlapping) ranges among taxa (Figure S2), and

214 interactions were therefore not fit. Study (s) was fit as a random effect, accounting for non-
215 independence among estimates within a paper. The same MA lines were sometimes analysed in
216 multiple studies, for different traits, or after further generations of accumulation (Table S3).
217 However, an additional random effect of shared lines (with study nested within this) did not account
218 for any variance, potentially a consequence of the unbalanced design, with only some studies sharing
219 lines. The residual error (ε) captured all unexplained variation among estimates. We used the
220 Satterthwaite approximation to correct the denominator degrees of freedom.

221 **Modified mutation accumulation experiment to address variability in mutational variance**

222 *Population history*

223 We further investigate variation in estimates of V_M via an empirical experiment in *Drosophila*
224 *serrata*. Offspring of an inseminated female, collected from Brisbane, Australia were subjected to
225 single-pair brother-sister inbreeding for 20 generations to establish a genetically homogeneous
226 isofemale line, as described in REDDIEX *et al.* (2018). After inbreeding, this line was maintained as a
227 mass-bred population (~300 individuals) for 96 generations. The population was then expanded over
228 three generations of random mating to allow 200 MA lines to be established. The MA lines were
229 maintained by single brother-sister pairs under standard conditions at (25°C, 12L:12D light regime,
230 sugar-yeast-agar media, as described in RUNDLE *et al.* (2005).

231 *Mutation accumulation with varying N_e treatments*

232 Starting from this set of MA lines after they had accumulated mutations over 20 generations (~5N_e
233 generations), we established a new experimental design consisting of two population size treatments,
234 a small population treatment (S; N = 24 flies), and a big population treatment (B; N = 288 flies). A
235 subset of 42 MA lines, randomly chosen from the surviving MA lines, were expanded over two (S) or
236 three (B) generations; for each of the 42 lines, the B treatment was founded first, and the S treatment

237 founded in the following generation by an independent sample of the original MA line. To ensure
238 similar environmental conditions, lines within each treatment consisted of a foundation experimental
239 unit of 12 male and 12 female parents per vial each generation, where 12 vials were included to make
240 up the B treatment. For the S treatment, a single vial was used to propagate each line, however, to
241 estimate within line variation, a replicate vial was established each generation but did not contribute
242 offspring to the next generation. For the B population, offspring from all 12 vials per line were
243 pooled before virgins were collected to found the next generation, allowing among-family natural
244 selection to act among vials. All lines were maintained for six generations under these conditions.

245 *Data collection*

246 Each generation, a wing (left or right, depending on condition) was collected from six, haphazardly
247 chosen male flies from two haphazardly chosen emergence vials for each B treatment line, and from
248 the focal and replicate vial for each S treatment line. Wings were mounted on microscope slides and
249 photographed using a Leica MZ6 microscope camera and the software LAS EZ v2.0.0 (Leica
250 Microsystems Ltd, Switzerland). A total of 5,135 wings were landmarked for nine positions, defined
251 by wing vein and margin intersections (Figure 2), using the software tpsDIG2 (ROHLF 2013). The
252 number of wings were evenly distributed across treatments (2,583 in S and 2,552 in B) and
253 generations (~8.3% per generation, per treatment). Landmarks were aligned using a General
254 Procrustes fit in tpsRelw (ROHLF 2007). Centroid size (CS), the square root of the sum of squared
255 deviations of the coordinates from the centroid (ROHLF 1999), was recorded as a metric of wing size.
256 The aligned X-Y coordinates for each landmark were then used to calculate ten inter-landmark
257 distances (ILDs) (Figure 2). ILDs scores were re-scaled prior to analysis (multiplied by 100) to aid
258 model convergence. Outliers >3.0 SD from the mean were removed for each of the 11 traits (10 ILD
259 and size) (329 measures across the 56,485 total measures).

260 *Analyses of trait means*

261 For non-fitness traits, such as wing traits, mutations are expected to have unbiased effects (JOHNSON
262 AND BARTON 2005). Therefore, we did not expect the two treatments, which differ in the opportunity
263 for mutation and selection, to differ in trait mean. However, micro-environmental variation within
264 laboratory experiments can generate both temporal trends and idiosyncratic changes among
265 generations (LANDE 1976; FALCONER AND MACKAY 1996). To test the null hypothesis that wing trait
266 mean remained constant across generations and treatments we implemented the following model
267 using PROC MIXED in SAS v9.4 (SAS Institute Inc., Cary, NC.), within a maximum-likelihood
268 framework:

269 $$y_{ijklm} = \mu + G_i + T_j + GT_{ij} + Line_{k(j)} + Vial_{l(ijk)} + \varepsilon_{ijklm} \quad (2)$$

270 where y_{ijkm} was the trait value for the *m*th wing, from the *l*th vial, within the *k*th line, for the *j*th
271 treatment in the *i*th generation. The mixed model fit the overall mean (μ) and the fixed effects of
272 generation (*G*), treatment (*T*), and their interaction (*GT*). The random effects of the model partitioned
273 variation among MA lines (*Line*) and between replicate rearing vials (*Vial*) within each line, while
274 the residual error (ε) contained unexplained variance among individual flies within a rearing vial.
275 Each line was sampled in all generations and we accounted for this repeated measures design, fitting
276 a reduced-rank factor analytic covariance structure (specifying TYPE=FA(1)), which estimated
277 generation-specific among-line variance (a diagonal matrix) and one dimension of among-generation
278 covariance (i.e., among-line variance shared across repeated measures of generations) (MEYER 2009;
279 MCGUIGAN AND BLOWS 2010). Preliminary analyses indicated that this FA(1) model captured all
280 among-line variance. Lastly, among-vial and residual variances were estimated for each generation
281 and treatment, accounting for any heterogeneity among generations and treatments in micro-
282 environmental effects, and segregating variation.

283 *Analyses of mutational variance*

To investigate the effects of mutation-selection-drift (captured by our two population size treatments) and micro-environment on estimates of mutational (among-line) variance, we first treated each generation, treatment and trait as independent realisations of the MA to characterise the variability in estimates (12 estimates per trait, 132 estimates in total). This was done by reducing model 2 to:

$$y_{klm} = \mu + Line_k + Vial_{l(k)} + \varepsilon_{klm} \quad (3)$$

fit using restricted-maximum likelihood (REML) in PROC MIXED in SAS v9.4 (SAS Institute Inc., Cary, NC.). We used REML-MVN sampling to place confidence intervals about our estimates, obtained by sampling from $N \sim (\hat{\theta}, \mathbf{V})$ 10,000 times, using the MASS package (VENABLES AND RIPLEY 2002) in R [v. 3.6.1], where $\hat{\theta}$ was the vector of observed covariance parameter estimates, and **V** was the asymptotic variance-covariance matrix, the inverse of the Fisher information matrix, $\boldsymbol{I}(\hat{\theta})^{-1}$ (MEYER AND HOULE 2013; HOULE AND MEYER 2015; SZTEPANACZ AND BLOWS 2017). We are interested here only in general patterns of variability among estimates, rather than whether the presence of mutational variance is statistically supported. We conduct further, more robust, statistical tests on these data, as described below. G-scale REML-MVN samples, which are not constrained to the parameter space (i.e., can be negative), allow inference of statistical support for variance when the lower 5% CI does not encompass zero (a one-tailed test); for individual variables, this approach provides the same conclusions as a Log Likelihood Ratio test (LRT) (DUGAND *et al.* 2021), but has the advantage of also providing a robust visualisation of the confidence in estimate values.

Next, to formally test whether mutational variance changed over time we again modified model 2, and used PROC MIXED and REML to fit a model with the intercept (μ), generation (G), *Line, Vial* and error, and included an additional random effect of the interaction between line and generation ($G \times Line$), estimating the variation in mutational effects among generations (GARCÍA-DORADO *et al.* 2000):

$$y_{iklm} = \mu + G_i + Line_k + (G \times Line)_{ik} + Vial_{l(ik)} + \varepsilon_{iklm} \quad (4)$$

308 This model was applied to each trait and population size treatment separately (22 models). Among-
309 vial and residual variances were estimated for each generation. To determine if the mutational
310 variance varied among generations, we applied log-likelihood ratio tests (LRT, 0.5 d.f.: SELF AND
311 LIANG 1987; LITTELL *et al.* 2006) to compare model 4 to a model that did not fit *G* x *Line*. We
312 similarly tested the hypothesis that there was significant among-line variance, using LRT to compare
313 model 4 to a reduced model that did not fit *Line*. We applied the Benjamini-Hochberg method
314 (BENJAMINI AND HOCHBERG 1995) to correct for multiple hypothesis testing (within each random
315 effect), using a conservative 5% false discovery rate (FDR).

316 To formally test the hypothesis that differences in the opportunity for mutation-drift-selection
317 between the Big and Small population size treatments affected the estimates of among-line variance,
318 or the among-line by generation variance, we ran the following model for each of the 11 traits, again
319 using PROC MIXED and REML, where we fit the intercept (μ), generation (G), treatment (T) and
320 the interaction between generation and treatment (GT), as well as the random effects of *Line, G*
321 $\times$ *Line, Vial* and error:

322 $$y_{ijklm} = \mu + G_i + T_j + GT_{ij} + Line_k + (G \times Line)_{ik} + Vial_{l(ijk)} + \varepsilon_{ijklm} \quad (5)$$

323 Although estimates of among-line by generation variance were typically not statistically
324 distinguishable from zero (see Results), we nonetheless modelled it here to ensure that small,
325 statistically undetectable, variation did not inflate the among-line variance of interest. We modelled
326 generation and treatment-specific among-vial and residual variances. We again used LRTs (1 d.f.) to
327 determine whether estimating among-line variance separately for the two population size treatments
328 (modelled using a Group statement in PROC MIXED) improved model fit over modelling a single,
329 pooled, estimate, again applying the Benjamini-Hochberg method (BENJAMINI AND HOCHBERG 1995)
330 to allow a 5% FDR. For the one trait with statistically supported among-line by generation variance
331 in at least one population size treatment (see Results), we again used a LRT (1 d.f.) to test the

332 hypothesis that Big and Small population treatments differed in the magnitude of among-line by
333 generation variance, after a 5% FDR correction.
334 Finally, to estimate the magnitude of mutational variance for each trait, we took the among-
335 line variance estimate from model (5), estimated from the total data, and calculated V_M as the among
336 line variance divided by 40 (twice the number of generations of mutation accumulation at the start of
337 our experiment) (LYNCH AND WALSH 1998). To place our estimates on a h^2_{M} scale, for each trait, we
338 calculated the average V_{E} (sum of among-vial and residual variances) across each generation and
339 treatment, while CV_{M} was estimated by dividing the V_{M} estimate by the trait mean, across all six
340 generations. Confidence intervals for our estimates of h^2_{M}, were estimated using REML-MVN
341 sampling (as described above), where, for each of the 10,000 REML-MVN samples, we calculated
342 V_{M} and the average V_E. Similarly, CV_{M} 95% CIs were estimated using the REML-MVN sample of
343 V_{M} and an estimate of sampling error of the mean, generated by obtaining 10,000 random samples
344 with the observed trait mean and trait variance, again using the MASS package (VENABLES AND
345 RIPLEY 2002). Estimates from the inverse Fisher Information matrix were unbounded, and ~0.09% of
346 MVN estimates of V_{L} fell below zero, where the square-root value could not be calculated, resulting
347 in CV_M CI based on <10,000 samples for six traits (CS, ILD1.9, ILD2.5, ILD3.8, ILD4.5 and
348 ILD4.8).
349

350 **RESULTS**

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

352 As predicted, CV_{M} varied among trait types ($F_{6,151}$ =15.89, P < 0.0001) where morphological traits
353 (predominantly size-related) had smaller CV_{M} than did life history traits (Figure 3B). In contrast,
354 heritability was similar across trait types ($F_{6,161}$ = 2.01, P= 0.0670; Figure 3A). The number of lines
355 sampled did not affect the magnitude of mutational variance (h^2_{M}: $F_{1,68.6}$ = 0.83, P = 0.3658;

356 CV_M: $F_{1, 70.6}$ = 0.27, P = 0.6076). Taxon significantly predicted the magnitude of mutational variance,
357 on both scales (h^2_M: $F_{4, 54.5}$ = 3.12, P = 0.0221; CV_M: $F_{4,56.5}$ =6.13, P = 0.0004), with plants having the
358 lowest model estimates of mutational variance and Nematodes the highest (Figure 4). The magnitude
359 of mutational variance significantly decreased with increasing number of generations of mutation
360 accumulation (h^2_M: $F_{1, 58.7}$ = 12.06, P = 0.0010, β = -0.622; CV_M: $F_{1, 65.6}$ = 21.05, P < 0.0001, β =
361 - 1.065). However, as plants and nematodes differed markedly in generation number (median of 17
362 and 250 generations, respectively; Figure S2), the effect of taxon versus generation number on the
363 magnitude of V_M cannot be clearly inferred. To further investigate the effect of the duration of
364 accumulation, we analysed V_M separately for each of the two best represented taxonomic groups,
365 Drosophila (Figure S2: 19 - 300 generations) and Nematode (50 – 370 generations; Figure S2). In
366 Drosophila, there was a significant negative relationship between number of generations and the
367 magnitude of mutational variance (h^2_M: $F_{1, 32}$ = 7.76, P = 0.0089, β = -0.49, N=56; CV_M: $F_{1,}$
368 $_{20.1}$=28.05, P < 0.0001, β = -1.299, N=73). In contrast, in the Nematode category, generation number
369 did not explain the magnitude of mutational variance (h^2_M: $F_{1, 14.1}$ = 1.38, P = 0.26, β = -0.34, N=86;
370 CV_M: $F_{1, 18.9}$ = 0.00, P = 0.97, β = -0.02, N = 76).

371 **Variation in estimates of V_M within the same taxon and trait type**

372 In the panel of modified *D. serrata* MA lines, we first investigated variation in trait means.
373 Directional trends through time, including increasing divergence between the Big and Small
374 population size treatments (i.e., generation by treatment interaction effect) would be consistent with
375 evolution via new mutation. While there was a statistically significant generation by treatment
376 interaction effect on most traits, the patterns did not correspond to increasing differentiation between
377 treatments (Figure 5). One trait differed in mean between the treatments (ILD4.8: $F_{1, 711}$ =25.93, P=
378 0.0001), but there was no evidence of increasing magnitude of divergence through time (Figure 5).
379 Overall, given the lack of unidirectional divergence through time, or increasing divergence between

380 treatments, these results are not consistent with evolution, being more consistent with effects of
381 micro-environmental variation among generations, or transient effects of segregating mutations that
382 are lost (fixed) via drift or selection.

383 Next, we investigated the variation in among-line variance, treating our data as representing
384 12 different experiments for each of our 11 traits. There was inconsistent statistical evidence for
385 among-line variance within each trait, where 60 out of 132 estimates (45.45%) were not significantly
386 different from zero (Figure 6). Given the variability in estimates for each trait (Figure 6), we could
387 draw very different conclusions about the magnitude of mutational variance and the statistical
388 support, depending on the generation-specific estimate (Figure 6). For example, among-line variance
389 for centroid size for the Big treatment was estimated to be more than twice as large in the first
390 generation compared to the third (3.43×10^{-4} and 1.38×10^{-4}, respectively; Figure 6). Marked
391 fluctuations in the magnitude of mutational variance remained apparent on the heritability and
392 coefficient of variance scales (Supplementary Figures S3 and S4, respectively).

393 Among-line variance did not exhibit any temporal trends (Figure 6), and variation in
394 mutational effects among generations (among-line by generation variance: model 4) was statistically
395 supported for only four of the 22 estimates (two treatments by eleven traits), with only two estimates
396 remaining significant at the 5% FDR (LRT for centroid size in the Big treatment: χ_1^2 = 13.9, P =
397 0.0001; LRT for ILD2.5 in the Big treatment: χ_1^2 = 11.1, P = 0.0005; Figure 7A). Consistent with the
398 largely overlapping CI of the two treatments (Figure 6), a difference between the Big and Small
399 population size treatments in among-line by generation variance was detected for only one trait
400 (centroid size), and this did not remain significant at the 5% FDR (Table S4B). Furthermore, there
401 was no statistical support for an effect of the population size treatment on the among-line variance for
402 any trait (Figure 7B; Table S4A). The variability in estimates (Figure 6) was thus typically within the
403 magnitude of the measurement error associated with each estimate.

404 Given the lack of evidence that mutation-selection-drift processes were contributing to
405 variability of among-line variance estimates, we utilised all data (six generations for two treatments)
406 to obtain a single, statistically robust, estimate of mutational variance for each trait (Figure 8). Unlike
407 per-generation estimates of among-line variance (Figure 6), mutational variance was statistically
408 supported for all traits (Figure 8). We had no *a priori* expectation of differences in magnitude of V_M
409 across the ten ILD wing shape traits yet estimates of h^2_M ranged from 0.001 to 0.005, while the CV_M
410 estimates ranged from 0.016 to 0.125 (Figure 8). While several traits consistently had the lowest or
411 highest estimates on both scales, most traits did not maintain their rank order across the two scales
412 (Figure 8). Wing size may have a greater mutational target size (for the rationale behind this, see
413 HOULE 1998), and thus be expected to have the largest estimate of mutational variance. Consistent
414 with this, size was one of the larger estimates on both scales, although was not the largest on either
415 scale (Figure 8).
416

417 **DISCUSSION**

418 Although numerous estimates of mutational variance (V_M) have been published and reviewed, it
419 remains unclear why estimates vary in magnitude. Our meta-analytic investigation revealed
420 substantial confoundment between potential causal factors, suggesting that further empirical data
421 must be collected to fully resolve the contribution of these general factors to variation in V_M.
422 Mutation is an inherently random process, but the expected variance about the true V_M due to
423 stochastic sampling of newly segregating mutations, or from experimental sampling error, is not
424 known. To address this, we conducted a new empirical experiment, designed to allow us to partition
425 the effects of micro-environmental variation and measurement error from the idiosyncrasies of the
426 mutation-drift process. Our data supported measurement error as accounting for most of the variation

427 in estimates. Below we discuss the specific outcomes and limitations of our meta-analyses and our
428 modified MA, and the implications our analyses have for future work.

429 *Effects of taxon, trait type and mutation accumulation duration on the magnitude of mutational*
430 *variance*

431 Genome size and per-nucleotide mutation rates vary among taxa (LYNCH *et al.* 1999; ELLIOTT AND
432 GREGORY 2015; LYNCH *et al.* 2016; KATJU AND BERGTHORSSON 2019), leading us to expect a
433 significant difference among taxa in the magnitude of V_M. Here, without detailed information on
434 taxon-specific mutation rates, we assumed more closely related taxa shared similar mutational
435 characteristics relative to other taxonomic groups. The plant taxon category differed substantially in
436 V_M relative to other taxa, particularly Nematode (on both CV_M and h^2_M scales). Notably, the per
437 nucleotide mutation rate estimated (from genome sequence data) for eight multicellular eukaryotes is
438 highest for *Arabidopsis* (the best-represented plant taxon in our data) and lowest for the nematode *C.*
439 *briggsae* (KATJU AND BERGTHORSSON 2019). However, MA in plant species had the shortest duration
440 of mutation accumulation (Figure S2), and the analysis supported a decline in V_M as the number of
441 MA generations increased. Thus, the apparent divergence in V_M for plants could be a consequence of
442 short duration, or the apparent decline in V_M over time could be an artifact, (partly) caused by the
443 high V_M in plants. Within well-sampled taxa, we observed a significant decline in V_M with generation
444 time in the Drosophila category, but not in Nematodes. Decline in V_M after a certain number of
445 generations of MA has been reported previously in *Drosophila melanogaster*, but were observed in
446 only some assayed traits (MACKAY *et al.* 1995; GARCÍA-DORADO *et al.* 2000). Therefore, while the
447 data suggest that taxon and MA duration could account for some variation in the magnitude of V_M,
448 further experiments are required to explicitly address this, decoupling the confounding of taxon and
449 duration apparent in the published literature. Explicitly estimating relevant parameters such as
450 genome size and per nucleotide mutation rates for more taxa, alongside classical MA could resolve

451 the relationship between the phenotypic variation generated by mutation (i.e., V_M) and mutation rate
452 (KATJU AND BERGTHORSSON 2019).

453 HOULE *et al.* (1996); (see also LYNCH AND WALSH 1998; LYNCH *et al.* 1999) strongly
454 concluded that, relative to morphological traits, life history traits had lower h^2_M but higher CV_M. Our
455 results were consistent with this for CV_M, but not for h^2_M, where estimates did not significantly vary
456 among trait types. Notably, in our meta-analysis, 76% of the estimates in the morphology category
457 were the size of either part of (e.g., wing length; SANTIAGO *et al.* 1992) or the whole organism (e.g.,
458 plant height; RUSSELL *et al.* 1963) due to paucity of estimates for other types of morphological traits.
459 HOULE *et al.* (1996) predicted that life history traits would have lower h^2_M due to their elevated
460 environmental variance. Size traits also exhibit considerable environmental sensitivity at both macro-
461 (e.g., MCGUIGAN 2009) and micro-environmental (e.g., SANTOS *et al.* 1994) scales. Size might also
462 be more similar to fitness traits in its underlying genetic complexity (and hence, mutational target
463 size; HOULE 1998). Investigation of diverse morphological traits, in conjunction with assessment of
464 environmental sensitivities, will help inform where the magnitude of h^2_M for size-related traits lies in
465 relationship to fitness versus morphological traits. We observed larger mutational variances (h^2_M and
466 CV_M) for fitness itself and in productivity than in survival. It is possible that this pattern reflects
467 differences in the size of the genomic basis of different life history traits. However, the MA design
468 could also allow less effective relaxation of selection on some fitness traits than others. Again,
469 combining fine-scale genomic analyses of MA lines could lead to further insights on why traits differ
470 in mutational variance (KATJU AND BERGTHORSSON 2019).

471 *The contributions of mutation-selection processes, micro-environmental variation, and measurement*
472 *error to variation in the magnitude of mutational variance*

473 Variance among MA lines increases by $2V_M$ per generation (LYNCH AND HILL 1986), but this
474 represents the average expectation (at least over some time window), where individual point

475 estimates might be influenced by other factors, including segregational variance and micro-
476 environment (LYNCH AND HILL 1986; LYNCH 1991). In the current study, we estimated among-line
477 variance in six sequential samples of a panel of 42 MA lines and observed substantial heterogeneity
478 in magnitude, leading to inconsistent statistical support for among-line variance for individual traits
479 (Figure 6). Below we consider the evidence of what caused this variability, and how general those
480 contributions might be over other studies.

481 In the only other study to investigate short-term (over two or three generations) variability in
482 mutational variance estimates, GARCÍA-DORADO *et al.* (2000) found that V_M fluctuated across
483 consecutive generations for one (sternopleural bristle count) of the four traits investigated (no
484 temporal variation in estimates of egg-to-adult viability, wing length or abdominal bristle count).
485 GARCÍA-DORADO *et al.* (2000) attributed the variation to (mutant) genotype-by-environment (GxE)
486 interactions, specifically to micro-environmental variation in rearing conditions altering the effect of
487 mutations on phenotype from generation to generation. We similarly found statistical support for
488 variation in mutational effects among generations for few (two of 11) traits, suggesting that micro-
489 environmental GxE contributed relatively little to the apparent heterogeneity of estimates. Wing
490 (centroid) size was one trait for which GxE was statistically supported; as discussed above, size
491 might be more dependent on micro-environmental variation in resources or energy requirements than
492 wing shape traits (CAVICCHI *et al.* 1985; BITNER-MATHÉ AND KLACZKO 1999). Thus, our results and
493 those of GARCÍA-DORADO *et al.* (2000) suggest that changes in mutational effects on phenotype with
494 micro-environmental variation may contribute to heterogeneity among point estimates of some traits,
495 which may reflect among trait differences in environmental sensitivity (i.e., size versus shape
496 morphology), or potentially covariation of environmental sensitivity and mutational effect size
497 (LYNCH *et al.* 1999; GARCÍA-DORADO *et al.* 2000).

498 The mutation-drift process can also cause departures from a constant per-generation increase
499 in among line variance, where chance sampling of either a newly-arisen mutation, or the loss of a

500 segregating mutation causes estimated among-line variance to fluctuate over time (LYNCH AND HILL
501 1986). However, our analyses did not support a contribution of this process to heterogeneity of
502 estimates as we did not observe differences in the magnitude of variation in among-line variance in
503 our big versus small population size treatments. The efficacy of selection depends on the effective
504 population size (N_e), where mutations will be deterministically removed when their effect on fitness
505 is greater than 1/2 N_e (WRIGHT 1931). Our population size treatments contrasted deterministic
506 evolution of mutations with relatively strong ($s > 0.029$: $N_e = 17$), versus weak ($s > 0.003$: $N_e = 202$)
507 fitness effects (effective population size estimated as 70% of census size, CROW AND MORTON 1955).
508 Thus, mutations contributing to observable wing variation, and with pleiotropic effects on fitness
509 resulting in $0.003 < s < 0.030$, may be relatively rare. ESTES *et al.* (2004) found that deleterious
510 mutations accumulated similarly in *C. elegans* when N=1 or N=5 ($s > 0.2$), whereas there was no
511 evidence of mutation accumulation when $N = 25$ ($s \sim 0.02$). Similar results have been obtained for
512 other studies in nematode (e.g., KATJU *et al.* 2015) and rotifers (LUIJCKX *et al.* 2018), suggesting
513 preponderance of mutations with large fitness effects. In the current study, divergence between the
514 big and small population size treatments might not have been detected due to small mutational effect
515 sizes on wing phenotypes, coupled with the short duration (six generations) limiting power. Notably,
516 KATJU *et al.* (2015) compared populations after 409 generations, for fitness traits for which strong
517 directional bias (toward lower fitness) is well supported and hence accumulation of mutations with
518 opposing (i.e., cancelling) effects of accumulating mutations are not expected. Nonetheless, this
519 conclusion of rarity of small effect mutations is at odds with the conclusions from comparative
520 genomic estimates of weak selection against new mutation (EYRE-WALKER AND KEIGHTLEY 2007;
521 BOYKO *et al.* 2008; KIM *et al.* 2017), and further work is required to resolve this.
522 Rejecting micro-environmental variation and transient segregation of variants as general
523 explanations of estimate heterogeneity in our data, we conclude that the observed variability is
524 largely the consequence of sampling (measurement) error. With a typical number of MA lines (42)

525 and samples per line (two replicate vials of six individuals each) our observation is likely typical for
526 MA studies. Our approach of using a sequential repeated-measures experimental design provided
527 statistical control over this experimental noise, allowing us to consistently detect statistically
528 significant mutational variance in all traits (Figure 8). Several studies have statistically accounted for
529 heterogeneity among estimates, irrespective of the cause, using a repeated-measures design and either
530 estimating the per-generation rate of increase in among line variance from the slope of the regression
531 of among-line variance against generation (VASSILIEVA *et al.* 2000; HOULE AND NUZHDIN 2004;
532 MCGUIGAN *et al.* 2011), or by including generation as a (fixed or random) factor in the model to
533 obtain a single estimate of among-line variance (MARTORELL *et al.* 1998; SHAW *et al.* 2000; BAER *et*
534 *al.* 2005). Our analysis particularly highlights the potential benefits of repeated measures of
535 sequential generations to improve estimate accuracy and power to detect small effects.

536 Using all data to obtain a robust estimate of mutational heritability for the *D. serrata* wing
537 shape traits, we observed h^2_{M} from 0.001–0.005 (Figure 8), which is similar to that reported for *D.*
538 *melanogaster* wing traits (0.001–0.003: SANTIAGO *et al.* 1992; GARCÍA-DORADO *et al.* 2000). Wing
539 (centroid) size did not have a larger estimate of mutational, despite the greater environmental
540 sensitivity (MCGUIGAN 2009) or potentially larger mutational target size, consisting of loci affecting
541 larval growth, pupation timing, metabolism and behaviour (ROFF 1981; HOULE 1998). While a
542 relatively narrow range, our results demonstrate that variability remains for estimates from a
543 completely controlled experiment (same taxon, trait type, macro- and micro-environment).
544 Establishing whether these differences are informative of the inherent genetic architecture for our
545 traits, or are merely a manifestation of the stochastic nature of mutation, would require repeating the
546 estimation using either the same or a different genetic background to determine if the traits remain in
547 the same rank order.

548 Understanding the contribution that mutations make to evolutionary and genetic phenomena
549 relies on better estimates of the phenotypic variance generated by new mutation. Our meta-analysis

of empirical estimates of mutational variance was unsuccessful in determining the relative contribution of trait type, taxon and experimental duration due to confoundment between these factors, illustrating how crucial disentangling the effects of these factors from one another is to clarify understanding. Our modified mutation accumulation design demonstrated that measurement error contributes substantially more to estimate variability than micro-environmental variation or sampling effects of segregating mutations. Use of repeated measures designs over few (e.g., sequential) generations therefore emerges as a relatively simple approach to improve estimate accuracy and inference. Overall, further empirical studies are needed to fully assess how both general and study specific factors affect V_M estimates, where improved precision and replicability in estimates will consequently advance broader evolutionary questions such as those addressing the maintenance of quantitative genetic variance (LYNCH 1988b; BARTON AND TURELLI 1989; JOHNSON AND BARTON 2005).

Data Availability

Data will be made publicly available via figshare upon publication, doi:10.6084/m9.figshare.14913051. For purposes of review: https://figshare.com/s/831df121e3452bcd83a4

Acknowledgments

We thank Adam Reddiex, Nicholas Appleton, Jack Price and Derek Sun for their help with maintaining, collecting, and digitising the mutation accumulation data. We also thank Stephen F. Chenoweth for providing us with the *Drosophila serrata* MA lines, and Jan Engelstädter for suggesting the combinational approach.

References

Azevedo, R. B., 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.

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. Proceedings of the National Academy of Sciences, USA 102**:** 5785-5790.

Barton, N. H., and P. D. Keightley, 2002 Understanding quantitative genetic variation. Nat Rev Genet 3**:** 11-21.

Barton, N. H., and M. Turelli, 1989 Evolutionary quantitative genetics: How little do we know? Annual Review of Genetics 23**:** 337-370.

Benjamini, Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 57**:** 289-300.

Bitner-Mathé, B. C., and L. B. Klaczko, 1999 Plasticity of *Drosophila melanogaster* wing morphology: effects of sex, temperature and density. Genetica 105**:** 203.

Box, G. E., and D. R. Cox, 1964 An analysis of transformations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 26**:** 211-243.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet 4**:** e1000083.

Cavicchi, S., D. Guerra, G. Giorgi and C. Pezzoli, 1985 Temperature-related divergence in experimental populations of *Drosophila melanogaster*. I. Genetic and developmental basis of wing size and shape variation. Genetics 109**:** 665-689.

Crow, J. F., and N. E. Morton, 1955 Measurement of gene frequency drift in small populations. Evolution 9: 202-214.

Dugand, R. J., J. D. Aguirre, E. Hine, M. W. Blows and K. McGuigan, 2021 The contribution of mutation and selection to multivariate quantitative genetic variance in an outbred population of *Drosophila serrata*. Proceedings of the National Academy of Sciences, USA.

Elliott, T. A., and T. R. Gregory, 2015 What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences 370: 20140331.

Estes, S., B. 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.

Estes, S., P. C. Phillips, D. R. Denver, W. K. Thomas and M. Lynch, 2004 Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in *Caenorhabditis elegans*. Genetics 166: 1269-1279.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat Rev Genet 8: 610-618.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to quantitative genetics*. Longman, Burnt Mill, England.

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.

Fry, J. D., S. L. Heinsohn and T. F. Mackay, 1996 The contribution of new mutations to genotype-environment interaction for fitness in *Drosophila melanogaster*. Evolution 50: 2316-2327.

García-Dorado, A., J. Fernández and C. López-Fanjul, 2000 Temporal uniformity of the spontaneous mutational variance of quantitative traits in *Drosophila melanogaster*. 75: 47-51.

622 Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in
623 evolutionary genetics. Annual Review of Ecology, Evolution and Systematics 40: 151-172.
624 Houle, D., 1991 Genetic covariance of fitness correlates: what genetic correlations are made of and
625 why it matters. Evolution 45: 630-648.
626 Houle, D., 1992 Comparing evolvability and variability of quantitative traits. Genetics 130: 195-204.
627 Houle, D., 1998 How should we explain variation in the genetic variance of traits? Genetica 102:
628 241-253.
629 Houle, D., G. H. Bolstad, K. van der Linde and T. F. Hansen, 2017 Mutation predicts 40 million
630 years of fly wing evolution. Nature 548: 447-450.
631 Houle, D., and K. Meyer, 2015 Estimating sampling error of evolutionary statistics based on genetic
632 covariance matrices using maximum likelihood. Journal of Evolutionary Biology 28: 1542-
633 1549.
634 Houle, D., B. Morikawa and M. Lynch, 1996 Comparing mutational variabilities. Genetics 143:
635 1467-1483.
636 Houle, D., and S. V. Nuzhdin, 2004 Mutation accumulation and the effect of *copia* insertions in
637 *Drosophila melanogaster*. Genetical Research 83: 7-18.
638 Huang, W., R. F. Lyman, R. A. Lyman, M. A. Carbone, S. T. Harbison *et al.*, 2016 Spontaneous
639 mutations and the origin and maintenance of quantitative genetic variation. Elife 5: e14625.
640 Johnson, T., and N. Barton, 2005 Theoretical models of selection and mutation on quantitative traits.
641 Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences
642 360: 1411-1425.
643 Katju, V., and U. Bergthorsson, 2019 Old trade, new tricks: insights into the spontaneous mutation
644 process from the partnering of classical mutation accumulation experiments with high-
645 throughput genomic approaches. Genome Biology and Evolution 11: 136-165.

646 Katju, V., L. B. Packard, L. Bu, P. D. Keightley and U. Bergthorsson, 2015 Fitness decline in
647 spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective
648 population sizes. Evolution 69**:** 104-116.
649 Kavanaugh, C. M., and R. G. Shaw, 2005 The contribution of spontaneous mutation to variation in
650 environmental responses of *Arabidopsis thaliana*: Responses to light. Evolution 59**:** 266-275.
651 Keightley, P. D., T. F. Mackay and A. Caballero, 1993 Accounting for bias in estimates of the rate of
652 polygenic mutation. Proceedings of the Royal Society of London, Series B: Biological
653 Sciences 253**:** 291-296.
654 Keightley, P. D., U. Trivedi, M. Thomson, F. Oliver, S. Kumar *et al.*, 2009 Analysis of the genome
655 sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines.
656 Genome Research 19**:** 1195-1201.
657 Kim, B. Y., C. D. Huber and K. E. Lohmueller, 2017 Inference of the distribution of selection
658 coefficients for new nonsynonymous mutations using large samples. Genetics 206**:** 345-361.
659 Kimura, M., 1965 A stochastic model concerning the maintenance of genetic variability in
660 quantitative characters. Proceedings of the National Academy of Sciences, USA 54**:** 731-736.
661 Kondrashov, A. S., and D. Houle, 1994 Genotype-environment interactions and the estimation of the
662 genomic mutation rate in *Drosophila melanogaster*. Proceedings of the Royal Society of
663 London, Series B: Biological Sciences 258**:** 221-227.
664 Lande, R., 1975 The maintenance of genetic variability by mutation in a polygenic character with
665 linked loci. Genetical Research 26**:** 221-235.
666 Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. Evolution 30**:**
667 314-334.
668 Latta, L. C., M. Peacock, D. J. Civitello, J. L. Dudycha, J. M. Meik *et al.*, 2015 The phenotypic
669 effects of spontaneous mutations in different environments. American Naturalist 185**:** 243-
670 252.

671 Latter, B. D., 1970 Selection in finite populations with multiple alleles. II. Centripetal selection,
672 mutation, and isoallelic variation. Genetics 66: 165-186.
673 Littell, R., G. Milliken, W. Stroup, R. Wolfinger and O. Schabenberger, 2006 *SAS for Mixed Models*.
674 SAS Institute, Cary, NC.
675 Luijckx, P., E. K. H. Ho, A. Stanic and A. F. Agrawal, 2018 Mutation accumulation in populations of
676 varying size: large effect mutations cause most mutational decline in the rotifer Brachionus
677 calyciflorus under UV-C radiation. Journal of Evolutionary Biology 31: 924-932.
678 Lynch, M., 1988a Design and analysis of experiments on random drift and inbreeding depression.
679 Genetics 120: 791-807.
680 Lynch, M., 1988b The rate of polygenic mutation. Genetical Research 51: 137-148.
681 Lynch, M., 1991 Methods for the analysis of comparative data in evolutionary biology. Evolution 45:
682 1065-1080.
683 Lynch, M., 2010 Evolution of the mutation rate. Trends in Genetics 26: 345-352.
684 Lynch, M., M. S. Ackerman, J.-F. Gout, H. Long, W. Sung *et al.*, 2016 Genetic drift, selection and
685 the evolution of the mutation rate. Nature Reviews: Genetics 17: 704-714.
686 Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz *et al.*, 1999 Perspective: spontaneous
687 deleterious mutation. Evolution 53: 645-663.
688 Lynch, M., and W. G. Hill, 1986 Phenotypic evolution by neutral mutation. Evolution 40: 915-935.
689 Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer, Sunderland,
690 MA.
691 Mackay, T., R. F. Lyman and W. G. Hill, 1995 Polygenic mutation in *Drosophila melanogaster*: non-
692 linear divergence among unselected strains. Genetics 139: 849-859.
693 Mackay, T. F. C., 2001 The genetic architecture of quantitative traits. Annual Review of Genetics 35:
694 303-339.

Martin, G., and T. J. E. Lenormand, 2006 The fitness effect of mutations across environments: a survey in light of fitness landscape models. 60: 2413-2427.

Martorell, C., M. Toro and C. Gallego, 1998 Spontaneous mutation for life-history traits in *Drosophila melanogaster*, pp. 315-324 in *Mutation and Evolution*. Springer.

McGuigan, K., 2009 Condition dependence varies with mating success in male *Drosophila bunnanda*. Journal of evolutionary biology 22: 1813-1825.

McGuigan, K., and M. W. Blows, 2010 Evolvability of individual traits in a multivariate context: partitioning the additive genetic variance into common and specific components. Evolution 64: 1899-1911.

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., 2009 Factor-analytic models for genotype x environment type problems and structured covariance matrices. Genetics Selection Evolution 41: 21.

Meyer, K., and D. Houle, 2013 Sampling based approximation of confidence intervals for functions of genetic covariance matrices. Proc. Assoc. Adv. Anim. Breed 20: 523-526.

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.

Mukai, T., 1964 Genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. Genetics 50: 1-19.

Ness, R. W., A. D. Morgan, N. Colegrave and P. D. Keightley, 2012 Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. Genetics 192: 1447-1454.

720 Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs *et al.*, 2007 Mutational bias for body size in
721 Rhabditid nematodes. 176**:** 1653-1661.
722 Reddiex, A. J., S. L. Allen and S. F. Chenoweth, 2018 A genomic reference panel for *Drosophila*
723 *serrata*. G3 8**:** 1335-1346.
724 Rockman, M. V., 2012 The QTN program and the alleles that matter for evolution: All that's gold
725 does not glitter. Evolution 66**:** 1-17.
726 Roff, D., 1981 On being the right size. American Naturalist 118**:** 405-422.
727 Rohlf, F., 2007 tpsRelw version 1.45. Department of Ecology and Evolution, State University of
728 New York, Stony Brook.
729 Rohlf, F. J., 1999 Shape statistics: Procrustes superimpositions and tangent spaces. J. Classif. 16**:**
730 197-223.
731 Rohlf, F. J., 2013 tpsDig, digitize landmarks and outlines, version 2.17. Department of Ecology and
732 Evolution, State University of New York at Stony Brook.
733 Rundle, H. D., S. F. Chenoweth, P. Doughty and M. W. Blows, 2005 Divergent selection and the
734 evolution of signal traits and mating preferences. PLoS Biol 3**:** e368.
735 Russell, W., G. Sprague and L. Penny, 1963 Mutations affecting quantitative characters in long-time
736 inbred lines of maize. Crop Science 3**:** 175-178.
737 Santiago, E., J. Albornoz, A. Domínguez, M. A. Toro and C. López-Fanjul, 1992 The distribution of
738 spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. Genetics
739 132**:** 771-781.
740 Santos, M., K. Fowler and L. Partridge, 1994 Gene-environment interaction for body size and larval
741 density in *Drosophila melanogaster:* an investigation of effects on development time, thorax
742 length and adult sex ratio. Heredity 72**:** 515-521.
743 Schrider, D. R., D. Houle, M. Lynch and M. W. J. G. Hahn, 2013 Rates and genomic consequences
744 of spontaneous mutational events in *Drosophila melanogaster*. 194**:** 937-954.

Self, S. G., and K.-Y. Liang, 1987 Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association 82: 605-610.

Shaw, R. G., D. L. Byers and E. Darmo, 2000 Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. Genetics 155: 369-378.

Sung, W., M. S. Ackerman, S. F. Miller, T. G. Doak and M. Lynch, 2012 Drift-barrier hypothesis and mutation-rate evolution. Proceedings of the National Academy of Sciences, USA 109: 18488-18492.

Sztepanacz, J. L., and M. W. Blows, 2017 Accounting for sampling error in genetic eigenvalues using random matrix theory. Genetics 206: 1271-1284.

Vassilieva, L. L., A. M. Hook and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution 54: 1234-1246.

Venables, W., and B. Ripley, 2002 *Modern Applied Statistics with S*. Springer, New York.

Walsh, B., and M. Lynch, 2018 *Evolution and selection of quantitative traits*. Oxford University Press.

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.

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97-159.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nature genetics 42: 565-569.

FIGURES AND TABLES

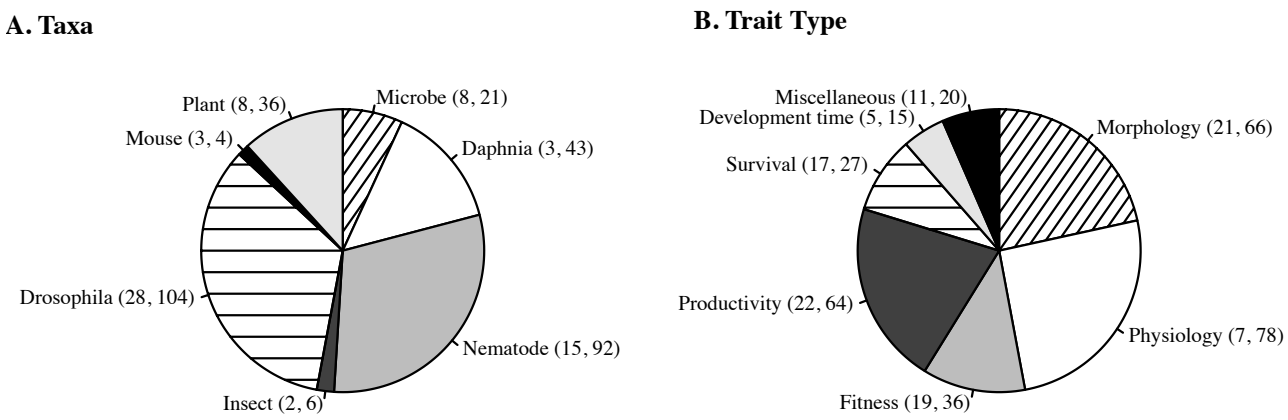

Figure 1. The distribution of published mutational variance estimates by taxon (A) and trait (B) categories. The number of studies per category are shown (first value in brackets), where many studies included multiple estimates (from different traits, life stages or sexes; second value in brackets). A. Seven taxonomic categories were considered, defined by true taxonomic considerations, as well as the distribution of available estimates. B. Seven trait categories were considered; for a full description of traits see Methods and Table S2.

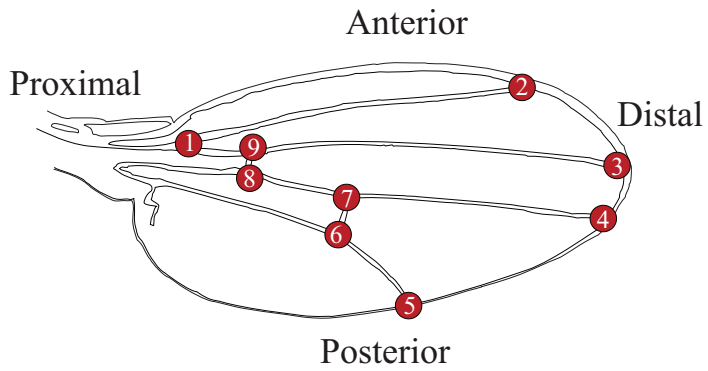

Figure 2. Positions of the nine landmarks used to characterise *Drosophila serrata* wing shape. Landmarks were: the proximal intersection of the radial vein (1), distal intersection of the radial vein (2), distal intersection of the medial vein (3), distal intersection of the cubital vein (4), distal intersection of the distal vein (5), the posterior (6) and the anterior (7) intersections of the posterior cross vein, and the posterior (8) and the anterior (9) intersections of the anterior cross-vein. Inter-landmark distance traits were described by their end-point landmarks (e.g., ILD1.2 was the distance between landmark1 and landmark 2).

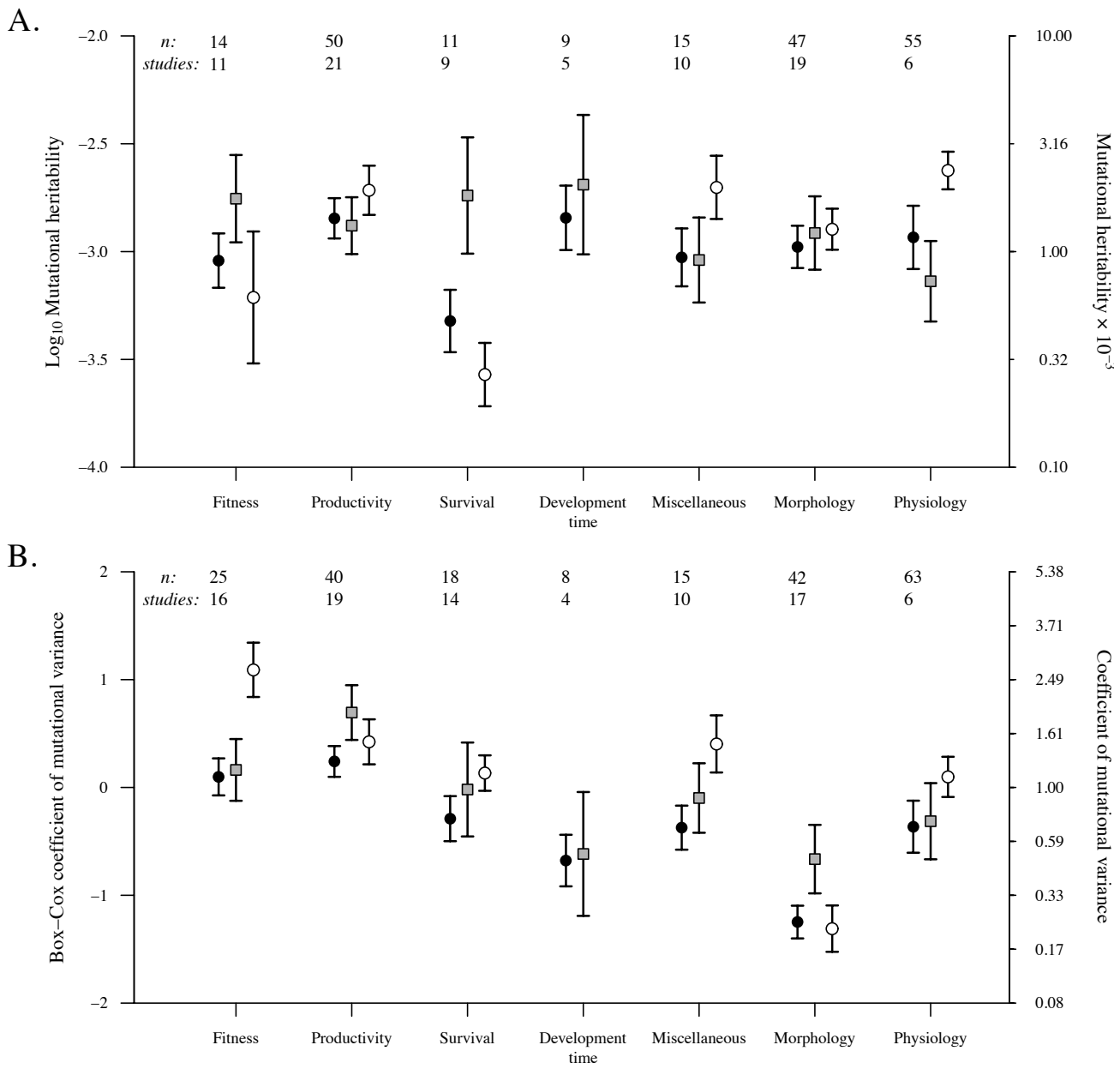

Figure 3. Variation in estimates of mutational heritability (A) and coefficient of mutational variance (B) across trait categories. For each trait type (x-axis, defined in Methods), least-squares means (± SE) were estimated from the analysis of model 1 for: all meta-analyses data (solid circles), Nematodes (grey squares) and Drosophila (open circles). The number of estimates, *n*, and studies for all taxa are shown across the top of each panel. For both A and B, the transformed (log10 or BoxCox) values, as input to the analysis, are shown on the left y-axis, and the untransformed values on the right y-axis. No estimates of development time mutational variance were available for Drosophila. Trait type significantly predicted the magnitude of CV_{M} in the total data (see main text), of h^2_{M} and

CV_{M} in Drosophila (h^2_{M}: $F_{5,38.9} = 7.25$, $P < 0.0001$; CV_{M}: $F_{5,26.9} = 11.87$, $P < 0.0001$), and CV_{M} in Nematodes (h^2_{M}: $F_{6,31.7} = 0.53$, $P = 0.78$; CV_{M}: $F_{6,62.8} = 3.20$, $P = 0.008$).

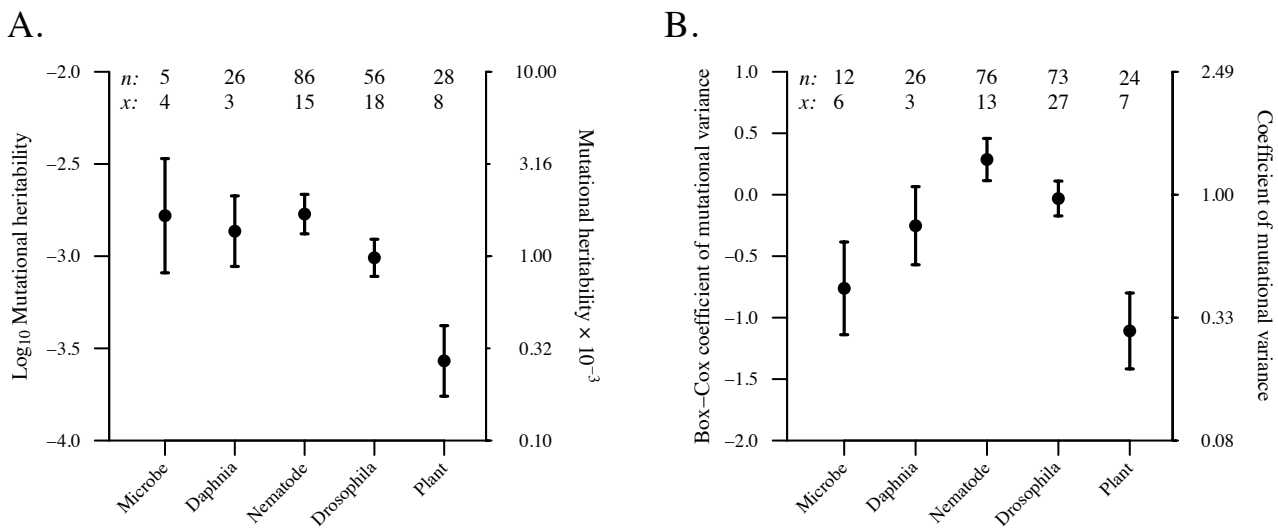

Figure 4. Variation in estimates of mutational heritability (A) and coefficient of mutational variance (B) across taxa categories. Point estimates are the least-squares mean estimate (± SE) from the analysis of model 1. The number of estimates (*n*) and of studies (*x*) are shown for each category. For both A and B, the transformed (log_{10} or Box-Cox) values input to the analysis are shown on the left y-axis, and the untransformed values on the right y-axis.

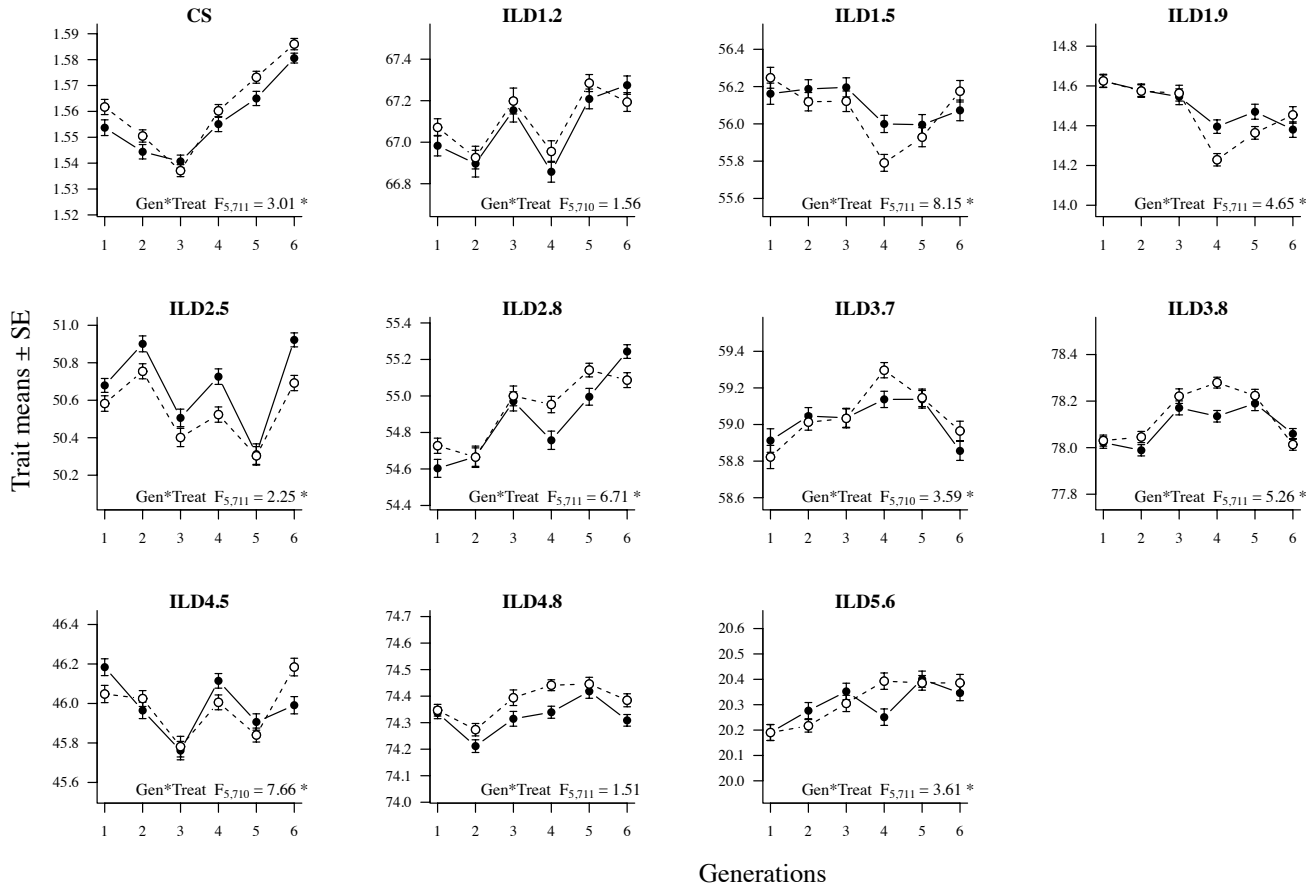

Figure 5. Variation in wing trait mean across six generations of a modified mutation accumulation experiment in *D. serrata*. For each of the 11 wing traits (panels), the least-squares mean (± SE) is plotted for each generation for the Small (solid circle, solid lines) and Big (open circles, dashed lines) 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 2 for trait definitions). The F-ratio (and degrees of freedom) for the generation by population size treatment effect for each trait, from model 2, is given, where the asterisk indicates $P \leq 0.05$.

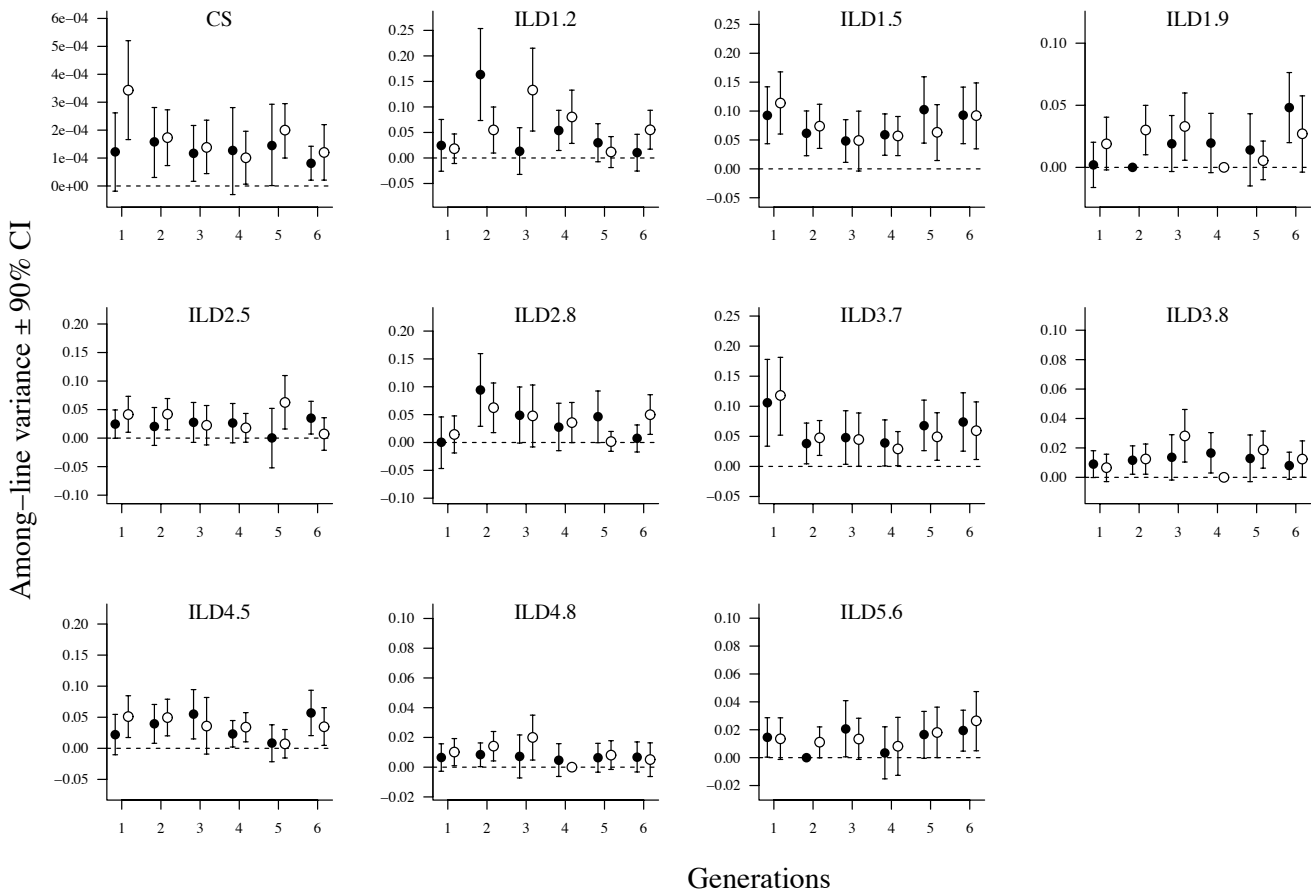

Figure 6. Among-line variance estimates across six generations in a modified mutation accumulation experiment in *D. serrata*. Variances were estimated independently for each trait (panel) in each generation (x-axis) for each of the two population size treatments (small: solid circles; big: open circles). Plotted are the REML point estimate, and the REML-MVN 90% confidence intervals (CI). The dashed horizontal line indicates zero; estimates for which the lower CI did not overlap zero were interpreted as statistically supported. Where REML estimates of among-line variance were zero, no CI are plotted.

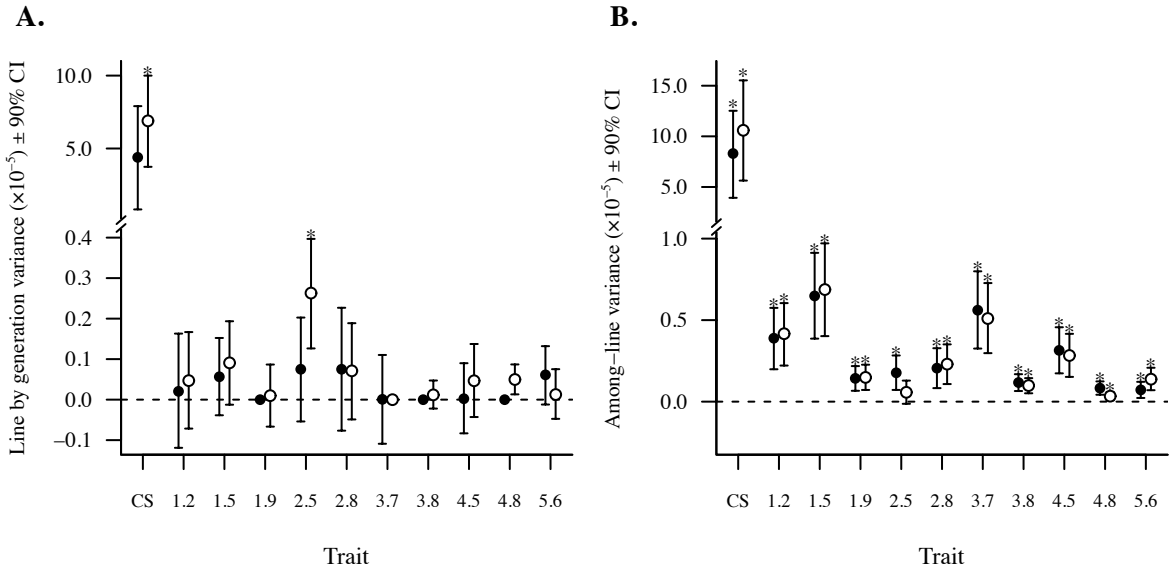

Figure 7. Estimates of variance from a modified mutation accumulation experiment in *D. serrata*. (A) Among-line by generation variance and (B) among-line variance estimated for 11 *D. serrata* wing traits (x-axis), in two different population size treatments (small: solid circles; big: open circles). Plotted are the REML point estimates (from model 4) and the REML-MVN 90% confidence intervals (CI) around this. The dashed horizontal line indicates zero; statistical significance was inferred where the lower 5% CI did not overlap zero. After applying a conservative 5% FDR correction, only two estimates in A) and 21 in B) remained significant (asterisk above CI).

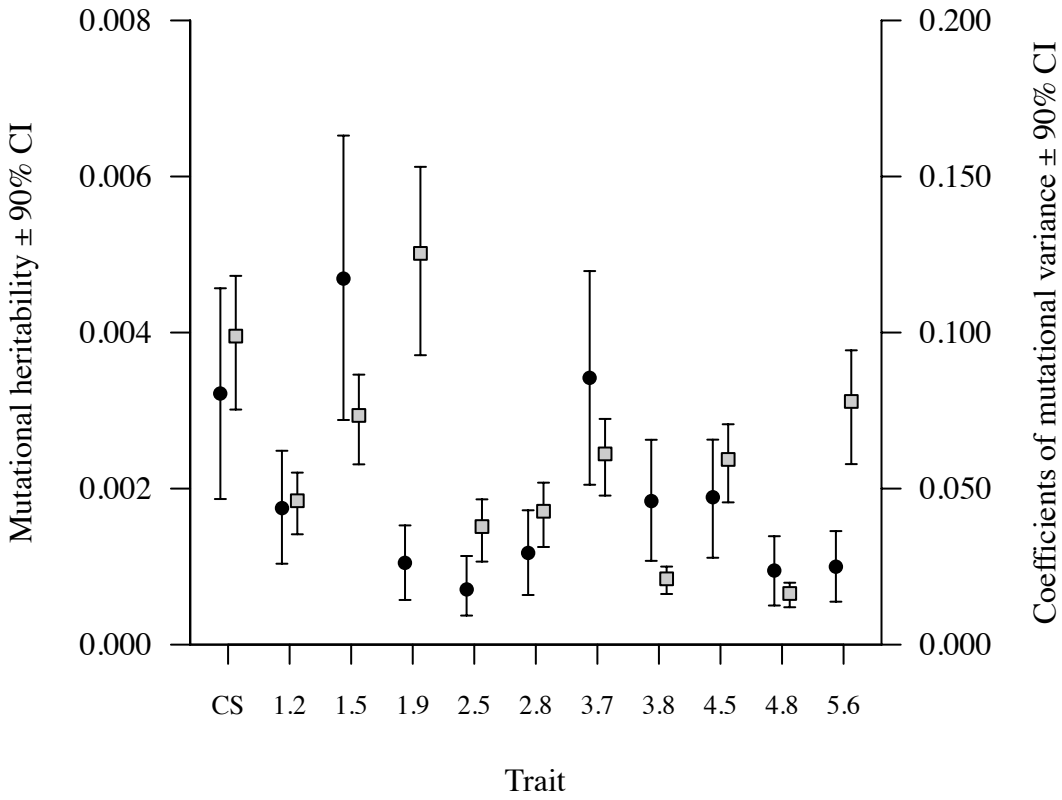

Figure 8. Mutational variances for eleven wing traits in *Drosophila serrata*. For each trait, mutational variance (V_M) was calculated (from among-line variance estimates from model 5), and then placed on either a mutational heritability (black circles, left y-axis) or coefficient of variance (grey squares, right y-axis) scale. The 90% confidence intervals were estimated as described in the Methods.

SUPPLEMENTARY INFORMATION

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 **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 **Mutational heritability estimates from a modified mutation accumulation experiment in *D. serrata*.**

Supplementary Figure S4 **Coefficients of mutational variance (CV_M) estimates from a modified mutation accumulation experiment 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). For each paper, the source (a previous review or Web of Science search) is indicated (see footnote for source key), along with information on where the standardised estimates of mutational variance (mutational heritability, $\boldsymbol{h_M^2}$ and, coefficient of mutational variance, $\boldsymbol{CV_M}$) were obtained from or, where relevant, the justification for exclusion from the meta-analyses. Note: some papers reported estimates that were included in the meta-analysis (A) and other estimates that were excluded (B). Estimates (papers) identified in B were not included in the summaries of data (Figure 1) or analyses. All estimates (A and B) are available in Supplementary File (*Figshare DOI to be supplied at acceptance*).

Reference	Source	Description of where estimates were obtained from
(A) ESTIMATES USED IN THE META-ANALYSIS		
Ajie, B. C., S. Estes, M. Lynch, and P. C. Phillips, 2005 Behavioural degradation under mutation accumulation in *Caenorhabditis elegans*. Genetics **170**: 655–660	1	Reference from in-text in Halligan & Keightley (2009). We calculated CV_M using the V_M (taking the number of generations as 307) from Estes *et al.* (2005) and the mean in Ajie *et al.* (2005)
Ávila, V. and A. García-Dorado, 2002 The effects of spontaneous mutation on competitive fitness in *Drosophila melanogaster*. J. Evol. Biol. **15**: 561–566	1	Halligan & Keightley (2009)'s Table 1.
Azevedo, R. B., 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	1	Table 1 of paper.
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	1	Halligan & Keightley (2009)'s Table 1.
Baer, C. F., N. Phillips, D. Ostrow, A. Avalos, D. Blanton, *et al.*, 2006 Cumulative effects of spontaneous mutations for fitness in Caenorhabditis: Role of genotype, environment and stress. Genetics **174**: 1387–1395	1	Table 1 of paper.
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	2	h^2_m estimates were from Table 2 of paper, insufficient information provided to calculate CV_M.
Bailey, D. W., 1959 Rates of subline divergence in highly inbred strains of mice. J. Hered **50**: 26–30	3	Houle *et al.* (1996)'s Table 1.
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	1	Halligan & Keightley (2009)'s Table 1.
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	1	CV_M calculated from in-text information, taking V_B from $\Delta V=V_B/(mean_c \times F_c)$; where F_c=203 and $mean_c$ = 0.625.
Casellas, J. and J. F. Medrano, 2008 Within-generation mutation variance for litter size in inbred mice. Genetics **179**: 2147–55	4	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.
Chang,S.andR.G.Shaw,2003Thecontributionofspontaneousmutationtovariationinenvironmental response in *Arabidopsis thaliana*: responses to nutrients. Evolution **57**: 984–994	4	Table 3 of paper.
Charlesworth, B., H. Borthwick, C. Bartolomé, and P. Pignatelli, 2004 Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. Genetics **167**: 815–826	1	Halligan & Keightley (2009)'s Table 1.
Clark, A. G., L. Wang, and T. Hulleberg, 1995 Spontaneous mutation rate of modifiers of metabolism in drosophila. Genetics **139**: 767–779	5	h^2_m and CV_M estimates taken from Lynch *et al.* (1999)'s Table 1, then CV_M converted to percentage scale.

Reference	Source	Description of where estimates were obtained from
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	1	Halligan & Keightley (2009)'s Table 1.
Cox, T., D. Cox, and K. Frey, 1987 Mutations for polygenic traits in barley under nutrient stress. Euphytica **36**: 823–829	3	Houle *et al.* (1996)'s Table 1.
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	2	Table S1 of paper.
Downie, D., 2003 Effects of short-term spontaneous mutation accumulation for life history traits in grape phylloxera, *Daktulosphaira vitifoliae*. Genetica **119**: 237–251	4	V_M and h^2_m were taken from Table 1. Line means were estimated from figure 3 (B =0.288, C = 0.235 for G12, and G24 survival, respectively). CV_M calculated from information provided.
Estes, S., B. 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	1	We calculated CV_M using V_M (taking the number of generations as 307) from Estes *et al.* (2005) and the mean from Ajie *et al.* (2005)
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	2	Table 1 of paper.
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	2	Trait means from Table 1 of paper, and variances from Table S4.
Fernández, J. and C. López-Fanjul, 1996 Spontaneous mutational variances and covariances for fitness- related traits in *Drosophila melanogaster*. Genetics **143**: 829–837	1	V_M and h^2_m taken from Table 3 & trait mean from Table1 of paper; CV_M calculated from those estimates.
Festing, M., 1973 A multivariate analysis of subline divergence in the shape of the mandible in c57bl/gr mice. Genet. Res. **21**: 121–132	3	Houle *et al.* (1996)'s Table 1.
Fry, J. D. and S. L. Heinsohn, 2002 Environment dependence of mutational parameters for viability in *Drosophila melanogaster*. Genetics **161**: 1155–1167	1	We calculated CV_M by dividing $\sqrt{V_M}/\text{mean}_{MA}\times 100$, where V_M was from Table 7, the mean_{MA} was from Table 1 of paper.
Fry, J. D., 2001 Rapid mutational declines of viability in *Drosophila*. Genet. Res. **77**: 53–60	1	Halligan & Keightley (2009)'s Table 1.
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	1	Halligan & Keightley (2009)'s Table 1.
García-Dorado, A., J. L. Monedero, and C. López-Fanjul, 1998 The mutation rate and the distribution of mutational effects of viability and fitness in *Drosophila melanogaster*. In *Mutation and Evolution,* pp. 255–265, Springer	1	Halligan & Keightley (2009)'s Table 1.
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	1	Table 2 of paper.
Hall, D. W., R. Mahmoudizad, A. W. Hurd, and S. B. Joseph, 2008 Spontaneous mutations in diploid *Saccharomyces cerevisiae*: Another thousand cell generations. Genet. Res. **90**: 229–241	4	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	5	CV_M was calculated from information in Table 3 of paper; among line variance was back-calculated using $4V_G$ (reported) and then divided by 2× the number of generations. h^2_m could not be calculated as no V_E estimates were reported.
Houle, D., D. Hoffmaster, S. Assimacopoulos, and B. Charlesworth, 1992 The genomic mutation rate for fitness in *Drosophila*. Nature **359**: 58–60	1	Halligan & Keightley (2009)'s Table 1.
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	3	Houle *et al.* (1996)'s Table 1.

Reference	Source	Description of where estimates were obtained from
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	2	h^2_m estimates were taken from supplementary information 1D. Estimates were averaged across sex. Could not calculate CV_M.
Joseph, S. B. and D. W. Hall, 2004 Spontaneous mutations in diploid *Saccharomyces cerevisiae* more beneficial than expected. Genetics **168**: 1817–1825	1	Halligan & Keightley (2009)'s Table 1.
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	2	Table 2 of paper.
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	2	Estimates of h^2_m (V_M/V_E) and CV_M (V_M/ mean for relative fitness) then taken as a percentage, were calculated using information from Table 1 of paper.
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	1	Estimates from information in Table 5 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	1	Estimates for productivity and *r* (intrinsic growth rate) from Halligan & Keightley (2009)'s Table 1. Lifespan estimates were taken from text on p. 3825 of paper.
Kibota, T. T. and M. Lynch, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature **381**: 694–696	1	Halligan & Keightley (2009)'s Table 1.
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	2	Table 1 of paper.
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	2	Authors did not specify exact generation of MA (taken as average of 93) and the reported estimates of V_M had not been divide by number of generations. V_M, h^2_m and CV_M provided in supplementary have all been recalculated using V_M/(2*93).
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	3	Houle *et al.* (1996)'s Table 1.
Lynch, M., 1985 Spontaneous mutations for life-history characters in an obligate parthenogen. Evolution pp. 804–818	3	Houle *et al.* (1996)'s Table 1.
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	4	Reference taken from in-text citations in Walsh & Lynch (2018); Estimates of h^2_m, V_M and V_E were taken from Table 1 of paper. MA mean calculated using z_0+R_m (Table 1), and used to calculate CV_M.
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	2	Table 9 of paper: Only the unselected lines were considered here, and we averaged estimates across sex.
Martorell, C., M. Toro, and C. Gallego, 1998 Spontaneous mutation for life-history traits in *Drosophila melanogaster*. In Mutation and Evolution, pp. 315–324, Springer	5	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	2	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 papaer.
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	2	Generation estimates from Table 2. Authors had repeated the fitness assays twice, where we report estimates of CV_M from the second attempt in Table 3. No estimates of V_E or h^2_M were reported.
Mukai, T., 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. spontaneous mutation rate of polygenes controlling viability. Genetics **50**: 1–19	3	Houle *et al.* (1996)'s Table 1.

Reference	Source	Description of where estimates were obtained from
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	1	Estimates taken from text of paper.
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	1	Halligan & Keightley (2009)'s Table 1.
Mukai,T., K. Harada,and H. Yoshimaru, 1984 Spontaneous mutations modifying the activity of alcohol dehydrogenase (adh) in *Drosophila melanogaster*. Genetics **106**: 73–84	3	Houle *et al.* (1996)'s Table 1.
Ohnishi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster* .II. homozygous effect of polygenic mutations. Genetics **87**: 529–545	1	Halligan & Keightley (2009)'s Table 1.
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	4	Table 2 of paper.
Pannebakker, B. 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	1	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.
Paxman, G., 1957 A study of spontaneous mutation in *Drosophila melanogaster*. Genetica **29**: 39–57	3	Houle *et al.* (1996)'s Table 1.
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	5	h^2_m, V_M and mean from Table 5 of paper and CV_M estimated from that information.
Roles, A. J., M. T. Rutter, I. Dworkin, C. B. 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	2	Table 2 of paper. Count traits were identified as outliers in preliminary analyses and excluded.
Russell, W., G. Sprague, and L. Penny, 1963 Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. **3**: 175–178	3	Houle *et al.* (1996)'s Table 1.
Rutter, M. T., F. H. Shaw, and C. B. Fenster, 2010 Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. Evolution **64**: 1825–35	2	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	3	Houle *et al.* (1996)'s Table 1.
Schoen, D. J., 2005 Deleterious mutation in related species of the plant genus *Amsinckia* with contrasting mating systems. Evolution **59**: 2370–2377	1	Halligan & Keightley (2009)'s Table 1.
Shaw, R. G., D. L. Byers, and E. Darmo, 2000 Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. Genetics **155**: 369–78	1	Estimates for number of fruits and seeds per fruit taken from Halligan & Keightley (2009)'s Table 1. Reproductive mass estimates were taken from Table 1 of paper.
Vassilieva, L. L., A. M. Hook, and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution **54**: 1234–1246	1	h^2_m was taken from Table 2. For CV_M, the mean of MA was calculated as the intercept plus the slope from Table 1, and V_M taken as the slope from Table 2, divided by 2 (diploid genome).
Vassilieva, L. L. and M. Lynch, 1999 The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics **151**: 119–129	1	h^2_M and V_M estimates were taken from Table 1 of paper. The MA mean calculated using $z_0 + R_m$ (Table 1), and used for calculating CV_M. We omitted CV_M estimate for Generation rate (1/T, the reciprocal of the mean age at reproduction) as estimates were an order of magnitude larger than any other estimate (MA_mean = 0.23, CV_M=93.796).
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	1	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	1	h^2_m estimates from Table 2 of paper. Insufficient information to calculate CV_M values.
Zeyl, C. and J. A. G. DeVisser, 2001 Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics **157**: 53–61	1	Halligan & Keightley (2009)'s Table 1.

Reference	Source	Description of where estimates were obtained from
(B) ESTIMATES REMOVED FROM META-ANALYSIS		
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); Removed *Caenorhabditis brenneri* heritability estimate for Productivity as the estimate was an outlier at 6.3 standard deviations from the mean.
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		As in (A); One estimate (for posterior centrosome oscillation frequency) was removed as the estimate for V_M was negative.
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 (2009)'s Table 1. However, these were an order of magnitude larger than any other CV_M estimate. Halligan & Keightley (2009) re-scaled theestimates to whole genome (× 5.5) and expressed as a proportion of the mean control, but nonetheless omit these values from their summary. We also chose to omit them here.
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.
Houle, D. and S. V. Nuzhdin, 2004 Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. Genet. Res. **83**: 7–18		Excluded from the meta-analysis as the estimate of mutational variance was derived from the slope of regression on generation, and the number of generations of mutation accumulation therefore undefined.
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		As in (A); Several estimates of mutational variance were negative, and these traits were excluded (Biomass, Bolting day, Fruit number, Height, Leaf length and Leaf number).
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		As in (A); Estimates for First instar duration, Spine length at maturity, Spine length at second adult instar, Third clutch size, and Total length at maturity excluded from analysis due to poor precision (all reported rounded to have the same value).
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		V_M and V_E estimates in Table 4 were pooled across generations of 10, 19 and 47, making the data unsuitable for meta-analysis of the effect of generation number. V_M was reported for each specific generation in Table 5, however corresponding estimates of V_E or trait mean were not reported.
Schultz, S. T., M. Lynch, and J. H. Willis, 1999 Spontaneous deleterious mutation in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. 96: 11393-11398		Estimates were taken from Table 3, where $\sqrt{CV_M}$ was calculated from V_B/z_0^2 reported in Table 3 as $\left(\sqrt{V_B}/\ z_0^2\right)/2$. Excluded from our analyses as the number of lines were a large outlier (8.7 SD above mean).

[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] Web of Science database. Journal article document types were searched for with the following search topics criteria: "mutation accumulation" and (varia* or "mutat* coefficient*"), on the 11/12/2019 at 4:38 PM AEST.

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

[4] Walsh, B. and M. Lynch, 2018 *Evolution and Selection of Quantitative Traits*. Oxford University Press

[5] Lynch, M., B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Sunderland, MA

Table S2. Broad and narrow trait categories. The trait description is a list of traits included within each category.

Broad Category	Trait	Trait Description
Life-history	Development time	Age at maturity, Age at reproduction, Growth rate, Instar duration, The reciprocal of the mean age at reproduction, Time to maturity
	Fitness	Colony size, Fitness, Competitive mating success, Phi (rate of convergence to the stable age distribution), Population growth, r (intrinsic rate of increase)
	Miscellaneous	Bolting day, Female longevity, Flowering day, Flowering timing, Lifespan, Longevity, Male longevity, Male mating, Sex ratio, Silking date
	Productivity	Clutch size, Fecundity, Grain yield, Litter size, Number of flowers, Ovariole number, Productivity, Reproductive mass, Tassel branches
	Survival	Age-specific survivorship, Germination, Pathogen susceptibility, Survival, Viability
Morphology	Morphology	Abdominal bristles, Biomass, Body length, Body size, Body width, Ear diameter, Ear length, Fruit size, Height, Leaf length, Leaf number, Limb lengths, Mandible length, Skull length, Spine length, Sternopleural bristles, Weight, Wing length
Physiology	Physiology	Day sleeping characteristics, Enzyme activity, Locomotor activity, Metabolites, Night sleeping characteristics, Velocity, Waking activity

Table S3. Mutation accumulation (MA) lines shared across studies. For each species, papers that share MA lines are grouped together, with the first published description of the MA lines indicated in bold, and non-bolded entries below that representing subsequent investigations of the same set of MA lines where either different traits were assayed, or the same trait after further accumulation.

Species name	Reference	Generation	Trait Description
Arabidopsis thaliana	**Shaw *et al.* (2000)**	17	Fecundity, Reproductive mass
	Chang & Shaw (2003)	17	Height, Leaf number, Fruit size
	Kavanaugh & Shaw (2005)	17	Bolting day, Flowering day, Fecundity, Height, Leaf length, Leaf number
	Rutter *et al.* (2010)	25	Fecundity
	Roles *et al.* (2016)	25	Fecundity, Germination, Age-specific survivorship, Fitness
	Rutter *et al.* (2018)	25	Biomass, Flowering timing, Fruit size, Germination, Survival
Caenorhabditis briggsae	**Baer *et al.* (2005)**	200	Productivity
	Baer *et al.* (2006)	220	Productivity
	Ostrow *et al.* (2007)	200	Body size
	Joyner-Matos *et al.* (2009)	250	Life span
Caenorhabditis elegans	**Baer *et al.* (2005)**	200	Productivity
	Baer *et al.* (2006)	220	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
	Farhadifar *et al.* (2016)	250	Mitotic cell division traits
	Keightley & Caballero (1997)	60	Lifespan, Productivity, Population growth parameters
	Azevedo *et al.* (2002)	60	Body size
	Vassilieva *et al.* (1999)	49.4	Longevity, Productivity, Survivorship, Population growth parameters
	Vassilieva *et al.* (2000)	214	Longevity, Productivity, Survivorship, Population growth parameters
	Azevedo *et al.* (2002)	152	Body size
	Ajie *et al.* (2005)	370	Locomotor activity, Productivity, Velocity
	Estes *et al.* (2005)	370	Population growth parameter, Body width
Drosophila melanogaster	**Houle *et al.* (1994)**	44	Fecundity, Fitness, Longevity, Male mating, Productivity
	Pletcher *et al.* (1998)	19	Female longevity, Male longevity
	Pletcher *et al.* (1999)	47	Female longevity, Male longevity
	Mackay & Lyman (1998)	223	Abdominal bristles, Sternopleural bristles
	Wayne & Mackay (1998)	228.5	Body size, Ovariole number
	Mukai (1964)	25	Viability
	Mukai & Yamazaki (1968)	60	Viability
	Santiago *et al.* (1992)	46	Sternopleural bristles, Wing dimensions
	Fernandez & Lopez-Fanjul (1996)	106	Viability, Fecundity
	Garcia-Dorado *et al.* (1998)	183	Fitness
	Martorell *et al.* (1998)	82	Productivity, Longevity
	Garcia-Dorado *et al.* (2000)	170	Wing length
		204	Abdominal bristles, Sternopleural bristles
		209	Viability
	Chavarrias *et al.* (2001)	250	Viability
		255	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
Saccharomyces cerevisiae	**Joseph & Hall (2004)**	1012	Fitness
	Hall *et al.* (2008)	2062	Fitness

Table S4. Statistical support for the effect of population size treatment on (A) among-line and (B) among-line by generation variance estimates. Model-fit information for model 5 (see main text) is reported per trait (defined in Figure 2), including: the number of parameters; the -2 log-likelihood score and; the Akaike information criterion (AIC). Model fit was assessed using a Likelihood Ratio Test (LRT) based on the difference in -2 log-likelihood scores between the model estimating treatment-specific variance, and the model estimating a single pooled estimate (i.e., the Chi-square statistic), where the degrees of freedom was the difference in the number of model parameters. Both the P-value from this LRT test (bold values <0.05), and the multiple-test corrected *P*-value (at a conservative 5% false discovery rate, FDR; significant tests in bold) are reported.

Trait	Model	Parameters	-2 Log likelihood	AIC	Chi-square statistic	*P*−value	5% FDR *P*−value
(A)	Model fit for estimating treatment-specific versus pooled *among-line* variance						
CS	Treatment-specific	28	-22247.9	-22191.9			
	Pooled	27	-22281.6	-22227.6	33.7	**0.000**	**0.000**
1.2	Treatment-specific	28	11381.7	11433.7			
	Pooled	27	11356	11406	25.7	**0.000**	**0.000**
1.5	Treatment-specific	28	9414.9	9466.9			
	Pooled	27	9338.7	9388.7	76.2	**0.000**	**0.000**
1.9	Treatment-specific	28	8626.1	8676.1			
	Pooled	27	8609.3	8657.3	16.8	**0.000**	**0.000**
2.5	Treatment-specific	28	10839.3	10887.3			
	Pooled	27	10818.2	10866.2	21.1	**0.000**	**0.000**
2.8	Treatment-specific	28	10598.6	10652.6			
	Pooled	27	10574.8	10626.8	23.8	**0.000**	**0.000**
3.7	Treatment-specific	28	9642.2	9694.2			
	Pooled	27	9603.5	9655.5	38.7	**0.000**	**0.000**
3.8	Treatment-specific	28	4803.1	4853.1			
	Pooled	27	4757.8	4805.8	45.3	**0.000**	**0.000**
4.5	Treatment-specific	28	9707.6	9755.6			
	Pooled	27	9668.3	9710.3	39.3	**0.000**	**0.000**
4.8	Treatment-specific	28	5102.4	5148.4			
	Pooled	27	5081	5125	21.4	**0.000**	**0.000**
5.6	Treatment-specific	28	7490.6	7540.6			
	Pooled	27	7480.9	7526.9	9.7	**0.002**	**0.002**
(B)	Model fit for estimating treatment-specific versus pooled *among-line by generation* variance						
CS	Treatment-specific	28	-22247.9	-22191.9			
	Pooled	27	-22242.4	-22188.4	5.5	**0.0190**	0.21
1.2	Treatment-specific	28	11381.7	11433.7			
	Pooled	27	11381.5	11431.5	0.2	0.6547	0.72
1.5	Treatment-specific	28	9414.9	9466.9			
	Pooled	27	9415.2	9465.2	0.3	0.5839	0.72
1.9	Treatment-specific	28	8626.1	8676.1			
	Pooled	27	8624.3	8672.3	1.8	0.1797	0.47
2.5	Treatment-specific	28	10839.3	10887.3			
	Pooled	27	10842.2	10890.2	2.9	0.0886	0.47
2.8	Treatment-specific	28	10598.6	10652.6			
	Pooled	27	10597.3	10649.3	1.3	0.2542	0.47
3.7	Treatment-specific	28	9642.2	9694.2			
	Pooled	27	9642.1	9694.1	0.1	0.7518	0.75
3.8	Treatment-specific	28	4803.1	4853.1			
	Pooled	27	4802.9	4852.9	0.2	0.6547	0.72
4.5	Treatment-specific	28	9707.6	9755.6			
	Pooled	27	9708.4	9754.4	0.8	0.3711	0.58

4.8	Treatment-specific	28	5102.4	5148.4			
	Pooled	27	5103.9	5149.9	1.5	0.2207	0.47
5.6	Treatment-specific	28	7490.6	7540.6			
	Pooled	27	7492.4	7538.4	1.8	0.1797	0.47

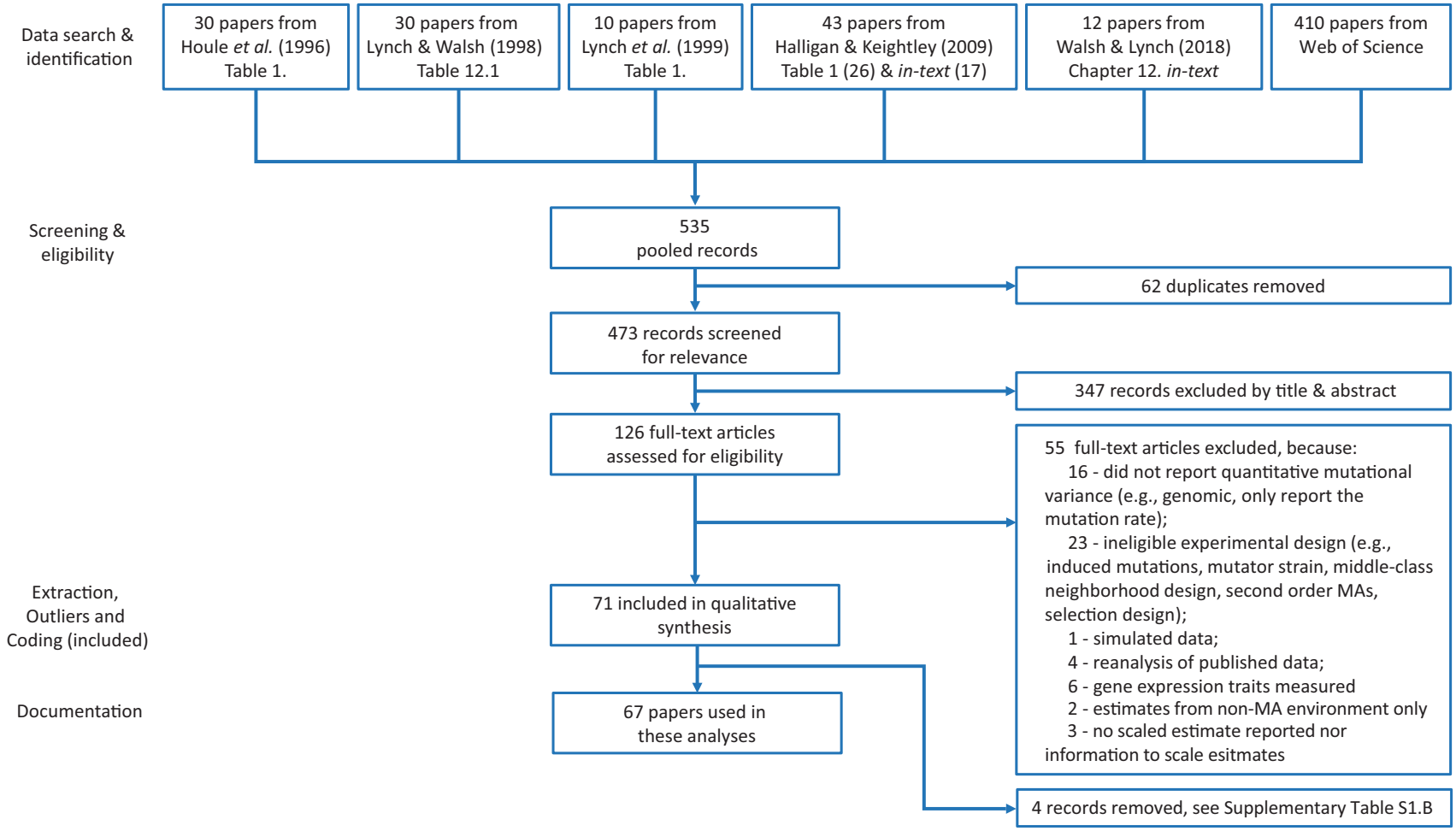

Figure S1. Diagram of systematic literature search for the meta-analysis of mutational variance estimates from mutation accumulation (MA) experiments. The below schematic follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis, MOHER *et al.* 2009) protocol. Further information, including mutational parameter estimates, available in Supplementary File (*Figshare DOI to be supplied at acceptance*).

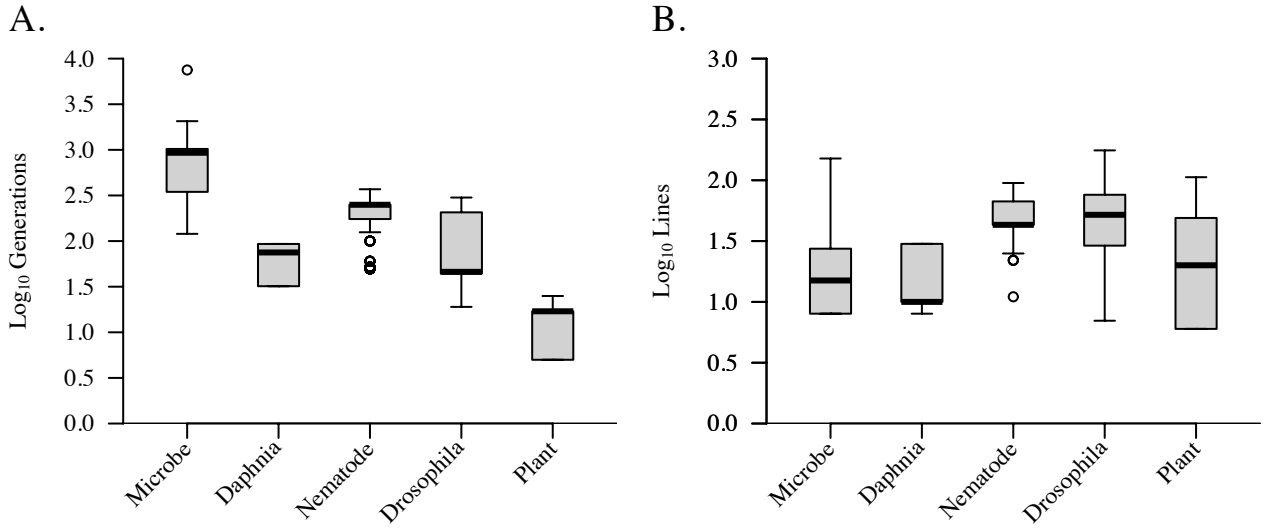

Figure S2. Distribution of the (log-transformed) number of generations (A) and the (log-transformed) number of lines (B) of mutation accumulation experiments in different taxon categories. For each box plot, the median of the distribution is shown as a black band, grey boxes represent the interquartile range, whiskers represent 1.5 times the interquartile range, and estimates that fell outside this range are designated with open circles.

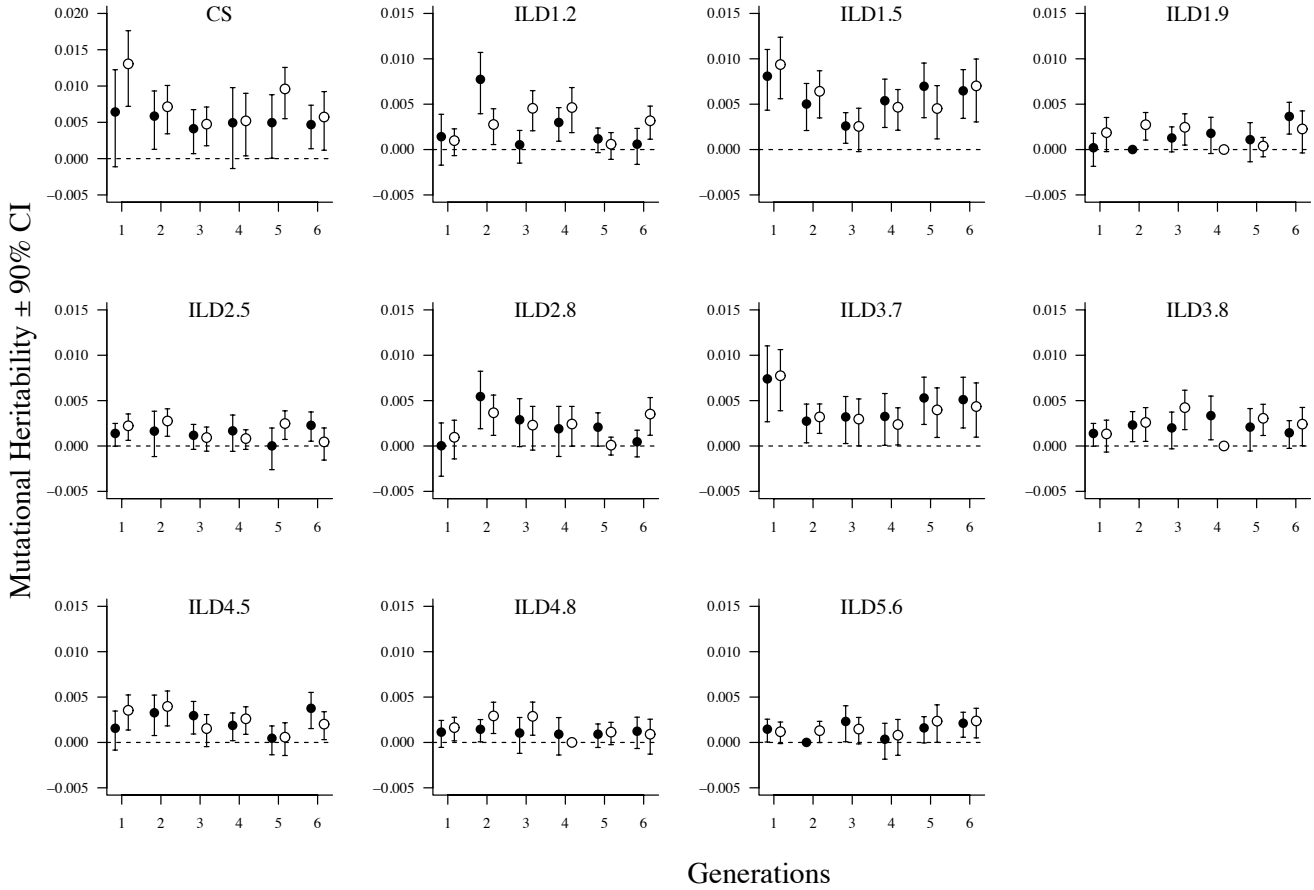

Figure S3. Mutational heritability estimates from a modified mutation accumulation experiment in *D. serrata*. Mutational heritability was estimated from the mutational variance ($\boldsymbol{V_M} = \frac{\boldsymbol{V_L}}{\boldsymbol{2t}}$, where V_L was the among-line estimate from the per trait per generation per treatment implementation of model (3), and t =20 generations of MA) divided by the corresponding estimate of residual error (from model (3)). The heritability estimate is plotted for each trait (panel; trait descriptions in Figure 2) in each generation (x-axis) for each of the two population size treatments (small: solid circles; big: open circles). Points are the scaled REML estimates, while the whiskers indicate the REML-MVN 90% confidence intervals (CI). The dashed horizontal line indicates zero: Estimates for which the lower CI did not overlap zero were interpreted as statistically supported.

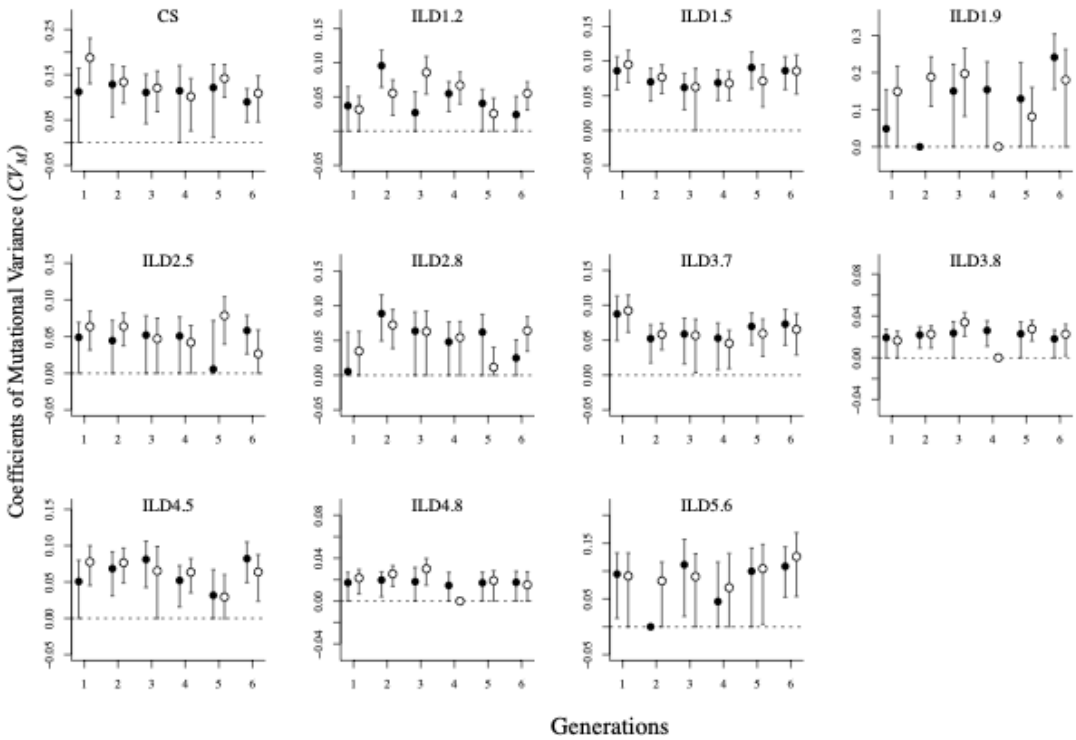

Figure S4. Coefficients of mutational variance ($\boldsymbol{CV_M}$) estimates from a modified mutation accumulation experiment in *D. serrata*. Among-line variances were estimated independently for each trait (panel; trait descriptions in Figure 2) in each generation (x-axis) for each of the two population size treatments (small: solid circles; big: open circles), and placed on a $\boldsymbol{CV_M}$ scale using the following formula: $\sqrt{\frac{V_M}{\bar{x}}} \times \mathbf{100}$ where V_M is as described in Figure S4, and $\bar{x}$ is the trait least squares mean (Figure 5). Plotted are the REML point estimate, and the 90% confidence intervals (CI) around this, estimated using 10,000 REML-MVN samples of $\boldsymbol{V_L}$ and the trait mean as described in the Methods. Some estimates of V_M were negative and $\boldsymbol{CV_M}$ could therefore not be calculated. The dashed horizontal line indicates zero.