1 Causes of variability in estimates of mutational variance from mutation

2 accumulation experiments

3

4 Cara Conradsen^{1*}, Mark W. Blows¹ and Katrina McGuigan¹

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

6

7 Running Title: Variability of mutational variance

8

9 Keywords: meta-analysis, Drosophila, life history, morphology, heritability, coefficient of variance,
10 wing, micro-environment, measurement error

11 *Corresponding Author: Cara Conradsen

12

13 School of Biological Sciences

14 The University of Queensland;

15 St. Lucia, QLD, 4072 AUSTRALIA

16 T: +61(0)450645318

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

18

19 ABSTRACT

20 Characteristics of the new phenotypic variation introduced via mutation have broad implications in
 21 evolutionary and medical genetics. Estimates of this mutational variance, V_M , span two orders of
 22 magnitude, but factors contributing to this range remain poorly resolved. We applied meta-analyses
 23 to ~200 estimates from 62 studies to assess how previously implicated factors affect the magnitude of
 24 V_M . While estimates are available from a range of taxa, and quantitative trait types over a range of
 25 timescales, these factors are confounded with one another. We call for further directly comparable
 26 empirical data to resolve these influences on variability of V_M . We also analysed data from
 27 *Drosophila serrata* lines, derived from a mutation accumulation experiment, observing substantial
 28 heterogeneity among replicated estimates of among-line variance for 11 morphological (wing) traits.
 29 Micro-environmental variation in mutational effects was implicated as causing variability in
 30 mutational variance for just two of 11 traits, suggesting estimation environment may contribute little
 31 to variability of published estimates. Varying the opportunity for evolutionary processes to increase
 32 within-line genetic variation did not affect among-line variance estimates, suggesting transient
 33 segregation of mutations is also an unlikely general explanation of variability in published estimates.
 34 We conclude that estimation error, expected to be relatively high due to the low differentiation
 35 among lines diverging solely through new mutation, is likely to contribute strongly to variability of
 36 estimates of mutational variance. Analyses of repeated measures of mutation accumulation lines can
 37 improve the precision of estimates, consequently expanding our understanding of the dynamics of
 38 mutations in natural populations.

39

40 INTRODUCTION

41 The magnitude of per-generational increase in genetic variance due to spontaneous mutations (V_M) is
 42 important for a wide range of genetic and evolutionary phenomena, including the maintenance of
 43 quantitative genetic variance (LYNCH 1988b; BARTON AND TURELLI 1989; JOHNSON AND BARTON
 44 2005). Much of our understanding of V_M comes from mutation accumulation (MA) experiments,
 45 where populations diverge phenotypically due solely to the fixation of new mutations (MUKAI 1964;
 46 HALLIGAN AND KEIGHTLEY 2009). Estimates from MA experiments in a wide range of traits and taxa
 47 indicate that mutation increases the variance in quantitative traits by 10^{-4} to 10^{-2} times the
 48 environmental variance of the trait per generation (HOULE *et al.* 1996; LYNCH AND WALSH 1998;
 49 LYNCH *et al.* 1999; HALLIGAN AND KEIGHTLEY 2009). This range of values (two orders of
 50 magnitude) suggests variation in either the rate or the average magnitude of effects of new mutation,
 51 potentially leading to differences in the magnitude of standing quantitative genetic variation and,
 52 ultimately, rates of phenotypic evolution (HOULE 1998; LYNCH *et al.* 1999; HOULE *et al.* 2017;
 53 WALSH AND LYNCH 2018). However, the causes of variation in estimates of V_M , and thus the
 54 evolutionary interpretation of this variability, are not well resolved.

55 A relatively simple potential cause of variation in V_M estimates is scale. HOULE *et al.* (1996)
 56 showed that life history traits have lower mutational heritability ($h_M^2 = V_M / V_E$, where V_E is the
 57 environmental variance) than morphological traits, but the reverse pattern holds when comparing
 58 coefficients of mutational variance (CV_M ; $100 \times \sqrt{V_M} / \bar{X}$, where $\bar{X}$ is the trait mean). While the
 59 pattern is reversed on the different scales, the variation in the magnitude of mutational variance
 60 between trait types nonetheless persists. Life-history traits have been hypothesised to be affected by
 61 more loci than morphological traits, with this larger target of mutation resulting in a higher mutation
 62 rate and greater mutational variance (HOULE 1991; HOULE 1992; HOULE *et al.* 1996; HOULE 1998;
 63 MERILÄ AND SHELDON 1999). However, the exact role mutational target size plays in generating

64 variability of V_M is difficult to assess, as a full catalogue of causal loci is unknown for most
 65 quantitative traits (BARTON AND KEIGHTLEY 2002; MACKAY *et al.* 2009; YANG *et al.* 2010;
 66 ROCKMAN 2012). Indeed, emerging evidence that diverse traits are all highly polygenic (YANG *et al.*
 67 2010; BOYLE *et al.* 2017) suggests that differences in the distribution of mutational effect sizes
 68 (SIMONS *et al.* 2018), rather than simply in number of contributing loci, might cause heterogeneity in
 69 estimates of V_M .

70 Mutation rate is known to vary widely among taxa (reviewed in KATJU AND BERGTHORSSON
 71 2019), which may reflect differences in ploidy levels, genome size, generation time, and/or effective
 72 population size (LYNCH *et al.* 1999; LYNCH 2010; SUNG *et al.* 2012). Marked variation in mutation
 73 rate has also been observed within species, both among replicated MA experiments (i.e., different
 74 founder genotypes: NESS *et al.* 2012; SUNG *et al.* 2012; SCHRIDER *et al.* 2013) and among lines
 75 within a single MA panel (HUANG *et al.* 2016). This heterogeneity may explain observed variation in
 76 V_M estimates for the same trait in the same taxon. For example, h_M^2 of body size in *Caenorhabditis*
 77 *elegans* ranged from 0.0035 to 0.0076 between two strains (genetic isolates) (OSTROW *et al.* 2007),
 78 and from 0.0011 to 0.0044 among four sets of MA lines established from the same strain (AZEVEDO
 79 *et al.* 2002; ESTES *et al.* 2005; OSTROW *et al.* 2007). Different combinations of rate and effect size
 80 have been inferred in different experiments despite highly similar h_M^2 values (Table 1 in HALLIGAN
 81 AND KEIGHTLEY 2009), again suggesting the need to disentangle the complex relationships between
 82 allelic effect sizes and genomic mutation rates to understand the causes of the observed heterogeneity
 83 in V_M .

84 Ultimately, the random nature of mutation-drift processes themselves might contribute to
 85 variability of estimates, where it is the distribution of estimates obtained from multiple experiments
 86 that is most informative of the expected input of mutational variance. The extent to which
 87 experimental error or biologically relevant sampling variance contribute to variability in estimates

might be revealed by further consideration of the MA experimental design. Firstly, sampling the mutation-drift process using a finite number of lines will yield variation in estimates, representing variation in the realisation of the mutation-drift process. LYNCH (1988a) recommended that the number of replicate lines (L) should be greater than 400 for selfing and 300 for full-sib lines (the two most common modes in bottlenecking in MA designs) to ensure less than 10% error in the estimate of among-line variance. However, maintaining a large number of lines simultaneously is often logistically prohibitive, and this recommendation is rarely met, with experiments typically considering far fewer (< 50) lines (see Methods).

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

Environmental context within which MA lines are assayed could also contribute variation to estimates of V_M . Several studies have considered the effect of macro-environmental variation (i.e., known, replicable, experimenter-imposed, variation), including temperature (WAYNE AND MACKAY 1998), light (KAVANAUGH AND SHAW 2005), density (FRY *et al.* 1996) or food availability (LATTA *et*

12 *al.* 2015). Although the magnitude of mutational variance often varies among environments, there is
 13 only weak evidence for predictable patterns, such as novel or stressful conditions increasing the
 14 magnitude of variance (KONDRASHOV AND HOULE 1994; MARTIN AND LENORMAND 2006). Even in
 15 carefully controlled laboratory experiments, micro-environmental conditions (due to uncontrolled,
 16 un-replicated, and typically unknown sources) contribute to phenotypic variation (LYNCH 1988a;
 17 FALCONER AND MACKAY 1996). If all MA lines respond similarly, micro-environmental variation
 18 will not [affect](#) among-line variance; if lines differ in their response, then this genotype by micro-
 19 environment variance could contribute undetected variability to estimates of V_M (LYNCH 1988a;
 20 GARCÍA-DORADO *et al.* 2000). The magnitude of inter-generation variation of among-line estimates
 21 has rarely been assessed (GARCÍA-DORADO *et al.* 2000) and the relative contributions to any such
 22 variation from micro-environment, measurement error and segregation remain unknown.

23 Here, we combined two approaches to investigate causes of variability in estimates of
 24 mutational variance. Given that it has been over twenty years since causes of variability in mutational
 25 variance have been broadly investigated (HOULE *et al.* 1996; HOULE 1998; LYNCH *et al.* 1999), we
 26 first conducted a meta-analysis to update tests of the previously implicated causal factors of taxon
 27 (LYNCH AND WALSH 1998; LYNCH *et al.* 1999; HALLIGAN AND KEIGHTLEY 2009), trait type (HOULE
 28 *et al.* 1996; HOULE 1998) and experimental duration (generation number) (LYNCH AND HILL 1986;
 29 MACKAY *et al.* 1995). Multiple estimates of V_M were available for different values of these factors,
 30 allowing us to test the hypothesised effects, although confounding across the factors resulted in little
 31 insight into how these factors affect estimates of V_M . Second, we conducted a new empirical
 32 experiment in *Drosophila serrata* to disentangle the effects of transient segregation, micro-
 33 environment and sampling error, overcoming the limitation on meta-analysis of too few studies of the
 34 same taxon, trait and macro-environment to address these potential causes of estimate variability. Our
 35 results suggest that sampling error is likely to contribute markedly to heterogeneity of published
 36 estimates, but that repeated measures can mitigate these effects.

37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59

METHODS

Meta-analysis of empirical estimates of mutational variance

Literature search

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

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

the number of mutation accumulation generations was not reported. After applying these criteria, 69 papers were retained.

Meta-analysis data collection

For each of the 69 papers, mutational parameter estimates were extracted, associated with the identity of the taxon and trait (Figure 1), number of generations of mutation accumulation, and the number of lines sampled (Figure S2). Where studies reported estimates from multiple environments, only estimates from the MA maintenance environment were extracted. Papers reported estimates of mutational variance for a median of three traits (range: 1–29). Where possible, we extracted (or calculated from provided information) both the coefficient of variance (CV_M ; $100 \times \sqrt{V_M}/\bar{X}$, where $\bar{X}$ is the trait mean) and mutational heritability (h_M^2 ; V_M/V_E) for each trait. As detailed in Table S1, after further inspection we excluded extreme estimates, and those reported with very low precision (Table S1.B).

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

Predictor variables for the meta-analysis

Estimates came from 22 species, and based on the distribution of estimates, we defined six taxon categories: *Drosophila* (*Drosophila melanogaster* [$n = 89$] and *D. serrata* [$n = 15$]), plant (*Arabidopsis thaliana* [five studies, $n = 15$]; cumulatively three studies and $n = 15$ from *Amsinckia douglasiana*, *A. gloriosa*, *Hordeum vulgare* and *Zea mays*), microbe (*Cryptococcus neoformans*,

83 *Escherichia coli*, *Chlamydomonas reinhardtii*, *Saccharomyces cerevisiae*, *Cystovirus pseudomonas*
84 *virus* $\phi 6$ and *Vesicular stomatitis virus*), nematode (*Caenorhabditis briggsae*, *C. brenneri*, *C.*
85 *elegans*, *C. remanei* and *Oscheius myriophila*), and *Daphnia* (*Daphnia pulex*) (Figure 1A). We
86 excluded from analyses estimates from two other insects (*Daktulosphaira vitifoliae* and *Nasonia*
87 *vitripennis*) for clarity of inference (insect versus *Drosophila*), and from *Mus musculus*, the only
88 vertebrate, due to low sample size (four estimates: Figure 1A).

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

:01 *Meta-analyses of mutational variance estimates*

:02 Our final meta-analysis data set consisted of 201 estimates of h_M^2 and 211 estimates of CV_M from 62
:03 studies. We implemented the following model (separately for h_M^2 and CV_M data) via PROC MIXED
:04 in SAS v.9.4 (SAS Institute Inc., Cary, NC.), using maximum likelihood:

$$:05 \quad y_{ijklmn} = \mu + tx_i + trt_j + g_k + l_l + s_m + \varepsilon_{ijklmn} \quad (1)$$

where y is the vector of published estimates, and μ was the overall mean h_M^2 or CV_M . We fit the categorical predictors taxon (tx , $i=1, 2, \dots, 5$) and trait (trt , $j=1, 2, \dots, 7$), and the continuous predictors generation number (g) and number of lines sampled (l) as fixed effects. The number of MA lines and generations spanned different (non-overlapping) ranges among taxa (Figure S2), and interactions were therefore not fit. Study (s) was fit as a random effect, accounting for non-independence among estimates within a paper. The same MA lines were sometimes analysed in multiple studies, for different traits, or after further generations of accumulation (Table S3). However, an additional random effect of shared lines (with study nested within this) did not account for any variance, potentially a consequence of the unbalanced design, with only some studies sharing lines. The residual error (ϵ) captured all unexplained variation among estimates. We used the Satterthwaite approximation to correct the denominator degrees of freedom.

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

Quantitative genetic parameters are associated with large sampling errors (KLEIN *et al.* 1973; KLEIN 1974; LYNCH AND WALSH 1998), and the relatively low signal (i.e., few genetic differences) among MA lines will make them particularly vulnerable to “noisy” estimation. We conducted a new empirical experiment to quantify the contribution of this estimation error relative to the effects of other sources of random variation, particularly the chance sampling of different genotypes within inbred lines (i.e., segregating mutations in the process of becoming fixed or lost), and micro-environmental variation due to the specific time and place of sampling. We conducted the experiment in *Drosophila serrata*, a member of the *montium* species group, endemic to Australia and Papua New Guinea, that has been extensively used in quantitative genetic research, including study of mutational variance (e.g., MCGUIGAN AND BLOWS 2013; MCGUIGAN *et al.* 2014; HINE *et al.* 2018; DUGAND *et al.* 2021).

Experimental design

A panel of 200 MA lines were founded from one of the *D. serrata* Reference Genome Panel (DsRGP) lines described in REDDIEX *et al.* (2018). The MA lines were each maintained by brother-sister inbreeding for 20 generations, following husbandry protocols previously established by MCGUIGAN *et al.* (2011) to minimise extinction and within-line selection. Given the relatively short duration of mutation accumulation, among-line variation may be inflated by segregating variation in the base population (LYNCH AND HILL 1986). Notably, genome-wide heterozygosity was very low (0.3%) in the DsRGP line that founded the MA lines, and among-line variance for wing phenotypes not statistically supported in the first generation of mutation accumulation (S. Chenoweth, pers. comm.). This panel of lines are therefore representative of the relatively low among-line variance characteristic of typical MA lines, and are well-suited to investigating the relative potential of estimation error versus other factors to generate variability in estimates of mutational variance.

We randomly chose 42 of the MA lines for further experimentation to partition putative causes of variation in estimates. Our experimental design had three main goals. Firstly, we aimed to retard ongoing mutation accumulation to maintain the low among-line variance, and achieved this by increased population size within each of the lines. The fate of a new mutation is determined by drift when its fitness effect is less than $1/2N_e$ (WRIGHT 1931; KIMURA 1983). Empirical evidence suggests that population sizes as low as ten may be sufficient to prevent fixation of mutations with deleterious fitness effects (ESTES *et al.* 2004; KATJU *et al.* 2015; LUIJCKX *et al.* 2018). We imposed a minimum population size of 12 males and 12 females (Figure 2A). This change in census population size complicates calculation of a per-generation rate of mutational increase in phenotypic variance; here, we instead focus on the among-line variance, V_L , the parameter estimated for calculating V_M (LYNCH AND WALSH 1998).

The second aim of our experimental design was to vary the opportunity for mutation-selection-drift operating within a line, and hence, the dynamics of segregating genetic variation; we achieved this by imposing two, substantially different, population sizes on sublines of each of the 42 MA lines ($N = 24$ versus 288, referred to hereafter as small, S, and large, L, population size treatments: Figure 2A). These treatments contrast deterministic evolution of mutations with relatively strong ($s > 0.029$: $N_e \sim 17$), versus weak ($s > 0.003$: $N_e \sim 202$) fitness effects (effective population size estimated as 70% of census size, CROW AND MORTON 1955), and thus a different opportunity of within-line variance to increase during the six generations, and for this segregating variance to influence estimates of among-line variance.

The final aim of the experimental design was to determine how variability in the effects of new mutations under different random micro-environments impact among-line variance estimates. Environmental variation contributes to phenotypes even under tightly controlled laboratory conditions (LANDE 1976; FALCONER AND MACKAY 1996); we randomly sampled the environmental conditions present within our laboratory by repeatedly sampling the lines (genotypes) over six consecutive generations. Thus, our experiment consisted of applying two population size treatments (S, L) to each of 42 lines (derived from a classical MA, with low among-line variation), where these 84 lines were maintained under the same micro-environment conditions (12 flies per sex per vial founding each generation, with S and L differing in the number of vials) for six generations (Figure 2A).

Data collection

Each generation, 12 males per subline (six per rearing vial) were randomly sampled for wing phenotypes (Figure 2). Wings were mounted on microscope slides and photographed using a Leica MZ6 microscope camera and the software LAS EZ v2.0.0 (Leica Microsystems Ltd, Switzerland). A total of 5,135 wings were landmarked for nine positions, defined by wing vein and margin

intersections (Figure 2B), using the software tpsDIG2 (ROHLF 2013). The number of wings were evenly distributed across treatments (2,583 in S and 2,552 in L) and generations (~8.3% per generation, per treatment). Landmarks were aligned using a General Procrustes fit in tpsRelw (ROHLF 2007). Centroid size (CS), the square root of the sum of squared deviations of the coordinates from the centroid (ROHLF 1999), was recorded as a metric of wing size. The aligned X-Y coordinates for each landmark were then used to calculate ten inter-landmark distances (ILDs) (Figure 2B). ILDs scores were re-scaled prior to analysis (multiplied by 100) to aid model convergence. Outliers >3.0 SD from the mean were removed for each of the 11 traits (10 ILD and size) (329 measures across the 56,485 total measures).

Analyses of trait means

Micro-environmental variation can generate both temporal trends and idiosyncratic changes among generations (LANDE 1976; FALCONER AND MACKAY 1996), but here is expected to have consistent effects on the two synchronously reared population size treatments. In contrast, genetic divergence, due to independent mutation and subsequent drift or selection within each of the 84 subpopulations could lead to increasing divergence between the treatment means if mutational effects on either the measured traits, or pleiotropic effects on fitness are biased toward a higher or lower value (JOHNSON AND BARTON 2005; MCGUIGAN AND BLOWS 2013). To test the null hypothesis that wing trait mean remained constant across generations and treatments we implemented the following model using PROC MIXED in SAS v9.4 (SAS Institute Inc., Cary, NC.), within a maximum-likelihood framework:

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

where y_{ijklm} was the trait value for the m th wing, from the l th vial, within the k th line, for the j th treatment in the i th generation. The model fits the overall mean (μ) and the fixed effects of generation

(G), treatment (T), and their interaction (GT), as well as random effects due to MA line (Line), replicate rearing vials (Vial) within each line, and residual error (ϵ), capturing unexplained variance among individual flies within a rearing vial. Each line was sampled in all generations and we accounted for this repeated measures design, fitting a reduced-rank factor analytic covariance structure (specifying TYPE=FA(1)), which estimated generation-specific among-line variance (a diagonal matrix) and one dimension of among-generation covariance (i.e., among-line variance shared across repeated measures of generations) (MEYER 2009; MCGUIGAN AND BLOWS 2010). Preliminary analyses indicated that this FA(1) model captured all among-line variance. Vial and residual variances were estimated for each generation and treatment, accounting for potential heterogeneity in micro-environmental effects, and segregating variation.

Analyses of among-line variance

To investigate the effects of evolutionary processes (captured by our two population size treatments) and micro-environment (captured by the six generations) on estimates of among-line variance, we first treated data from each generation and treatment as independent experimental estimates for each trait, allowing us to characterise the variability in estimates (12 estimates per trait, 132 estimates in total). This was done by reducing model 2 to:

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

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 $\mathbf{V}$ was the asymptotic variance-covariance matrix, the inverse of the Fisher information matrix, $\mathbf{I}(\hat{\theta})^{-1}$ (MEYER AND HOULE 2013; HOULE AND MEYER 2015; SZTEPANACZ AND BLOWS 2017). 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. We are interested here only in general patterns of variability among estimates and do not correct for multiple testing.

Next, to formally test whether among-line variance differed among generations 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 ($G \times \text{Line}$), estimating the variation in genetic effects among generations (GARCÍA-DORADO *et al.* 2000):

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

This model was applied to each trait and population size treatment separately (22 models), with vial and residual variances estimated separately for each generation. Statistical support for $G \times \text{Line}$ and for Line was determined using log-likelihood ratio tests (LRT, 0.5 d.f.: SELF AND LIANG 1987; LITTELL *et al.* 2006) to compare model 4 to a reduced models that did not fit $G \times \text{Line}$, or did not fit Line. We applied the Benjamini-Hochberg method (BENJAMINI AND HOCHBERG 1995) to correct for multiple hypothesis testing (within each random effect), using a conservative 5% false discovery rate (FDR).

To formally test the hypothesis that differences in the opportunity for mutation-drift-selection between the Large and Small population size treatments affected the estimates of among-line variance, or the among-line by generation variance, we ran the following model for each of the 11 traits, again using PROC MIXED and REML, where we fit the intercept (μ), generation (G),

46 treatment (T) and the interaction between generation and treatment (GT), as well as the random
47 effects of Line, $G \times \text{Line}$, Vial and error:

$$48 \quad y_{ijklm} = \mu + G_i + T_j + GT_{ij} + \text{Line}_k + (G \times \text{Line})_{ik} + \text{Vial}_{l(ijk)} + \varepsilon_{ijklm} \quad (5)$$

49 Although estimates of among-line by generation variance were typically not statistically
50 distinguishable from zero (see Results), we nonetheless modelled it here to ensure that small,
51 statistically undetectable, variation did not inflate the among-line variance of interest. Generation and
52 treatment-specific vial and residual variances were estimated. We used LRTs to determine whether
53 estimating among-line variance separately for the two population size treatments (modelled using a
54 Group statement in PROC MIXED) improved model fit over modelling a single, pooled, estimate,
55 again applying the Benjamini-Hochberg method (BENJAMINI AND HOCHBERG 1995) to allow a 5%
56 FDR. For the one trait with statistically supported $G \times \text{Line}$ variance in at least one population size
57 treatment (see Results), we again used a LRT to test the hypothesis that Large and Small population
58 treatments differed in the magnitude of this $G \times \text{Line}$ variance; a 5% FDR correction was applied
59 assuming 11 traits were tested.

60 Finally, to infer whether the magnitude of among-line (V_L) variance differed among the 10
61 different wing shape traits, we took the V_L estimates from model (5), estimated from the data from
62 both treatments, and obtained REML-MVN confidence intervals as described above. We regressed
63 the among-line variance on phenotypic variance (V_P : sum of among-line, among-vial and residual)
64 for the REML estimate and each of the 10,000 REML-MVNA samples. Similarly, we regressed the
65 REML V_L estimates on the trait mean (estimated over all generations and treatments), and the REML-
66 MVN samples of V_L on 10,000 samples of the trait mean taken from a MVN distribution with the
67 observed mean and standard deviation, sampled using the MASS package (VENABLES AND RIPLEY
68 2002). Where 95% CI of the regression slope did not include zero, we concluded the parameter (V_P or
69 mean) significantly predicted V_L .

70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89

RESULTS

Meta-analysis of published mutational variance estimates

Unexpectedly, we observed mutational heritability to have a similar magnitude across all trait types ($F_{6,161} = 2.01$, $P = 0.0670$; Figure 3A). In contrast, CV_M significantly varied among trait types ($F_{6,151} = 15.89$, $P < 0.0001$), following the predicted pattern of morphological traits having smaller magnitude CV_M than life history traits (Figure 3B). The number of lines sampled did not affect the magnitude of mutational variance (h_M^2 : $F_{1,68.6} = 0.83$, $P = 0.3658$; CV_M : $F_{1,70.6} = 0.27$, $P = 0.6076$). Taxon significantly predicted the magnitude of mutational variance, on both scales (h_M^2 : $F_{4,54.5} = 3.12$, $P = 0.0221$; CV_M : $F_{4,56.5} = 6.13$, $P = 0.0004$), with plants having the lowest model estimates of mutational variance and Nematodes the highest (Figure 4). The magnitude of mutational variance significantly decreased with increasing number of generations of mutation accumulation (h_M^2 : $F_{1,58.7} = 12.06$, $P = 0.0010$, $\beta = -0.622$; CV_M : $F_{1,65.6} = 21.05$, $P < 0.0001$, $\beta = -1.065$). However, as plants and nematodes differed markedly in generation number (median of 17 and 250 generations, respectively; Figure S2A), the effect of taxon versus generation number on the magnitude of V_M cannot be clearly inferred. To further investigate this, we analysed V_M separately for each of the two best represented taxonomic groups, *Drosophila* (19 – 300 generations; Figure S2A) and Nematode (50 – 370 generations; Figure S2A). In *Drosophila*, there was a significant negative relationship between number of generations and the magnitude of mutational variance (h_M^2 : $F_{1,32} = 7.76$, $P = 0.0089$, $\beta = -0.486$, $N = 56$; CV_M : $F_{1,20.1} = 28.05$, $P < 0.0001$, $\beta = -1.299$, $N = 73$). In contrast,

generation number did not explain the magnitude of mutational variance in Nematodes (h_M^2 : $F_{1,14.1} = 1.38$, $P = 0.2600$, $\beta = -0.337$, $N = 86$; CV_M : $F_{1,18.9} < 0.00$, $P = 0.9708$, $\beta = -0.023$, $N = 76$).

Variation in estimates of among-line variance within the same taxon, trait type and duration:

An experiment in *Drosophila serrata*

While there was statistically significant generation by treatment effects on most *D. serrata* wing traits, these appeared as idiosyncratic changes in direction and magnitude of differences between population size treatments (Figure 5), more consistent with effects of micro-environmental fluctuations than evolving phenotypic differentiation. One trait differed in mean between the treatments (ILD4.8: $F_{1,711} = 25.93$, $P = 0.0001$), but again there was no evidence of increasing magnitude of divergence through time (Figure 5), expected under evolution.

Next, we investigated the variation in among-line variance, treating our data as representing 12 different "experiments" for each of our 11 traits. There was inconsistent statistical evidence for among-line variance within each trait, where 60 out of 132 estimates (45.45%) were not significantly different from zero (Figure 6). We could draw very different conclusions about the magnitude of among-line variance, depending on which "experiment" we had conducted (Figure 6). For example, among-line variance was estimated to be >2 times higher in the first than third generation in the analyses of wing (centroid) size (CS) in the large population size treatment (3.43×10^{-4} and 1.38×10^{-4} , respectively; Figure 6). Marked fluctuations in the magnitude of among-line variance persisted when accounting for heterogeneity among generations in the magnitude of phenotypic variance or trait mean (Supplementary Figures S3 and S4, respectively).

Among-line variance did not exhibit a directional trend over time for any trait (Figure 6), and variation in mutational effects among generations (among-line by generation variance: model 4) was statistically supported for only four of the 22 estimates (two treatments by eleven traits), with only two estimates remaining significant at a 5% FDR (LRT for CS in L: $\chi_1^2 = 13.9$, $P = 0.0001$; LRT for

ILD2.5 in L: $\chi^2_1 = 11.1$, $P = 0.0005$; Figure 7A). Consistent with the largely overlapping CI of the two treatments (Figure 6), a difference between the Small and Large population size treatments in among-line by generation variance was detected for only one trait (centroid size), and this did not remain significant at the 5% FDR (Table S4B). Furthermore, there was no statistical support for an effect of the population size treatment on the among-line variance for any trait (Figure 7B; Table S4A). The variability in estimates (Figure 6) was thus typically within the magnitude of the error associated with each estimate.

Given these results, we utilised all data (six generations for two treatments) to obtain a single, statistically supported, estimate of among-line variance for each trait (Figure 8A). Utilising these robust estimates (and excluding size), we found that the total phenotypic variance ($\beta = 0.075$, 95% CI: 0.050, 0.106), or the magnitude of the trait mean ($\beta = 1.689 \times 10^{-6}$, 95% CI: 0.406×10^{-6} , 3.026×10^{-6}) significantly predicted the observed magnitude of among-line variance (Figure 8B). Nonetheless, substantial heterogeneity in estimates remained unexplained (Figure 8B; for V_L on V_P , $R^2 = 0.286$ [95% CI: 0.141-0.477]; for V_L on mean, $R^2 = 0.029$ [95% CI: 0.002-0.076]), suggesting either inherent stochasticity of the mutation process, or differences in genetic architecture of the traits affected estimates of among-line variance.

DISCUSSION

Although numerous estimates of mutational variance (V_M) have been published, it remains unclear why they vary in magnitude. Our meta-analytic investigation revealed substantial confoundment between potential causal factors, suggesting that further empirical data must be collected to fully resolve the contribution of these general factors to variation in V_M . Mutation is an inherently random process, but the expected variance about the true V_M due to stochastic sampling of newly segregating mutations, from micro-environment specific effects of mutations, or from estimation error, is not

known. Analyses of data from a manipulative experiment in *D. serrata* suggests that estimation error may contribute most generally to variation in estimates. Below we discuss the specific outcomes and limitations of both approaches, and the implications our analyses have for future work characterising mutational input to quantitative genetic variation.

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

Among eight multicellular eukaryotes, the reported per nucleotide mutation rate is highest for *Arabidopsis* and lowest for the nematode *C. briggsae* (KATJU AND BERGTHORSSON 2019). Consistent with this, we observed a significant effect on both CV_M and h_M^2 of taxon category, with plant (predominantly estimates from *Arabidopsis*) and Nematode (predominantly estimates from *Caenorhabditis* species) differing most in the magnitude of estimates. However, experiments in plants had the shortest duration of mutation accumulation (Figure S2A), where our analyses also supported a significant decline in V_M as the number of MA generations increased. Thus, the apparent elevation of V_M for plants could be a consequence of short duration, or the apparent decline in V_M over time could be an artifact, (partly) caused by the high V_M in plants. We observed a significant decline in the magnitude of V_M with number of MA generations in *Drosophila*, but not in Nematodes. Decline in V_M over time has been reported previously in *D. melanogaster*, but only for some assayed traits (MACKAY *et al.* 1995; GARCÍA-DORADO *et al.* 2000). Therefore, while the data suggest that taxon and MA duration could account for some variation in the magnitude of V_M , further experiments are required to decouple their effects. Advances in accessibility of genome data for a broad range of taxa provide substantial scope for such experiments to explicitly estimate relevant genomic parameters (e.g., frequency spectra for different types of mutations) alongside the phenotypic variation generated by those mutations (i.e., V_M) (KATJU AND BERGTHORSSON 2019). Furthermore, given evidence that epigenetic mutations arise more frequently than genetic mutations (e.g., VAN DER

GRAAF *et al.* 2015) the potential contribution of epimutation to patterns of heterogeneity of V_M should be explicitly assessed in future studies.

HOULE *et al.* (1996); (see also LYNCH AND WALSH 1998; LYNCH *et al.* 1999) strongly concluded that, relative to morphological traits, life history traits had lower h_M^2 but higher CV_M . Our results were consistent with this for CV_M , but not for h_M^2 , where estimates did not significantly vary among trait types. Notably, 76% of estimates in the morphology category for the meta-analysis were the size of either part of (e.g., wing length; SANTIAGO *et al.* 1992) or the whole organism (e.g., plant height; RUSSELL *et al.* 1963), reflecting bias in the traits studied. In contrast to other aspects of morphology, size may share the same heightened sensitivity to environmental variance (and hence depressed h_M^2) expected for life-history traits (SANTOS *et al.* 1994; HOULE *et al.* 1996). Size might also be more similar to fitness traits in its underlying genetic complexity (and hence, mutational target size; HOULE 1998). Investigation of diverse morphological traits, in conjunction with assessment of environmental sensitivities, will help inform where the magnitude of h_M^2 for size-related traits lies in relationship to fitness versus morphological traits.

Notably, life-history categories varied in h_M^2 and CV_M (Figure 3), which may reflect differences in the genomic basis of different life history traits. Alternatively, the MA design may allow less effective relaxation of selection on some fitness traits (e.g., survival) than others (e.g., productivity). Within the *D. serrata* experiment, variation among wing-shape traits in the magnitude of among-line variance was not fully accounted for by differences in the magnitude of phenotypic variance or trait mean (Figure 8). Establishing whether these differences are informative of the inherent genetic architecture, or are a manifestation of the stochastic nature of mutation, requires repeating the estimation using either the same or a different genetic background to determine if consistent differences among traits persist. Again, combining phenotypic investigation with fine-

scale genomic analyses of MA lines could lead to further insights on why traits differ in mutational variance (KATJU AND BERGTHORSSON 2019).

The contributions of mutation-drift-selection processes, micro-environmental variation, and estimation error to variation in the magnitude of mutational variance

Individual point estimates of among-line variance, and thus of V_M , might deviate from theoretical expectations due to factors including transient segregation of mutations, and interactions with micro-environmental variation, as well as simple estimation error (LYNCH AND HILL 1986; LYNCH 1991). Here, we observed substantial heterogeneity in the magnitude of, and statistical support for, estimates of among-line variance of 11 wing traits measured in *D. serrata* (Figure 6). In the only other study to investigate short-term (over two or three generations) variability of estimates, GARCÍA-DORADO *et al.* (2000) found that V_M varied across consecutive generations for one (sternopleural bristle count) of four traits investigated (no temporal variation in estimates of egg-to-adult viability, wing length or abdominal bristle count). GARCÍA-DORADO *et al.* (2000) attributed the variation to (mutant) genotype-by-environment (GxE) interactions, specifically to micro-environmental variation in rearing conditions altering the phenotypic effects of mutations from generation to generation. We similarly found statistical support for variation in mutational effects among generations for few (two of 11) traits. Notably, wing (centroid) size exhibited significant generation by among-line variance; as discussed above, size might be more dependent on micro-environmental variation in resources or energy requirements than wing shape traits (CAVICCHI *et al.* 1985; BITNER-MATHÉ AND KLACZKO 1999). Thus, our results and those of GARCÍA-DORADO *et al.* (2000) suggest that changes in mutational effects with micro-environment may contribute to heterogeneity among point estimates of some traits, which may reflect differences in trait environmental sensitivity, or potentially covariation of environmental sensitivity and mutational effect size (LYNCH *et al.* 1999; GARCÍA-DORADO *et al.* 2000). Further data on whether phenotypic traits differ in the frequency or effect size of epi-

09 mutations may also help to inform how and why traits differ in temporal heterogeneity or micro-
10 environmental effects on among-line variance.

11 The mutation-drift process itself can also cause heterogeneity in among-line variance due to
12 chance sampling of a newly-arisen mutation, segregating within lines (LYNCH AND HILL 1986).
13 Transient segregation of mutation could contribute to estimation error (increasing within relative to
14 among-line variance), as well as causing bounce in estimates if a line's breeding value is extreme due
15 to the presence of a segregating variant that is later lost (or not sampled in subsequent phenotype
16 assays). In the current study, statistical analyses did not support any effect of the population size
17 treatments on estimates of among-line variance. Given the different opportunity for mutation-drift-
18 selection afforded by the ~order of magnitude difference in census size in these paired sets of lines,
19 this suggests that segregating variation may have little effect on variability of among-line variance
20 estimates. Notably, populations as small as five might be sufficiently large for selection to prevent
21 declines in fitness due to new mutations (ESTES *et al.* 2004; KATJU *et al.* 2015; LUIJCKX *et al.* 2018).

22 Here, pleiotropic effects of mutations on wing traits and fitness (MCGUIGAN AND BLOWS
23 2013) may have resulted in little difference between Small and Large treatments in the segregating
24 frequencies of new mutation, or consequently within-line variance. Inter-generational changes in the
25 phenotype of a line, contributing to temporal heterogeneity in among-line variance, will depend on
26 dominance coefficients of mutations. Mutations influencing fitness may typically be at least partially
27 recessive, with the dominance coefficient positively correlated with the selection coefficient
28 (AGRAWAL AND WHITLOCK 2011). In contrast, the pleiotropic effects of these mutations on wing
29 traits (which may be typical of other non-fitness phenotypes) appear to typically be additive (HOULE
30 AND FIERST 2013; DUGAND *et al.* 2021). Thus, our focus on morphological traits may have under-
31 estimated the temporal heterogeneity that transient segregation of mutations may cause.

Rejecting micro-environmental variation and transient segregation of variants as general

explanations of estimate heterogeneity in our data, we conclude that the observed variability is largely the consequence of estimation error, and suggest that estimation error is likely to contribute to heterogeneity of published estimates. We aimed to capture typical levels of estimation error by considering a low, but typical number of MA lines (42; see Supplementary Figure S2B), and balancing the relatively low among line variance (due to relatively few generations of mutation accumulation) by considering traits that were expected to have strong genetic signal relative to residual error (ROFF AND MOUSSEAU 1987; HOULE *et al.* 1996).

The sequential repeated-measures experimental design provided statistical control over the experimental noise, allowing us to consistently detect statistically significant mutational variance in all traits (Figure 8A). Several previous studies have analysed repeated measures of MA lines to estimate V_M as the slope of the regression of among-line variance on generation (VASSILIEVA *et al.* 2000; HOULE AND NUZHIDIN 2004; MCGUIGAN *et al.* 2011), or included generation as a (fixed or random) factor in the model to similarly obtain a single estimate of among-line variance (MARTORELL *et al.* 1998; SHAW *et al.* 2000; BAER *et al.* 2005). Our analysis particularly highlights the potential benefits of short-term repeated measures (sequential generations) to improve estimate precision, and power to detect small effects.

Understanding the contribution that mutations make to evolutionary and genetic phenomena 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, an issue requiring further data to resolve. Our manipulative experiment suggested that estimation error may contribute substantially to estimate variability, and demonstrated that repeated measures over few (e.g., sequential) generations provides a simple but effective approach to address

this and improve 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 (BARTON AND TURELLI 1989; JOHNSON AND BARTON 2005; WALSH AND LYNCH 2018).

Data Availability

All analysed data are available at: <https://doi.org/10.6084/m9.figshare.14913051.v2>

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. This manuscript was improved by the input of two anonymous reviewers and the AE.

References

- Agrawal, A. F., and M. C. Whitlock, 2011 Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* 187: 553-566.
- 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.
- Boyle, E. A., Y. I. Li and J. K. Pritchard, 2017 An expanded view of complex traits: From polygenic to omnigenic. *Cell* 169: 1177-1186.
- 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* 118: e2026217118.
- 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.
- 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.
- Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annual Review of Ecology, Evolution and Systematics* 40: 151-172.
- Hine, E., D. E. Runcie, K. McGuigan and M. W. Blows, 2018 Uneven distribution of mutational variance across the transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. *Genetics* 209: 1319-1328.
- Houle, D., 1991 Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. *Evolution* 45: 630-648.
- Houle, D., 1992 Comparing evolvability and variability of quantitative traits. *Genetics* 130: 195-204.
- Houle, D., 1998 How should we explain variation in the genetic variance of traits? *Genetica* 102: 241-253.
- Houle, D., G. H. Bolstad, K. van der Linde and T. F. Hansen, 2017 Mutation predicts 40 million years of fly wing evolution. *Nature* 548: 447-450.
- Houle, D., and J. Fierst, 2013 Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. *Evolution* 67: 1116-1130.
- Houle, D., and K. Meyer, 2015 Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. *Journal of Evolutionary Biology* 28: 1542-1549.
- Houle, D., B. Morikawa and M. Lynch, 1996 Comparing mutational variabilities. *Genetics* 143: 1467-1483.

- Houle, D., and S. V. Nuzhdin, 2004 Mutation accumulation and the effect of *copia* insertions in *Drosophila melanogaster*. *Genetical Research* 83: 7-18.
- Huang, W., R. F. Lyman, R. A. Lyman, M. A. Carbone, S. T. Harbison *et al.*, 2016 Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *Elife* 5: e14625.
- Johnson, T., and N. Barton, 2005 Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* 360: 1411-1425.
- Katju, V., and U. Bergthorsson, 2019 Old trade, new tricks: insights into the spontaneous mutation process from the partnering of classical mutation accumulation experiments with high-throughput genomic approaches. *Genome Biology and Evolution* 11: 136-165.
- Katju, V., L. B. Packard, L. Bu, P. D. Keightley and U. Bergthorsson, 2015 Fitness decline in spontaneous mutation accumulation lines of *Caenorhabditis elegans* with varying effective population sizes. *Evolution* 69: 104-116.
- Kavanaugh, C. M., and R. G. Shaw, 2005 The contribution of spontaneous mutation to variation in environmental responses of *Arabidopsis thaliana*: Responses to light. *Evolution* 59: 266-275.
- Keightley, P. D., T. F. Mackay and A. Caballero, 1993 Accounting for bias in estimates of the rate of polygenic mutation. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 253: 291-296.
- Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, New York.
- Klein, T. W., 1974 Heritability and genetic correlation: Statistical power, population comparisons and sample size. *Behavior Genetics* 4: 171-189.
- Klein, T. W., J. C. De Vries and C. T. Finkbeiner, 1973 Heritability and genetic correlation: Standard errors of estimates and sample size. *Behavior Genetics* 3: 355-364.
- Kondrashov, A. S., and D. Houle, 1994 Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 258: 221-227.
- Lande, R., 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314-334.
- 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. *American Naturalist* 185: 243-252.
- Littell, R., G. Milliken, W. Stroup, R. Wolfinger and O. Schabenberger, 2006 *SAS for Mixed Models*. SAS Institute, Cary, NC.
- Luijckx, P., E. K. H. Ho, A. Stanic and A. F. Agrawal, 2018 Mutation accumulation in populations of varying size: Large effect mutations cause most mutational decline in the rotifer *Brachionus calyciflorus* under UV-C radiation. *Journal of Evolutionary Biology* 31: 924-932.
- Lynch, M., 1988a Design and analysis of experiments on random drift and inbreeding depression. *Genetics* 120: 791-807.
- Lynch, M., 1988b The rate of polygenic mutation. *Genetical Research* 51: 137-148.
- Lynch, M., 1991 Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45: 1065-1080.
- Lynch, M., 2010 Evolution of the mutation rate. *Trends in Genetics* 26: 345-352.
- Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz *et al.*, 1999 Perspective: spontaneous deleterious mutation. *Evolution* 53: 645-663.
- Lynch, M., and W. G. Hill, 1986 Phenotypic evolution by neutral mutation. *Evolution* 40: 915-935.
- Lynch, M., and B. Walsh, 1998 *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.
- Mackay, T., R. F. Lyman and W. G. Hill, 1995 Polygenic mutation in *Drosophila melanogaster*: non-linear divergence among unselected strains. *Genetics* 139: 849-859.
- Mackay, T. F. C., E. A. Stone and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10: 565-577.
- Martin, G., and T. J. E. Lenormand, 2006 The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* 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., 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., J. M. Collet, S. L. Allen, S. F. Chenoweth and M. W. Blows, 2014 Pleiotropic mutations are subject to strong stabilizing selection. *Genetics* 197: 1051-1062.
- 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.
- Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs *et al.*, 2007 Mutational bias for body size in Rhabditid nematodes. *Evolution* 61: 1653-1661.
- Reddix, A. J., S. L. Allen and S. F. Chenoweth, 2018 A genomic reference panel for *Drosophila serrata*. *G3* 8: 1335-1346.
- Rockman, M. V., 2012 The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66: 1-17.
- Roff, D. A., and T. A. Mousseau, 1987 Quantitative genetics and fitness: Lessons from *Drosophila*. *Heredity* 58: 103-118.
- Rohlf, F., 2007 tpsRelw version 1.45. Department of Ecology and Evolution, State University of New York, Stony Brook.
- Rohlf, F. J., 1999 Shape statistics: Procrustes superimpositions and tangent spaces. *J. Classif.* 16: 197-223.
- Rohlf, F. J., 2013 tpsDig, digitize landmarks and outlines, version 2.17. Department of Ecology and Evolution, State University of New York at Stony Brook.
- Russell, W., G. Sprague and L. Penny, 1963 Mutations affecting quantitative characters in long-time inbred lines of maize. *Crop Science* 3: 175-178.
- Santiago, E., J. Albornoz, A. Domínguez, M. A. Toro and C. López-Fanjul, 1992 The distribution of spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. *Genetics* 132: 771-781.
- Santos, M., K. Fowler and L. Partridge, 1994 Gene-environment interaction for body size and larval density in *Drosophila melanogaster*: an investigation of effects on development time, thorax length and adult sex ratio. *Heredity* 72: 515-521.
- Schrider, D. R., D. Houle, M. Lynch and M. W. J. G. Hahn, 2013 Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster*. *Genetics* 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. Darms, 2000 Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* 155: 369-378.
- Simons, Y. B., K. Bullaughey, R. R. Hudson and G. Sella, 2018 A population genetic interpretation of GWAS findings for human quantitative traits. *Plos Biology* 16: e2002985.
- 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.
- van der Graaf, A., R. Wardenaar, D. A. Neumann, A. Taudt, R. G. Shaw *et al.*, 2015 Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences of the United States of America* 112: 6676-6681.
- 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

A. Taxa

B. Trait Type

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). See Methods (and Table S2) for details of the categories.

Figure 2 Schematic of design (A) and phenotypes (B) from a manipulative experiment in *Drosophila serrata*. (A) 42 MA lines (evolved through 20 generations of brother-sister mating) each founded two sublines: Small (S; 24 flies) and Large (L; 288 flies distributed evenly among 12 vials). These 84 lines (S and L subline per 42 MA lines) were maintained at these census population sizes for six generations (only two shown here). Each generation, all emergent flies from the 12 vials per L subline were pooled prior to virgin collection. For S sublines, two vials were established each generation; the focal vial contributed offspring to the next generation, while the replicate vial (grey shaded) did not. Each generation, wings were randomly sampled for six males from each of two vials per line (focal and replicate vials for S; randomly chosen two for L). (B) Wing size and shape were characterised from landmarks recorded on each image: proximal (1) and distal intersections of the radial vein (2); distal intersections of medial (3), cubital vein (4), and distal veins (5) and; the posterior (6, 7) and anterior (8, 9) cross-veins. Inter-landmark distance traits were described by their end-point landmarks (e.g., ILD1.2 was the distance between landmark1 and landmark 2).

Figure 3 Variation in estimates of mutational heritability (A) and coefficient of mutational variance (B) across trait categories. For each trait type (x-axis, defined in Methods), least-squares means ($\pm$ 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 ($\log_{10}$ or BoxCox) values, as input to the analysis, are shown on the left-hand y-axis, and the untransformed values on the right-hand y-axis. No estimates of mutational variance in development time were

available for *Drosophila*. Trait type significantly predicted the magnitude of CV_M in the total data (see main text), of h_M^2 and CV_M in *Drosophila* (h_M^2 : $F_{5,38.9} = 7.25$, $P < 0.0001$; CV_M : $F_{5,26.9} = 11.87$, $P < 0.0001$), and CV_M in Nematodes (h_M^2 : $F_{6,31.7} = 0.53$, $P = 0.7780$; CV_M : $F_{6,62.8} = 3.20$, $P = 0.0083$).

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 ($\pm$ 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-hand y-axis, and the untransformed values on the right-hand y-axis.

Figure 5 Variation in wing trait mean across six generations of an experiment in *Drosophila serrata*.

For each of the 11 wing traits (panels), the least-squares mean ($\pm$ SE) is plotted for each generation for the Small (solid circle, solid lines) and Large (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 2B 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 an experiment in *Drosophila 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; Large: 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 an experiment in *Drosophila 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; Large: 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 Among-line variances for eleven wing traits in *Drosophila serrata*. (A) Among-line variance, V_L , was estimate per trait (model 5), and 90% CI estimated from the 10,000 REML-MVN samples as described in the Methods. (B) REML estimates of among line variance (points in panel A) were plotted against the total phenotypic variance (black circles, bottom x-axis) or mean (grey squares, top x-axis) of the trait.

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

Supplementary Figure S4 **Contribution of variability in trait mean to variability of among-line variance estimates 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 (L). 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, h_M^2 and, coefficient of mutational variance, 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 at: <https://doi.org/10.6084/m9.figshare.14913051.v2>.

Reference	Source	Description of where estimates were obtained from
(A) ESTIMATES USED IN THE META-ANALYSIS		
Ajé, L. C., S. Estes, M. Lynch, and P. C. Phillips. 2005 Behavioural degradation under mutation accumulation in <i>Caenorhabditis elegans</i> . <i>Genetics</i> 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 <i>et al.</i> (2005) and the mean in Ajé <i>et al.</i> (2005)
Ávila, V. and A. García-Dorado. 2002 The effects of spontaneous mutation on competitive fitness in <i>Drosophila melanogaster</i> . <i>J. Evol. Biol.</i> 15 : 561–566	1	Halligan & Keightley (2009) 's Table 1.
Azevedo, R. L., P. D. Keightley, C. Lauren-Matäta, L. L. Vassileva, M. Lynch, <i>et al.</i> , 2002 Spontaneous mutational variation for body size in <i>Caenorhabditis elegans</i> . <i>Genetics</i> 162 : 755–765	1	Table 1 of paper.
Baer, C. F., F. Shaw, C. Steading, M. Baumgartner, A. Hawkins, <i>et al.</i> , 2005 Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. <i>Proc. Natl. Acad. Sci.</i> 102 : 5785–5790	1	Halligan & Keightley (2009) 's Table 1.
Baer, C. F., N. Phillips, D. Ostrow, A. Avalos, D. Blanton, <i>et al.</i> , 2006 Cumulative effects of spontaneous mutations for fitness in <i>Caenorhabditis</i> . Role of genotype, environment and stress. <i>Genetics</i> 174 : 1387–1395	1	Table 1 of paper.
Baer, C. F., J. Joyner-Matos, D. Ostrow, V. Grigalichik, M. P. Salomon, <i>et al.</i> , 2010 Rapid decline in fitness of mutation accumulation lines of gonochoristic (outcrossing) <i>Caenorhabditis nematodes</i> . <i>Evolution</i> 64 : 3242–3253	2	h_m^2 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. <i>J. Hered</i> 50 : 26–30	3	Houle <i>et al.</i> (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 $\phi 6$. <i>Genetics</i> 176 : 467–476	1	Halligan & Keightley (2009) 's Table 1.
Caballero, A., E. Cusi, C. Garcia, and A. Garcia-Dorado. 2002 Accumulation of deleterious mutations: Additional <i>Drosophila melanogaster</i> estimates and a simulation of the effects of selection. <i>Evolution</i> 56 : 1150–1159	1	CV_M calculated from in-text information, taking V_B from $\Delta V = V_B(\text{mean}_e \times F_2)$, where $F_2 = 203$ and $\text{mean}_e = 0.625$.
Casellas, J. and J. F. Medrano. 2008 Within-generation mutation variance for litter size in inbred mice. <i>Genetics</i> 179 : 2147–55	4	MA litter mean was taken for generation 46 in Table 1. The V_M and h_m^2 estimates were taken from text and Table 3. CV_M was calculated from reported information.
Chang, S. and R. G. Shaw. 2003 The contribution of spontaneous mutation to variation in environmental response in <i>Arabidopsis thaliana</i> : responses to nutrients. <i>Evolution</i> 57 : 984–994	4	Table 3 of paper.
Charlesworth, L., H. Borthwick, C. Bartolomé, and P. Pignatelli. 2004 Estimates of the genomic mutation rate for detrimental alleles in <i>Drosophila melanogaster</i> . <i>Genetics</i> 167 : 815–826	1	Halligan & Keightley (2009) 's Table 1.
Clark, A. G., L. Wang, and T. Hultberg. 1995 Spontaneous mutation rate of modifiers of metabolism in <i>Drosophila</i> . <i>Genetics</i> 139 : 767–779	5	h_m^2 and CV_M estimates taken from Lynch <i>et al.</i> (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, <i>et al.</i> , 1993 Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. <i>J. Virol.</i> 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. <i>Euphytica</i> 36 : 823–829	3	Houle <i>et al.</i> (1996)'s Table 1.
Davies, S. K., A. Leroy, A. Burt, J. G. Bundy, and C. F. Baer, 2016 The mutational structure of metabolism in <i>Caenorhabditis elegans</i> . <i>Evolution</i> 70 : 2239–2246	2	Table S1 of paper.
Downie, D., 2003 Effects of short-term spontaneous mutation accumulation for life history traits in grape phytophthora, <i>Daktulosphaira vitifoliae</i> . <i>Genetica</i> 119 : 237–251	4	V_M and h^2_M were taken from Table 1. Line means were estimated from figure 3 ($L=0.288$, $C=0.235$ for G12, and G24 survival, respectively). CV_M calculated from information provided.
Estes, S., L. C. Ajie, M. Lynch, and P. C. Phillips, 2005 Spontaneous mutational correlations for life-history, morphological and behavioral characters in <i>Caenorhabditis elegans</i> . <i>Genetics</i> 170 : 645–653	1	We calculated CV_M using V_M (taking the number of generations as 307) from Estes <i>et al.</i> (2005) and the mean from Ajie <i>et al.</i> (2005)
Etienne, V., E. C. Andersen, J. M. Ponciano, D. Blanton, A. Cadavid, <i>et al.</i> , 2015 The red death meets the abdominal bristle: polygenic mutation for susceptibility to a bacterial pathogen in <i>Caenorhabditis elegans</i> . <i>Evolution</i> 69 : 508–519	2	Table 1 of paper.
Farhadi, 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 <i>Caenorhabditis elegans</i> . <i>Genetics</i> 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 <i>Drosophila melanogaster</i> . <i>Genetics</i> 143 : 829–837	1	V_M and h^2_M taken from Table 3 & trait mean from Table 1 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. <i>Genet. Res.</i> 21 : 121–132	3	Houle <i>et al.</i> (1996)'s Table 1.
Fry, J. D. and S. L. Heinsohn, 2002 Environment dependence of mutational parameters for viability in <i>Drosophila melanogaster</i> . <i>Genetics</i> 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 <i>Drosophila</i> . <i>Genet. Res.</i> 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 <i>Drosophila melanogaster</i> . <i>Proc. Natl. Acad. Sci.</i> 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 <i>Drosophila melanogaster</i> . In <i>Mutation and Evolution</i> , 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 <i>Drosophila melanogaster</i> . <i>Genet. Res.</i> 75 : 47–51	1	Table 2 of paper.
Hall, D. W., R. Mahmoudizad, A. W. Hurd, and S. L. Joseph, 2008 Spontaneous mutations in diploid <i>Saccharomyces cerevisiae</i> : Another thousand cell generations. <i>Genet. Res.</i> 90 : 229–241	4	Table 2 of paper.
Harada, K., 1995 A quantitative analysis of modifier mutations which occur in mutation accumulation lines in <i>Drosophila melanogaster</i> . <i>Heredity</i> 75 : 589–598	5	CV_M was calculated from information in Table 3 of paper; among line variance was back-calculated using $4V_E$ (reported) and then divided by $2 \times$ 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 L. Charlesworth, 1992 The genomic mutation rate for fitness in <i>Drosophila</i> . <i>Nature</i> 359 : 58–60	1	Halligan & Keightley (2009)'s Table 1.
Houle, D., K. A. Hughes, D. K. Hoffmaster, J. Ihara, S. Assimacopoulos, <i>et al.</i> , 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. <i>Genetics</i> 138 : 773–785	3	Houle <i>et al.</i> (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, <i>et al.</i> , 2016 Spontaneous and the origin and maintenance of quantitative genetic variation. <i>Elife</i> 5 : e14625	2	h_m^2 estimates were taken from supplementary information 1D. Estimates were averaged across sex. Could not calculate CV_M .
Joseph, S. L. and D. W. Hall, 2004 Spontaneous mutations in diploid <i>Saccharomyces cerevisiae</i> more beneficial than expected. <i>Genetics</i> 168 : 1817–1825	1	Halligan & Keightley (2009)'s Table 1.
Joyner-Matos, J., A. Upadhyay, M. P. Salomon, V. Grigalovich, and C. F. Baer, 2009 Genetic (co)variation for life span in rhabditid nematodes: Role of mutation, selection, and history. <i>J. Gerontol. A Biol. Sci. Med. Sci.</i> 64 : 1134–1145	2	Table 2 of paper.
Joyner-Matos, J., L. C. Bean, H. L. Richardson, T. Sammel, and C. F. Baer, 2011 No evidence of elevated germline mutation accumulation under oxidative stress in <i>Caenorhabditis elegans</i> . <i>Genetics</i> 189 : 1439–1447	2	Estimates of $h_m^2(V_M/V_E)$ and $CV_M(V_M/V_E)$ 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 <i>Arabidopsis thaliana</i> : Responses to light. <i>Evolution</i> 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 <i>Caenorhabditis elegans</i> . <i>Proc. Natl. Acad. Sci.</i> 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 <i>E. coli</i> . <i>Nature</i> 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 <i>Drosophila serrata</i> . <i>Evolution</i> 68 : 1824–1837	2	Table 1 of paper.
Latta, L. C., M. Peacock, D. J. Civitello, J. L. Dudycha, J. M. Meik, <i>et al.</i> , 2015 The phenotypic effects of spontaneous mutations in different environments. <i>Am. Nat.</i> 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_m^2 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 <i>Drosophila melanogaster</i> . II. Distribution of mutant effects on the trait and fitness. <i>Genet. Res.</i> 61 : 117–126	3	Houle <i>et al.</i> (1996)'s Table 1.
Lynch, M., 1985 Spontaneous mutations for life-history characters in an obligate parthenogen. <i>Evolution</i> pp. 804–818	3	Houle <i>et al.</i> (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. <i>Evolution</i> 52 : 727–733	4	Reference taken from in-text citations in Walsh & Lynch (2018): Estimates of h_m^2 , V_M and V_E were taken from Table 1 of paper. MA mean calculated using $z_{0+}R_{m0}$ (Table 1), and used to calculate CV_M .
Mackay, T. F. and R. F. Lyman, 1998 Polygenic mutation in <i>Drosophila melanogaster</i> : Genotype $\times$ environment interaction for spontaneous mutations affecting bristle number. <i>Genetica</i> 102 : 199	2	Table 9 of paper: Only the unselected lines were considered here, and we averaged estimates across sex.
Marorell, C., M. Toro, and C. Gallego, 1998 Spontaneous mutation for life-history traits in <i>Drosophila melanogaster</i> . In <i>Mutation and Evolution</i> , 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. <i>Evolution</i> 65 : 2816–2829	2	h_m^2 was taken from text. CV_M was calculated by determining V_M as $(\text{mean}(V_E) \times h_m^2)$, where V_E and MA means were taken from Tables S1 and S2 of paper.
Morgan, A. D., R. W. Ness, P. D. Keightley, and N. Colegrave, 2014 Spontaneous mutation accumulation in multiple strains of the green alga, <i>Chlamydomonas reinhardtii</i> . <i>Evolution</i> 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_m^2 were reported.
Mukai, T., 1964 The genetic structure of natural populations of <i>Drosophila melanogaster</i> . I. spontaneous mutation rate of polygenes controlling viability. <i>Genetics</i> 50 : 1–19	3	Houle <i>et al.</i> (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 <i>Drosophila melanogaster</i> . V. coupling-repulsion effect of spontaneous mutant polygenes controlling viability. <i>Genetics</i> 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 <i>Drosophila melanogaster</i> . <i>Genetics</i> 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 <i>Drosophila melanogaster</i> . <i>Genetics</i> 106 : 73–84	3	Houle <i>et al.</i> (1996)'s Table 1.
Ohtshi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in <i>Drosophila melanogaster</i> . II. homozygous effect of polygenic mutations. <i>Genetics</i> 87 : 529–545	1	Halligan & Keightley (2009)'s Table 1.
Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs, <i>et al.</i> , 2007 Mutational bias for body size in <i>Rhabditid</i> nematodes. <i>Genetics</i> 176 : 1653–1661	4	Table 2 of paper.
Pamela, L. A., D. L. Halligan, K. T. Reynolds, G. A. Ballantyne, D. M. Shuker, <i>et al.</i> , 2008 Effects of spontaneous mutation accumulation on sex ratio traits in a parasitoid wasp. <i>Evolution</i> 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 <i>Drosophila melanogaster</i> . <i>Genetica</i> 29 : 39–57	3	Houle <i>et al.</i> (1996)'s Table 1.
Pletcher, S. D., D. Houle, and J. W. Curnutt, 1998 Age-specific properties of spontaneous mutations affecting mortality in <i>Drosophila melanogaster</i> . <i>Genetics</i> 148 : 287–303	5	h_m^2 , 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. L. Fenster, and J. K. Conner, 2016 Field measurements of genotype by environment interaction for fitness caused by spontaneous mutations in <i>Arabidopsis thaliana</i> . <i>Evolution</i> 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. <i>Crop Sci.</i> 3 : 175–178	3	Houle <i>et al.</i> (1996)'s Table 1.
Rutter, M. T., F. H. Shaw, and C. L. Fenster, 2010 Spontaneous mutation parameters for <i>Arabidopsis thaliana</i> measured in the wild. <i>Evolution</i> 64 : 1825–35	2	Table 2 of paper.
Santiago, E., J. Albornoz, A. Dominguez, M. Toro, and C. López-Fanjul, 1992 The distribution of spontaneous mutations on quantitative traits and fitness in <i>Drosophila melanogaster</i> . <i>Genetics</i> 132 : 771–781	3	Houle <i>et al.</i> (1996)'s Table 1.
Schoen, D. J., 2005 Deleterious mutation in related species of the plant genus <i>Amsinckia</i> with contrasting mating systems. <i>Evolution</i> 59 : 2370–2377	1	Halligan & Keightley (2009)'s Table 1.
Shaw, R. G., D. L. Byers, and E. Darmon, 2000 Spontaneous mutational effects on reproductive traits of <i>Arabidopsis thaliana</i> . <i>Genetics</i> 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.
Vassiljeva, L. L., A. M. Hook, and M. Lynch, 2000 The fitness effects of spontaneous mutations in <i>Caenorhabditis elegans</i> . <i>Evolution</i> 54 : 1234–1246	1	h_m^2 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).
Vassiljeva, L. L., and M. Lynch, 1999 The rate of spontaneous mutation for life-history traits in <i>Caenorhabditis elegans</i> . <i>Genetics</i> 151 : 119–129	1	h_m^2 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 <i>Drosophila melanogaster</i> . II. Mutational variation and genotype-environment interaction. <i>Genetics</i> 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 <i>Cryptococcus neoformans</i> . <i>Genetics</i> 168 : 1177–1188	1	h_m^2 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 <i>Saccharomyces cerevisiae</i> . <i>Genetics</i> 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. Grigalchik, M. P. Salomon, <i>et al.</i> , 2010 Rapid decline in fitness of mutation accumulation lines of gonochoristic (outcrossing) <i>Caenorhabditis nematodes</i> . <i>Evolution</i> 64 : 3242–3253	As in (A). Removed <i>Caenorhabditis brenneri</i> 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 <i>Caenorhabditis elegans</i> . <i>Genetics</i> 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 <i>Drosophila melanogaster</i> using concomitant sibling controls. <i>Biol. Lett.</i> 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 these estimates to whole genome ($\times 5.5$) 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 <i>Drosophila melanogaster</i> . <i>Heredity</i> 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 <i>Drosophila melanogaster</i> . <i>Genet. Res.</i> 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 <i>Arabidopsis thaliana</i> : Responses to light. <i>Evolution</i> 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. Meli, <i>et al.</i> , 2015 The phenotypic effects of spontaneous mutations in different environments. <i>Am. Nat.</i> 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. Cutsinger, 1999 The evolution of age-specific mortality rates in <i>Drosophila melanogaster</i> : Genetic divergence among unselected lines. <i>Genetics</i> 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 <i>Arabidopsis thaliana</i> . <i>Proc. Natl. Acad. Sci.</i> 96 : 11393–11398	Estimates were taken from Table 3, where $\chi^2 CV_M$ was calculated from V_M/z_0^2 reported in Table 3 as $(\sqrt{V_M}/z_0^2)/2$. 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. <i>Annu. Rev. Ecol. Syst.</i> 40 : 151–172		
² Web of Science database. Journal article document types were searched for with the following search topics criteria: "mutation accumulation" and (vari* or "mutat* coefficient*"), on the 11/12/2019 at 4:38 PM AEST.		
³ Houle, D., L. Morikawa, and M. Lynch, 1996 Comparing mutational variabilities. <i>Genetics</i> 143 : 1467–1483		
⁴ Walsh, L. and M. Lynch, 2018 <i>Evolution and Selection of Quantitative Traits</i> . Oxford University Press		
⁵ Lynch, M., L. Walsh, 1998 <i>Genetics and Analysis of Quantitative Traits</i> . 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. The first published description of a set of MA lines is indicated in bold; the following non-bolded entries represent subsequent investigations of the same set of MA lines. See Table S1 for citation details.

Species name	Reference	Generation	Trait Description
<i>Arabidopsis thaliana</i>	Shaw <i>et al.</i> (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 <i>et al.</i> (2010)	25	Fecundity
	Roles <i>et al.</i> (2016)	25	Fecundity, Germination, Age-specific survivorship, Fitness
	Rutter <i>et al.</i> (2018)	25	Biomass, Flowering timing, Fruit size, Germination, Survival
<i>Caenorhabditis briggsae</i>	Baer <i>et al.</i> (2005)	200	Productivity
	Baer <i>et al.</i> (2006)	220	Productivity
	Ostrow <i>et al.</i> (2007)	200	Body size
	Joyner-Matos <i>et al.</i> (2009)	250	Life span
<i>Caenorhabditis elegans</i>	Baer <i>et al.</i> (2005)	200	Productivity
	Baer <i>et al.</i> (2006)	220	Productivity
	Ostrow <i>et al.</i> (2007)	200	Body size
	Joyner-Matos <i>et al.</i> (2009)	250	Life span
	Etienne <i>et al.</i> (2015)	250	Pathogen susceptibility
	Davies <i>et al.</i> (2016)	250	Metabolites
	Farhadifar <i>et al.</i> (2016)	250	Mitotic cell division traits
	Keightley & Caballero (1997)	60	Lifespan, Productivity, Population growth
	Azevedo <i>et al.</i> (2002)	60	Body size
	Vassilieva <i>et al.</i> (1999)	49.4	Longevity, Productivity, Survivorship, Population growth
	Vassilieva <i>et al.</i> (2000)	214	Longevity, Productivity, Survivorship, Population growth
	Azevedo <i>et al.</i> (2002)	152	Body size
	Ajie <i>et al.</i> (2005)	370	Locomotor activity, Productivity, Velocity
	Estes <i>et al.</i> (2005)	370	Population growth parameter, Body width
<i>Drosophila melanogaster</i>	Houle <i>et al.</i> (1994)	44	Fecundity, Fitness, Longevity, Male mating, Productivity
	Pletcher <i>et al.</i> (1998)	19	Female longevity, Male longevity
	Pletcher <i>et al.</i> (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 <i>et al.</i> (1992)	46	Sternopleural bristles, Wing dimensions
	Fernandez & Lopez-Fanjul (1996)	106	Viability, Fecundity
	Garcia-Dorado <i>et al.</i> (1998)	183	Fitness
	Martorell <i>et al.</i> (1998)	82	Productivity, Longevity
	Garcia-Dorado <i>et al.</i> (2000)	170	Wing length
		204	Abdominal bristles, Sternopleural bristles
		209	Viability
	Chavarrias <i>et al.</i> (2001)	250	Viability
		255	Viability
	Caballero <i>et al.</i> (2002)	210	Viability
	Avila & Garca-Dorado (2002)	288	Fitness
<i>Drosophila serrata</i>	McGuigan <i>et al.</i> (2011a)	23	Mating Success
		26	Productivity
	Latimer <i>et al.</i> (2014)	32	Locomotor activity
<i>Oscieus myriophila</i>	Baer <i>et al.</i> (2005)	200	Productivity
	Ostrow <i>et al.</i> (2007)	200	Body size
	Joseph & Hall (2004)	1012	Fitness

<i>Saccharomyces cerevisiae</i>	Hall <i>et al.</i> (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 2B), 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 <i>among-line</i> 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 <i>among-line by generation</i> 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 at: <https://doi.org/10.6084/m9.figshare.14913051.v2>.

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. The data are plotted on the log scale, but values reported on the y-axes correspond to the raw scale.

Figure S3. Contribution of variability in phenotypic variance to estimates of among-line variance in *D. serrata*. The among-line estimate from the per trait per generation per treatment implementation of model (3) (Figure 6) was divided by the corresponding estimate of total phenotypic variance (summed among-line, among-vial and residual variance from model (3)). Plotted are these variance-scaled REML estimates (and REML-MVN 90% CI) for each trait (panel; trait descriptions in Figure 2B) in each generation (x-axis) for each of the two population size treatments (Small: solid circles; Large: open circles). The dashed horizontal line indicates zero.

Figure S4. Contribution of variability in trait mean to variability of among-line variance estimates in *D. serrata*. Among-line variance estimates (V_L , Figure 6) were placed on a coefficient of variance scale to account for the influence of trait mean on magnitude of estimates, using the following formula: $\sqrt{\frac{V_L}{\bar{x}}} \times 100$ where $\bar{x}$ is the trait least squares mean (Figure 5). Plotted are the REML point estimates (and 90% CI) for each trait (panel; trait descriptions in Figure 2B) in each generation (x-axis) for each of the two population size treatments (Small: solid circles; Large: open circles). The 90% confidence intervals were estimated using 10,000 REML-MVN samples of V_L and the trait mean as described in the Methods. Some MVN sample estimates of V_L were negative and the coefficient of variance was thus undefined for those samples. The dashed horizontal line indicates zero. Where the REML estimate of among-line variance was zero, CI were not estimated.