Identification of multivariate phenotypes most influenced by mutation: *Drosophila serrata* wings as a case study

Cara Conradsen[1,2] and Katrina McGuigan[1,3]

1. School of the Environment
   The University of Queensland;
   St. Lucia, QLD, 4072 AUSTRALIA

2. European Centre for Environment and Human Health,
   University of Exeter Medical School,
   Environment and Sustainability Institute,
   Treliever Road, Penryn Campus,
   Penryn, TR10 9FE, United Kingdom

   Environment and Sustainability Institute,
   University of Exeter, Treliever Road,
   Penryn Campus, Penryn, TR10 9FE, United Kingdom

3. k.mcguigan1@uq.edu.au

All data analysed in this paper is published: (Conradsen *et al.* 2021)

Running Title: Multivariate mutational variance

Keywords: mutational variance; subspace; size; morphology; DsGRP

ABSTRACT

The distribution of mutational effects, especially pleiotropic effects, significantly impacts phenotypic evolution and population persistence. However, our empirical understanding of these effects remains quite limited. Here, we interrogated multivariate wing shape data from 12 replicated samples of the same mutation accumulation experiment in *Drosophila serrata* to characterize both the variation in wing shape introduced by mutations, and the among-sample heterogeneity in estimates of that variation. Consistent with previous studies and expectations from statistical theory, there was considerable heterogeneity among replicate estimates of the same parameter (median ratio of largest to smallest estimate of mutational variance per trait: 7.6; estimates of the same pairwise trait correlation ranged up to a maximum of the theoretical limits: -1 to 1). However, multivariate analyses identified a trait combination associated with relatively similar magnitude of mutational variance in all population samples (ratio of largest to smallest estimate: 1.96). This major axis of mutational variance aligned closely with the major axis of among-line variance within most of the population samples, suggesting that each sample adequately estimated the predominant pattern of correlated mutational effects, despite the marked heterogeneity of individual (co)variance parameters. Mutation predominantly introduced variation along an axis of shorter, wider vs. longer, narrower wing tips. This was qualitatively the same mutational wing shape variance identified in independent studies of *Drosophila* wings, using different experimental designs and analytical approaches. Our study suggests that while estimation error can independently influence variances of individual traits and covariances among traits, multivariate analyses may be effective at revealing clear signals higher dimensional patterns of variation.

46 **INTRODUCTION**

47 How spontaneous mutations affect the phenotype is central to a range of evolutionary genetic
48 phenomena, including the maintenance of quantitative genetic variance (Bürger 2000; Johnson
49 and Barton 2005), and the nature of complex human diseases (Eyre-Walker 2010). Reflecting
50 this broad significance, substantial effort has been devoted to characterising the input of new
51 mutation to individual quantitative traits (Houle *et al.* 1996; Halligan and Keightley 2009;
52 Walsh and Lynch 2018; Conradsen *et al.* 2022). However, this individual trait focus is at odds
53 with the long-standing recognition that selection does not act on individual traits in isolation
54 and that the shared genetic basis of traits influences the response to selection (Walsh and Blows
55 2009; Svensson *et al.* 2021). Mutational correlation among traits has been implicated as
56 determining the stability of patterns of standing genetic covariance (Jones *et al.* 2003),
57 highlighting the potential for mutation to exert long-term influence on the direction and rate of
58 multivariate phenotypic evolution. Available data suggests mutations frequently have
59 correlated effects across traits (e.g., Estes *et al.* 2005; McGuigan *et al.* 2014b; Mallard *et al.*
60 2023; Miller *et al.* 2023). However, how closely patterns of trait covariation generated by
61 mutation match that observed in standing genetic variation ranges widely, from markedly
62 different (e.g., Latimer *et al.* 2014) to highly concordant (e.g., Houle *et al.* 2017). Further
63 empirical data are required to advance our understanding of the mutational correlations across
64 sets of traits, the factors that contribute to variability in patterns of correlations, and their
65 persistence to standing genetic variation.

66 High estimation error is a well-known challenge to studying genetic covariances (Klein
67 *et al.* 1973; Klein 1974; Lynch and Walsh 1998). Random sampling error can generate strong
68 spurious covariance among variables, particularly when the number of independent samples is
69 small relative to the number of variables (traits) (Johnstone 2007; Blows and McGuigan 2015;

70 Sztepanacz and Blows 2017). Relative to genetic covariances among highly polygenic traits,
71 this statistical problem is likely to be exacerbated for studies of mutation: few mutations
72 affecting the traits are sampled, resulting in small magnitudes for covariances (i.e., small
73 statistical effect size). This limitation can be mitigated over hundreds of generations of
74 mutation-drift evolution, with accumulated mutational variance approaching the magnitude of
75 standing genetic variation (Lynch *et al.* 1999; Johnson and Barton 2005). However, phenotype
76 variation in such experiments is likely to increasingly reflect the combined effects of multiple
77 mutations per genotype (homozygous genetic line), which will limit insight into the distribution
78 of mutational effects. Furthermore, due to logistical limitations, long-term maintenance might
79 typically trade-off with sample size (e.g., the number of mutationally divergent genotypes
80 assayed), opposing the statistical gain from higher variance. Thus, mutational covariance
81 estimates are expected to typically have low accuracy and precision, reducing our ability to
82 characterize mutational effects, and to establish general patterns or causes of heterogeneity in
83 trait covariances.

84 The converse to this perspective of statistical limits on our capacity to acquire
85 knowledge is that multivariate analyses may substantially improve the statistical power to
86 estimate effects by leveraging the shared information among covarying traits (Schmitz *et al.*
87 1998). For example, growing availability of extensive phenotypic data in human genome-wide
88 association studies (Bycroft *et al.* 2018) has been exploited via multi-trait frameworks that
89 demonstrate greatly improved variant detection (Qi and Chatterjee 2018; Turley *et al.* 2018;
90 Baselmans *et al.* 2019; Wu 2020). Several studies have sought to understand mutational
91 covariance patterns via eigenanalyses of multi-trait mutational variance-covariance matrices
92 (e.g., Latimer *et al.* 2014; Houle *et al.* 2017; Dugand *et al.* 2021; Mallard *et al.* 2023). The first
93 eigenvector (g_{max}) of the additive genetic (co)variance matrix, **G**, reflects the dominant pattern
94 of shared standing genetic variation among the traits, and is of particular interest in

95 evolutionary genetics because it describes the phenotype that can evolve most rapidly (Lande
96 1979; Schluter 1996). Similarly, the leading eigenvector of the mutational (co)variance matrix,
97 m_{max}, informs how mutation may most strongly influence **G**, and future evolutionary potential.
98 Sampling error is well known to inflate the estimated magnitude of variance along these
99 eigenvectors, and statistical tools have been developed to account for this (Blows and
100 McGuigan 2015; Sztepanacz and Blows 2017). However, how well experiments might
101 typically capture the multivariate effects of mutation, as opposed to sampling error and
102 biological factors such as environmentally dependent mutational effects, has received little
103 attention.

104 Drosophila wings have emerged as a model for studying genetics and evolution (Mezey
105 and Houle 2005; Carreira *et al.* 2011; Houle *et al.* 2017; Dugand *et al.* 2023). Wing veins (the
106 relative positions of which define wing shape traits) have a shared developmental basis,
107 controlled by diffusion of patterning genes (reviewed in Blair 2007; Matamoro-Vidal *et al.*
108 2015), suggesting that covariance among shape traits should be pervasive. Consistent with this
109 prediction, empirical evidence indicates substantial genetic (Mezey and Houle 2005;
110 McGuigan and Blows 2007) and mutational (Houle *et al.* 2017; Dugand *et al.* 2021) covariance
111 among parts of the wing. Furthermore, specific alleles causing broadscale changes across the
112 wing (i.e., in multiple shape traits) have been identified as segregating in natural populations
113 (e.g., Pitchers *et al.* 2019). Here, we seek to exploit the shared information expected to arise
114 through pleiotropic mutation on wing shape to determine the statistical robustness of estimates
115 of multivariate mutational variance in *Drosophila serrata* wing shape.

116 Our experimental design is based on a classical mutation accumulation experiment, in
117 which an inbred ancestor is subdivided into daughter populations that diverge from one another
118 via chance fixation of spontaneous mutations (Halligan and Keightley 2009). Following 20

119 generations of brother-sister inbreeding (and mutation accumulation), each of the 42 MA lines
120 were subdivided and maintained over six generations as two sublines, differing in population
121 size, with wing phenotypes sampled every generation (Conradsen *et al.* 2022; Figure 1). This
122 design, in which effects of mutations fixed during the 20 generations of mutation accumulation
123 are measured repeatedly, allowed us to characterize how experimental factors might influence
124 estimates of among-line (mutational) covariance. Specifically, the 12 experimental samples
125 contain the same set of fixed mutations, but potentially vary due to the influence of micro-
126 environment, segregating mutations, measurement and estimation errors (as detailed in
127 Conradsen *et al.* 2022). Through comparative analyses of the 12 mutational covariance
128 matrices, we seek to both determine how these factors affect estimation, and to robustly
129 estimate the major patterns of multivariate wing shape variation introduced by new mutation
130 in *D. serrata*.

131 **METHODS**

132 **Population history and wing data collection**

133 The experimental design was described in detail in Conradsen *et al.* (2022). Briefly, we used a
134 highly inbred *D. serrata* Reference Genome Panel (DsRGP) line (Reddiex *et al.* 2018),
135 DsGRP-226, as a progenitor to establish 200 MA lines that were maintained for 20 generations
136 by brother-sister inbreeding (Kannan *et al.* 2023). From these MA lines, 42 were randomly
137 selected and expanded to found two sublines (84 lines in total) under two population size
138 treatments (Small and Large: Figure 1a) (Conradsen *et al.* 2022). For each of the 84 sublines,
139 wings were sampled from up to six male flies in each of two replicate rearing vials for six
140 sequential generations: a total of 5,135 wings were available for analysis (Conradsen *et al.*
141 2021).

142 Nine wing landmark positions (Figure 1b) were recorded from images, aligned using a
143 generalised Procrustes least squares superimposition, and centroid size (CS, the square root of
144 the sum of the squared distances between each land-mark and their centroid) determined (Rohlf
145 2007). Inter-landmark distances (ILDs, in units of centroid size) were calculated from the
146 aligned coordinates. Using the model described below, we attempted to estimate the among-
147 line variance-covariance matrix for the 11 traits previously analysed individually by Conradsen
148 *et al.* (2022). However, this model failed to converge – a likely consequence of the relatively
149 few lines (42) and the small magnitude of among-line variance accumulated after 20
150 generations of mutation-drift evolution. We therefore report on a subset of those 11 traits, and
151 consider further in the Discussion how our pragmatic, but non-random, trait selection may
152 influence conclusions. To arrive at the subset of five ILDs, along with CS (Figure 1b), that
153 were retained for multivariate analysis, four of the original 11 traits (ILD1.9, ILD3.8, ILD4.8
154 and ILD5.6) were excluded because at least one of the 12 (two treatments measured in six
155 generations) among-line variance estimates was zero (see Figure 5 of Conradsen *et al.* 2022),
156 negatively impacting on multivariate model convergence. A further trait (ILD4.5) was
157 excluded as models containing this trait returned a non-positive definite asymptotic matrix
158 (indicating poor model convergence) in several samples.

159 **Estimating mutational variance-covariance matrices**

160 Prior to analysis, Mahalanobis distance was used to identify multivariate outliers (critical value
161 of χ^2 = 22.46, df = 6, α = 0.01), which were excluded (55 outliers across the 5135 flies, <
162 1.1%). To facilitate convergence in multivariate models, data for each trait was variance
163 standardised (mean = 0, standard deviation = 1) across all generations and treatments. We
164 estimated the 12 among-line (**M**) covariance matrices for our six traits, using restricted-

165 maximum likelihood (REML) in PROC MIXED in SAS v9.4 (SAS Institute Inc., Cary, NC.)
166 to fit:

167 $$y_{ijk} = \mu + Line_i + Vial_{j(i)} + \varepsilon_{ijk} \quad (1)$$

168 where y_{ijk} was the vector of six trait values for the kth wing (k = 1…6), for the jth vial (j = 1,2),
169 within the ith line (i = 1...42); μ was the vector of global means per trait; the random effects of
170 *Line*, *Vial* and the residual error (ε) estimated the 6 x 6 unstructured covariance matrices
171 corresponding to among-line (mutational), among replicate vial and among individual flies
172 within a rearing vial, respectively.

173 To determine the number of statistically supported independent axes of among-line
174 variation, a factor-analytic approach (Hine and Blows 2006) was applied, using likelihood ratio
175 tests (LRT) to compare sequential models constraining among-line variance from zero up to
176 six dimensions. The LRT degrees of freedom were defined by the difference between models
177 in the number of parameters estimated. Factor analytic modelling is conservative but represents
178 the most robust approach for determining statistical support for the dimensionality of
179 covariance matrices (Sztepanacz and Blows 2017).

180 To place confidence intervals about estimates of the variance components for result
181 visualisation, we applied the REML-MVN method developed by Meyer and Houle (2013;
182 Houle and Meyer 2015). To quantify uncertainty, 10,000 random samples were drawn from
183 the distribution $N \sim (\hat{\boldsymbol{\theta}}, \mathbf{V})$, where $\hat{\boldsymbol{\theta}}$ was the vector of observed (REML) covariance parameter
184 estimates, and **V** was the asymptotic variance-covariance matrix. These samples were used to
185 calculate confidence intervals for both model-estimated and derived parameters (detailed
186 below).

187 **Constructing null model variance-covariance matrices**

188 The overarching null hypothesis of the current study is that we have 12 repeated measures of
189 the same mutational (co)variance of wing traits, differing from one another only through
190 transient segregation of recent mutations, micro-environmental effects and estimation error.
191 We constructed a null distribution reflecting this by randomly sampling MA lines (with
192 replacement) to create 12 experimental populations (two treatments by six generations) of 42
193 lines. By shuffling lines among population size treatments and generations, while retaining the
194 within-line information (i.e., the set of replicate vials and sampled individual flies within each
195 vial were collectively assigned their new population identity), we maintained the signal of
196 mutation in the populations, focusing our null on the potential for sampling error to generate
197 spurious divergence among 12 repeated measures. We generated 1000 (randomised) null
198 datasets, and again implemented model 1 to estimate 12 **M**. Of the 12,000 null data models
199 (1000 randomised datasets, each analysed to estimate 12 population sample **M**), nine models
200 failed to converge, and those **M** were omitted from analyses. Interpretation of the null is thus
201 based on all 1000 randomised datasets where nine datasets had only 11 populations for which
202 all **M** models converged. The null matrices were taken through the same analytical pipeline
203 (detailed below) as the observed data.

204 **Comparing variance-covariance matrices**

205 We used Krzanowski's subspace method to determine the extent to which the 12 **M** estimated
206 the same multivariate axes of mutational variance. This approach is described in detail
207 elsewhere (Krzanowski 1979; Blows *et al.* 2004; Aguirre *et al.* 2014). Briefly, R (R Core Team
208 2024) was used to diagonalise each **M**, and to estimate the Krzanowski subspace matrix, **H**:

209 $$\mathbf{H} = \sum_{t=1}^{p} \mathbf{L}_t^{\mathrm{T}} \mathbf{L}_t \quad (2)$$

210 where $\mathbf{L}_t$ contains the subset of the jth largest eigenvectors, as rows, of the tth **M** matrix, (and
211 $\mathbf{L}_t^{\mathrm{T}}$ its transpose containing eigenvectors as columns) and $p = 12$ is the number of populations
212 to be compared. Spectral decomposition of **H** describes the degree of shared structure across
213 the matrices, with the eigenvalues ranging from 0 to p. Here, an eigenvalue of 12 would indicate
214 that all 12 **M** have variance in the direction of the associated eigenvector of **H**. The
215 interpretation of **H**'s eigenvalues depends on how j is defined (Krzanowski 1979; Blows *et al.*
216 2004; Aguirre *et al.* 2014). When $j > \frac{n}{2}$ (where n is the number of traits; here, $n = 6$, setting $j >$
217 3), the analysis is constrained to recover common subspaces of two matrices. Conversely, if j
218 is allowed to vary among populations, above $j = \min(j_1, j_2, \ldots j_p)$ dimensions are constrained to
219 be orthogonal (i.e., not contain a common subspace) for at least one population — the one with
220 the lowest j. Here, we followed Aguirre *et al.* (2014) in defining j for the tth **M** matrix as the
221 number of eigenvectors required to explain at least 90% of the variance; for each of our 12
222 sampled matrices, this included all dimensions with statistically supported variation (Table 1,
223 Table S2). The 90% criterion ensured that we could interpret at least the first two eigenvalues
224 of **H** (Table 1). For four **M**, $j = 2$, indicating that these matrices were constrained to share no
225 more than two common dimensions with any other matrix; for seven of the 12 **M**, $j = 3$, and
226 these **M** may share all three dimensions. Notably, only one **M** had $j > 3$, indicating that no two
227 matrices were constrained to share common subspaces.

228 To place confidence intervals on the eigenvalues of **H**, we applied equation 2 and
229 extracted the eigenvalues of the resulting matrix for each of the 10,000 REML samples of the
230 12 matrices. To determine which of the 12 matrices had variance in which parts of the subspace
231 captured by **H**, and whether matrices differed in the magnitude of among-line variance in
232 shared (common) subspaces, we projected the eigenvectors of **H** through each **M** (Aguirre *et*
233 *al.* 2014). To place confidence intervals on these estimates of common-subspace variance,
234 eigenvectors of the observed REML **H** matrix were projected through each of the 10,000

235 REML-MVN samples of the matrix. The null datasets were also analysed, with equation 2
236 applied to estimate **H** for each of the 100 null datasets (we used $j = 3$ eigenvectors for all null
237 datasets).

238 **RESULTS**

239 **Among-line variation in the 12 population samples**

240 Consistent with previous analyses of each trait individually (Conradsen *et al.* 2022), we
241 observed the 12 population estimates of among-line (mutational) variance to vary in both
242 magnitude and statistical support for all traits (Figure 2a,b; Table S1). There were no patterns
243 of consistent change over time or divergence between the population size treatments (Figure
244 2a), suggesting the observed heterogeneity was not due to mutation-drift-selection processes.
245 There was statistical support for 46 (63.9%) of the 72 estimates (6 traits in 12 population
246 samples), with statistical support across all 12 samples for only one trait (ILD1.5, capturing
247 proximo-posterior wing width: Figure 1b) (Figure 2a,b). Particularly for traits with fewer
248 statistically supported estimates (Figure 2b) there were large differences in magnitude of
249 among-line variance (ratio of largest to smallest ranged from 1.93 up to 34.4, median 7.6;
250 Figure 2a). Nonetheless, there was little statistical support for among-line variance to differ
251 between the population samples - estimates typically fell well within each other's 90% CI, with
252 only a few exceptions for traits ILD1.2 (two largest versus five smallest estimates) and ILD2.8
253 (largest versus smallest) (Figure 2a).

254 There was considerably less statistical support for mutational covariance among traits,
255 with only 15% (27 of 180) of pairwise covariances statistically distinct from zero (Figure 2a,c).
256 Nonetheless, the average absolute correlation was 0.51 (although this value is upwardly biased
257 by estimates at the statistical boundary of -1 or 1: Table S1). This pattern of frequent strong
258 association between traits, and consistency of sign (particularly for statistically supported

259 estimates: Figure 2a,c) suggests that the relatively large sampling error of covariance estimates,
260 rather than a lack of correlated mutational effects among traits, might contribute to the limited
261 statistical evidence for mutational covariance. Binomial tests of the null hypothesis of an equal
262 number of positive and negative values rejects the null for six (of 15, 40%) trait pairs that had
263 consistent sign for at least 10 estimates ($P = 0.0193$) (Figure 2a,c).

264 The distribution of eigenvalues of each population sample's among-line covariance
265 matrix was consistent with mutation generating trait covariance, and with variation in
266 mutational effects among traits. Specifically, the first eigenvector accounted for substantially
267 more variance than any individual trait (Figure 3; Table 1 vs Table S1), consistent with shared
268 variance among some traits, while robust statistical support for two axes of mutational variance
269 in most population samples (Table S2) suggested that not all mutations had concordant effects
270 across all traits. These general conclusions are well supported by consistent evidence of at least
271 some statistically supported trait covariances in each population sample (median = 2 of the 15
272 covariances per population sample: Figure 2a), and consistent patterns of eigenvalue
273 distributions (Table 1; Figure 3). Notably, relative to the individual traits, there was markedly
274 less heterogeneity among the 12 samples in the magnitude of the total variance (ratio of largest
275 to smallest trace: 2.16) and eigenvalues (ratio of largest to smallest: 1.92, 2.90 and 3.08 for m_1,
276 m_2 and m_3, respectively: Table 1). Thus, traits might be independently affected by sampling
277 error in each population sample, as suggested by the absence of general patterns across traits
278 in the variation among the 12 population sample estimates (Figure 2a; see also Conradsen *et*
279 *al.* 2022). The multivariate analysis may be relatively effective at partitioning out independent
280 (uncorrelated) spurious signals affecting individual traits, resulting in more consistent
281 conclusions about the magnitude of variance. Nonetheless, given the heterogeneity among the
282 12 estimates of variance per trait and covariance per trait pair (Figure 2a), the specific

283 multivariate wing shapes associated with among-line variance (Table S3) may differ among
284 population samples. Below, we present further analyses aimed at addressing this.

285 **Common subspace and major axis of mutational variation**

286 There was strong support for consistent orientation of multivariate among-line variance across
287 the 12 population samples. Because sampling error is expected to impact each of the 12 **M**
288 independently, shared patterns of among-line variance provide strong evidence of the nature of
289 mutational input to wing shape. The major axis of the Krzanowski common subspace had an
290 eigenvalue of 11.68, close to the maximum value (12) attainable when the vector is present in
291 the sampled subspace of each of the 12 populations (Table 2). Further, the eigenvalue was
292 within the confidence bounds estimated from the null datasets simulating a single population
293 (Table 2). The eigenvalues of the other two interpretable common subspace axes (8.9 and 6.6:
294 Table 2) implicated further axes of mutational variance, but suggesting that other factors, such
295 as estimation error or micro-environmental heterogeneity, may limit the robust estimation of
296 this variation.

297 As suggested by the large eigenvalue of h_1, all 12 population samples had statistically
298 supported variance for this multivariate wing shape trait (Figure 4a). This common axis of wing
299 shape variation, representing the best estimate of the multivariate wing shape most affected by
300 mutation in the current experiment, reflected an axis of expansion versus contraction of the
301 proximal versus distal wing length (strongest, opposing, loadings on h_1 from ILD1.5 and
302 ILD3.7: Table 2). Notably, these two traits had the most widespread statistical support for
303 among-line variance, the least heterogeneity in magnitude among repeated estimates, and were
304 significantly negatively correlated in all but one population sample (average bounded
305 correlation coefficient = -0.87: Figure 2; Table S1).

306 The major axis of multivariate mutational input, h_1, was typically captured by the
307 leading eigenvector of a population sample's **M** (i.e., was aligned with m_1). This is evidenced
308 by: i) the angle between h_1 and eigenvectors of each population's **M** being closest to zero
309 (indicating identical orientation) for m_1 in 10 populations (11 – 21 degrees, median = 14,
310 excluding S2 and L1: Table S4); ii) the magnitude of among-line variance being greater for h_1
311 than for h_2 or h_3 in 11 populations (Figure 4a,b) and; iii) h_1 typically accounting for almost as
312 much variance as m_1 did (median = 86% of m_1: Figure 4d). In population S2, h_1 was more
313 closely aligned with m_2 (h_1 and m_1: 37 degrees and h_1 and m_2: 15 degrees: Table S4), reflected
314 in less variance along h_1 than h_2 (0.30 versus 0.36: Figure 4a,b) and h_1 accounting for only 60%
315 as much variance as m_1 (lower right point in Figure 4d). Nonetheless, even for this relatively
316 incongruent population sample, m_1 did capture some signal of the major axis of mutational
317 variance (h_1) (Figure 4a,d), suggesting that a single population sample might, therefore,
318 typically be adequate to estimate the multivariate phenotypic trait that was most strongly
319 influenced by mutation. Along h_1, the largest estimate of among-line variance (in population
320 L3) was 1.96 times greater than the smallest (L4), with estimates typically well within the CI
321 of all other populations (Figure 4a).

322 The magnitude of variance along h_1 reflected the magnitude of variance along m_1, but
323 was consistently lower (Figure 4d). Leading eigenvalues of genetic (and mutational)
324 covariance matrices are particularly vulnerable to inflation by sampling error (Hill and
325 Thompson 1978; McGuigan *et al.* 2014a), but the Krzanowski common subspace analysis finds
326 the shared pattern among population samples, rather than being strongly influenced by
327 independent sampling error within each population sample. Comparing variance along h_1 and
328 population-specific m_1 confirms the expectation that individual experiments will be prone to
329 overestimating multivariate mutational variance. There also remains, however, heterogeneity
330 among replicated estimates in the estimated magnitude of mutational variance (i.e., along h_1)

331 (Figure 4a). A pattern of greater variance in earlier and later generations for both population
332 size treatments (with L3 estimate a notable exception) may suggest that heterogeneity in micro-
333 environmental effects among generations weakly influenced the magnitude of among-line
334 variance.

335 There were no clear patterns in the magnitude of variance along h_2 or h_3, although there
336 was a weak trend for the first two generation estimates to be larger than later in time, and for
337 the presence of among-line (mutational) variance to be more robustly supported along h_2 in the
338 large populations and along h_3 in the small populations (Figure 4b,c). Despite this pattern in
339 statistical support, there was little difference between the population size treatments in the
340 magnitude of variance along h_2 (median = 0.148 vs 0.141 in small vs large population
341 treatment, respectively; Figure 4b) or in the propensity for this axis to be most closely aligned
342 to m_2 within each population (median angle = 0.28 vs 0.27 in small vs large population
343 treatment, respectively; Table S4). For h_3, there was higher median variance in the small
344 population size treatment (0.094 versus 0.059; Figure 4c) and a greater propensity for h_3 to be
345 most closely aligned with lower eigenvectors (2x m_6, 1x m_5; Table S4) of the among-line
346 variance matrix in the large than small population size treatment.

347 Notably, h_2 primarily reflects mutational variance in wing size (CS), while h_3 is most
348 strongly influenced by ILD2.5 (Table 2): these are the only two traits for which individual trait
349 analyses presented by Conradsen *et al.* (2022) robustly rejected the null hypothesis that among-
350 line variance was homogenous over time. ILD1.2 and ILD2.8 also contributed to h_2 (larger
351 wings had disproportionally longer anterior margins, while smaller wings had shorter anterior
352 margin: Table 2, Figure 1b), and were the two traits with the greatest heterogeneity in variances
353 and covariance among the 12 **M** (Figure 2). Importantly, despite this evidence that specific
354 micro-environmental context may influence the observed magnitude of variance along these

355 axes, the presence of among-line variance along h_2 and h_3 in multiple independent samples of
356 the mutation accumulation lines suggests that mutation did contribute variation to these aspects
357 of wing phenotype.

358 **DISCUSSION**

359 Mutations contribute, on average, 0.24% to the phenotypic variance or 1.57% of the trait mean
360 per generation (Houle *et al.* 1996; Lynch *et al.* 1999; Conradsen *et al.* 2022). Accurately
361 estimating these small effects on each trait represents a clear challenge to understanding how
362 trait type, genetic background and environmental context might each influence mutational
363 characteristics, with differences in the magnitude of mutational variance reported among
364 studies comparable to the magnitude of differences expected from sampling error alone
365 (Conradsen *et al.* 2022). Here, focusing on a biased subset of traits that Conradsen *et al.* (2022)
366 found to have the most statistically robust estimates, we nonetheless re-derive their inference
367 of large sampling error effects. There was substantial heterogeneity among the 12 estimates of
368 among-line variance per trait (median ratio of largest to smallest estimate was 7.57) but no
369 apparent temporal or between-treatment divergence in magnitude (Figure 2a; Table S1),
370 consistent with sampling error, rather than mutation-drift-selection process driving divergence
371 in the magnitude of among-line variance.

372 The challenge posed by estimation error is greatly exacerbated for covariances (Klein
373 *et al.* 1973; Klein 1974; Lynch and Walsh 1998). Shared development (Matamoro-Vidal *et al.*
374 2015), known pleiotropic alleles (Pitchers *et al.* 2019) and prior estimation of mutational
375 covariance (Carreira *et al.* 2011; Houle *et al.* 2017; Dugand *et al.* 2021) all led us to expected
376 strong among-line trait covariances for wing shape traits. The magnitude of estimates (median
377 absolute correlation ~0.5) was consistent with this, but only one covariance (between ILD1.5
378 and ILD3.7; Figure 2a,c; Table S1) was robustly statistically supported. Each of our population

379 samples was designed to approximate the sample size of a typical mutation accumulation
380 experiment (Conradsen *et al.* 2022). From this perspective, for nine of the 15 trait pairs, we
381 would have alternatively concluded that mutations did, or did not, have correlated effects,
382 depending on which of the 12 replicate samples we considered. In addition to contributing wide
383 confidence intervals, sampling error may generate strong (significant) spurious associations
384 between variables. Here, in the sign of covariance estimates were consistent for some trait pairs,
385 suggesting that mutation generated covariance, but for other trait pairs with inconsistently
386 estimated direction of covariance despite some statistically supported estimate(s) (e.g., CS:
387 ILD2.5; Figure 2a,c) it remains unknown if the association is spurious.

388 Relatively few specific traits have been investigated across more than one mutational
389 study leaving it relatively unknown whether mutation consistently causes certain traits to
390 covary. Life history traits (e.g., early versus late fecundity, productivity, longevity) are the best
391 studied, with independent studies in several species concordantly reporting statistically
392 supported positive (>0.6) mutational covariances (Houle *et al.* 1994; Estes *et al.* 2005). Positive
393 covariance of life history traits is predicted by their joint dependence on energetic resources
394 (i.e., condition dependence) (Houle 1991; Rowe and Houle 1996). However, positive
395 covariance can also arise in mutation accumulation lines (and during mutagenesis) via linkage
396 (all mutations within a line are in complete linkage with one another), particularly where
397 mutations have directionally biased effects on the traits (Estes *et al.* 2005). When mutational
398 effects are not biased to predominantly increase (or decrease) trait values, different
399 (independent) mutation accumulation experiments are expected to report inconsistent strength
400 and direction of covariance of a specific trait pair due to the chance linkage of different,
401 independently acting, mutations. As discussed further below, replicated patterns of the same
402 trait covariance in independent studies thus provides support for pleiotropic mutation, rather
403 than linkage, as causal of observed trait covariances among wing traits. Although we were

404 unable to assess it in the current study, investigations in *D. melanogaster* have found no
405 evidence of directional bias in mutations affecting wing shape (Santiago *et al.* 1992; Houle and
406 Fierst 2013). Traits under stabilising selection are predicted to lack directional bias in
407 mutational effects (Johnson and Barton 2005), with some evidence that *Drosophila* wing shape
408 is under stabilizing selection (McGuigan *et al.* 2011; Houle and Fierst 2013).

409 **Multivariate mutational variance in *Drosophila* wings**

410 Despite the substantial variability of most individual variance and covariance parameter
411 estimates, the multivariate metrics (trace, eigenvalues) were notably similar among the 12
412 population samples (ratio of largest to smallest ~2: Table 1, Figure 4). Further, the Krzanowski
413 common subspace analysis suggested that each population sample was reasonably effective at
414 estimating the major axes of mutational variance (Figure 4). This indicates that multivariate
415 analyses may be effective at partitioning out estimation error independently affecting
416 individual traits, leveraging shared mutational information among traits to improve the
417 precision of estimates.

418 With the Krzanowski shared subspace analysis, we uncovered a single major axis, h_1,
419 along which there was statistically supported mutational (among-line) variance in all 12
420 population samples. Trait loadings on this h_1 vector (Table 2) suggest that mutation
421 predominantly generated variation along an axis of shorter, wider versus longer, narrower wing
422 tips. Intriguingly, this was qualitatively the same mutational wing shape variance identified in
423 independent studies of *Drosophila* wings, using different experimental designs and analytical
424 approaches. Dugand *et al.* (2021) estimated mutational (co)variance in an outbred, randomly
425 mating, *D. serrata* population founded by the DsGRP lines (including the ancestor line of the
426 mutation accumulation lines considered in the current study). In that outbred population
427 analysis, m_1 was characterized by lengthening ILD1.5 and ILD5.7 while shortening along

428 ILD4.6 and ILD3.9 (and vice versa: see their Table 3). Similarly, Houle *et al.* (2017) estimated
429 the major axis of mutational variance using inbred mutation accumulation lines of *D.*
430 *melanogaster* (see their Extended Data Figure 1), finding that mutation pushed out the
431 landmarks corresponding to LM5 and LM7 in the current study (increasing ILD1.5 and
432 ILD1.7), while pulling closer together landmarks corresponding to LM3,4 and LM9,10
433 (shortening distances between those landmarks, including ILD3.7, ILD3.9 and ILD4.6).

434 The consistency of mutational covariance patterns across these different studies of
435 *Drosophila* wings suggests bias in either the frequency or effect sizes of pleiotropic mutations,
436 driving persistent bias in the phenotypic variation generated by new mutation. Houle *et al.*
437 (2017) found that the distribution of new wing shape variation introduced by mutation in *D.*
438 *melanogaster* strongly predicted the pattern of divergence among Drosophilidae species over
439 40 million years (see also Houle *et al.* 2020). Rohner and Berger (2023) extended upon this,
440 demonstrating that developmental processes bias the frequency distribution of wing shape
441 variation in another family of flies, Sepsidae, in a manner that was strongly correlated with the
442 distribution of mutational covariance reported by Houle *et al.* (2017). The similar patterns of
443 mutational variance in *D. serrata* from this study and Dugand *et al.* (2021) support the
444 inference from Houle *et al.* (2017) and Rohner and Berger (2023) by providing evidence that
445 the frequency distribution of phenotypes introduced by new mutation is conserved among
446 genotypes. We note this evidence of conserved effects is qualitative rather than quantitative.
447 This reflects, in part, methodological challenges that prevent exact duplication of trait
448 measurement (i.e., wing imaging methods capture different landmarks). Factors influencing
449 model convergence (such as the ratio of the number of inbred lines to the number of traits) can
450 also preclude matched analyses of the same trait sets. Importantly however, variability in the
451 presence of statistically supported variance for each trait can also limit direct comparability.

452 Such inconsistency in detection of variance may reflect random sampling from the same
453 underlying distribution of effects, but further data is required to assess this.

454 The factor analytic modelling and Krzanowski shared subspace analysis revealed robust
455 statistical support for additional dimensions of mutational co-variance, although variability
456 among the 12 population samples limited insight into the exact nature of these mutational
457 effects. Wing size contributed strongly to the second axis of the common subspace, h_2 (Table
458 2), which was associated with statistically supported among-line variance in 58% (7 of 12) of
459 the population samples (Figure 4b). Mutational (and genetic) variance in size may vary due to
460 sensitivity of growth to resource availability (Cavicchi *et al.* 1985; Bitner-Mathé and Klaczko
461 1999). In their analysis of individual traits, Conradsen *et al.* (2022) reported significant change
462 in mutational effects on size among the six assay generations. Here, we found some marginal
463 statistical support for the direction of covariance of size with other traits (ILD2.5 and ILD3.7)
464 to reverse sign among some population samples (Figure 2a: non-overlapping CI among
465 estimates with CI either not including or marginally including zero). Kannan *et al.* (2023)
466 investigated the same mutation accumulation lines, maintained via brother sister mating for a
467 further 10 generations (30 in total). Rearing flies under different diet and temperature
468 conditions, they found evidence of reduced mutational variance in male wing size when
469 nutrition was limited (Kannan *et al.* 2023), suggesting the potential that unintended
470 heterogeneity in nutrition among generations may have contributed to the observed
471 heterogeneity in variance along this second common axis of mutational variance.

472 Overall, our results suggest that even relatively modest mutation accumulation
473 experiments may provide valuable information on the multivariate axes of traits space to which
474 mutation contributes the greatest new variance each generation. Such insights are valuable,
475 particularly for deepening our understanding of how mutation and selection interact to shape

476 standing genetic variance, and long-term responses to selection – for example, does mutation
477 predominantly introduce variance to traits with relatively high standing genetic variation (as
478 observed for Drosophildae wings: Houle *et al.* 2017; Dugand *et al.* 2021) or is there a mismatch
479 between mutational and standing genetic variance implicating selection acting against new
480 mutations (for example, as observed for thermal performance traits: Latimer *et al.* 2014). The
481 current study, prior studies of *Drosophila* wings (Houle and Fierst 2013; Houle *et al.* 2017;
482 Dugand *et al.* 2021), and studies of other trait types and species (e.g., Estes *et al.* 2005; Latimer
483 *et al.* 2014; Hine *et al.* 2018; Mallard *et al.* 2023; Miller *et al.* 2023) have all highlighted the
484 challenge of detecting mutational variance for much of the trait space, with (near) zero
485 eigenvalues (so-called nearly null spaces: Gomulkiewicz and Houle 2009) observed in all
486 studies. While it will remain challenging to characterize the properties of such nearly null
487 phenotypes, determining whether they reflect chance sampling of mutations in a single
488 experiment, or persistently low input of mutation, is an important first step to improving our
489 understanding of how mutation might limit evovable variation.

490 In conclusion, we have shown substantial variability in the estimated magnitude (and
491 confidence range) of mutational (co)variances of traits, but also that this heterogeneity, driven
492 by uncontrollable micro-environmental variation and estimation error, can be partitioned out
493 in multivariate analysis. The variability we observe is likely at the lower end of the range given
494 our selection of mutationally variable traits (i.e., exclusion of any trait with a zero estimate of
495 among-line variance in any of the 12 samples) coupled with our focus on morphological traits,
496 which might typically have lower non-genetic contributions to observed variance than other
497 types of traits (Houle *et al.* 1996). The heterogeneity among estimates suggests that we will
498 have low confidence in the magnitude of individual mutational variances and pairwise
499 mutational covariances. Yet, uncorrelated non-mutational contributions to trait variability
500 might allow multivariate analyses to generate relatively robust insight into general

501 characteristics of mutational covariances. Qualitatively similar patterns of multivariate
502 mutational effects on *Drosophila* wing shape have been reported in different species, consistent
503 with relatively conserved pleiotropic effects of mutations. Drosophiladae wings, whose
504 development is programmed by highly conserved developmental patterning genes (Neto-Silva
505 *et al.* 2009; Matamoro-Vidal *et al.* 2015) and which evolve slowly (Houle *et al.* 2017) may be
506 atypical in having such conserved patterns of mutational covariance. Nonetheless, the
507 experimental approaches of replicated estimation and leveraging multivariate analyses have
508 promise for improving confidence in estimates of mutational covariance, and of detecting any
509 true differences in patterns among genetic backgrounds.

DATA AVAILABILITY

All data analysed in this paper are available at doi 10.6084/m9.figshare.14913051.v3

ACKNOWLEDGMENTS

The authors thank Stephen F. Chenoweth for providing us with the D. serrata MA lines, and Adam Reddiex, Nicholas Appleton, Jack Price and Derek Sun for their help with maintaining the flies and collecting the data. The authors also thank Emma Hine for helpful discussions about analyses.

CONFLICTS OF INTEREST

None declared.

522 **LITERATURE CITED**

523 Aguirre, J. D., E. Hine, K. McGuigan and M. W. Blows, 2014 Comparing G: multivariate
524 analysis of genetic variation in multiple populations. Heredity 112: 21-29,
525 10.1038/hdy.2013.12.

526 Baselmans, B. M., R. Jansen, H. F. Ip, J. van Dongen, A. Abdellaoui, M. P. van de Weijer, . .
527 . G. Willemsen, 2019 Multivariate genome-wide analyses of the well-being spectrum.
528 Nature Genetics 51: 445-451.

529 Bitner-Mathé, B. C., and L. B. Klaczko, 1999 Plasticity of *Drosophila melanogaster* wing
530 morphology: effects of sex, temperature and density. Genetica 105: 203-210.

531 Blair, S. S., 2007 Wing vein Patterning in *Drosophila* and the analysis of intercellular signaling.
532 Annual Review of Cell and Developmental Biology 23: 293-319,
533 10.1146/annurev.cellbio.23.090506.123606.

534 Blows, M. W., S. F. Chenoweth and E. Hine, 2004 Orientation of the genetic variance-
535 covariance matrix and the fitness surface for multiple male sexually selected traits. Am
536 Nat 163: 329-340, 10.1086/381941.

537 Blows, M. W., and K. McGuigan, 2015 The distribution of genetic variance across phenotypic
538 space and the response to selection. Molecular Ecology 24: 2056-2072,
539 10.1111/mec.13023.

540 Bürger, R., 2000 *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley,
541 Chichester.

542 Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, . . . J. O'Connell, 2018
543 The UK Biobank resource with deep phenotyping and genomic data. Nature 562: 203-
544 209.

545 Carreira, V. P., I. M. Soto, J. Mensch and J. J. Fanara, 2011 Genetic basis of wing
546 morphogenesis in *Drosophila*: Sexual dimorphism and non-allometric effects of shape
547 variation. BMC Developmental Biology 11: 32, 10.1186/1471-213x-11-32.

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

551 Conradsen, C., M. W. Blows and K. McGuigan, 2021 Variability of Mutational Variance
552 Estimates: MA_dserrata_wings. Figshare
553 https://doi.org/10.6084/m9.figshare.14913051.v3,
554 https://doi.org/10.6084/m9.figshare.14913051.v3.

555 Conradsen, C., M. W. Blows and K. McGuigan, 2022 Causes of variability in estimates of
556 mutational variance from mutation accumulation experiments. Genetics 221: iyac060,
557 10.1093/genetics/iyac060.

558 Dugand, R. J., J. D. Aguirre, E. Hine, M. W. Blows and K. McGuigan, 2021 The contribution
559 of mutation and selection to multivariate quantitative genetic variance in an outbred
560 population of *Drosophila serrata*. Proceedings of the National Academy of Sciences
561 118: e2026217118.

562 Dugand, R. J., M. W. Blows and K. McGuigan, 2023 Using inbreeding to test the contribution
563 of non-additive genetic effects to additive genetic variance: A case study in *Drosophila*

serrata. Proceedings of the Royal Society B-Biological Sciences 290, 10.1098/rspb.2022.2111.

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.

Eyre-Walker, A., 2010 Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proceedings of the National Academy of Sciences of the United States of America 107**:** 1752-1756, 10.1073/pnas.0906182107.

Gomulkiewicz, R., and D. Houle, 2009 Demographic and genetic constraints on evolution. The American Naturalist 174**:** E218-E229, 10.1086/645086.

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, DOI 10.1146/annurev.ecolsys.39.110707.173437.

Hill, W. G., and R. Thompson, 1978 Probabilities of non-positive definite between-group or genetic covariance matrices. Biometrics 34**:** 429-439.

Hine, E., and M. W. Blows, 2006 Determining the effective dimensionality of the genetic variance–covariance matrix. Genetics 173**:** 1135-1144.

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, 10.1534/genetics.118.300757.

Houle, D., 1991 Genetic covariance of fitness correlates: What genetic correlations are made of and why it matters. Evolution 45**:** 630-648.

Houle, D., G. H. Bolstad and T. F. Hansen, 2020 Fly wing evolutionary rate is a near-isometric function of mutational variation. bioRxiv**:** 2020.2008.2027.268938, 10.1101/2020.08.27.268938.

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, 10.1111/j.1558-5646.2012.01838.x.

Houle, D., K. A. Hughes, D. K. Hoffmaster, J. Ihara, S. Assimacopoulos and B. Charlesworth, 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. Genetics 138**:** 773-785.

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.

Johnson, T., and N. Barton, 2005 Theoretical models of selection and mutation on quantitative traits. Philosophical Transactions of the Royal Society B: Biological Sciences 360**:** 1411-1425, DOI 10.1098/rstb.2005.1667.

Johnstone, I. M., 2007 High dimensional statistical inference and random matrices. Proceedings of the International Congress of Mathematics: 307-333.

Jones, A. G., S. J. Arnold, R. Bürger and D. Houle, 2003 Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution 57: 1747-1760.

Kannan, A., R. J. Dugand, N. C. Appleton, S. F. Chenoweth, C. M. Sgro and K. McGuigan, 2023 Environmental dependence of mutational (co)variances of adaptive traits. Evolution 77: 2341–2351, 10.1093/evolut/qpad154.

Klein, T. W., 1974 Heritability and genetic correlation: statistical power, population comparisons, and sample size. Behavior Genetics 4: 171-189.

Klein, T. W., J. DeFries and C. T. Finkbeiner, 1973 Heritability and genetic correlation: standard errors of estimates and sample size. Behavior Genetics 3: 355-364.

Krzanowski, W., 1979 Between-groups comparison of principal components. Journal of the American Statistical Association 74: 703-707.

Lande, R., 1979 Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402-416.

Latimer, C. A. L., K. McGuigan, R. S. Wilson, M. W. Blows and S. F. Chenoweth, 2014 The contribution of spontaneous mutations to thermal sensitivity curve variation in *Drosophila serrata*. Evolution 68: 1824-1837, Doi 10.1111/Evo.12392.

Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva and J. Willis, 1999 Spontaneous deleterious mutation. Evolution 53: 645-663.

Lynch, M., and B. Walsh, 1998 Correlations between characters, pp. 629-656 in *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.

Mallard, F., L. Noble, C. F. Baer and H. Teotónio, 2023 Variation in mutational (co) variances. G3: Genes, Genomes, Genetics 13: jkac335.

Matamoro-Vidal, A., I. Salazar-Ciudad and D. Houle, 2015 Making quantitative morphological variation from basic developmental processes: Where are we? The case of the *Drosophila* wing. Developmental Dynamics 244: 1058-1073, 10.1002/dvdy.24255.

McGuigan, K., and M. W. Blows, 2007 The phenotypic and genetic covariance structure of Drosophilid wings. Evolution 61: 902-911.

McGuigan, K., J. M. Collet, S. L. Allen, S. F. Chenoweth and M. W. Blows, 2014a Pleiotropic mutations are subject to strong stabilizing selection. Genetics 197: 1051-1062.

McGuigan, K., J. M. Collet, E. A. McGraw, Y. X. H. Ye, S. L. Allen, S. F. Chenoweth and M. W. Blows, 2014b The nature and extent of mutational pleiotropy in gene expression of male *Drosophila serrata*. Genetics 196: 911-921, 10.1534/genetics.114.161232.

McGuigan, K., L. Rowe and M. W. Blows, 2011 Pleiotropy, apparent stabilizing selection and uncovering fitness optima. Trends in Ecology & Evolution 26: 22-29, DOI 10.1016/j.tree.2010.10.008.

Meyer, K., and D. Houle, 2013 Sampling based approximation of confidence intervals for functions of genetic covariance matrices. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 523-526.

Mezey, J. G., and D. Houle, 2005 The dimensionality of genetic variation for wing shape in *Drosophila melanogaster* Evolution 59: 1027-1038, 10.1554/04-491.

Miller, C. L., D. R. Sun, L. H. Thornton and K. McGuigan, 2023 The contribution of mutation to variation in temperature-dependent sprint speed in zebrafish, *Danio rerio*. American Naturalist 202: 519-533, 10.1086/726011.

Neto-Silva, R. M., B. S. Wells and L. A. Johnston, 2009 Mechanisms of growth and homeostasis in the *Drosophila* wing. Annual Review of Cell and Developmental Biology 25: 197-220.

Pitchers, W., J. Nye, E. J. Márquez, A. Kowalski, I. Dworkin and D. Houle, 2019 A multivariate Genome-Wide Association Study of wing shape in *Drosophila melanogaster*. Genetics 211: 1429-1447, 10.1534/genetics.118.301342.

Qi, G., and N. Chatterjee, 2018 Heritability informed power optimization (HIPO) leads to enhanced detection of genetic associations across multiple traits. PLoS Genetics 14: e1007549.

R Core Team, 2024 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reddiex, A. J., S. L. Allen and S. F. Chenoweth, 2018 A genomic reference panel for *Drosophila serrata*. G3: Genes, Genomes, Genetics 8: 1335-1346.

Rohlf, F., 2007 tpsRelw version 1.45. Department of Ecology and Evolution, State University of New York, Stony Brook.

Rohner, P. T., and D. Berger, 2023 Developmental bias predicts 60 million years of wing shape evolution. Proceedings of the National Academy of Sciences of the United States of America 120: e2211210120, 10.1073/pnas.2211210120.

Rowe, L., and D. Houle, 1996 The lek paradox and the capture of genetic variance by condition dependent traits. Proceedings of the Royal Society B-Biological Sciences 263: 1415-1421.

Santiago, E., J. Albornoz, A. Dominguez, M. A. Toro and C. Lopez-Fanjul, 1992 The distribution of spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. Genetics 132: 771-781.

Schluter, D., 1996 Adaptive radiation along genetic lines of least resistance. Evolution 50: 1766-1774.

Schmitz, S., S. S. Cherny and D. W. Fulker, 1998 Increase in power through multivariate analyses. Behavior Genetics 28: 357-363, 10.1023/a:1021669602220.

Svensson, E. I., S. J. Arnold, R. Burger, K. Csillery, J. Draghi, J. M. Henshaw, . . . A. Runemark, 2021 Correlational selection in the age of genomics. Nature Ecology & Evolution 5: 562-573, 10.1038/s41559-021-01413-3.

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

Turley, P., R. K. Walters, O. Maghzian, A. Okbay, J. J. Lee, M. A. Fontana, . . . N. A. Furlotte, 2018 Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature Genetics 50: 229-237.

Walsh, B., and M. W. Blows, 2009 Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. Annual Review of Ecology Evolution and Systematics 40: 41-59, 10.1146/annurev.ecolsys.110308.120232.

Walsh, B., and M. Lynch, 2018 Maintenance of quantitative genetic variation, pp. 1015–1078 in *Evolution and Selection of Quantitative Traits*. Oxford University Press.

Wu, C., 2020 Multi-trait genome-wide analyses of the brain imaging phenotypes in UK Biobank. Genetics 215**:** 947-958.

TABLES AND FIGURES

Table 1. The variance associated with independent axes of trait mutational variance. Eigenanalyses were performed separately for **M** in each population size treatment (Treat: Small or Large) in each generation (Gen: 1 through 6). The eigenvalues (i.e., λ) and their 90% CIs* are reported, along with the number of eigenvectors (*j*) accounting for at least 90% of the variation. These *j* eigenvectors were included in the estimation of the Krzanowski subspace. The estimated **M** and the eigenvector loadings are reported in Tables S1 and S3, respectively.

Treat	Gen	m_1	m_2	m_3	m_4	m_5	m_6	j
	1	0.495	0.102	0.084	0.027	0.001	-0.024	3
		0.193; 0.798	0.014; 0.191	0.004; 0.165	-0.036; 0.092	-0.025; 0.027	-0.073; 0.026	
	2	0.506	0.270	0.066	0.015	-0.014	-0.050	2
		0.206; 0.804	0.133; 0.402	-0.005; 0.136	-0.013; 0.043	-0.040; 0.012	-0.091; -0.009	
	3	0.276	0.116	0.037	0.016	-0.011	-0.056	2
Small		0.079; 0.472	0.032; 0.197	-0.028; 0.102	-0.039; 0.071	-0.042; 0.020	-0.138; 0.028	
	4	0.279	0.205	0.051	0.019	0.000	-0.013	3
		0.087; 0.471	0.029; 0.375	-0.021; 0.121	-0.008; 0.045	-0.025; 0.025	-0.044; 0.017	
	5	0.429	0.159	0.114	0.029	0.010	-0.07	3
		0.213; 0.648	0.023; 0.297	0.008; 0.219	-0.001; 0.060	-0.015; 0.034	-0.147; 0.007	
	6	0.439	0.093	0.047	0.021	-0.004	-0.037	2
		0.238; 0.641	0.020; 0.165	-0.001; 0.094	-0.006; 0.049	-0.044; 0.034	-0.099; 0.025	
	1	0.446	0.267	0.070	0.033	0.008	-0.007	3
		0.211; 0.676	0.131; 0.406	0.007; 0.136	-0.017; 0.082	-0.013; 0.030	-0.040; 0.026	
	2	0.313	0.215	0.077	0.052	0.025	0.000	4
		0.164; 0.460	0.073; 0.354	0.015; 0.138	-0.006; 0.110	-0.000; 0.050	-0.020; 0.019	
	3	0.529	0.144	0.070	0.020	-0.020	-0.038	2
		0.272; 0.792	-0.014; 0.300	0.004; 0.137	-0.009; 0.051	-0.069; 0.029	-0.115; 0.039	
Large	4	0.303	0.160	0.101	0.020	-0.006	-0.029	3

	0.144; 0.462	0.046; 0.273	0.030; 0.170	-0.016; 0.058	-0.025; 0.014	-0.069; 0.010	
5	0.304	0.163	0.059	0.023	0.001	0.000	3
	0.151; 0.461	0.067; 0.257	-0.019; 0.136	-0.015; 0.060	-0.054; 0.057	-0.021; 0.020	
6	0.398	0.129	0.112	0.022	0.004	-0.055	3
	0.194; 0.602	0.050; 0.209	0.010; 0.214	0.004; 0.039	-0.035; 0.042	-0.111; 0.002	

* CIs were determined by projecting each eigenvector (e_1: e_6 as reported in Table S3) through the 10,000 REML-MVN samples of the respective population's **M**. Although a positive lower CI suggests statistical support for non-zero variance, REML-MVN (and Bayesian MCMC posterior) CI for eigenvalues are known to be biased, leading to false positive inference for leading eigenvalues (Sztepanacz and Blows 2017). We rely instead on the robust Factor Analytic modelling (Table S2) to infer statistical support.

Table 2. Eigenvalues and eigenvectors of the Krzanowski subspace matrix, H. The 95% CI (below estimate) for the eigenvalues (λ) of the observed data **H** were determined from Krzanowski subspace analyses applied to each of the 10,000 REML-MVN samples of each of the 12 **M***. For the null data **H**, the reported eigenvalue is the mean (and 95% CI the 5% quantiles) of the Krzanowski subspace analyses applied to the 12 **M** estimated from the 1000 randomised datasets.

	h_1	h_2	h_3
λ	11.62	8.86	6.62
	10.35; 11.49	6.87; 9.27	5.16; 7.14
Null λ	11.38	10.93	8.56
	10.61; 11.79	9.35; 11.66	5.77; 10.88
CS	-0.37	0.76	-0.32
ILD1.2	0.27	0.41	0.36
ILD1.5	0.65	-0.05	-0.46
ILD2.5	0.20	-0.03	0.64
ILD2.8	0.16	0.42	0.34
ILD3.7	-0.55	-0.27	0.19

* Eigenvalues of **H** are bounded statistics (ranging from 0 to p – here, 0 to 12). In conjunction with the nature of sampling error, which independently affects each estimated **M** and thus contributes dissimilarity, this has resulted in the eigenvalue of h_1 falling above its upper REML-MVN CI. Importantly, the observed λ_1 is within the range observed for the null λ_1, which estimates the error in h_1 when the 12 samples of **M** are all drawn from the same population.

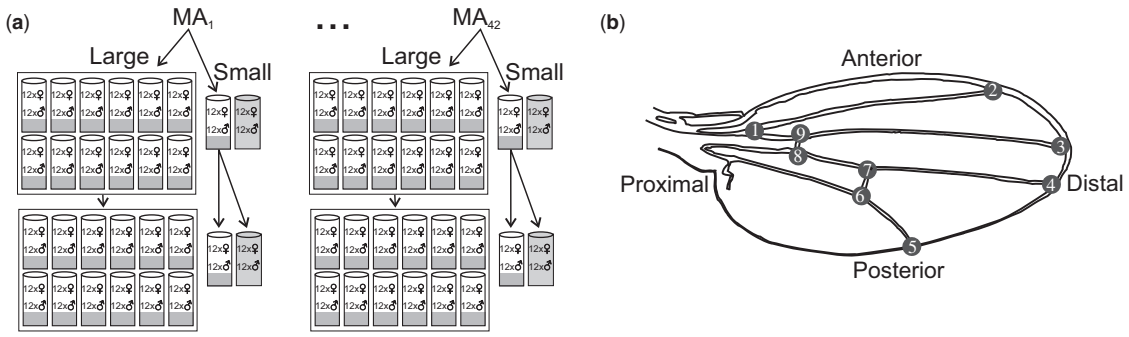

Figure 1. Schematic of (a) experimental design and (b) *Drosophila* wing traits. (a) 42 MA lines each founded two sublines: Small (S; 12 virgin males and 12 virgin females) and Large (L; 144 virgin males and 144 virgin females, distributed evenly among 12 vials). These 84 lines (S and L subline per 42 MA lines) were maintained at these census population sizes for six generations (first two shown here). Each generation, all emergent flies from the 12 vials per L subline were pooled prior to virgin collection. For S sublines, the focal vial contributed offspring to the next generation, while a second, replicate, vial (grey shaded) was used for wing collection only. Each generation, six randomly chosen males were sampled from each of two vials per line (focal and replicate vials for S; randomly chosen two for L). (b) Positions of nine landmarks were recorded and aligned. Six traits were analysed, five inter-landmark distances (named by their end landmarks: ILD1.2, ILD1.5, ILD2.5, ILD2.8, ILD3.7) and wing size, characterised as centroid size (CS).

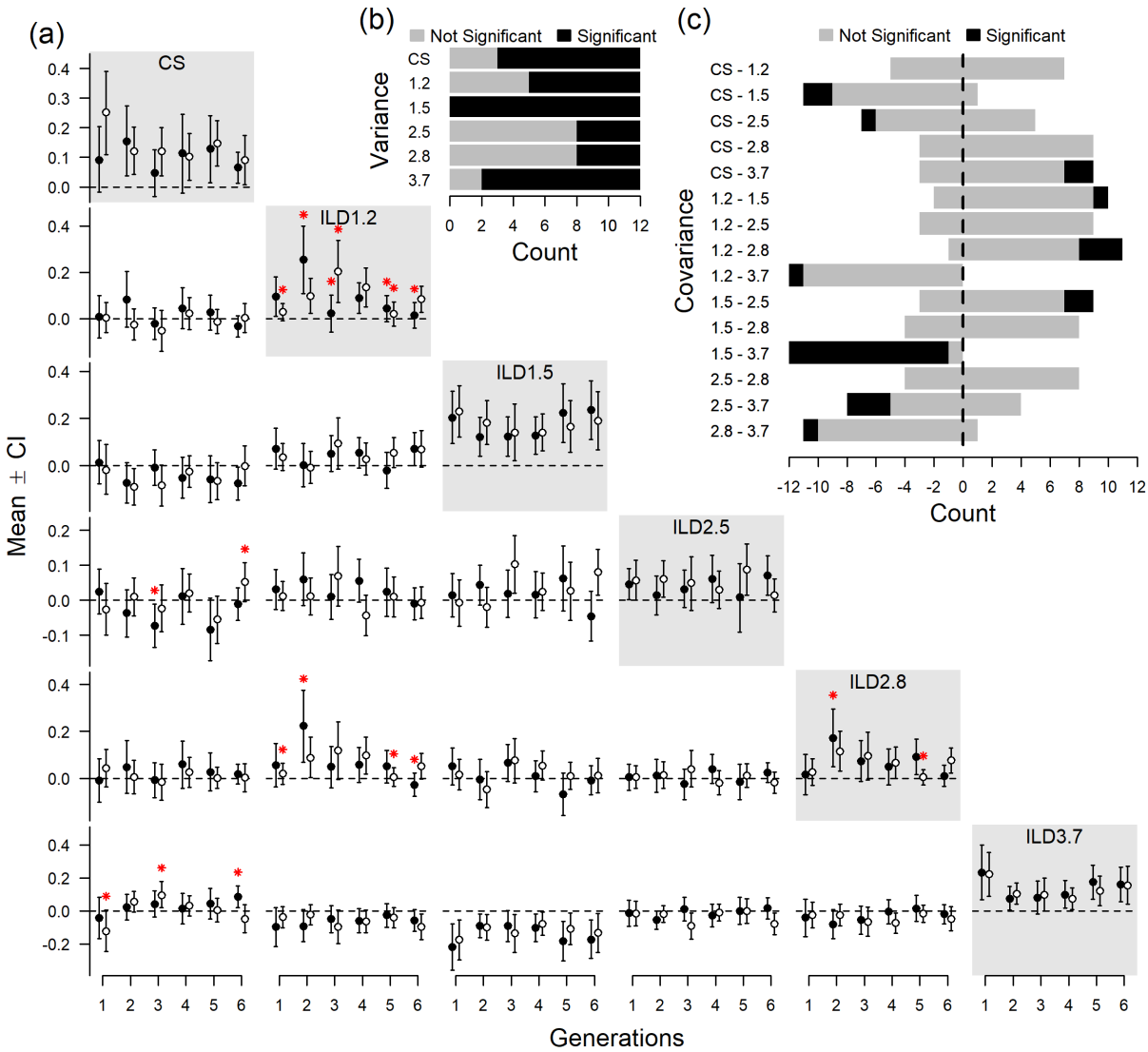

Figure 2. The variability among the 12 population sample estimates of among-line variance and covariance. (a) The REML estimate (point) and REML-MVN bootstrap confidence intervals of among-line variance (on diagonal, grey boxes; 90% CI) and covariances (below the diagonal; 95% CI) in each population size treatment (Small or Large, closed and open symbols, respectively) in each of six generations. The horizontal dashed line indicates zero; statistical support was concluded when CIs did not include zero (a one-tailed test for variances and two-tailed for covariances). Red stars highlight estimates where CI did not overlap with at least one other estimate of the same parameter. Inset plots are the counts of the 12 estimates of (b) variance for each trait or (c) pairwise covariances between traits that were (black) or were not (grey) statistically supported as different from zero. For the covariances,

the counts of estimates are further plotted by sign (negative or positive). See Figure 1b for trait definitions and Table S1 for the parameter estimates. NB: two estimates of covariance between ILD2.5 and ILD3.7 were exactly 0; these were classified as positive.

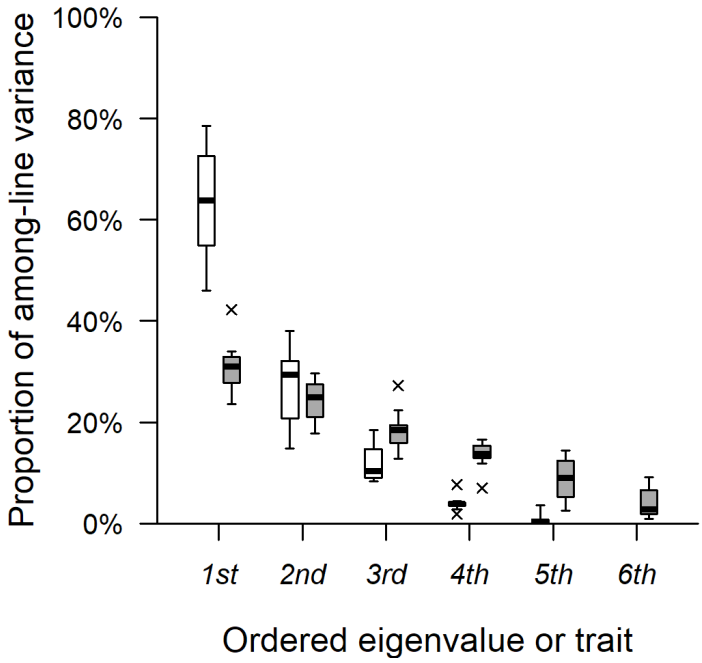

Figure 3. The distribution of among-line variation of six wing traits in 12 population samples. Open boxplots represent distribution across the 12 **M** of the ranked eigenvalues (largest to smallest, x-axis 1st to 6th), plotted as the proportion of the variance in their respective **M**. For comparison, among-line variances of individual traits (diagonals of Table S1) were ranked from largest to smallest within each population sample, and their distribution also plotted (grey boxplots). Note, the combination of the six traits (i.e., eigenvector) associated with a specific ordered eigenvalue (1st through 6th), and the specific trait ranked 1st through 6th may differ among the 12 matrices. For each boxplot, the median of the distribution is shown as a black band, boxes represent the interquartile range, whiskers represent 1.5 times the interquartile range, and estimates that fell outside this range are designated with crosses.

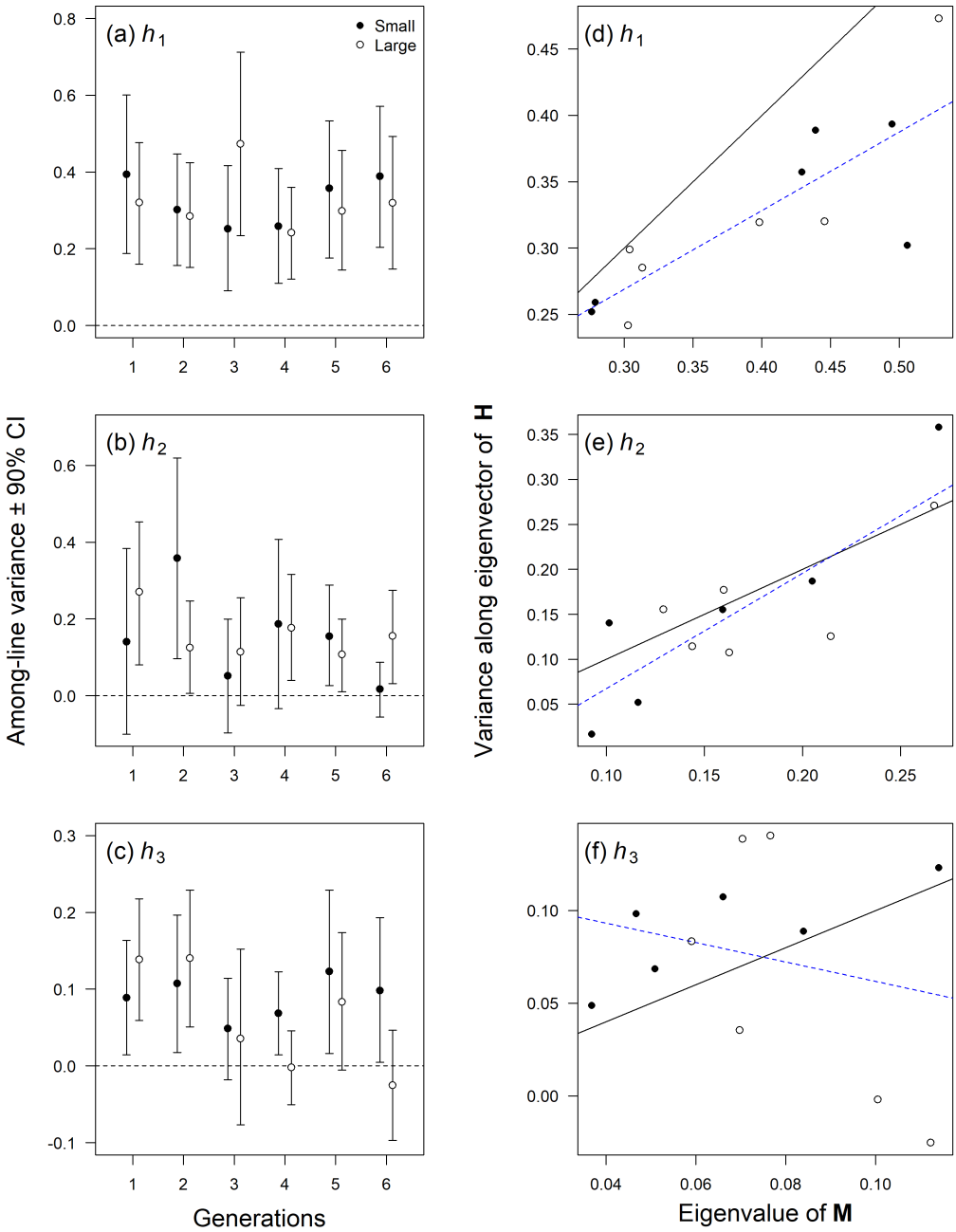

Figure 4. The among-line variance associated with each of the three eigenvectors of the Krzanowski's subspace. Left-hand panels: The magnitude of the among-line variance (90% REML-MVN CI) is plotted for (a) h_1, (b) h_2 and (c) h_3. Within each plot, the populations are ordered by generation (x-axis), then by population size treatments (Small or Large, closed and open symbols, respectively). The dashed lines indicate zero. Note that the y-axis range differs among panels, reflecting decreasing total variance associated with each successive common subspace vector (a to c). Right-hand panels: the magnitude of variance along eigenvectors of

H are plotted against the same-rank eigenvalue of each population sample's respective **M**. (a) h_1 vs m_1, (b) h_2 vs m_2 and (c) h_3 vs m_3. Population size treatment is again indicated by symbol (Small or Large, closed and open symbols, respectively). The dashed blue line indicates the regression slope; solid grey line is the 1:1 line indicating equal magnitude along x and y axes.