doi:10.1111/j.1558-5646.2012.01838.x

# PROPERTIES OF SPONTANEOUS MUTATIONAL VARIANCE AND COVARIANCE FOR WING SIZE AND SHAPE IN *DROSOPHILA MELANOGASTER*

**David Houle[1,2] and Janna Fierst[1,3]**

[1] *Department of Biological Science, Florida State University, Tallahassee, Florida 32306*

[2] *E-mail: dhoule@bio.fsu.edu*

[3] *University of Oregon Institute of Ecology and Evolutionary Biology Eugene, Oregon 97403*

Received May 10, 2012
Accepted October 6, 2012
*Data Archived: Dryad doi:10.5061/dryad.3b7g5*

**We estimated mutational variance–covariance matrices, M, for wing shape and size in two genotypes of *Drosophila melanogaster* after 192 generations of mutation accumulation. We characterized 21 potentially independent aspects of wing shape and size using geometric morphometrics, and analyzed the data using a likelihood-based factor-analytic approach. We implement a previously unused analysis that describes those directions with the greatest difference in evolvability between pairs of matrices. There are significant mutational effects on 19 of 21 possible aspects of wing form, consistent with the high dimensionality of standing genetic variation for wing shape previously identified in *D. melanogaster*. Mutations have partially recessive effects, consistent with average dominance around 0.25. Sex-specific matrices are relatively similar, although male-specific matrices are slightly larger, as expected due to dosage compensation on the X chromosome. Genotype-specific matrices are quite different. Matrices may differ both because of sampling error based on small samples of mutations with large phenotypic effects, and because of the mutational properties of the genotypes. Genotypic differences are likely to be involved, as the two genotypes have different molecular mutation rates and properties.**

**KEY WORDS:** **Covariance matrix, dimensionality, dominance, evolvability, genetic architecture, geometric morphometrics.**

That mutation is critical to evolution is a truth universally acknowledged, yet until very recently nearly every paper on mutation could ruefully note how understudied mutation was. In one key respect this situation has changed: The drop in sequencing costs and increase in accuracy has made it relatively easy to estimate the base pair mutation rate by sequencing genotypes with known evolutionary histories (Schrider et al. 2011). The importance of mutation for evolution, however, is not principally in its molecular details. The phenotypic and fitness effects of mutations and the rate at which they arise are ultimately most important for the study of phenotypic variation, adaptation, and diversification. Although the study of genotypic effects of mutation has now been relocated to the well-lit upper floors of biology, mutation's phenotypic effect is still in the "dingy basement" where Hermann Muller found it in 1921 (Muller 1973).

The reasons that the phenotypic effects of mutation are less well studied are easy to discern: measuring phenotypes is time consuming, and particularly large sample sizes are necessary for estimates of genetic parameters. Although the investment in genomics has made the study of sequence variation orders of magnitude easier, there has been no comparable investment in phenomics, the high-throughput measurement of phenotypes (Houle et al. 2010). Phenotypic effects of individual mutations have been measured in a few simple organisms (e.g., Sanjuán et al. 2004), and attempts to connect specific spontaneous mutations with phenotypic effects are beginning to be made in multicellular






EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

organisms (e.g., Rutter et al. 2012), but serve mainly to highlight the difficulties of connecting genotype and phenotype.

What we can study are the aggregate effects of mutations accumulated over may generations in mutation–accumulation experiments (Houle and Kondrashov 2006; Halligan and Keightley 2009). The mutational covariance matrix, **M**, summarizes the properties of such collections of mutations. Theory (Lande 1980; Turelli 1985; Zhang 2012) and simulations (Jones et al. 2003, 2004; Jones et al. 2007) make clear that the details of the **M** matrix can affect a wide variety of population properties, including evolvability—the capability of the genetic system to produce potentially adaptive variants— (Hansen 2006), segregating genetic variation, the stability of the variation across space and time, and the predictability of evolution.

Despite the importance of these questions, estimates of **M** matrices are few (Houle et al. 1996; Keightley and Halligan 2009). The number of mutation accumulation (MA) studies that have measured more than one phenotype is rather small. These studies have resulted in estimates of **M** matrices for as many as 11 traits (Lynch 1985), but typically many fewer (e.g., Santiago et al. 1992; Houle et al. 1994; Shaw et al. 2000; Joyner-Matos et al. 2009). Most studies have been directed at measurement of fitness and its components (e.g., Rutter et al. 2010), either to understand the maintenance of genetic variance (Eyre-Walker and Keightley 2007; Mackay 2010) or the evolution of life histories (e.g., Houle et al. 1994; Joyner-Matos et al. 2009). Although fitness components are of great importance, they are also difficult to measure well, so these estimates are usually plagued by large uncertainties (e.g., Kavanaugh and Shaw 2005). Available **M** matrices are based on a wide variety of unrelated traits, what I have called extensive characterization of the phenotype (Houle 2010; Houle et al. 2010). The alternative, intensive characterization of mutational effects on one related set of phenotypes, has not yet been attempted. A previous study of the effect of spontaneous mutations on *Drosophila melanogaster* wings used just two traits—wing length and width (Santiago et al. 1992).

We present and compare estimates of the **M** matrices for wing shape and size in *D. melanogaster* based on 192 generations of MA in two genotypes. We measured the location of 12 vein intersections in 12,000 flies. We address two classes of questions about **M**. The first is the genetic architecture of mutational effects (Hansen 2006; Keightley and Halligan 2009); properties such as the amount of new phenotypic variation produced, the degree of dominance and epistatic interactions. The second is the "quality" of mutational variation (Hansen and Houle 2004; Arnold et al. 2008), the pleiotropic effects of mutations and their distribution in phenotype space. Previous analyses of this MA experiment have considered the mutational bias and variance in fitness, transposable element insertions (Houle and Nuzhdin 2004), gene expression (Rifkin et al. 2005) and nuclear (Haag-Liautard et al. 2007; Schrider et al. unpubl. ms.) and mitochondrial (Haag-Liautard et al. 2008) mutations.

# Methods

## MUTATION ACCUMULATION

The MA portion of the experiment is detailed in Houle and Nuzhdin (2004) and in the Supplementary Material, section 1.1 (SI-1.1). Each of two inbred lines, IV*e*-33 and IV*e*-39, were used to found a population of 75 MA sublines. Thereafter these sublines were maintained by full-sib mating of single pairs of flies of virgin flies as much as possible to minimize natural selection.

## WING MEASUREMENT

We measured the spatial locations of 12 vein intersections on wings of live flies using an automated image-analysis system (Wingmachine; Houle et al. 2003). The data were aligned by generalized Procrustes least squares superimposition (Rohlf and Slice 1990), which scales point configurations to unit centroid size, translates all centroids to the same coordinates, and rotates the scaled coordinates to minimize the sum of the squared distances between configurations. The result of the superimposition is that each wing is represented by the x- and y-coordinates of the displacement of each landmark from the centroid, measured in units of centroid size. The raw data consist of 24 coordinates, but three degrees of freedom are lost due to fitting the nuisance translation and rotation variables, although information about size is transferred from the coordinate data to centroid size. Additional measurement and analysis details are summarized in SI-1.2 and SI-1.3.

## EXPERIMENTS

Flies were measured in four different experiments during the MA process (SI-1.2, Tables S1 and S2). In the 1998 experiment, we measured wings of male flies descended from control genotypes and homozygous IV*e*-39 sublines after 73 generations of MA. In the 2002 experiment, homozygous MA flies of both sexes were imaged from all surviving sublines of both lines IV*e*-33 and IV*e*-39 between generations 166 and 173. The Cross 2002 experiment consists of two partial diallel crosses of four sublines each performed at generation 170 with sublines from IV*e*-39. Finally, in the Cross 2003 experiment, the surviving sublines were used to perform a series of full diallel crosses each utilizing four sublines.

## MUTATION MODELS

A total of $n$ sites are capable of influencing a trait, and the $i$th site has a haploid mutation rate $u_i$. A mutation causes deviation $a_i$ from the mean in heterozygous condition and $2a_i + d_i = g_i$ in homozygous condition, where $d_i$ is the dominance deviation at

site $i$. The probability that each individual carries a new mutation at the $i$th site is $2u_i$. Lynch and Hill (1986) defined two different mutational variances—one based on the effects of mutations once they are fixed, which we call the homozyous model, and the other on the effects of mutant alleles before they can segregate into homozygotes, which is the nonsegregational model. The homozygous model defines the mutational variance, $V_M$, as

$$\begin{aligned} V_M &= 2\sum_i u_i \left(\frac{g_i}{2}\right)^2 = \frac{1}{2}U\mathrm{E}[g^2] \\ &= U\left(2\mathrm{E}[a^2] + \frac{1}{2}\mathrm{E}[d^2] + \mathrm{COV}[a,d]\right), \end{aligned} \tag{1}$$

where $U = \sum_i \mu_i = n\bar{\mu}$ is the total haploid mutation rate, and $E[x^2]$ is the expected squared mutational effect $x_i$ over all loci. The nonsegregational mutational variance, $V_{Mns}$, is

$$V_{Mns} = 2\sum_i u_i a_i^2 = 2U\,E[a^2]. \tag{2}$$

In the case where effects are additive, $g_i = 2a_i$, and $V_M = V_{Mns}$. The difference between $V_M$ and $V_{Mns}$ gives an indication of the importance of dominance and epistasis among new mutations.

In SI-1.6 we derive the interpretation of covariances in terms of these two models. The resulting matrix of heterozygous line effects, **B**, is used to estimate $V_{Mns}$, whereas the among homozygous subline covariance matrix, **S**, is used to estimate $V_M$.

## ANALYSES

Quantitative genetic analyses were carried out in Wombat (Meyer 2007, 2010). Wombat implements mixed model analyses of continuous data by restricted maximum likelihood. It handles datasets with large numbers of traits, and fits reduced rank models to the variance component matrices (Kirkpatrick and Meyer 2004; Meyer and Kirkpatrick 2005, 2008). The default approach to estimating genetic covariance matrices in Wombat assumes that the individuals at the head of the pedigree are outbred, which is not the case with our data. We directly calculated the relationship matrix using the coefficients derived in SI-1.4. Sex, block, and line were treated as fixed effects.

For each combination of sex and genotype, we analyzed the combined heterozygous and homozygous genotypes to estimate the covariance matrix **B** and the nonsegregational mutation matrix $\mathbf{M}_{ns} = 2\mathbf{B}/192$. In addition, we analyzed the homozygous data alone to estimate the subline covariance matrices **S** and the total mutation matrix $\mathbf{M} = \mathbf{S}/(2 \times 192)$. We compare **M** matrices based on their evolvabilities, $e$ (Hansen and Houle 2008). Evolvability is the predicted response to a unit-length selection gradient in the direction of that gradient, and is equivalent to the variance in the direction of the gradient. Size evolvability was calculated by dividing the mutational variance of wing centroid size by the square of mean centroid size. This standardization places size and shape on comparable scales (Mitteroecker et al. 2004).

Other statistical analyses were carried out in SAS (SAS Institute 2004). Additional details are in SI-1.5.

## COMMON SUBSPACES AND EVOLVABILITY RATIO ANALYSIS

The **M** matrices estimated are all less than full rank, which complicates comparisons among them. One key aspect of such comparisons is to characterize the overlap in the subspaces covered by pairs of matrices, call them $\mathbf{M}_A$ and $\mathbf{M}_B$. Phenotypic space can first be divided into the common or "non-null" subspace where both $\mathbf{M}_A$ and $\mathbf{M}_B$ possess statistically significant variation versus the complementary "nearly-null" space (Gomulkiewicz and Houle 2009). The nearly-null space can be further partitioned into the "doubly nearly-null" space, the space in which neither matrix possesses significant variation, and two "singly nearly-null" spaces, in which one matrix has significant variation, whereas the other does not.

Flury (1983) showed that eigenanalysis of $\mathbf{C} = \mathbf{M}_A^{-1}\mathbf{M}_B$ estimates the directions that differ most in the variance of one covariance matrix relative to another. We exploited this result to partition phenotype space into the categories of non-null, singly null and doubly-null, described in SI-1.7. The eigenvectors and eigenvalues of **C** calculated in the common subspace of the two matrices are also useful for describing what is different about $\mathbf{M}_A$ and $\mathbf{M}_B$. The eigenvectors, $\mathbf{V}_C$, estimate the directions that have most extreme ratios of variances in $\mathbf{M}_B$ relative to $\mathbf{M}_A$. The eigenvectors therefore locate directions with the largest and smallest ratios of evolvabilities in genotype B relative to A. The eigenvalues, $\Lambda_C$, estimate the ratio of evolvabilities in those directions. Mitteroecker and Bookstein (2009) derived a distance metric for covariance matrices based on $\mathbf{V}_C$ that we use to represent the relationship between our estimates of **M**.

## DATA ARCHIVED

Data for the results in this article are archived in the Dryad repository: doi:10.5061/dryad.3b7g5.

# Results

## DO MUTATIONS HAVE DIRECTIONAL EFFECTS?

We could directly test for directional mutational bias using the 1998 experiment data, which include IV$e$-39 control populations, as well as IV$e$-39 MA sublines. Variance component analyses in Wombat and the SAS Mixed procedure (SAS Institute 2004) showed no significant variance among the replicates of the control population on the first five principal components of the data, but highly significant variance for scores on these axes among

sublines in the IV*e*-39 MA population. The overall size of the estimated genetic matrices reflect this (control replicate matrix trace 2520, MA subline trace 6010). MANOVA showed no significant mean differences between control and MA genotypes (Table S4). Any mutational bias is too small to be detectable.

Unfortunately, control populations were not available for direct comparison with MA sublines later in the experiment. Multivariate multivariable regressions of shape and size on generation of MA showed significant effects of generation, plus all interactions of generation with sex and genotype, although the effects are of modest size. These significant effects could be due to environmental factors, or to mutational bias. Mutational bias is by definition consistent in rate and direction over time, at least within genotypes, whereas environmental effects would be more likely to be variable over time. This signature of consistent bias should be detectable in males of genotype IV*e*-39, which were measured in 1998, 2002, and 2003. The per-generation rate of change between the 1998 and 2002 experiments was 1.33 shape units/generation, whereas the rate from 2002 to 2003 was 7.00 units/generation. The angle between the directions of change over the two periods is 81°, not the small angle expected if there is a consistent mutational bias. The directions of change within all sex and genotype combinations between the 2002 and 2003 experiments were similar (within 30° of each other, with an average angle of 24.2°), and significantly less than a sample of 10,000 random angles in 21 dimensional space (median 81°, with a lower 95% confidence limit of 65°). The inconsistency between directions and magnitudes of change between 1998 and 2002 and between 2002 and 2003, coupled with the similarities in change over the second interval seem best explained by environmental effects, rather than mutational bias.

## HETEROZYGOUS VERSUS HOMOZYGOUS MEAN

The Cross 2003 experiment allows us to test for directional dominance, which would impart a consistent direction to the difference between inbred and outbred means. Genotypic means on the first two within-genotype canonical axes from four representative blocks of the 2003 experiment are shown in Figure 1. Means of inbred and outbred genotypes do not differ in a consistent direction. Inbred means tend to lie on the periphery of each ordination, and outbred genotypes have means intermediate between the parental inbreds. Both of these observations are consistent with an additive component to allelic effects. Multivariate analyses of variance (Supplementary Results 2.1) showed evidence for dominance in the highly significant interactions of inbreeding status, sex and genotype, but no consistent main effect of inbreeding, lending no support to the hypothesis of directional dominance.

## OVERALL PROPERTIES OF COVARIANCE MATRICES

We estimated covariance matrices for the nonsegregational and homozygous models for the nine different partitions based on genotype (IV*e*33 and IV*e*39) and sex. For each model and partition of the data (e.g., homozygous females for IV*e* 33), we fit a series of models varying the rank of the mutational matrix. We used the corrected Akaike information criterion (AICc) score to judge the best fitting model, then conducted likelihood ratio tests of this best model against similar models to judge the range of models supported (Tables S5 and S6). Table 1 presents the rank of the best-fitting model, the rank of models not significantly different from that best model, estimates of matrix sizes, evolvability, and mutational heritability.

Mutational effects of heterozygous mutations significantly affected at least $k = 7$ phenotypic dimensions. Models of lower rank always fit less well by more than 323 AICc units ($P < 0.0001$ in all cases, Table S5). Models of higher rank returned eigenvalues of all higher rank eigenvectors than those shown in Table 1 as 0, but paradoxically, these models usually had better likelihoods and AICc scores. This behavior is known to occur in Wombat for reduced-rank analyses of some datasets (Meyer and Kirkpatrick 2008; Supplementary Results 2.2). We report results only for models that had only non-0 eigenvalues.

The homozygous model converged at all matrix ranks tested, although the estimated likelihood of higher rank models were usually lower than that of the best model (Table S6). This should not occur, and thus again indicates numerical problems of some sort. In three cases, the best model was not significantly better than the next best model. The ranks of these best-fit homozygous models were all 11 or more. Rank rises as more inclusive partitions of the data are analyzed, supporting up to rank 19 of 21 possible for the pooled dataset including both sexes and genotypes. This pattern of greater inferred rank based on larger datasets is expected due to greater power in the larger datasets.

The estimated mutational and residual covariance matrices for the best-fitting models are presented in Table S7. Models of rank $k$ in Wombat are fit in a $k$-dimensional space, so the **M** matrices have precisely rank $k$, even when these are rotated back into the original 25 dimensional phenotype space, as in Table S7.

## COMPARISONS OF COVARIANCE MATRICES

Table 1 gives the size of the genetic and residual matrices calculated as their traces, the sum of the elements on the main diagonal. Size is a key property of an **M** matrix, because it is linearly related to evolvability, $e$ (Hansen and Houle 2008). **M** was substantially larger in line 39 than in line 33, predicting higher evolvability in a population founded by IV*e*39. **B** was 2.7 times as large, and **S** was 1.6 times as large for the combined sex analyses. The residual variances in line 39 were also a bit higher, 6% higher in the nonsegregational model and 18% higher in the homozygous model. Male mutational variances were 1.4 times higher than female variances in the nonsegregational analysis.

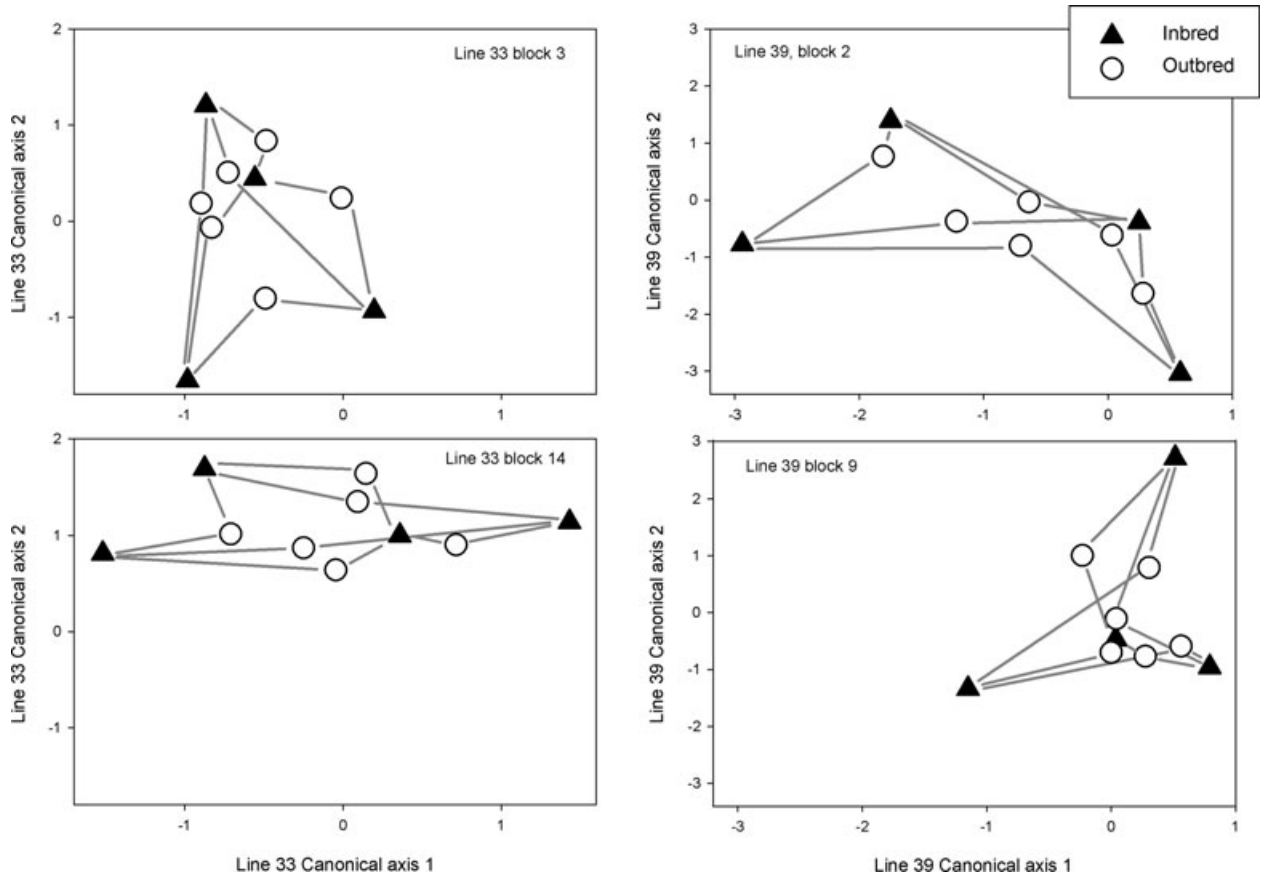


**Figure 1.** **Genotypic means on the first two canonical axes within MA genotypes for representative blocks of the 2003 experiment. Genotypes share the same two parental sublines, regardless of which was used as the maternal parent. Grey lines connect each outbred genotype with its two parental inbred sub-lines. Canonical discriminant analyses were carried out on genotype means within each genotype. The first two axes explain 23% and 14% of the variation in line 33, and 24% and 19% of the variation in line 39.**

For comparison with previous estimates of mutational variance, we obtained an average mutational heritability over all traits, by standardizing each **M** size estimate by its corresponding **R** matrix size. The nonsegregational mutational heritability is a bit greater than 0.1%, near the average of previous estimates for a wide variety of traits in many different organisms (Lynch 1988; Houle et al. 1996) including Drosophila wings (Santiago et al. 1992). The homozygous mutational heritability is approximately two to three times larger. We caution that these mutational heritabilities are not related to the ability of traits to evolve (Hansen et al. 2011).

Comparisons between the sums of partition-specific AICc scores and the AICc scores for corresponding combined analyses from Table S6 show that there are very substantial differences between the properties of mutations in the two genotypes, and in the effect of mutations on the two sexes. All of the combined analyses fit at least 1092 AICc units less well than the more parameter-rich models fit separately to each genotype–sex combination. For example, the sum of the four separate analyses of sex/genotype combinations fits 11212 AICc units better than the combined analysis using the homozygous model. Despite these differences, we regard the combined analyses as the best measures of the average properties of mutational effects because mutations arise in a diversity of genotypes, and are expressed and selected on in both sexes.

To test whether the differences in evolvability between homozygous and nonsegregational models evident in Table 1 is due to the lower dimensionality of the nonsegregational models, we compared evolvability in the first seven dimensions defined by each **M** matrix. The results (Table S8) show that the same differences evident in Table 1 persist in the subspaces where each **M** is full rank.

## AVERAGE DOMINANCE

The lower nonsegregational mutational variances relative to homozygous mutational variances must be due to departures from an additive model. Under the simplifying assumption of perfectly correlated $a$ and $d$, and no epistasis, we can calculate the dominance, $h$, using the parameterization where the phenotypes of the reference homozygote, the heterozygous, and homozygous

**Table 1.** Rank of best fitting mutational models and scalar measures of variance over the entire phenotypic space.

| | | Nonsegregational model | | | Homozygous model | | | Mutational evolvability | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Matrix size[2] | | | Matrix size[2] | | Shape[3] ($\times 10^3$) | | Centroid size[4] ($\times 10^6$) | | Mutational heritability[5] (%) | |
| | | Rank of **B**[1] | **B** | **R** | Rank of **S**[1] | **S** | $\mathbf{R}_T$ | $\bar{e}_{Mns}$ | $\bar{e}_M$ | $e_{Mns}$ | $e_M$ | $h^2_{Mns}$ | $h^2_{MT}$ |
| ♂ | 33 | 7 | 1613 | 27330 | 11(12) | 21156 | 27739 | 0.65 | 2.16 | 1.19 | 3.38 | 0.06 | 0.20 |
| ♂ | 39 | 7 | 4358 | 32503 | 12(13) | 33801 | 36599 | 1.80 | 3.48 | 2.32 | 4.75 | 0.14 | 0.24 |
| ♀ | 33 | 7 | 1409 | 29619 | 12 | 21512 | 30397 | 0.57 | 2.25 | 0.70 | 1.59 | 0.05 | 0.18 |
| ♀ | 39 | 7 | 3678 | 26260 | 12 | 38884 | 29049 | 1.55 | 4.07 | 0.81 | 2.83 | 0.15 | 0.35 |
| ♂ | 33 and 39 | 8 | 2895 | 30825 | 15 | 29996 | 32803 | 1.19 | 3.09 | 1.69 | 3.92 | 0.10 | 0.24 |
| ♀ | 33 and 39 | 7 | 2048 | 28819 | 16 | 30263 | 29975 | 0.85 | 3.15 | 0.76 | 2.53 | 0.07 | 0.26 |
| Both | 33 | 7 | 1804 | 28899 | 15 | 22341 | 29475 | 0.73 | 2.29 | 1.10 | 2.91 | 0.07 | 0.20 |
| Both | 39 | 7 | 4587 | 30772 | 14(15) | 34654 | 34776 | 1.89 | 3.58 | 2.32 | 4.05 | 0.16 | 0.26 |
| Both | 33 and 39 | 7 | 2879 | 30872 | 19 | 31309 | 32266 | 1.18 | 3.24 | 1.58 | 3.61 | 0.10 | 0.25 |

[1]Rank of the best-fitting model. Ranks in parentheses are model not significantly different from the best model by a log likelihood test (Table S6).
[2]Matrix size is the trace of the matrix, or the sum of its eigenvalues.
[3]Average eigenvalue of the shape portion of the M matrix in centroid size units.
[4]Mutational variance of centroid size, standardized to mean 1.
[5]Mutational heritability calculated as the size of the M matrix, divided by the size of the corresponding R matrix.

mutant phenotypes are 0, $hx = a$ and $x = 2a + d$, respectively (SI-1.6). Mutations are on average partially recessive, with $h = 0.26$ for shape and 0.29 for size in the combined dataset. Dominance is quite stable across data partitions, ranging from 0.22 and 0.33 (Table S9). This quantifies the trends shown in Fig. 1, where outbred genotypes are generally less than half as far from the mean as their inbred parents.

## DIFFERENCES IN MUTATIONAL PROPERTIES

We have strong evidence that the **M** matrices estimated from different data partitions are different. This raises the question of what those differences are. We focus on the four independent comparisons of sex-specific estimates pooled over genotypes, and of genotype-specific estimates pooled over sexes, as these are better estimated than comparisons within particular genotypes or sexes.

One simple way in which covariance matrices may differ is by being proportional to each other. We tested whether proportional differences were necessary and sufficient to explain matrix differences by estimating the proportionality constant that, when multiplied by the best estimate of the mutation matrix for the first data partition, gave mutation parameter estimates that maximized the likelihood of the data from the second partition.

For all four comparisons, fitting a proportionality constant significantly increased the fit of one partition's estimates to the other. For the genotype-specific estimates, a proportionality constant of 2.04 for the line 33 estimates improved the fit to the line 39 data by 52.9 log-likelihood units in the nonsegregational case. In the homozygous case a proportionality constant of 2.14 for the line 33 estimates improved the fit to the line 39 data by 73.5 log-likelihood units. For the sex-specific estimates, a proportionality constant of 1.17 for the female estimates improved the fit to the male data by 4.6 log-likelihood units in the nonsegregational case. In the homozygous case a proportionality constant of 1.24 for the female estimates improved the fit to the male data by 10.4 log-likelihood units.

These results demonstrate that proportional differences are a component of the differences between matrices. However, the best fitting proportionality constants always yielded fits that were greatly inferior to the best estimates (1253 and 736 logL units for male–female differences; 4772 and 6278 units for genotype differences in nonsegregational and homozygote models, respectively). There are real differences in matrix size, particularly between genotypes 33 and 39, but the nonproportional differences in matrix structure are quantitatively far more important.

A useful measure of matrix difference is the angle between the responses to the same selection gradients, or "random skewers" (Cheverud 1996). For each comparison we probed each pair of 24-dimensional shape matrices with the same 10,000 unit-length vectors of random direction. Table 2 shows the median angles of response vectors between the matrices pooled over sexes and/or genotypes above the diagonal. All of the comparisons in Table 2 are significantly less than random expectation of 82° (calculated by simulation) by a one-sided test, with the exception of the nonsegregational matrices for line 33 and line 39, where

**Table 2.** Random skewers angles above diagonal, and matrix correlations below the diagonal. Bold-faced entries are independent planned comparisons. Italicized entries compare nonsegregational and homozygous matrices for the same partitions. Random skewers entries are the median angle between the 10,000 responses, and in parentheses, the 2.5% and 97.5% quantiles of the angles.

| | | | Nonsegregational | | | | | Homozygous | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | ♂ pool | ♀ pool | Both 33 | Both 39 | Both Pool | ♂ pool | ♀ pool | Both 33 | Both 39 | Both Pool |
| | ♂ | pool | | **23 (12–55)** | 40 (22–73) | 27 (13–61) | 19 (8–45) | *19 (8–46)* | 29 (13–59) | 43 (23–76) | 23 (12–56) | 23 (9–48) |
| | ♀ | pool | **0.89** | | 41 (24–77) | 23 (10–56) | 19 (9–46) | 28 (15–65) | *20 (9–47)* | 44 (26–79) | 28 (13–63) | 26 (13–58) |
| Nonsegregational | Both | 33 | 0.64 | 0.63 | | **53 (32–86)** | 36 (21–64) | 38 (21–67) | 35 (20–63) | *20 (10–45)* | 50 (28–84) | 34 (20–62) |
| | Both | 39 | 0.84 | 0.92 | **0.46** | | 28 (12–65) | 31 (16–68) | 31 (14–67) | 53 (31–85) | *20 (8–51)* | 30 (14–63) |
| | Both | pool | 0.92 | 0.93 | 0.69 | 0.86 | | 27 (12–60) | 21 (12–49) | 38 (23–71) | 30 (12–65) | *24 (10–55)* |
| | ♂ | pool | *0.92* | 0.84 | 0.67 | 0.81 | 0.85 | | **25 (15–52)** | 37 (21–64) | 23 (11–51) | 12 (6–26) |
| | ♀ | pool | 0.85 | *0.92* | 0.72 | 0.85 | 0.91 | **0.87** | | 31 (18–59) | 30 (13–60) | 18 (9–36) |
| Homozygous | Both | 33 | 0.61 | 0.58 | *0.90* | 0.46 | 0.66 | 0.69 | 0.76 | | **49 (27–82)** | 31 (18–55) |
| | Both | 39 | 0.89 | 0.86 | 0.51 | *0.92* | 0.84 | 0.89 | 0.85 | **0.52** | | 22 (9–47) |
| | Both | pool | 0.91 | 0.87 | 0.72 | 0.84 | *0.89* | 0.96 | 0.94 | 0.76 | 0.91 | |

the upper 95% quantile is 82.1°. Only four of these comparisons, shown in bold, are independent. The angles between genotype-specific comparisons are relatively large (49–53°), whereas sex-specific comparisons have smaller angles (23–25°). The structure of the nonsegregational and homozygous matrices seems to be rather similar, as the median angles between the nonsegregational and homozygous matrices, shown in italics, average just 21°, despite their large differences in $\bar{e}$. Table 2 also shows matrix correlations between the 24-dimensional shape matrices, which gives the same qualitative picture of matrix similarity as random skewers.

Mitteroecker and Bookstein (2009) showed that an appropriate measure of distances among full-rank covariance matrices is

$$\|\mathbf{M}_A, \mathbf{M}_B\| = \sqrt{\sum_{i=1}^{p} \log(\lambda_{C\cdot i})^2},$$

where $\lambda_{C\cdot i}$ are the eigenvalues of $\mathbf{C} = \mathbf{M}_A^{-1}\mathbf{M}_B$. We calculated these distances in the space of the first five eigenvectors of the covariance matrix of the pooled genotype, combined sex **M** matrix. Multidimensional scaling (MDS) on the distance matrix (SAS Institute 2004) in two dimensions fit very well. MDS on the matrices uncorrected for size separated homozygous and nonsegregational matrices in coordinate space (Fig. S1). MDS on matrices standardized by their traces is shown in Figure 2. The corresponding homozygous and nonsegregational estimates tend to be near neighbors. Matrices estimated from larger subsets of the overall

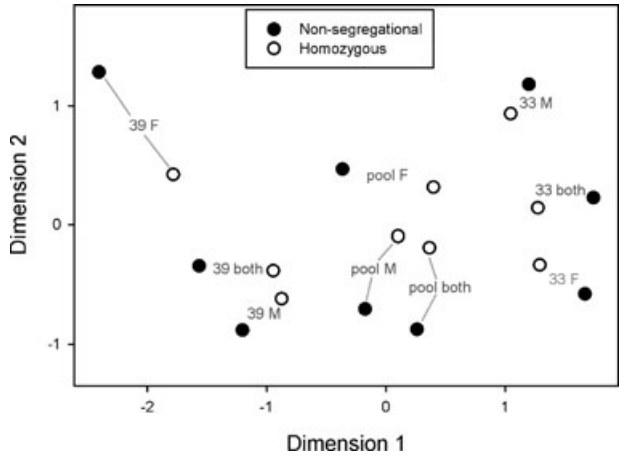


**Figure 2.** Multidimensional scaling of the distance matrix among the matrix size standardized M matrices. Distances calculated in the subspace defined by the first 5 eigenvectors of the pooled genotype and sex M matrix. Labels and lines connect the corresponding nonsegregational and homozygous estimates.

data are more central than those estimated with subsets of the data. Genotype-specific differences define dimension 1, whereas sex-by-genotype interactions largely define dimension 2.

To perform pairwise comparisons among **M** matrices, we partitioned phenotype space into the common, non-null region in which both matrices possessed significant genetic variance, versus those in which one or both matrices did not have significant

**Table 3.** Scalar measures of matrix similarity calculated in the common space of the shape matrices. Conditional evolvability, $\bar{c}$, autonomy, $\bar{a}$, respondability, $\bar{r}$, and response difference, $\bar{d}$ defined in Hansen and Houle (2008).

| | Matrices compared | | | |
|---|---|---|---|---|
| | Nonsegregational | | Homozygous | |
| Statistic | $\mathbf{M}_A$: pool M<br>$\mathbf{M}_B$: pool F | $\mathbf{M}_A$: 33 both<br>$\mathbf{M}_B$: 39 both | $\mathbf{M}_A$: pool M<br>$\mathbf{M}_B$: pool F | $\mathbf{M}_A$: 33 both<br>$\mathbf{M}_B$: 39 both |
| Non-null dimensions | 5 | 5 | 12 | 10 |
| Singly nearly-null dimensions | 2 | 2 | 0 | 2 |
| Mean $\bar{e}$ in non-null subspace | 4.37 | 5.44 | 6.16 | 6.78 |
| Mean $e$ in singly null subspace[1] | 0.68 | 0.31 | — | 0.11 |
| $e_B/e_A$ on common $\boldsymbol{e}_{max}$ | 0.78 | 5.44 | 1.08 | 2.90 |
| $\bar{e}_B/\bar{e}_A$ | 0.75 | 2.59 | 1.01 | 1.59 |
| $\bar{c}_B/\bar{c}_A$ | 0.70 | 2.32 | 0.82 | 1.35 |
| $\bar{a}_B/\bar{a}_A$ | 0.98 | 0.83 | 0.84 | 1.04 |
| $\bar{r}_B/\bar{r}_A$ | 0.73 | 2.83 | 1.04 | 1.66 |
| $\bar{d}$ | 2.05 | 6.85 | 3.47 | 7.16 |
| Angle between $\boldsymbol{e}_{maxA}$ and $\boldsymbol{e}_{maxB}$ | 19.1° | 51.8° | 22.7° | 57.8° |
| Median (95%) skewer angle | 17.7° (34.1) | 45.6° (72.6) | 24.0° (41.6) | 48.4° (73.3) |
| Median random angle | 70.0° | 70.0° | 78.3° | 77.2° |
| Matrix correlation | 0.12 | −0.19 | −0.38 | −0.56 |

[1]Mean of the maximum $e$ in the comparison M matrices along singly nearly-null vectors of $C_r = M_{Ar}^{-1}M_{Br}$.

genetic variance (the singly- or doubly-null spaces) using the algorithm described in the Methods section. Table 3 gives measures of matrix similarity based on this common subspace analysis. Matrices from the nonsegregational analysis share a five-dimensional non-null subspace, and each comparison reveals two singly null dimensions. The maximum possible dimensionality of these shape data is 20, as four of 24 degrees of freedom are used to register the forms before analysis. The nonsegregational matrices are thus doubly nearly-null for 13 of the 20 dimensions. In the homozygous model, sex-specific matrices share a 12 dimensional non-null subspace. Genotype-specific matrices share a 10 dimensional non-null subspace, and two singly null dimensions. There are eight doubly nearly-null dimensions, although the analysis of the complete pooled dataset showed 19 significant non-null dimensions (Table 1). One interesting question is whether the singly nearly-null dimensions ever have a large amount of variance in one of the two comparison matrices. We addressed this by comparing mean $\bar{e}$ in the common subspace, shown in the third line of the Table, with the mean of the maximum of the two evolvabilities along eigenvectors of **C** judged to lie in the singly nearly-null subspace, shown in the fourth line of the Table. In all cases, the singly nearly-null evolvabilities are much smaller than the average of evolvability of the non-null subspace, and within a factor of 2 of the minimum eigenvalue statistically supported for each **M** matrix. The singly-nearly null dimensions are those where the amount of variance in both **M** matrices is near the threshold of statistical significance, with one happening to fall above that threshold, and one below. The low proportion of singly null dimensions suggests that despite the differences among M matrices, they have substantial similarities in which directions in phenotype space have the lowest genetic variances.

The remaining entries in Table 3 give a variety of scalar measures of matrix similarity calculated in the common non-null subspace. The vector $\mathbf{e}_{max}$ is the direction with the greatest variance, and we show the ratio of evolvabilities on this vector in the common subspace, and the angle between $\mathbf{e}_{max}$ vectors in the comparison matrices. Evolvabilities, respondabilities and conditional evolvabilities (Hansen and Houle 2008) of sex-specific matrices are within a factor of 1.4, whereas genotype-specific matrices generally differ by larger factors, up to the 5.4× difference in evolvability along $\mathbf{e}_{max}$ for the nonsegregational model. Mean autonomies are relatively similar. Response differences are in shape units that are difficult to interpret, but the ratio of response differences is more than twice as large for genotype comparisons as for sex comparisons.

Table 3 also shows median skewer angle (Cheverud 1996) calculated from 10,000 random selection gradients, and the 95% quantile of angle as a one-sided confidence interval. The expected median angle between random vectors is shown for comparison. Sex-specific matrices have small response angles that are highly

**Table 4.** Descriptions of directions with different variances in the non-null subspace defined by comparison matrices. See text for explanation. Bold-faced vectors are represented in Fig. 3.

| Comparison matrices: **C** vector | Nonsegregational | | | | Homozygous | | | |
|---|---|---|---|---|---|---|---|---|
| | $\mathbf{M}_A$: pool M<br>$\mathbf{M}_B$: pool F | | $\mathbf{M}_A$: 33 both<br>$\mathbf{M}_B$: 39 both | | $\mathbf{M}_A$: pool F<br>$\mathbf{M}_B$: pool M | | $\mathbf{M}_A$: 33 both<br>$\mathbf{M}_B$: 39 both | |
| | $e_B/e_A$ | $\frac{e_A+e_B}{2}$ | $e_B/e_A$ | $\frac{e_A+e_B}{2}$ | $e_B/e_A$ | $\frac{e_A+e_B}{2}$ | $e_B/e_A$ | $\frac{e_A+e_B}{2}$ |
| 1 | 1.44 | 2.61 | **16.32** | **3.72** | 5.98 | 1.16 | **38.98** | **2.61** |
| 2 | 1.17 | 2.25 | **5.93** | **10.02** | 3.80 | 1.50 | 12.16 | 2.25 |
| 3 | 0.73 | 2.63 | 2.05 | 3.84 | **2.55** | **2.82** | 9.62 | 2.30 |
| 4 | 0.50 | 1.59 | 0.63 | 1.73 | 1.95 | 2.43 | 5.01 | 1.42 |
| 5 | **0.39** | **2.20** | 0.40 | 2.96 | 1.26 | 1.62 | 3.10 | 1.66 |
| 6 | | | | | 1.22 | 2.05 | 1.38 | 2.38 |
| 7 | | | | | 0.77 | 1.89 | 1.07 | 1.95 |
| 8 | | | | | 0.51 | 1.62 | 0.56 | 3.49 |
| 9 | | | | | 0.47 | 1.68 | 0.26 | 2.54 |
| 10 | | | | | 0.37 | 1.43 | **0.01** | **5.45** |
| 11 | | | | | 0.27 | 1.78 | | |
| 12 | | | | | 0.13 | 0.95 | | |
| mean | 0.85 | 2.25 | 5.07 | 4.45 | 1.61 | 1.74 | 7.22 | 2.61 |
| std. dev. | 0.45 | 0.42 | 6.67 | 3.22 | 1.76 | 0.52 | 11.91 | 1.15 |
| Median ratio[1] | 1.44 | | 2.52 | | 2.33 | | 4.46 | |

[1]Median of the greater of $e_B/e_A$ and $e_A/e_B$ in each column.

significantly different from the angle between random vectors. Genotype-specific response angles are much larger, but nonetheless generally less than random angles. The genotype-specific responses predicted from the nonsegregational matrices are not significantly less than the random vectors, although the other three comparisons differ significantly from random. We also show the matrix correlations between comparison matrices. Matrix correlations are mostly negative, despite the similarities of matrix structure revealed by other measures of similarity.

To further investigate the nature of the differences between **M** matrices, we characterized those directions with the greatest ratios of variances and evolvabilities, as described in the Methods section. Table 4 shows the results of these evolvability ratio analyses. The first- and last-ranked vectors are those for which the ratios of evolvabilities are most extreme, as shown in the first column within each comparison. Note that disparity of evolvabilities is greatest for both the largest and smallest values. A more interpretable measure of overall difference is the larger of the two ratios $e_B/e_A$ or $e_A/e_B$, and the median value of these is shown in the last row of the Table.

The interest of each of these vectors is also dependent on the amount of variance along those vectors, measured in the second column of each comparison by the average evolvability in the two matrices along that vector. The distinction is striking in the comparison between genotype-specific matrices in the nonsegregational analysis: the first vector is notable because line 39 has 16 times the variance found in line 33; the second vector is notable both because line 39 has six times the variance in line 33, and an average evolvability of 10.02, far above the mean $\bar{e}$ of 5.44 in the common non-null subspace (Table 3), and close to $e_{max}$ of 11.5 for the pooled matrix (Table S8). For the sex-specific comparisons, all vectors have evolvabilities less than the mean evolvability in the non-null subspace. The same is true of the genotype-specific comparison in the homozygous analysis, although vector 10, which has the greatest difference (108 times more variance in line 33 than line 39) also has an average evolvability of 5.45, only slightly less than the mean $\bar{e}$ of 6.78 in the common non-null subspace.

The directions and magnitudes of some of the more notable vectors from Table 4 are shown as deviations from a reference wing in Figure 3. None of these vectors show similar patterns of landmark differences, an observation confirmed by the high angles between these vectors (results not shown).

## Discussion

We emphasize three findings. First we have detected mutational genetic variation for almost every aspect of wing shape and size that we studied. Second, sex-specific **M**s are more similar than the genotype-specific **M**s. Third, mutational effects are partially recessive. Before discussing these conclusions, we justify a few other aspects of our analysis.

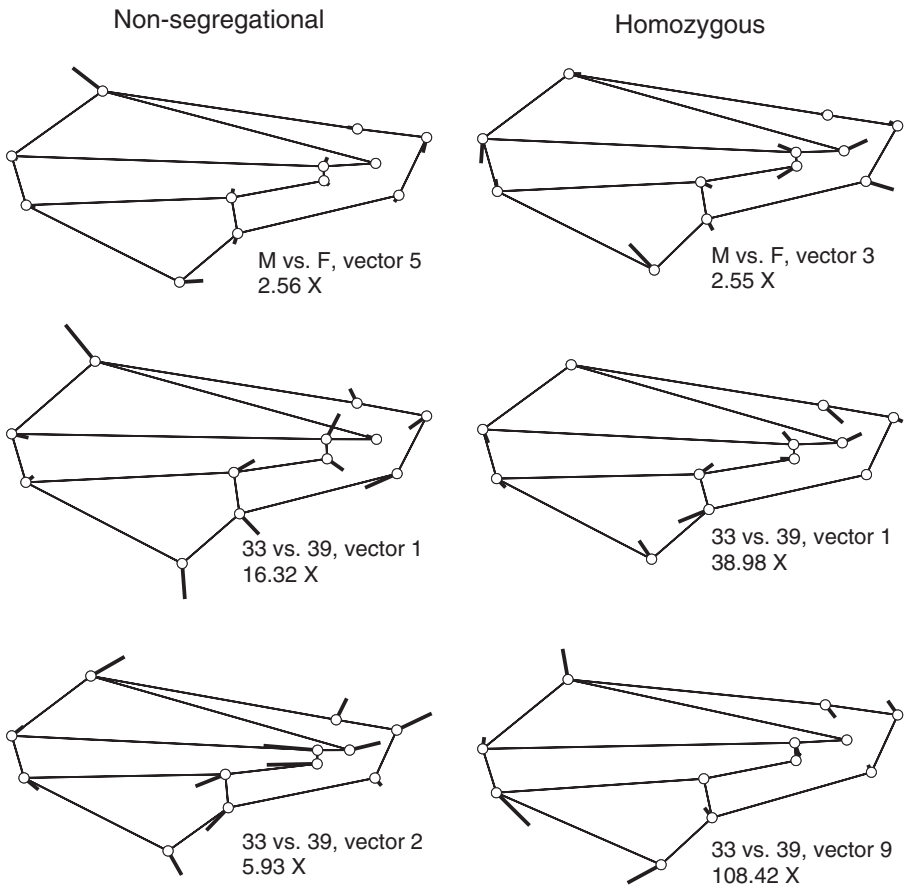


**Figure 3. Vectors identified as notably different in the comparisons in the common space of sex-specific and genotype-specific matrices in Table 4. Circles denote the position of the 12 landmarks, and thick lines the magnitude and direction of the mutational variance along each vector. The units are 1000 × the average mutational evolvability along that vector in the matrices compared, except for the lower left vector, which has units of 500 × evolvability. Thin lines connect landmarks that are connected by a vein. The direction of vectors has been reversed if necessary to minimize overlap in the deviations from the reference. The reference phenotype is for males. Numbering of vectors follows that in Table 4.**

## ANALYSIS PHILOSOPHY

Previous analyses of covariance matrices have emphasized hypothesis testing, even when the hypotheses are not biologically motivated. We emphasize instead the evolutionary properties of **M** matrices, even if those measures have no available statistical test, because evolutionary prediction is the motivation for studying covariance matrices (Houle et al. 2011). Hansen and Houle (2008) suggested a family of measures of genetic covariance matrix properties related to the amount of response to directional selection they predict. Evolvability, $e$, is the predicted response to unit strength selection in the direction of the selection gradient in the absence of stabilizing selection. Conditional evolvability, $c$, is the response to unit strength selection when stabilizing selection around the selected direction is infinitely strong. Evolvability predicts the ability to evolve in a given direction in the ideal case when there is no selection on other directions in phenotype space, whereas conditional evolvability predicts evolvability in the case of maximal conflict between directional and nonlinear selection. The random skewers statistic (Cheverud 1996) measures the mean angle between response vectors predicted by two different covariance matrices, and is thus is a complementary measure of matrix similarity that depends only on the structure of the covariance matrices, and not their sizes. It is critical for the interpretability of all these measures that traits are either measured in the same units, or are standardized to make them comparable, for example by mean-standardization (Hansen and Houle 2008; Huttegger and Mitteroecker 2011). Our shape variables are in the same units, which are directly comparable to mean-standardized centroid size (Mitteroecker et al. 2004).

We are not opposed to statistical testing, and have employed it whenever we have a biologically relevant hypothesis for which we know how to construct a test. For some parameters, such as matrix rank, statistical testing is an integral part of the process of arriving at a well-supported, conservative conclusion.

We have, however, eschewed several commonly used analyses that offer tests of hypotheses that are not biologically

interpretable; these analyses include matrix correlations and common principal components analysis (Houle et al. 2002). Matrix correlation is a commonly used scalar measure of matrix similarity but it has no necessary relationship to predictions of evolution. Correlations are standardized and centered measures of relationship and consequently discard information about the mean values of covariances, and disparities of the covariances within matrices. Consequently, it is easy to construct pairs of matrices that either predict extremely different patterns of evolution, yet have perfect matrix correlations, or that predict virtually the same pattern of evolution yet are uncorrelated, or even negatively correlated. Our results provide an example. Although the matrix correlations in the 20 dimensional space of wing shapes are positive (Table 3), consistent with matrix similarities, in the common five-dimensional spaces of sex- and genotype-specific comparisons, matrix correlations are near 0 or negative (Table 4). This contradicts the evolutionarily motivated analyses that show great similarities of matrix structure.

## MUTATIONAL BIAS

An important question about the effects of new mutation is whether they are biased in direction. Bias would have both biological (Keightley and Lynch 2003; Braendle et al. 2010) and statistical (Keightley et al. 2000) consequences. Bias is expected for traits under persistent directional selection, such as fitness or its components. The majority of previous analyses of mutational effects on fitness components and life history traits have found the expected mutational bias (Keightley and Lynch 2003; Shaw et al. 2003), although there are exceptions (Shaw et al. 2000; 2002; Rutter et al. 2010). For Drosophila wings, a previous mutation-accumulation experiment found no significant bias in wing length or width (Santiago et al. 1992). In addition, wing form is remarkably conservative in the genus Drosophila (Houle et al. 2003), even though the evolvability of many aspects of wing form is quite high (Weber 1990, 1992; Houle et al. 2003; Le Rouzic et al. 2011), suggesting that wing trait means are at naturally-selected optima. For traits under stabilizing selection there is no a priori expectation of mutational bias. Our data shows no evidence for a mutational bias in wing form, although these tests were fairly weak.

## GENETIC ARCHITECTURE OF MUTATIONAL VARIANCE

Mutational variance, $V_M$, is defined as $V_M = UE[g^2]/2$ (Lynch and Hill 1986), where $U$ is the haploid genomic mutation rate and $E[g^2]$ is the expected squared homozygous effect of a mutation. We applied two models that estimate $V_M$. The nonsegregational model yields estimates of the rate of increase in additive genetic variance in an initially homozygous population when mutant alleles are introduced in heterozygous condition. Twice this quantity estimates $V_M$ in the special case where genotype–phenotype map is linear or additive. The second model, which we call the homozygous model, directly estimates $UE[g^2]$ and thus has a more direct relationship to $V_M$ as defined by Lynch and Hill (1986).

When the genotype–phenotype map is not linear, neither the nonsegregational nor the homozygous model directly determine the additive genetic variance maintained by mutation-selection balance. Additive genetic variance is a population parameter that is determined both by the genotype–phenotype map and by the frequency that alleles have in the population (Hansen 2006). If mutant alleles have deleterious fitness effects that ensure that they remain rare, the nonsegregational $V_M$ will better predict the amount of additive genetic variance in the population. If mutant alleles are neutral, $V_M$ estimated by the homozygous model will be the better predictor, as alleles will ultimately come to be both rare and common. The evidence that wings are under stabilizing selection reviewed above makes it unlikely that mutations affecting wing shape are all neutral. The relevant $V_M$ for wing traits probably lies somewhere between our homozygous and nonsegregational estimates.

There is a substantial difference between the nonsegregational and the homozygous estimates of $V_M$, which is most readily explained if mutant alleles have partially recessive effects. More complex alternative explanations could also involve epistatic effects. Assuming that dominance is the cause of the difference, mutations in heterozygous condition have about 25% of the effect that they do in homozygotes. Partial recessivity of spontaneous mutations is a common finding, although most estimates are for life history traits (Halligan and Keightley 2009). The only previous estimates for wing traits in Drosophila show additivity of mutational effects (Santiago et al. 1992). Our finding of partially recessive mutational effects also contrasts with the additivity of effects estimated in a natural population of *D. melanogaster* (Mezey and Houle 2005) for the same geometric morphometric traits we used here. This can be explained if spontaneous mutations with recessive effects have more deleterious fitness effects than those with more additive functional interactions. Previous studies of dominance and fitness of spontaneous mutations for non-wing traits support this hypothesis (Simmons and Crow 1977; López and López-Fanjul 1993).

## DIMENSIONALITY

Our previous study of standing genetic variation found evidence for additive genetic variation in nearly all of the 20 aspects of wing shape studied (Mezey and Houle 2005). This result has been criticized based on the hypothesis that one of our methods for determining dimensionality was too liberal (Hine and Blows 2006). We have reanalyzed our natural population data using the factor-analytical approach in Wombat, and found that the reanalysis supports a 20-dimensional model (not shown), supporting our original contention that our estimates of dimension were conservative.

Our analysis of the full mutation-accumulation experiment using the homozygous model shows mutational variance in at least 19 of 21 possible dimensions in **M**. Homozygous analyses of other partitions of the data lead to lower estimates of dimensionality, although this is likely to be due to the smaller sample of mutations within genotypes and lower statistical power. The sublines descended from the two inbred genotypes in this experiment have very different patterns of mutational effects, suggesting the possibility that each genotype engenders a different mutational subspace. The nonsegregational model estimated dimensionality at either 7 or 8 dimensions, depending on the data partition. We suspect that this is artificially low, as models with higher dimensionalities did not properly converge, so that the likelihood of more complex models could not be assessed.

One possible explanation for the high dimensionality of genetic effects is that they lie on a curved surface or manifold of lower dimensionality embedded in the space of all possible wing shapes. If this is so, then describing that manifold would provide a simpler picture of genetic variation in wing shape. We have applied variants of the Isomap technique for discovering manifolds in high-dimensional data (Tenenbaum et al. 2000), but have so far found no evidence for low-dimensional curved surfaces (Bendich and Houle, unpubl. ms.). Alternatively, dimensionality may truly be high. Many genes affect wing shape (Weber et al. 1999; 2001, 2005; Mezey et al. 2005) and each may explore somewhat different aspects of the phenotypic space. Some studies of the dimensionality of genetic variance in other species support rather low dimensional models (Kirkpatrick and Lofsvold 1992; Hine and Blows 2006; McGuigan and Blows 2007). One possible explanation for this discrepancy is that geometric morphometric data are different because they intensively characterize variation within one part of an organism such as a wing, skull or jaw. Clearly studies of the dimensionality of variation in a wide variety of phenotypes are needed.

### M MATRIX COMPARISONS

We have very strong evidence that the **M** matrices are significantly different for male and female flies (sex-specific matrices) and the two base inbred genotypes (genotype-specific matrices). The sex-specific differences are due to differences in expression of the same set of mutations, whereas the two genotypes independently accumulated mutations that are almost certainly different. On the other hand, there are also substantial similarities among matrices, as indicated, for example, by random skewers results that show intermediate angles between predicted responses, and the fact that a minority of dimensions are singly-null, having significant genetic variation in one partition of the data, but not in its partner partition.

The sex-specific matrices are different in size, with males having more total variance than females in the nonsegregational model and for centroid size in both analyses. The modest size of these differences (17% more variation in males in the nonsegregational model and 24% in the homozygous model) is consistent with dosage compensation on the X chromosome in males. Dosage compensation is expected to magnify the X-specific variance by a factor of 2, and the X chromosome constitutes about 20% of the genome (Mezey et al. 2005). Unfortunately, models that specifically estimated the X-chromosome variance as well as the autosomal variances did not converge. Differences in matrix size only accounted for a small proportion of the differences in the sex-specific matrices implicating other differences in the sex-specific genotype–phenotype map.

The average angles between predicted responses to selection (random skewers) in sex-specific matrices only differed by about 20° over the whole space (Tables 2 and 3). Those directions with the most different evolvabilities in the common subspace differed by less than 2.6× in the nonsegregational analysis, and up 6× in the homozygous analysis (Table 4). The directions with the biggest disparities in evolvablity had rather small average evolvabilities, suggesting that some of these differences are due to error in estimating the differences between matrices.

The genotype-specific matrices of the two base inbred lines differ from each other considerably, but the interpretation of these differences is less clear. It may indicate genetic differences in the number or nature of mutations between genotypes. However, if the number of mutations with large effects on wings is modest, sampling would cause replicate experiments to produce rather different realized **M** matrices, even if they are drawn from the same population of effects, especially given the high number of dimensions in which those mutation vectors can fall.

One indication that some of the genotype-specific differences may be due to genetic differences in mutation is that the line 39 **M** matrix is substantially larger than the line 33 matrix, by a factor of greater than 2 over the whole phenotype space. The random skewers angles are large, around 50°. The leading eigenvectors are at an angle of more than 45°, and there are substantial evolvability differences on the leading eigenvectors of the common matrices. The evolvability ratio analysis (Table 4 and Fig. 3) shows some spectacularly large differences in evolvabilities in particular directions.

One possible explanation for these differences that we can rule out is the presence of a small number of sublines with very atypical phenotypes. Inspection of plots of genotype means in the first few dimensions of discriminant space, such as those in Figure 1, and more generally the distribution of Mahalanobis distances between genotype means, does not reveal any sublines in either genotype with extreme values. The differences seem to be distributed across many sublines, rather than due to just a few atypical values.

The differences in matrix size could be explained by an overall difference in mutation rate. Our estimates of the base-pair

nucleotide mutation rate, however, show that line 33 has a mutation rate that is at least twice that in line 39 (Haag-Liautard et al. 2007; Schrider, Houle, Lynch and Hahn unpubl. ms.). The mutation rate differences between genotypes are principally due to a fivefold higher rate of G:C→A:T transitions in line 33. The fact that the phenotypic variance of mutation on wings in line 39 is about twice as large despite this lower mutation rate gives increased confidence that the nature of phenotypic effects is truly different between the two genotypes. There are many possible explanations, including differences in mutational hotspots between genotypes, differences in GC content of genes that influence wing shape, and differences in phenotypic robustness.

## PROSPECTS

The lack of detailed data on the **M** matrix has been a key limitation of empirical analyses of the role of quantitative genetic variation in evolution (Arnold et al. 2008). With these data in hand, we can go on to compare mutational variance to standing variation (Mezey and Houle 2005) and to the pattern of diversification among species (Houle et al. 2003). Such analyses can help to answer a variety of unsolved questions in evolutionary genetics, such as: Is mutation limiting the rate of evolution? Are the phenotypes that do not evolve forbidden by an absence of variation, or by low fitness? We encourage the gathering of data on mutational variation in other high-dimensional suites of traits, so that we can begin to build solid conclusions about these fundamental questions.

## ACKNOWLEDGMENTS

The authors thank Eladio Marquez and Thomas Hansen for discussions, Karin Meyer for patient assistance with Wombat, two anonymous reviewers for their comments, Jeff Birdsley for managing the data collection, and Paul Galpern, Fiona Smyth, Yunily Ng, and Liv Carpenter for imaging wings. This research was supported by National Science Foundation grants DEB-0129219 and DEB-00950002.

## LITERATURE CITED

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the **G** matrix. Evolution 62:2451–2461.

Braendle, C., C. F. Baer, and M. A. Felix. 2010. Bias and evolution of the mutationally accessible phenotypic space in a developmental system. PLoS Genet. 6:e1000877.

Cheverud, J. M. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. J. Evol. Biol. 9:5–42.

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

Gomulkiewicz, R., and D. Houle. 2009. Joint demographic and genetic constraints on evolution. Am. Nat. 174:E218–E2229.

Haag-Liautard, C., N. Coffey, D. Houle, M. Lynch, B. Charlesworth, and P. D. Keightley. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol. 6:1706–1714.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, D. Houle, B. Charlesworth, and P. D. Keightley. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature 445:82–85.

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

Hansen, T. F. 2006. The evolution of genetic architecture. Ann. Rev. Ecol. Evol. Syst. 37:123–157.

Hansen, T. F., and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pp. 130–150 *in* M. Pigliucci, and K. Preston, eds. The evolutionary biology of complex phenotypes. Oxford University Press, Oxford, U. K.

———. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21:1201–1219.

Hansen, T. F., C. Pelabon, and D. Houle. 2011. Heritability is not evolvability. Evol. Biol. 38:258–277.

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

Houle, D. 2010. Numbering the hairs on our heads: the shared challenge and promise of phenomics Proc. Natl. Acad. Sci. USA 107:1793–1799.

Houle, D., and A. Kondrashov. 2006. Mutation. Pp. 32–48 *in* C. W. Fox, and J. B. Wolf, eds. Evolutionary genetics: concepts and case studies. Oxford University Press, Oxford, U. K.

Houle, D., and S. V. Nuzhdin. 2004. Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. Genet. Res. 83:7–18.

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

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

Houle, D., J. Mezey, and P. Galpern. 2002. Interpreting the results of common principal component analyses. Evolution 56:433–440.

Houle, D., J. Mezey, P. Galpern, and A. Carter. 2003. Automated measurement of Drosophila wings. BMC Evol. Biol. 3:25.

Houle, D., D. R. Govindaraju, and S. W. Omholt. 2010. Phenomics: the next challenge. Nat. Rev. Genet. 11:855–866.

Houle, D., C. Pelabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. Q. Rev. Biol. 86:3–34.

Huttegger, S. M., and P. Mitteroecker. 2011. Invariance and meaningfulness in phenotype spaces. Evol. Biol. 38:335–351.

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

———. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution 58:1639–1654.

———. 2007. The mutation matrix and the evolution of evolvability. Evolution 61:727–745.

Joyner-Matos, J., A. Upadhyay, M. P. Salomon, V. Grigaltchik, and C. F. Baer. 2009. Genetic (Co)variation for life span in Rhabditid Nematodes: Role of mutation, selection, and history. J. Gerontol. Ser. A-Biol. Sci. Med. Sci. 64:1134–1145.

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., and D. L. Halligan. 2009. Analysis and implications of mutational variation. Genetica 136:359–369.

Keightley, P. D., and M. Lynch. 2003. Toward a realistic model of mutations affecting fitness. Evolution 57:683–685.

Keightley, P. D., E. K. Davies, A. D. Peters, and R. G. Shaw. 2000. Properties of ethylmethane sulfonate-induced mutation affecting life-history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. Genetics 156:143–154.

Kirkpatrick, M., and D. Lofsvold. 1992. Measuring selection and constraint in the evolution of growth. Evolution 46:954–971.

Kirkpatrick, M., and K. Meyer. 2004. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. Genetics 168:2295–2306.

Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94:203–215.

Le Rouzic, A., D. Houle, and T. F. Hansen. 2011. A modelling framework for the analysis of artificial-selection time series. Genet. Res. 93: 155–173.

López, M. A., and C. López-Fanjul. 1993. Spontaneous mutation for a quantitative trait in *Drosophila melanogaster*. II. Distribution of mutant effects on the trait and fitness. Genet. Res. 61:117–126.

Lynch, M. 1988. The rate of polygenic mutation. Genet. Res. 51:137–148.

Lynch, M. 1985. Spontaneous mutations for life-history characters in an obligate parthenogen. Evolution 39:804–818.

Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. Evolution 40:915–935.

Mackay, T. F. C. 2010. Mutations and quantitative genetic variation: lessons from Drosophila. Phil. Trans. R. Soc. Lond., Ser. B: Biol. Sci. 365:1229–1239.

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

Meyer, K. 2007. WOMBAT—a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. (Sci. B) 8:815–821.

Meyer, K. 2010. Wombat: A program for mixed model analyses by restricted maximum likelihood 1.0. Animal Genetics and Breeding Unit, University of New England, Armidale, NSW. Available via http://agbu.une.edu.au/∼kmeyer/wombat.html

Meyer, K., and M. Kirkpatrick. 2005. Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. Genet. Sel. Evol. 37:1–30.

———. 2008. Perils of parsimony: Properties of reduced-rank estimates of genetic covariance matrices. Genetics 180:1153–1166.

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

Mezey, J. G., D. Houle, and S. V. Nuzhdin. 2005. Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics 169:2101–2113.

Mitteroecker, P., and F. Bookstein. 2009. The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. Evolution 63:727–737.

Mitteroecker, P., P. Gunz, M. Bernhard, K. Schaefer, and F. L. Bookstein. 2004. Comparison of cranial ontogenetic trajectories among great apes and humans. J. Hum. Evol. 46:679–698.

Muller, H. J. 1973. The modern concept of nature. SUNY Press, Albany, NY.

Rifkin, S. A., D. Houle, J. Kim, and K. P. White. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. Nature 438:220–223.

Rohlf, F. J., and D. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39:40–59.

Rutter, M. T., F. H. Shaw, and C. B. Fenster. 2010. Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. Evolution 64:1825–1835.

Rutter, M. T., A. Roles, J. K. Conner, R. G. Shaw, F. H. Shaw, K. Schneeberger, S. Ossowski, D. Weigel, and C. B. Fenster. 2012. Fitness of *Arabidopsis thaliana* mutation accumulation lines whose spontaneous mutations are known. Evolution 66:2335–2339.

Sanjuán, R., A. Moya, and S. F. Elena. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA 101:8396–8401.

Santiago, E., J. Albornoz, A. Dominguez, 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.

SAS Institute, Inc. 2004. The SAS System for Windows Release 9.2. SAS Institute, Inc., Cary, NC.

Schrider, D. R., Jonathan N. Hourmozdi, and Matthew W. Hahn. 2011. Pervasive multinucleotide mutational events in eukaryotes. Curr. Biol. 21:1051–1054.

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

Shaw, F. H., C. J. Geyer, and R. G. Shaw. 2002. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. Evolution 56:453–463.

Shaw, R. G., F. H. Shaw, and C. Geyer. 2003. What fraction of mutations reduces fitness? A reply to Keightley and Lynch. Evolution 57:686–689.

Simmons, M. J., and J. F. Crow. 1977. Mutations affecting fitness in Drosophila populations. Ann. Rev. Genet. 11:49–78.

Tenenbaum, J. B., V. de Silva, and J. C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323.

Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics 111:165–195.

Weber, K., R. Eisman, L. Morey, A. Patty, J. Sparks, M. Tausek, and Z.-B. Zeng. 1999. An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. Genetics 153:773–786.

Weber, K., R. Eisman, S. Higgins, L. Morey, A. Patty, M. Tausek, and Z.-B. Zeng. 2001. An analysis of polygenes affecting wing shape on chromosome 2 in *Drosophila melanogaster*. Genetics 159:1045–1057.

Weber, K., N. Johnson, D. Champlin, and A. Patty. 2005. Many P-element insertions affect wing shape in *Drosophila melanogaster*. Genetics 169:1461–1475.

Weber, K. E. 1990. Selection on wing allometry in *Drosophila melanogaster*. Genetics 126:975–989.

Weber, K. E. 1992. How small are the smallest selectable domains of form? Genetics 130:345–353.

Zhang, X.-S. 2012. Fisher's geometrical model of fitness landscape and variance in fitness within a changing environment. Evolution 66:2350–2368.

Associate Editor: R. Bürger

## *Supporting Information*

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Table S1.** Wing data obtained.

**Table S2.** Means for all variables in each data partition. See attached file "Table S2 MAmeans.csv."

**Table S3.** Coefficients of covariance relationships between genotypes.

**Table S4.** MANOVA testing for control and MA differences within the 1998 experiment.

**Table S5.** Likelihood ratio tests of the best-fitting nonsegregational model by the AICC criterion, relative to the next simplest model.

**Table S6.** Likelihood ratio tests of the best-fitting homozygous model by the AICC criterion, relative to similar models. Likelihoods should increase with number of parameters, but more complex models often returned lower likelihoods. The *P* values for these cases are denoted "NA" as no likelihood ratio test could be performed in those cases.

**Table S7.** Estimated mutational and residual variance-covariance matrices. Table is in separate file 'Table S7 All M R matrices.csv'

**Table S8.** Scalar evolvability measures calculated on the first 7 dimensions of each **M** matrix. Measures are evolvability, *e*, conditional evolvability, *c*, autonomy, *a*, and respondability, *r*, defined in Hansen and Houle (2008).

**Table S9.** Average dominance of mutations.

**Figure S1.** Multidimensional scaling of the distance matrix among the unstandardized **M** matrices. Distances calculated in the subspace defined by the first 5 eigenvectors of the pooled genotype and sex **M** matrix. The six points to the upper left are from genotype IV*e*-39, those to the lower right are from IV*e*-33, whereas the six intermediate points are for the pooled genotype estimates.

**Supplementary References**