Supporting Information for

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

David Houle and Janna Fierst

Supplementary Materials and Methods

1.1 MUTATION ACCUMULATION

The IV*e* population (Houle and Rowe 2003) was the base for this experiment. IV*e* was derived from the outbred IV population that has been maintained under similar conditions in the laboratory since 1975 (Charlesworth and Charlesworth 1985). It is maintained under crowded conditions in ten bottles at 25ºC, and probably has an effective size of at least a few hundred individuals. IV*e* is marked with an *ebony* allele that spontaneously occurred in the IV population, and was subsequently backcrossed into a sample of the IV population to yield a population homozygous for *ebony*. Inbred lines were derived by full-sib mating of single pairs of flies for 40 generations, and checked for residual heterozygosity at generation 35 using transposable element insertion sites (Houle and Nuzhdin 2004). Two of the surviving inbred lines showing low TE heterozgosity and high fitness, IV*e*-33 and IV*e*-39, were chosen as base genotypes for mutation accumulation.

IV*e*-33 and IV*e*-39 were each used to found 75 MA sublines, and thereafter these were maintained by full-sib mating of virgin flies, except in rare instances. We minimized selection within sublines by allowing single pairs to lay eggs for a short time, and collecting all offspring that eclosed within 13 days. Parents to initiate the next generation were chosen at random with respect to development time. Individual pairs often failed to produce offspring, so we set up four vials for each subline. Two vials (A and B) had a single pair of parents, and two (C and D) had four pairs. These were ordered a priori, so that if several males and females eclosed from the A vial, we set up the next generation from the A vial. If A did not produce the 12 flies necessary to set up A-D vials, we would use B vial offspring instead. Finally, if both A and B did not produce sufficient offspring, we pooled flies from vials A, B and C, preferentially choosing flies from A, then B, then C. If these three vials still did not produce sufficient flies, we would resort to flies from the D vials as well. Every other generation, we set up an E vial with four pairs of parents, which was reared at 17ºC to provide an additional backup to replace the subline if all the vials A-D failed. The two lines differed in their failure rates. Individual IV*e*-33 pairs (A and B vials) failed to produce sufficient offspring 14.1% of the time, while IV*e*-39 pairs failed 28.5%

of the time. In IV*e*-33 we founded a generation from one of the four-pair vials 6% of the time, and for IV*e*-39 14% of the time (Houle and Nuzhdin 2004).

MA sublines were gradually lost over generations. The typical course for subline loss would be that all available vials did not produce sufficient offspring at some generation, so that the full complement of vials could not be set up. Those sublines that were later lost would never subsequently produce enough flies to increase the number of vials, and eventually the number of potential parents would dwindle until flies of only one sex were obtained.

1.2 WING MEASUREMENT

We measured wings of live flies using an automated image-analysis system (Wingmachine, Houle et al. 2003). We imaged the upper surface of the left wing, except in rare instances when that wing was damaged, when we substituted the right wing. Prior to wing measurement, subline populations were expanded to a larger size in two or more generations. We extracted the locations of 12 vein intersections from cubic B-splines (Lu and Milios 1994) fit to the vein structure of each wing (see Fig. 1 in Mezey and Houle 2005). Outliers were manually checked against the images, and spline position corrected when necessary. Splining and error correction was accomplished in the Java program Wings (Van der Linde 2004-2008). Throughout the paper, coordinate data has units of centroid size in μm × 10, and centroid size has units of 1 μm.

One grain of live baker's yeast was added to these vials to improve fecundity and production. We assumed that no fixation of mutations took place during the subline populations during this expansion phase, so the number of generations of mutation accumulation was the number of generations of inbred propagation. Flies whose wings were measured were reared in vials with six female parents and a similar number of male parents. Parental flies were discarded after four days, and offspring collected without anesthesia to avoid interfering with wing expansion. Flies were measured in four different experiments during the mutation accumulation process, summarized in Table S1.

Control genotypes were available in the first of these experiments, conducted in 1998 at generation 73 of mutation accumulation. Control populations were derived from embryos cryopreserved at generation 0 of the mutation accumulation experiment (Houle and Nuzhdin 2004). Five pairs of IV*e*39 flies were recovered at generation 47 of mutation accumulation, and each was used to found an independently maintained population at a size of ten vials/generation. At generation 73 the control populations thus had undergone 26 generations of evolution, including mutation and natural selection, while the MA sublines had experienced 73 generations of mutation under conditions minimizing the strength of natural selection. The increased population size of these control populations relative the MA sublines enabled selection to reduce the impact of new mutations on the mean phenotype. Unfortunately, cryopreserved IV*e*33 embryos did not survive to generation 47 (Houle and Nuzhdin 2004). In addition, we had to revive all the remaining cryopreserved IV*e*39 embryos at generation 47 to obtain the ten control flies. The IV*e*39 control populations were lost before subsequent experiments.

1.3 PREPARING DATA FOR ANALYSIS

Extraction of data from wing images and error correction was accomplished in the Java program Wings (Van der Linde 2004-2008). Once a batch of wings is splined, Wings uses the minimum-volume ellipsoid (MVE) approach (Rousseeuw and van Zomeren 1990) to identify outlier wings, and overlays the fitted spline model onto the wing image, allowing the user to evaluate the fitted splines. Outliers that were due to poor fit by the automated splining algorithm were corrected manually using an interface in Wings that allows the user to alter the control points of the spline model until it fits the image well. Outliers due to a flawed wing image (a fold or tear in the wing, or a failure to properly position the wing in the field of view) or a wing with a major developmental defect were discarded before analysis.

The data we analyzed consisted of the locations of 12 vein intersections. These were aligned by generalized Procrustes least squares superimposition (Rohlf and Slice 1990). The resulting data gives the x- and y- coordinates of the displacement of each landmark from the centroid, measured in units of centroid size. Throughout the paper, coordinate data has units of centroid size in μm × 10, and centroid size has units of 1 μm.

We did a second round of outlier detection on the residuals from ANOVAs on each coordinate separately, and on scores along the first 22 principal components derived from the data, with gender and subline genotype as main effects. Configurations more than 4.5 S.D. from on any coordinate or score were checked manually for fit of the splines to the image using the editing procedure in Wings. Splining and manual adjustments were performed by two observers (DH and JF), but the identity of the editor introduced negligible variance and was ignored in the subsequent analyses. Even following correction, 185 wings had values on some traits or principal components that lay more than 4.5 S.D. from the mean. These wings were genuinely atypical, and not due to splining errors. They were approximately randomly distributed over genotypes. Analyses including and omitting these observations produced very similar estimates of genetic components. We report only analyses with outliers omitted.

The distribution of both the data and the residuals following the analyses outlined below was symmetrical and approximately normal. The distribution of residuals along each of the principal components of the shape data and centroid size departed significantly from normality ($P<0.01$ by Cramer-vonMises' test) due to heavy tails (median kurtosis 0.88). Skewness was generally small (median -0.06). Centroid size had high kurtosis (7.7 in males, 5.9 females), and positive skew in males (0.38 due to rare males that achieved sizes typical of females), and negatively skewed in females (-1.27) due to a long tail of flies that have male-typical sizes. Both departures are due to the sensitivity of size to nutritional conditions.

Canonical variates and discriminant analyses on subline means revealed that that the homozygotes of subline IV*e*-39-64 had a phenotype typical of an IV*e*-33 subline in each of the three experiments it was measured in. We included homozygous crosses with IV*e*-39-64 in our analyses as an IV*e*-33 subline, but deleted all crosses where IV*e*-39-64 was mated with other IV*e*-39 sublines. The residuals of all variables in the final data set differed significantly from normality, but visual inspection of q-q plots suggests that the departures are quite small.

1.4 MUTATION MODELS

We define a mutation model for a single trait. Within each line, IV*e*-33 and IV*e*-39, accumulation begins with a single homozygous reference genotype with sex-specific phenotypic means. A total of n loci are capable of influencing a trait, and the ith site has a haploid mutation rate u_i. We assume that mutations are rare enough that we can neglect the possibility that these will change the population grand mean in the short term, and that mutation rates are low, so that each mutation that occurs during the duration of the mutation-accumulation experiment is unique. In analyses including both sexes, we assume that the mutational effects are the same in each sex. A mutation at a single site 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. More generally, both homozygous and heterozygous effects may be altered by epistatic deviations when in mutationally altered background, rather than the original inbred line genotype, which we treat as a reference genotype. We assume that there is no bias to the epistatic effects, so that the expectation of the epistatic deviations remains 0 throughout the experiment. The probability that each individual carries a new mutation at the ith site is $2u_i$. .

For mutations with selection coefficients much less than the inverse of the effective population size, the per generation rate of fixation in a subline is u_i. After t generations of mutation accumulation, the among subline variance due to site i will be $t\mu_i \mathsf{E}\left[g_i^2 \right]$ in homozygotes and $t\mu_i \mathsf{E}\left[a_i^2 \right]$ in heterozygotes. Over all n loci affecting the trait the total genetic variance among the full set of homozygous sublines is $V_S = tU\mathsf{E}\left[g^2 \right]$, assuming no epistatic interactions. Among heterozygous crosses between sublines the variance is $V_b = tU\,\mathrm{E}\left[a^2 \right]$, again assuming no directional epistasis. From this, and equations (1) and (2) in the Methods section, we can see that $V_M = V_s/2t$, and that $V_{Mns} = 2V_b/t$.

Wing phenotypes were measured in both homozygous and heterozygous genotypes. The phenotypes of homozygotes of the jth subline at time t have components

$$Y_{jjtr} = \mu + 2b_{jt} + c_{jt} + e_{r(jjt)} = \mu + 2\sum_i A_{ijt} + c_{jt} + e_{r(jjt)}$$

where b_{jt} is the heterozygous effect of the autosomal genotype of the jth subline, c denotes departures from the additive model due to a combination of dominance and epistatic interactions, and $e_{r(jj)}$ the unique effects on individual r. The total heterozygous effect $b_j = \sum_i A_{ij}$, where A_{ij} is an indicator variable that is 0 if no mutation was fixed at site i in line j, and a_i if one was fixed. The phenotype of heterozygotes with maternal parent subline j and paternal parent subline k at time t has components

$$Y_{jktr} = \mu + b_{jt} + b_{kt} + \varepsilon_{jkt} + e_{r(jkt)} = \mu + \sum_i A_{ijt} + \sum_i A_{ikt} + \varepsilon_{jkt} + e_{r(jkt)}$$

where ε_{jk} represents the epistatic deviations from the additive model. Assuming that each mutation is unique, heterozygous genotypes formed by crossing sublines are never homozygous for any new mutations, and all departures from an additive model in the heterozygotes must be due to epistatic interactions. The homozygous genotypes may deviate from the additive expectation due to a combination of epistasis and dominance, so the terms *c* and ε and their variances are not comparable.

The covariances of relatives are derived from these phenotypic effects by defining each effect as independent of all other effects, so that only terms that involve squares of each of the components have non-zero expectations. For example, the covariance between maternal half sibs, the *r*th offspring of cross $j \times k$ at time *t* with that of the *s*th offspring of cross $j \times m$ at time *x* where $x>t$, is

$$\mathrm{E}\left(\left(Y_{jktr}-\mu\right)\left(Y_{jmxs}-\mu\right)\right)=\mathrm{E}\left[\left(b_{jt}+b_{kt}+\varepsilon_{jkt}+e_{r(jkt)}\right)\times\left(b_{jx}+b_{mx}+\varepsilon_{jmx}+e_{s(jmx)}\right)\right]$$
$$=\mathrm{E}\left[{b_{jt}}^{2}\right]=V_b$$

as subline *j* is the only source of common effects. Ignoring the time dimension, there are five degrees of relationships within the complete diallel crosses that generate unique covariances between relatives. The covariances generated by each relationship are given in Table S3 in the form of the coefficients of the variances of the causal variances of the *b, c* and ε effects.

The covariances associated with the additive effect of a subline can be interpreted in terms of the mutation model as

$$V_b=\mathrm{E}\left[{b_{jt}}^{2}\right]=\mathrm{E}\left[\left(\sum_i A_{ijt}\right)^2\right]=tU\,\mathrm{E}\left[a^2\right]={tV_{Mns}}/{2}\,.$$

Similarly, the among sub-line variance can be used to estimate V_M as

$$V_S=4V_b+V_c=4tU\,\mathrm{E}\left[a^2\right]+V_c=tU\,\mathrm{E}\left[g^2\right]=2tV_M$$

Consequently we expect that $V_M \geq V_{Mns}$ because $V_S/2t \geq 2V_b/t$.

Complete diallel data also offer the opportunity to estimate many additional variance components, including sex-specific, X-linked additive, interaction effects, and parental effects. Unfortunately, analyses of more complex models did not converge.

1.5 ANALYSES

To make inferences about fixed effects, we used multivariate analyses of variance and covariance implemented in the GLM procedure in SAS (SAS Institute 2004). Effect sizes and consistency of effects were judged using least squares means or parameter estimates. We attempted to implement mixed model multivariate analyses of variance using maximum likelihood in the Mixed procedure in SAS, but in most cases multivariate models with more than a few traits did not converge.

The animal model program Wombat (Meyer 2007, 2010) used for our quantitative genetic analyses assumes that individuals at the head of the pedigree are outbred, so we directly

calculated the relationship matrix using the coefficients in Table S3. Covariances were scaled to the 192 generations of the experiment by multiplying each relationship coefficient by the proportion of the 192 generations that the two genotypes shared accumulated mutations. For example, the covariances for individuals from generation 167 and 192 were multiplied by 167/192. We assumed that each subline was completely inbred as the expected within-subline variation due to unfixed mutations is approximately $4V_M$ (Lynch and Hill 1986), negligible relative to the variances accumulated over the 192 generations of the experiment. We then calculated the generalized inverse of this relationship matrix using SAS IML (SAS Institute 2004), and furnished this matrix to Wombat using the GIN option.

Following suggestions of Meyer (Meyer 2008),we used the parameter expanded expectation maximization (PX-EM) algorithm, which converges relatively rapidly from starting values far from a maximum, in combination with the average information (AI) algorithm which converges quadratically when close to a maximum, but poorly far from the optimum. Experimentation suggested that cycling between the two was a useful generic method, as it was difficult to anticipate whether initial values were poor or not. In most cases we used 10 rounds of PX-EM, followed by 10 rounds of AI. If convergence had still not occurred we repeated this cycle as many times as necessary. For the first analysis of each data partition, the initial parameter estimates were calculated as a proportion of the phenotypic variance-covariance matrix, while later analyses of variant models used estimates from previous WOMBAT runs as initial values. Starting covariance matrices were checked to determine if they were singular. If they were, the values were 'bent' by reducing the largest eigenvalues, while increasing the smallest ones by an equal total amount until full rank was achieved (Hayes and Hill 1981).

1.6 ESTIMATING AVERAGE DOMINANCE

To estimate the dominance implied by our data, we solved

$$\frac{V_M - V_{Mns}}{V_{Mns}} = \frac{U\left(2\mathrm{E}\left[a^2\right] + \frac{1}{2}\mathrm{E}\left[d^2\right] + COV\left[a,d\right]\right) - 2U\,\mathrm{E}\left[a^2\right]}{2U\,\mathrm{E}\left[a^2\right]}$$

$$= \frac{1}{2}\frac{\mathrm{E}\left[d^2\right]}{2\mathrm{E}\left[a^2\right]} + \frac{COV\left[a,d\right]}{2\mathrm{E}\left[a^2\right]}$$

Under the limiting condition of a perfect correlation between *d* and *a*, $d = xa$ and $COV\left(a,d\right) = x\mathrm{E}\left[a^2\right]$. Then

$$\frac{1}{2}\frac{x^2\,\mathrm{E}\left[a^2\right]}{2\mathrm{E}\left[a^2\right]} + \frac{x\,\mathrm{E}\left[a^2\right]}{2\mathrm{E}\left[a^2\right]} - \frac{V_M - V_{Mns}}{V_{Mns}}$$

$$= \frac{x^2}{4} + \frac{x}{2} - \frac{V_M - V_{Mns}}{V_{Mns}} = 0$$

and

$$x = \frac{-\frac{1}{2} \pm \sqrt{\frac{1}{4} + \frac{V_M - V_{Mns}}{V_{Mns}}}}{\frac{1}{2}}$$
$$= 2\sqrt{\frac{1}{4} + \frac{V_M - V_{Mns}}{V_{Mns}}} - 1$$

The most familiar parameterization of dominance denotes the dominance, *h*, as the proportion of the homozygous effect seen in heterozygotes. Then under the assumption that *d*=*xa*,

$$h(2a + xa) = a$$
$$h = \frac{1}{2 + x}$$

Simulations suggest that neither variance in *x* nor variance in *d* independent of *a* affect this result. Estimates of dominance are shown in Table S9.

1.7 COMMON SUBSPACE ANALYSIS

Consider two populations A and B with covariance matrices $\mathbf{M}_A$ and $\mathbf{M}_B$. A naive algorithm to estimate the common subspace is to assess the variances in $\mathbf{M}_A$ and $\mathbf{M}_B$ along the eigenvectors of the matrix derived from the pooled data $\mathbf{M}_{A+B}$. If the variance in either $\mathbf{M}_A$ and $\mathbf{M}_B$ is less than the significance threshold, that eigenvector is declared to be in the corresponding nearly null space, and dropped from the common subspace. The flaw in this algorithm is that the eigenvectors of $\mathbf{M}_{A+B}$ do not necessarily find directions with minimal variance in $\mathbf{M}_A$ and $\mathbf{M}_B$. To improve on this algorithm we exploited a result from Flury (1983) that exploits 'generalized' eigenanalysis of **C** the product one covariance matrix ($\mathbf{M}_B$) premultiplied by the inverse of another covariance matrix ($\mathbf{M}_A$). The matrix $\mathbf{C} = \mathbf{M}_A^{-1}\mathbf{M}_B$ will approach the identity matrix if $\mathbf{M}_A$=$\mathbf{M}_B$. As $\mathbf{M}_A$ and $\mathbf{M}_B$ become more unequal, the diagonal elements of **C** will become more different from 1, being larger when the corresponding direction in B has more variance than that in A, and smaller in the converse case where A has more variance than B. Flury (1983) showed that the eigenvectors of **C**, $\mathbf{V}_C$, estimate directions that have most extreme ratios of variances in $\mathbf{M}_B$ relative to $\mathbf{M}_A$, and therefore the largest and smallest ratios of variances in population B relative to A. The eigenvectors, $\mathbf{V}_C$, are not necessarily orthogonal, although the scores on the vectors within populations are constrained to be uncorrelated. The eigenvalues, $\mathbf{\Lambda}_C$, estimate the ratio of variances in those directions. Thus, the largest element of $\mathbf{\Lambda}_C$, $\lambda_{\max}$ identifies the direction where ratio of the variance in population B to variance in population A is greatest. The smallest eigenvalue $\lambda_{\min}$ correspondingly estimates that direction where the variance in A is maximal relative to that in B. This operation is also used in canonical variates analysis, where it is used to find the directions in a space that maximize the ratio of between group variance to

within group variance. Our subspace-finding algorithm is based on the idea that $\mathbf{V}_C$ is more likely to identify directions with non-significant variation in one matrix than the eigenvectors of $\mathbf{M}_{A+B}$.

The modified subspace partitioning algorithm for matrices $\mathbf{M}_A$ and $\mathbf{M}_B$ then proceeds as follows. The eigenvectors of the matrix $\mathbf{M}_{A+B}$ without significant variation identify most axes of the doubly-null subspace. $\mathbf{M}_A$ and $\mathbf{M}_{B,}$ are projected into a space of reduced dimension, based on the eigenvectors of the pooled matrix $\mathbf{M}_{A+B}$ with significant genetic variance. We then calculated $\mathbf{C}_r = \mathbf{M}_{Ar}{}^{-1}\mathbf{M}_{Br}$, decomposed $\mathbf{C}_r$ into its eigenvectors, $\mathbf{V}_C$, then evaluated the evolvability of $\mathbf{M}_{Ar}$ and $\mathbf{M}_{Br}$ along each of these eigenvectors. If any variances were less than the minimum eigenvalue statistically supported for the corresponding matrix ($\mathbf{M}_A$ or $\mathbf{M}_B$) we reduced rank *r* by one, and repeated the analysis until all variances on the eigenvectors of $\mathbf{C}_r$ were greater than the minimal supported eigenvalues for $\mathbf{M}_A$ and for $\mathbf{M}_B$. In most cases, the first or last eigenvectors were the nearly-null directions, although for a few comparisons the only nearly-null variances were found on one of the interior eigenvectors of $\mathbf{C}_r$. To project a $p \times p$ covariance matrix $\mathbf{M}_X$ into *r*-dimensional space we use the first *r* columns of the eigenvector matrix of $\mathbf{M}_{A+B}$, $\mathbf{V}_{A+B}$

$$\mathbf{M}_{Xr} = \left(\left[\mathbf{V}_{A+B}\right]_{n\times r}\right)^T \mathbf{M}_{Xr}\left[\mathbf{V}_{A+B}\right]_{n\times r}$$

where the brackets indicate a submatrix and ^{T} indicates transpose.

Supplementary Results

2.1 TESTING FOR INBREEDING DEPRESSION AND DIRECTIONAL DOMINANCE

The Cross 2003 experiment simultaneously reared inbred and outbred genotypes, so we could test for directional dominance and inbreeding depression. Multivariate analyses of variance were carried out with main effects inbreeding status, sex, and genotype, block and male and female parental subline nested in genotypes, and wing size as a covariate. Subline effects were treated as random, while other effects were fixed. The results show highly significant three-way interactions between inbred status, sex and genotype, but ambiguous evidence for a main effect of inbreeding status (results not shown). The main effect of inbreeding status was highly significant in MANOVAs within each sex and genotype combination. We estimated the vector difference between inbred and outbred flies within each combination of sex and genotype, and no two of these vectors were similar in direction (range of angles 60º to 87º). The average angle between all six pairs of these vectors is 75º, which is close to the 80º difference expected between random vectors. These results suggest that the complex interactions with inbreeding status are due to sex- and genotype-specific differences between outbred and inbreds, rather than a consistent mean difference due to inbreeding. Such differences are consistent with the pattern of variation shown in Fig. 1, as inbreds are more extreme in their phenotypes, and thus can be expected to show mean differences of random direction from outbreds depending on the small sample of sub-lines sampled in each block.

2.2 CONVERGENCE OF THE NON-SEGREGATIONAL MODEL

The non-segregational model invariably showed convergence problems in Wombat at ranks higher than 7 or 8. Convergence of higher rank models came only after the average information (AI) and parameter-expanded (PX) EM algorithms had failed in some iterations, and Wombat carried a single round of EM maximization which was deemed to have converged. Simulations show that this behavior appears in WOMBAT when the eigenvectors of the between-group covariance matrix diverge sufficiently from that of the variance component matrix being fit, and when the effective sample size is small (Meyer and Kirkpatrick 2008; K. Meyer, pers. comm.).

2.3 EVOLVABILITY IN THE FIRST 7 DIMENSIONS OF EACH DATA PARTITION

The lower rank of non-segregational matrices may explain some of the differences in evolvability evident in Table 1. Table S8 gives scalar measures of evolvability calculated over the space defined by each matrix's first seven eigenvectors, in which each matrix is full rank. Note that the phenotypic spaces defined by each partition may be different. The maximum possible evolvability, e_{max}, is the eigenvalue associated with the first eigenvector of each matrix (Schluter 1996). This is a useful standard to compare with evolvabilities in other directions (Kirkpatrick 2009). The remaining measures of evolvability in Table S8 are those suggested by Hansen and Houle (2008): evolvability, conditional evolvability ($\bar{c}$), autonomy ($\bar{a}$), and respondability ($\bar{r}$). The interpretation of these measures is briefly covered in the Analysis

Philosophy section of the Discussion. Comparison of $e_{\max,}$ $\overline{e}$ and $\overline{r}$ between non-segregational and homozygous models shows that each of these is about 2.5 times larger in the homozygous analysis than the non-segregational analysis, similar to the 2.75 ratio in the 24 dimensional shape space; the higher homozygous evolvabilities are not an artifact of the lower estimated dimensionality in the non-segregational model. Respondability is only modestly larger than $\overline{e}$, suggesting that responses are not deflected very far away from the selected direction.

To interpret the value of $\overline{c}$ and $\overline{a}$, it is important to realize that conditional evolvability is highly dependent on the phenotypic space considered. For matrices of less than full rank, $\overline{c} = 0$, so the shape conditional evolvabilities of all **M** matrices is predicted to be 0. In general, $\overline{c}$ will be smaller the more traits that are considered. In seven phenotypic dimensions, $\overline{c}$ is 40% of $\overline{e}$ in the non-segregational model, and 65% of $\overline{e}$ in the homozygous model. This implies that the deviations from additive gene action explore a higher-dimensional space than the additive effects themselves.

Supplementary Tables

Table S1. Wing data obtained.

Experiment	Generation of MA	Genotypes	Line	Sublines	Wings measured ♂	Wings measured ♀
1998	74	Homozygotes	33	1	16	0
1998	73	Homozygotes	39	27	989	0
1998	0	Control homozygotes	39	5	313	0
2002	167-173	Homozygotes	33	29	1470	1575
2002	166-172	Homozygotes	39	20	970	981
Diallel 2002	170	2 partial diallels	39	8	565	573
Diallel 2003	192	7 complete diallels	33	27	1347	1339
Diallel 2003	191	6 complete diallels*	39	17	973	964
Totals					6653	5432

*Subline IV*e*-39 64 had a phenotype typical of an IV*e*-33 each time it was measured. Consequently, we analyzed it as an IV*e*-33 subline. IV*e*-39 64 was a parent in two diallel crosses in 2003, and all heterozygotes with IV*e*-39 64 as a parent were excluded from analyses.

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.

Cross 1	Cross 2	*Relationship*	V_b	V_c	V_ε
i × i	i × i	Inbreds with inbreds	4	1	0
i × i	i × j	Inbreds with half sibs	2	0	0
i × j	i × j	Full sibs	2	0	1
i × j	i × k	Half sibs	1	0	0
i × j	k × l	Unrelated	0	0	0

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

Effect	Random?	Error SSCP	df Num.	df Den.	F	Wilk's λ	P
Treatment	Fixed	Genotype in treatment	22	7	0.31	0.509	0.98
Genotype in treatment	Random	Error	616	22113	7.71	0.034	<0.0001
Block	Fixed	Error	22	1270	2.32	0.961	0.0005

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

Sex	Genotype	Rank	AICc	parameters	df*	D†	P
♂	33	7	467842	386	16	517.5	<0.0001
♂	39	7	583579	386	16	555.42	<0.0001
♀	33	7	482952	386	16	353.44	<0.0001
♀	39	7	405297	386	16	355.96	<0.0001
♂	33 and 39	8	1058986	401	15	866.32	<0.0001
♀	33 and 39	7	896047	386	16	833.04	<0.0001
Both	33	7	952639	386	16	1040.14	<0.0001
Both	39	7	993126	386	16	1100.12	<0.0001
Both	33 and 39	7	1961351	386	16	1901.6	<0.0001

*df =parameters of best model –parameters of next best model

†D=−2 ln (likelihood of best model/likelihood of next best 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.

Sex	Genotype	Rank	AICc	params	df*	D†	P
♂	33	10	304569	428	12	-83.78	<<0.0001
♂	33	11	304510	440	0	0	
♂	33	12	304516	451	11	15.7	0.15
♂	33	13	304554	461	21	-1.66	NA
♂	39	11	375239	440	11	-37.32	0.0001
♂	39	12	375224	451	0	0	
♂	39	13	375244	461	10	0.7	1.00
♂	39	14	375276	470	19	-13.54	NA
♀	33	11	319968	440	11	-48.44	<<0.0001
♀	33	12	319942	451	0	0	
♀	33	13	319950	461	10	12.2	0.27
♀	33	14	319985	470	19	-4.1	NA
♀	39	11	202536	440	11	-50.32	<<0.0001
♀	39	12	202508	451	0	0	
♀	39	13	202553	461	10	-24.22	NA
♂	33 and 39	14	685050	470	8	-53.7	<<0.0001
♂	33 and 39	15	685013	478	0	0	
♂	33 and 39	16	685027	485	7	-0.26	NA
♀	33 and 39	15	526308	478	7	-43.8	<<0.0001
♀	33 and 39	16	526279	485	0	0	
♀	33 and 39	17	526306	491	6	-15	NA
Both	33	13	625571	461	17	-61.1	<<0.0001
Both	33	14	625545	470	8	-16.36	0.04
Both	33	15	625545	478	0	0	
Both	33	16	625562	485	7	-3.18	NA
Both	39	13	580087	461	9	-77.48	<<0.0001
Both	39	14	580027	470	0	0	
Both	39	15	580037	478	8	6.12	0.63
Both	39	17	580120	491	21	-50.58	NA
Both	33 and 39	17	1213501	491	9	-122.76	<<0.0001
Both	33 and 39	18	1213406	496	4	-17.7	0.0014
Both	33 and 39	19	1213396	500	0	0	
Both	33 and 39	20	1213425	503	3	-22.9	NA

*df =|parameters of best model –parameters of comparison model|

†D= 2 ln (likelihood of comparison model/likelihood of best model)

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, $\overline{e}$, conditional evolvability, $\overline{c}$, autonomy, $\overline{a}$, and respondability, $\overline{r}$, defined in Hansen and Houle (2008).

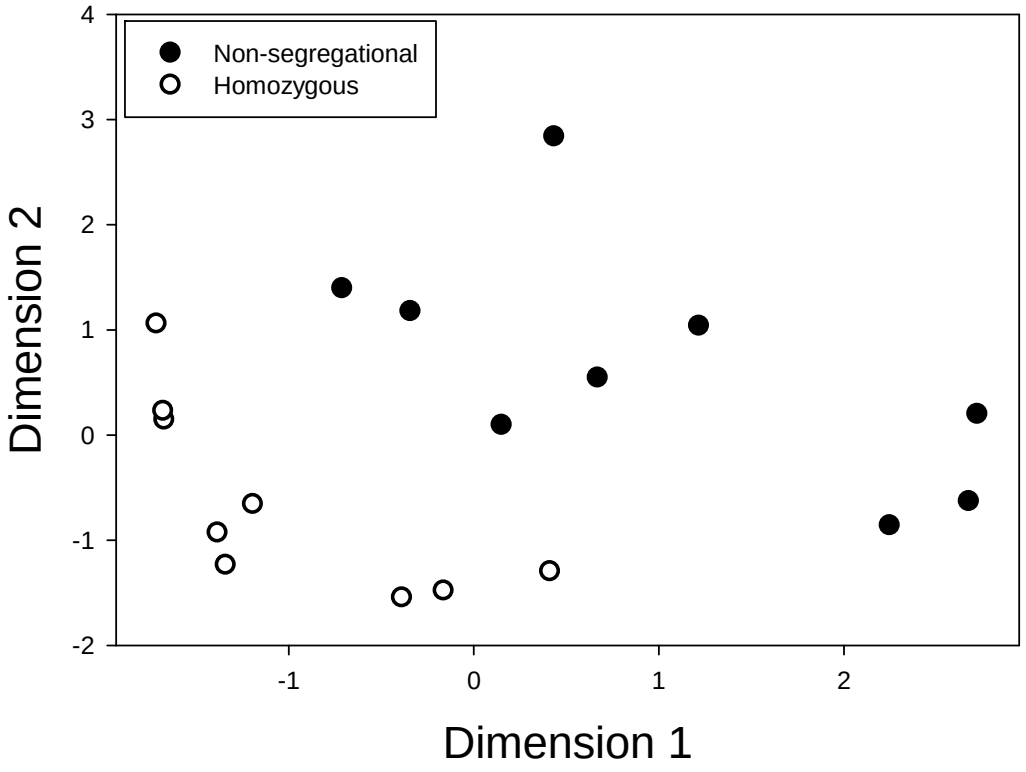
		Non-segregational model					Homozygous model				
Sex	Genotype	e_{max}	$\overline{e}$	$\overline{c}$	$\overline{a}$	$\overline{r}$	e_{max}	$\overline{e}$	$\overline{c}$	$\overline{a}$	$\overline{r}$
♂	33	6.9	2.23	0.84	0.39	2.93	19.9	7.19	3.37	0.49	8.86
♂	39	19.3	6.16	2.07	0.37	8.18	33.3	11.51	5.42	0.49	14.57
♀	33	4.5	1.97	0.70	0.38	2.46	17.9	7.48	3.37	0.46	9.13
♀	39	18.2	5.32	0.94	0.19	7.86	51.6	13.50	4.66	0.36	18.87
♂	33 and 39	11.9	4.01	1.72	0.45	5.26	24.9	9.97	5.90	0.61	11.90
♀	33 and 39	9.4	2.91	1.05	0.38	3.84	27.2	10.14	5.26	0.54	12.39
Both	33	6.2	2.51	1.06	0.44	3.10	15.3	7.54	3.43	0.48	8.94
Both	39	21.3	6.47	1.81	0.31	8.71	36.0	11.75	5.69	0.50	15.00
Both	33 and 39	11.5	4.04	1.53	0.40	5.11	24.3	10.21	6.48	0.65	11.83

Table S9. Average dominance of mutations.

		h	
Sex	Genotype	Shape	Size
♂	33	0.24	0.26
♂	39	0.31	0.31
♀	33	0.22	0.29
♀	39	0.27	0.23
♂	33 and 39	0.27	0.29
♀	33 and 39	0.23	0.24
Both	33	0.24	0.27
Both	39	0.32	0.33
Both	33 and 39	0.26	0.29

Supplementary Figure

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, while the six intermediate points are for the pooled genotype estimates.



References to Supplementary Information

Charlesworth, B., and D. Charlesworth. 1985. Genetic variation in recombination in Drosophila. I. Responses to selection and preliminary genetic analysis. Heredity 54:71-83.

Hayes, J. F., and W. G. Hill. 1981. Modification of estimates of parameters in the construction of genetic selection indices (`bending'). Biometrics 37:483-493.

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

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., and L. Rowe. 2003. Natural selection in a bottle. Am. Nat. 161:50-67.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271-284.

Lu, F., and E. E. Milios. 1994. Optimal spline fitting to planar shape. Signal Processing 37:129-140.

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

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. 2008. Parameter expansion for estimation of reduced rank covariance matrices. Genet. Sel. Evol. 40:3-24.

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 at http://agbu.une.edu.au/~kmeyer/wombat.html

Meyer, K., and M. Kirkpatrick. 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.

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

Rousseeuw, P. J., and B. C. van Zomeren. 1990. Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. 85:633-639.

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

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

Van der Linde, K. 2004-2008. Wings 3.7. Available at http://bio.fsu.edu/~dhoule/Software/