Evolution, 59(5), 2005, pp. 1027–1038

THE DIMENSIONALITY OF GENETIC VARIATION FOR WING SHAPE IN *DROSOPHILA MELANOGASTER*

JASON G. MEZEY[1,2] AND DAVID HOULE[1]
[1]*Department of Biological Science, Florida State University, Tallahassee, Florida 32306-1100*

Abstract.—Absolute constraints are limitations on genetic variation that preclude evolutionary change in some aspect of the phenotype. Absolute constraints may reflect complete absence of variation, lack of genetic variation that extends the range of phenotypes beyond some limit, or lack of additive genetic variation. This last type of absolute constraint is bidirectional, because the mean cannot evolve to be larger or smaller. Most traits do possess genetic variation, so bidirectional absolute constraints are most likely to be detected in a multivariate context, where they would reflect combinations of traits, or dimensions in phenotype space that cannot evolve. A bidirectional absolute constraint will cause the additive genetic covariance matrix (**G**) to have a rank less than the number of traits studied. In this study, we estimate the rank of the **G**-matrix for 20 aspects of wing shape in *Drosophila melanogaster*. Our best estimates of matrix rank are 20 in both sexes. Lower 95% confidence intervals of rank are 17 for females and 18 for males. We therefore find little evidence of bidirectional absolute constraints. We discuss the importance of this result for resolving the relative roles of selection and drift processes versus constraints in the evolution of wing shape in *Drosophila*.

Key words.—Absolute constraint, dimensionality, *Drosophila*, evolutionary constraints, **G**-matrix, wing shape.

Received August 9, 2004. Accepted February 3, 2005.

Heritable variation is the raw material for evolution. The pattern of heritable variation and covariation determines the probability that a given phenotype will evolve under drift (Lande 1976; Arnold et al. 2001), can deflect the response to directional selection away from the direction of selection (Lande 1979; Phillips and Arnold 1989), and can potentially alter the equilibrium state reached on a multipeaked fitness surface (Price et al. 1993; Steppan et al. 2002). Genetic covariation is usually an evolutionary constraint, because it restricts the rate of evolution in most directions in the space of phenotypes (Arnold 1992). Therefore, one of the goals of evolutionary quantitative genetics is to quantify the opportunities and constraints inherent in patterns of genetic covariation among traits (e.g., Roff and Mousseau 1999; Steppan et al. 2002). Much of this work has focused on quantitative constraints (Houle 2001): the numerical values of genetic covariances and correlations or derived measures such as heritability. Over relatively short time scales, quantitative constraints can profoundly affect the state of a population (Steppan et al. 2002). Conversely, if selection for a particular optimum state persists long enough, quantitative constraints will be overcome (Via and Lande 1985; Zeng 1988).

In contrast, absolute constraints exist where a lack of genetic variation precludes a response to selection altogether (Via and Lande 1985; Houle 1991, 2001; Kirkpatrick and Lofsvold 1992). Such constraints are frequently invoked to explain the limited variety of realized biological forms (e.g., Gould and Lewontin 1979; Brooks and McLennan 1991), but are rarely the subject of direct study at the genetic level. A few authors have outlined programs for identifying such constraints (e.g., Maynard Smith et al. 1985; Kirkpatrick and Lofsvold 1992), but there have been few rigorous attempts to put these programs into practice. The extent of such absolute constraints and their importance in evolution is therefore unknown. This is a pity, because absolute constraints allow us to make strong macroevolutionary predictions concerning what cannot be attained by natural and artificial selection. A lack of variation for a phenotype would provide important evidence for the prevalence of constraints in evolution. Conversely, rigorous demonstration that absolute constraints do not exist indicates that other factors, such as natural selection, must be found to explain limitations on phenotypic evolution.

We consider there to be three categories of absolute evolutionary constraints: (1) complete lack of phenotypic variation; (2) unidirectional absolute constraints; and (3) bidirectional absolute constraints. If a trait lacks phenotypic variation, there is by definition no heritable variation, and further evolution is impossible. These constraints have been invoked when there are clear developmental or mechanistic reasons why phenotypes exist in one lineage but cannot in a related lineage (Alberch and Gale 1983; Alberch 1985) and in some rare cases have been quantified directly (Bradshaw 1991). Unidirectional constraints are caused by a lack of alleles that extend the range of phenotypes beyond some limit. Evolution in one direction is still possible. Such a constraint could explain cases where persistent artificial selection reaches a plateau where further response does not occur (Falconer and Mackay 1996), but reverse selection often creates a response in the opposite direction. The existence of trade-offs among fitness components are often assumed to be a consequence of a limit reached where there is no heritable variation for further increases in fitness (Lande 1982; Houle 1991; de Jong and van Noordwijk 1992). Finally, a bidirectional absolute constraint occurs when a phenotypically variable trait entirely lacks additive genetic variation (Via and Lande 1985). Because additive genetic variation is a prerequisite for a trait to respond to selection, there can be no response to selection in either direction. If such constraints are common and persist through time, the existence of variation for a trait is not a strong predictor of its ability to evolve. Conversely, if there are no bidirectional constraints on a phenotype, this makes

[2] Present address: Center for Population Biology, University of California, Davis, California, 95616; E-mail: jgmezey@ucdavis.edu.

a similarly strong prediction: Where variation can be measured, evolution can proceed.

The possibility of bidirectional absolute constraints may appear to be at odds with the robust result of quantitative genetics that nearly all features of organisms possess additive genetic variation (Mousseau and Roff 1987; Roff and Mousseau 1987; Houle 1992; Lynch and Walsh 1998). This is not the case, as the pattern of covariation may mean that some trait combinations lack genetic variation. This possibility becomes more likely the larger the number of traits considered. For example, it may not be possible to increase one trait while holding others constant. It is therefore essential to test for constraints in a multivariate context.

To test for bidirectional constraints, we need to consider the geometry of variation. Each trait that is measured defines a direction in what we call ''phenotype space,'' which has as many dimensions as the number of traits, say *n*. The mean phenotype defines a point in that *n*-dimensional space. The question we want to ask is whether the breeding values of the genotypes studied extend in all directions from this central point in phenotype space, that is, fill out all *n* possible dimensions. Alternatively, there may be some directions in phenotype space where no breeding values different from the mean exist. An easy-to-visualize example is in a three-trait phenotype space where a bidirectional absolute constraint exists if all the breeding values fall on a plane in the space, rather than the usual three-dimensional ellipsoid. In this case, the variation has a dimensionality of two. In linear algebra, dimension is captured in the rank of the *n*-dimensional variance-covariance matrix that summarizes the data, where rank is the number of linearly independent rows in the matrix (Strang 1988). Whenever a particular direction lacks variation, the value of one row of the matrix will be predictable from the remaining rows. Thus, bidirectional absolute constraints cause an $n \times n$ trait **G**-matrix to have rank $< n$, and may therefore be quantified by studying rank of a **G**-matrix. This is the goal of our study—to quantify the rank of a **G**-matrix in enough detail to meaningfully explore the possibility of bidirectional absolute constraints.

It is important to note at the outset that one can never demonstrate absence in a statistical context. Instead, what one can do is design powerful studies for rejecting the hypothesis of less-than-full rank **G**-matrices. Demonstrating a full-rank **G**-matrix would then constitute powerful evidence against the sort of simple constraint hypothesis we have outlined. Failure to demonstrate a full-rank matrix would allow for the possibility of bidirectional constraints. The more phenotypic dimensions the null hypothesis cannot be rejected for, the more attractive a bidirectional constraint hypothesis would become.

This framework puts a substantial burden on the design of an experiment to look for constraints. To be informative, the experiment must first examine a large number of traits. A bidirectional constraint can only be detected when the number of traits analyzed is greater than the true rank of **G**. Second, the chance of finding constraints would be maximized by exhaustively sampling a restricted aspect of the phenotype, rather than sampling many diverse phenotypes such as size, a behavior, a life-history trait, and so on. Third, statistical power for finding variation in any one trait should be high. Failure to find variation in traits with high error or realization variances could readily be attributed to a lack of power, rather than an absence of variation. Fourth, the overall sample size of the study must be much larger than *n*. Otherwise, it is likely that the observations will by chance fall in a space of less than *n* dimensions. This is known as the ''curse of dimensionality.''

There has been speculation that bidirectional constraints are common based on a number of studies that have found **G**-matrices with rank less than *n* (Bailey 1956; Leamy 1977; Atchley et al. 1981). However, many of these studies used relatively small sample sizes. None of them present confidence intervals on the rank of **G**. Without a confidence interval, it is difficult to assess whether an estimated absolute constraint may simply reflect low power. One exception is the study of Kirkpatrick and Lofsvold (1992). The authors used published **G**-matrices from studies of growth in mice, chickens, and two populations of sheep to estimate developmental trajectories and confidence intervals for the ranks of these matrices. For each of these four cases, the authors estimated the 95% confidence interval for rank as substantially less than *n*. They conclude that there are considerable constraints for growth trajectories and some trajectories are evolutionarily forbidden. Because *n* ranged from four to nine in these analyses, the results are consistent with the supposition that the level of absolute constraints can be quite high.

To address these issues, we have undertaken a study of genetic variation in the shape of the *Drosophila melanogaster* wing. Wing shape is particularly interesting because it is both highly conserved in the family Drosophilidae (Houle et al. 2003) and shows additive genetic variation within *D. melanogaster* (e.g., Whitlock and Fowler 1999). A number of indirect lines of evidence indicate that the level of absolute constraint on wing shape may be low. Artificial selection on many specific wing shape indices has consistently generated large evolutionary responses (Weber 1990, 1992; Houle et al. 2003). Populations and inbred lines are readily distinguished by shape (e.g., Cavicchi et al. 1991; Imasheva et al. 1995; Birdsall et al. 2000; Gilchrist et al. 2000). In addition, there are several characterized genetic pathways involving many genes that could affect aspects of wing shape (Bier 2000; Mohler et al. 2000; Held 2002; Lunde et al. 2003; Cook et al. 2004). Several independent quantitative trait loci (QTL) analyses (Weber et al. 1999; Zimmerman et al. 2000; Weber et al. 2001; Mezey et al. 2005) have indicated that the number of loci responsible for observed variation in wing shape is quite large. However, none of these lines of evidence directly addresses the existence of bidirectional constraints, as the effects of the variation observed may be concentrated in less than the full number of phenotypic dimensions.

Our aim in this study is to directly estimate the dimensionality of genetic variation in the wing, and therefore to see if we can reject the possibility that bidirectional absolute constraints exist for the *Drosophila* wing.

Materials and Methods

Stocks and Breeding Design

The laboratory population was founded from approximately 140 isofemale lines derived from flies captured on

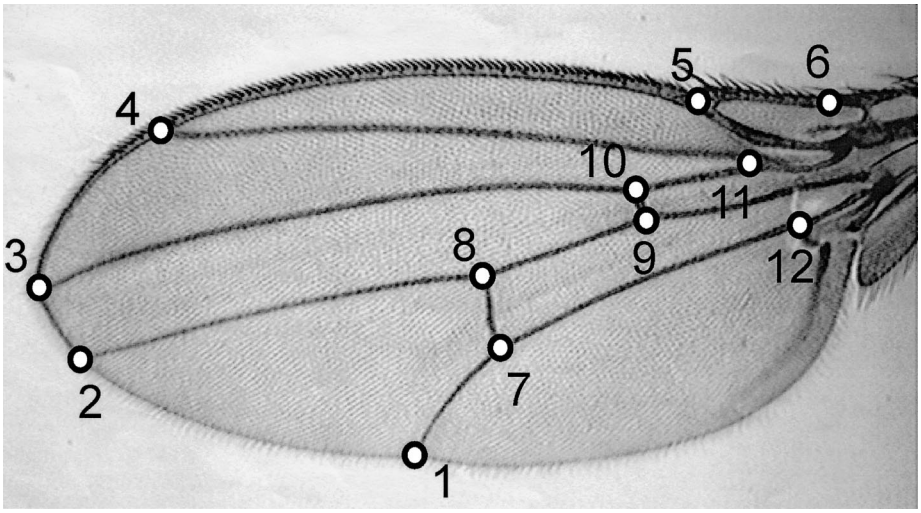

FIG. 1. Identity of the landmarks used in this study.

March 12, 2002, at Wabasso, Indian River County, Florida. The stock was maintained on cornmeal, sucrose, and brewer's yeast medium in bottles at 25°C with alternating 12-h light and dark cycles. The flies were transferred to fresh containers every two weeks. Initially, and through the first half of the study, the stock was maintained as isofemale lines. These were later pooled into a single breeding population, which was maintained in 10 bottles at a density of approximately 40 parental flies/bottle. At each transfer, adults from different bottles were mixed.

A half-sib breeding design was carried out in 36 temporal blocks, each consisting of four to six half-sib families. For the blocks 1–17, males were chosen at random from isofemale lines and were mated to four or five virgin females, each from a different isofemale line. Isofemale lines should become progressively more inbred over time. To check whether inbreeding might have affected the estimates from this part of the experiment, we regressed the sire component of variance from sibling data on block number. There was no evidence of a relationship between these (analysis not shown), suggesting that inbreeding remained low. For blocks 18–36, a male was mated to four or five virgin females, where each individual was selected at random from the pooled population. Comparison of data from the two parts of the experiment similarly suggested that there were no differences between them.

In the first part of the experiment, parents were reared in vials, and in the second part in bottles. Virgin parents were collected within 10 h of eclosion, aged for two to six days, then mated for three days. Female parents were allowed to lay eggs for two days in each of two replicate vials. One wing from each parent and from approximately five offspring of each sex from each vial were measured. The upper surface of left wings was measured whenever possible. When the left wing was damaged, the upper right wing was recorded instead.

Wing Measurement and Morphometrics

Wing measurement was performed using WingMachine, an automated image analysis system, the details of which are described elsewhere (Houle et al. 2003). WingMachine consists of a suction device that holds the wing of an anesthetized fly between a slide and a cover slip to allow a video image of the wing to be captured. The image of each wing plus two landmarks provided by a human observer were passed to an image processing system which fits B-splines (Lu and Milios 1994) to all wing veins posterior to the humeral break, plus the outline of the wing. Twelve intersections of these splines define the landmark coordinates used in these analyses (Fig. 1).

The data were aligned by generalized Procrustes least squares superimposition (Rohlf and Slice 1990) implemented in the tpsRegr program (Rohlf 1998b). In this approach, wings are first scaled to unit centroid size, where centroid size is the square root of the sum of squared distances of each landmark from the centroid of each wing. The centroid is the mean of the landmark coordinates for that specimen. Scaling by centroid size is optimal for minimizing the differences in wing size when variances are equal at each landmark. After scaling, the wings were aligned by finding the minimum squared distance between the landmarks that can be achieved by translating each landmark polygon to a common centroid position and then rotating the wings with respect to one another. Separate superimpositions were run on males and females because there is a difference in shape between the sexes. 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.

Following alignment, a robust covariance matrix was fit to the data using the minimum-volume ellipsoid approach (Rousseeuw and van Zomeren 1990), as implemented in the S-Plus program cov.mve (Insightful Corporation 2001). This covariance matrix was then used to identify potential outliers, which were then checked by a human observer and corrected when necessary using the digitization program TPSdig (Rohlf 1998a). Manual adjustments were performed by two observers. Landmark adjustment introduced negligible error and was ignored in the subsequent analysis. Throughout this pa-

per coordinate data has units of centroid size in mm $\times 10^4$, and centroid size has units of 10^{-3} mm or 1 μm.

Parameter Estimation

Approximately 20% of the *D. melanogaster* genome is sex linked, which makes it important to take sex linkage into account during analysis. Furthermore, sex limited effects are commonly identified, making it potentially important to separate the genetic effects in males and females. To provide the necessary degrees of freedom to undertake such analyses, we combined information on parent-offspring covariances with information from the standard nested sibling design.

For parent-offspring analyses we estimated the covariance between parent and the mean of their same-sex offspring directly (Fry 2004) using restricted maximum likelihood (REML) implemented in the SAS Mixed procedure (SAS ver. 8.01, SAS Institute, Inc., Cary.), taking into account block and generations as fixed effects. For the offspring data, the *x*- or *y*-landmark displacement *z* was decomposed as follows:

$$z_{i.j.k.l.m} = \mu + b_i + s_{i.j} + d_{i.j.k} + v_{i.j.k.l} + e_{i.j.k.l.m}, \qquad (1)$$

where μ is the population mean; *b, s, d, v* are the block, sire, dam, and vial effects; and *e* is the residual error. With the current half-sib design, every level is nested within the upper level. All effects were treated as random except for block effects, which were treated as fixed. The variances associated with each component of the model were estimated using REML in SAS Mixed (SAS ver. 8.01). We attempted to estimate variance and covariance terms in a single analysis (Fry 2004), but convergence problems, plus long run times, led us to estimate the variance for each trait separately. Covariance terms were calculated by estimating variance components for the sum of two traits and then subtracting the variance associated with each trait and dividing by two.

We estimated the causal components of (co)variance using weighted least-squares (Cowley et al. 1986). A vector of observational components, **Y**, was formed with (in order) the parent-offspring covariances for females and males ($C_{po\cdot f}$, $C_{po\cdot m}$) and sire, dam, vial, and error variances for females ($V_{s\cdot f}$, $V_{d\cdot f}$, $V_{v\cdot f}$, $V_{e\cdot f}$), then the corresponding terms for males ($V_{s\cdot m}$, $V_{d\cdot m}$, $V_{v\cdot m}$, $V_{e\cdot m}$). We formed the weight matrix **W** as a block diagonal consisting of the estimated variances for $C_{po\cdot f}$, $C_{po\cdot m}$, followed by the 4×4 matrices from the female and male sibling analyses. Exploratory analyses suggested that the differences in error and nonadditive variances between the sexes were relatively insignificant for the coordinate data, so we fit a vector of causal effects **E** that (in order) were an additive genetic autosomal effect for each sex ($V_{A\cdot f}$, $V_{A\cdot m}$), an X-linked additive effect for each sex ($V_{Ax\cdot f}$, $V_{Ax\cdot m}$), and common residual variances. These residual variances are (in order) maternal (V_M), vial (V_V), and error (V_E). The columns of the design matrix for the additive terms were twice the coancestries of the individuals whose covariances were estimated. With the observational and causal vectors in the order given, the design matrix was

$$\mathbf{X} = \begin{pmatrix} 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & 0.25 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0.25 & 0 & 0 & 0 & 1 \\ 0 & 0.25 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0.5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 & 1 \end{pmatrix}. \qquad (2)$$

The causal (co)variances were then estimated as $\mathbf{E} = (\mathbf{X^T W^{-1} X})^{-1} \mathbf{X^T W^{-1} Y}$ using the IML procedure in SAS (ver. 8.01). In a few cases, the vial variances and the corresponding values in the weight matrix were estimated as zero. We incorporated these zeroes in our estimates by setting the corresponding diagonal in the weight matrix to the average value of the vial variance from other traits. When elements of **E** were estimated as negative, they were set to zero, and the remainder of **E** estimated again with the corresponding column of **X** deleted. The adequacy of the overall model was tested comparing the weighted residual sums of squares to a χ^2 distribution with 10 estimates -7 parameters $= 3$ df (Lynch and Walsh 1998). The significance of each term was assessed by refitting the model with that term deleted and comparing the change in sums of squares to a χ^2 distribution with 1 df.

Estimation of Matrix Rank

The number of variable dimensions in wing shape space at each level of causal hierarchy is equal to the rank of the corresponding covariance matrix. The rank of a symmetric matrix is equal to the number of nonzero eigenvalues, so to estimate matrix rank, we used the distribution of the eigenvalues from principal components analysis (PCA).

To account for sampling error in this estimate, we also estimated a confidence quantile for matrix rank by bootstrapping. Because our method for estimating causal components of variation uses all of the observational components of variation (see Parameter Estimation above), we had to choose a bootstrap method that did not bias the variation at any observational level (parents and offspring, sires, dams, or error [co] variances). Simulations showed that conventional bootstrapping of the lower levels of the experiment (sires on down) biased estimates at the lowest level bootstrapped downward, with a corresponding bias of the next level upward (results not shown). For example, bootstrapping sires only biases the sire estimate downward. Bootstrapping at every random level (sire, dam, flies within families) biases the error variance downward and the dam level upward. These biases lead to very inaccurate estimation of the causal components. Our solution was to resample with replacement at the level of blocks, preserving the structure of the families within each block. The matrices of interest were then estimated for each bootstrap sample and subjected to PCA. Eigenvalue quantiles were estimated from the distribution of 1000 bootstraps. To reduce run times, we fit a model without the vial variances.

TABLE 1. Dimensionalities of simulated data. Each entry in the table corresponds to a simulated dataset that was bootstrapped 100 times. The number of the eigenvector where > 95% of the bootstrapped estimates are positive is the dimension of that dataset. The numbers of simulated datasets in each row are unequal due to the long run times of these simulations, plus the fact that each was run on a different processor.

Expected eigenvalue[1]		Inferred dimensionality		
9th eigenvector	10th eigenvector	8	9	10
0.00530	0.00140	0	2	21
0.00530	0.00042	0	24	2
0.00530	0.00014	0	27	0
0.00530	0.00004	0	22	0
0.00530	0.00001	0	24	0
0.00530	0.0	0	18	0
0.00176	0.00042	16	12	0
0.00053	0.00014	12	2	0
0.00017	0.00004	13	1	0
0.0	0.0	23	2	0

[1] Expressed as a proportion of the expected variance accounted for by that eigenvector.

This should affect the results very little, as the vial variances were in most cases not significantly different from zero (see Results). Our bootstrap estimator of rank is the number of eigenvalues where the lower 5% quantile of bootstrap estimates was greater than zero.

To assess the performance of our bootstrap estimator, we analyzed simulated data. We simulated phenotypic values, *z*, according to the following model:

$$z_{i.j.k} = \mu + b_i + g_{i.j} + e_{i.j.k}, \quad (3)$$

where μ is the population mean; *b*, *g*, and *e* are the block, genotypic, and environmental effects. To provide a relevant comparison to our empirical results, we used the covariance matrices at the block (**B**), additive genetic (**G**), and error (**E**) levels estimated from our female data for the first five landmarks (providing 10 traits) as parameters for these simulations. Each matrix was decomposed into eigenvectors and eigenvalues. At each level (**B, G, E**), 10 normal deviates were drawn with the eigenvalues as variances, then this vector rotated back into the phenotypic space using the appropriate eigenvectors to obtain the vector of observations. To generate **G**-matrices of less than full rank, the 10th or ninth and 10th eigenvalues of the **G**-matrix were reduced or set equal to zero. For the simulations reported, we simulated a half-sib design with 40 blocks, four sires per block, four dams per sire, and, four individuals per dam. Variance components were estimated in Proc Mixed with block as a fixed effect. Each simulated dataset was then bootstrapped 100 times at the level of blocks, and the rank was estimated as described above.

The results of these simulations are shown in Table 1. The first line of Table 1 shows results where the full **G**-matrix was simulated. For this parameter set, the simulated datasets were large enough that our power to detect variation in all 10 dimensions was quite high. Other entries in the table show cases where the 10th or ninth and 10th eigenvalues of the **G**-matrix were reduced or eliminated. As the magnitude of the smallest eigenvalues decreases, the bootstrap rank estimates also drop, even when there was in fact some variation present. When two eigenvalues are set to zero, two of our 25 bootstraps showed significant variation in the ninth eigenvector. This is a bit higher than the expected number based on the 5% criterion used. Further analysis shows that these are cases where the ninth eigenvalue was estimated to be large and positive. Figure 2 shows that the proportion of bootstrap estimates with positive eigenvalues is essentially determined by the size of the estimate of the corresponding eigenvalue in the full analysis of each dataset, rather than the parameter value used to simulate the data. When the bootstrap estimates are compared directly with the highest ranked positive eigenvalue from the full analysis of each simulated dataset, none of the bootstrap estimates were higher than the best estimate and 39% were one or two ranks lower.

Rejected Methods for Estimating Matrix Rank

We also assessed the performance of other methods for estimating matrix rank on simulated data (results not shown). These methods performed poorly, either dramatically over- or underestimating the true matrix rank. A number of heuristic methods have been suggested for estimating the number of eigenvalues that are not reflecting variation that is entirely due to error (Jackson 1993). Of these, the only method with any formal statistical justification is Bartlett's test (Lawley 1956). This is a test of eigenvalue evenness and assumes a model of independent and equal errors in the estimation of each landmark dimension, which is incorrect for our dataset. There is no biological reason that eigenvalues cannot be more even than random, and therefore Bartlett's test does not test an interesting null hypothesis.

A second method is to test for significant mean sire family variation along each of the PCs of the **G**-matrices using separate ANOVAs adjusted for multiple comparisons. Those PCs for which the hypothesis of no significant sire variation can be rejected indicate evidence for additive genetic variation. The number of these PCs would therefore reflect the rank of **G**. Analyses of simulated data indicated that this approach tends to detect variation in all eigenvectors, even when the true matrix is not of full rank. This is presumably because of errors in the estimated direction of the PCs of **G**.

We also assessed the two methods proposed by Kirkpatrick et al. (1990, their appendix C). Both use conventional estimates of estimation error in the elements of the **G**-matrix to test for significant variation in each dimension. For a simulated **G**-matrix with a true rank of 20 and a distribution of eigenvalues close to that we observed, both of these methods generally showed fewer than 10 eigenvalues with confidence intervals that did not overlap zero, dramatically underestimating rank. One potential source of the bias we observed is that in expectation eigenvalues are constrained to be nonnegative and so have an asymmetrical distribution that the Kirkpatrick approaches do not take into account.

RESULTS

Basic Information

The final dataset consisted of 175 half-sib families and 790 full-sib families. Some parents were lost or their wings damaged before they could be measured, so wing data was ob-

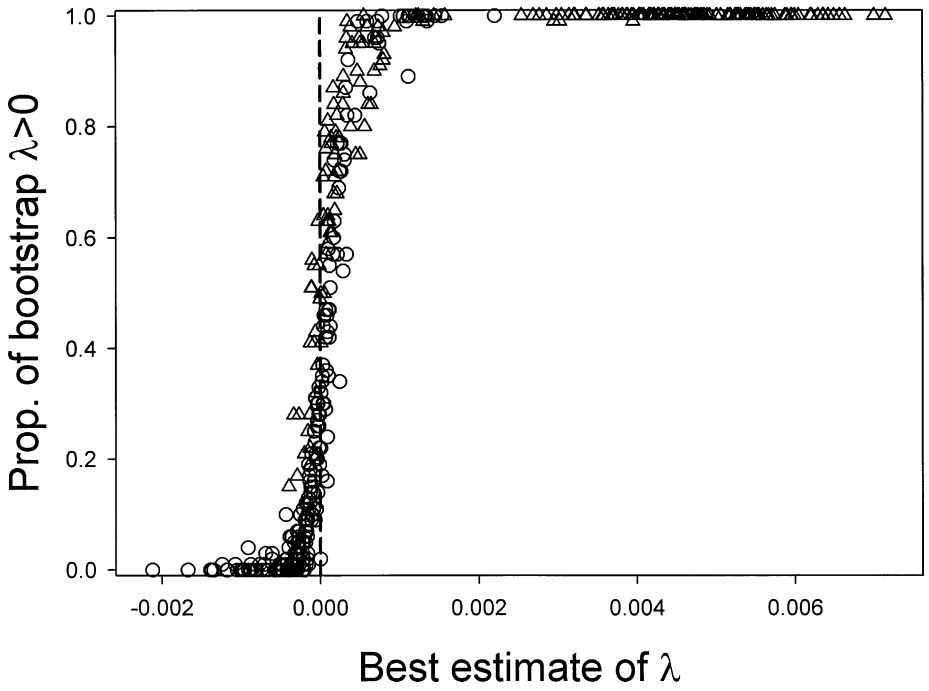

FIG. 2. Proportion of bootstrap estimates with positive eigenvalues (λ) of a given rank, as a function of the estimate of corresponding eigenvalue for the full simulated dataset, expressed as the proportion of variance accounted for by that eigenvector. Circles, λ for the 10th eigenvector; triangles, λ for the ninth eigenvector.

tained for 149 sires and 567 dams. A total of 8254 male and 8361 female offspring were measured.

Parents were smaller and more variable in size than offspring. Centroid size of male parents averaged 2.036 mm (SD = 0.071) and male offspring averaged 2.077 mm (SD = 0.056), which is a highly significant difference. Female parents averaged 2.355 mm (SD = 0.091) and female offspring averaged 2.411 mm (SD = 0.061), again a highly significant difference in mean size. This difference in size is probably due to the different rearing conditions of parents and offspring. Females are also significantly larger than males.

The parent sample had slightly and significantly higher variance overall than the offspring sample, suggesting that genetic parameters might need to be adjusted to reflect these differences. The ratios of parental to offspring variances, symbolized *k,* is shown in the online supplemental material (Table S1; available online only at http://dx.doi.org/10.1554/04-491.1.s1). The mean *k* for coordinate data is 1.13 for females and 1.05 for males. Regressions of offspring means on parents were significantly different from zero ($P < 0.0001$) for each of the 24 coordinates and centroid size (see online Table S1). We calculated heritabilities using the conservative assumption that both additive and phenotypic variances in parents are *k* times the corresponding variances of offspring. In this case, the regression estimates $h^2/(2\sqrt{k})$. The average heritability of coordinate position was 0.54 for females and 0.51 for males.

These high heritabilities are not explained by an allometric relationship between coordinates and size. The average R^2 value for the regression of the family means of each coordinate on parent centroid size is only 2.2% (median 1.1%), with a maximum of 9.8%, suggesting that most variation is independent of size. Similarly, at the phenotypic level the regression of coordinates on individual size within the offspring generation explains an average of 1.6% (median = 0.6%) of the phenotypic variation within sex.

Choosing a Causal Model to Fit to the Data

Supplemental Tables S2 and S3 (available online only at http:dx.doi.org/10.1559/04-491.1.s1) show the observational components of variance. Most observational components are much larger than their standard errors. The exception is vial variances, which are generally small and rarely significantly different from zero. The error variances for the coordinate data are quite similar between the sexes for most traits. Overall, the average ratio of female to male error variance is 0.95. A discrepancy of this magnitude is expected if there are X-linked additive effects, as a larger proportion of the male X variance is found in the error term, compared to female X variance. To examine whether the differences in error variances are significant, we fit a causal model by least squares that constrained the X-linked (the male contribution scaled for dosage compensation; Cowley et al. 1986) and additive variances to be the same in the two sexes, but fit different parameters for the error and maternal variances. Significantly better fit was obtained with different error variances in just six of the 24 coordinates, the *x* and *y* for points 5, 6, and 11. For point 5, the error variance was significantly higher for females, while it was significantly lower in females for points 6 and 11. Overall, we regard an assumption of equal error variances in the sexes over landmarks to be adequate, and use this to obtain estimates of the causal variances below.

TABLE 2. Causal components of variance. *P*-values shown for each term were obtained from the change in sums of squares when each term is omitted from the model. The error variance was significant at $P < 0.0001$ in every case. $V_{P.f}$ and $V_{P.m}$ are the sex-specific phenotypic variances, calculated as the sums of the appropriate sex-specific additive terms, plus V_M, V_V, and V_E. Units for coordinates are centroid size $\times 10^4$, while for centroid size the unit is log(mm $\times 10^3$).

Trait	$V_{Aa.f}$	$V_{Aa.m}$	$V_{Ax.f}$	$V_{Ax.m}$	V_M	V_V	V_E	$V_{P.f}$	$V_{P.m}$	Model fit[1]
X1	2907****	3224****	512	185	96	39	860	4413	4403	ns
Y1	1204****	1091****	52	282*	61	26*	735	2079	2196	ns
X2	978****	1027****	99	190	67	1	348	1492	1632	ns
Y2	428****	655****	138	15	46*	18**	408	1037	1142	ns
X3	1030****	962****	43	359**	52	19*	425	1568	1816	ns
Y3	514****	690****	156	91	29	30****	433	1161	1273	ns
X4	4400****	4574****	0	968*	80	25	1697	6202	7344	ns
Y4	1040****	1128****	100	96	36	49****	534	1760	1843	ns
X5	889****	400****	0	206	0	44	1040	1972	1690	***
Y5	472****	273****	0	47	0	25	713	1210	1058	****
X6	381***	564****	134	20	52*	32	468	1067	1136	ns
Y6	265***	320****	70	72	14	21****	416	785	842	**
X7	2919****	2557****	238	163	149	0	941	4248	3811	ns
Y7	910****	983****	126	100	11	22**	440	1509	1555	ns
X8	2381****	2516****	327	279	80	0	1407	4194	4282	ns
Y8	259****	242****	13	47	7	13****	250	542	559	ns
X9	1774****	1698****	43	0	44	16	710	2587	2468	ns
Y9	38	125****	50**	4	5	2	144	240	280	*
X10	2066****	1846****	16	26	24	32*	735	2874	2664	ns
Y10	175****	149****	16	31	0	5**	131	327	316	ns
X11	307***	407****	68	69	24	32****	549	980	1081	*
Y11	102	188****	45	0	11	3	246	406	448	****
X12	350**	526****	159	122	42	39***	818	1408	1547	ns
Y12	141**	182****	46	50	8	19	348	563	608	**
Log(size)	236****	351****	0	37	23	32****	220	511	663	ns

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.
[1] Significance of the lack of fit of the causal model.

Centroid size data was significantly better fit by a model that allowed sex-specific error variances, with the female error variance estimated as a substantial 1.39 times that in males. After log-transformation, the female and male error variances were within 0.4% of each other, and fitting a model with separate error variances no longer significantly increased the fit of the model.

If X-linked effects are present, we expect that the female parent-offspring covariance will be higher than that for males, as fathers and sons do not share an X chromosome, while mothers and daughters do. On average this effect is borne out, as the female have on average a 19% higher parent-offspring covariance than males. Similarly, paternal half-brothers do not share an X, while paternal half-sisters do, again leading to the expectation that the female half-sib variance will exceed that of the male half-sib variance. This is again borne out, with female components exceeding the male component by an average of 13%. Conversely, the dam components of females are smaller by 9%, as expected based on the same reasoning. These findings all justify fitting separate X-linked additive effects when obtaining the causal components of variation.

As outlined above, phenotypic variances for most traits were higher in the parental generation than in the offspring generation. If this were due to general increase in all sources of phenotypic variation, then the parent-offspring covariances would be overestimated by a factor $\sqrt{k}$, where k is the factor by which parental variances exceed offspring variances. Alternatively, the excess variation in the parents could be entirely due to nongenetic causes, and no correction would be appropriate. We fit the causal model with and without variance corrections. In most cases, the fit of the model was worse with the variance correction, suggesting that the second explanation may be closer to the truth. This was particularly striking for log centroid size, which had the biggest departures of k from one. Correction of parent-offspring covariance actually increased the residual sum of squares from 4.54 to 12.06. Consequently, no corrections for differences in variance were applied.

Causal Components

The causal components estimated are shown in Table 2. In most cases the causal model fit well. Significant lack of fit was usually associated with traits identified above as having unequal error variances in the two sexes. The exception is coordinate Y12, where the female parent-offspring covariance was the major contributor to the error variance.

In all cases, the male additive autosomal component was highly significantly different from zero. Twenty-two of the 24 female additive autosomal components were also significantly different from zero. Contrary to the pattern of differences in sex-specific terms discussed above, additive X-linked terms were rarely significant. This likely indicates insufficient power to detect terms of the magnitude found. Vial variances were sometimes significant. The general lack of significant V_M terms means that there is no compelling evidence for any maternal or dominance effects on these traits.

We also calculated the overall additive genetic variance in

TABLE 3. Causal components of variance as percentages of phenotypic variance. The h^2 estimate is based on the sum of autosomal and X-linked additive variation. The maternal, vial, and error variances are expressed as a percentage of the total variance in females.

	Female		Male				
Trait	h^2	% X-linked variation	h^2	% X-linked variation	Maternal	Vial	Error
X1	77	15	77	5	2	1	19
Y1	60	4	63	21	3	1	35
X2	72	9	75	16	4	0	23
Y2	55	24	59	2	4	2	39
X3	68	4	73	27	3	1	27
Y3	58	23	61	12	2	3	37
X4	71	0	75	17	1	0	27
Y4	65	9	66	8	2	3	30
X5	45	0	36	34	0	2	53
Y5	39	0	30	15	0	2	59
X6	48	26	51	3	5	3	44
Y6	43	21	47	18	2	3	53
X7	74	8	71	6	4	0	22
Y7	69	12	70	9	1	1	29
X8	65	12	65	10	2	0	34
Y8	50	5	52	16	1	2	46
X9	70	2	69	0	2	1	27
Y9	37	57	46	3	2	1	60
X10	72	1	70	1	1	1	26
Y10	58	8	57	17	0	2	40
X11	38	18	44	14	2	3	56
Y11	36	30	42	0	3	1	61
X12	36	31	42	19	3	3	58
Y12	33	25	38	22	1	3	62
Average	57	14	58	12	2	2	39
Log(size)	46	0	59	10	5	6	43

males ($V_{A\cdot m}$) and females ($V_{A\cdot f}$) as the sum of the autosomal and X effects. Table 3 summarizes the relative importance of the various components of variation using this sum as a baseline. On average, the heritabilities of each trait are quite high, and very similar in the sexes. The X-linked components are somewhat less than expected given that the X chromosome makes up 20% of the *D. melanogaster* genome.

In addition to the analyses shown in Tables 2 and 3, we also estimated all of the covariance components at each level of the design and used these to estimate the causal variance-covariance matrices. Of these, the **G**-(combined X and autosomal) matrices for males and females, **E**- and **P**-matrices are available as online supplementary Tables S4–S7 (available online only at http:dx.doi.org/10.1554/04-491.1.s1).

G-Matrix Rank

Our estimates of the rank of the causal component matrices for both landmark coordinate data and a distance dataset are shown in Table 4. Although we have coordinate data from 12 landmarks, and therefore 24 variates, four degrees of freedom are lost during the alignment process. Thus, the maximum expected rank for each landmark- or distance-based matrix is 20. Results from the full coordinate analysis suggest that there is significant additive genetic variation in all 20 possible dimensions of wing shape. For a few matrices, there were more than 20 positive eigenvalues, which presumably reflects numerical inaccuracies in the estimation procedures.

Table 4 also shows the median, 5%, and 1% quantiles for

TABLE 4. Dimensionality of each matrix of causal components. Bootstrap results based on analysis of 1023 resampled datasets.

	Coordinate data					
			Bootstrap			
Matrix	Full analysis	No vial	Median	5%	1%	Distance data
$V_{Aa.f}$	16	17	15	14	13	16
$V_{Aa.m}$	20	20	18	16	16	20
$V_{Ax.f}$	13	13	11	8	7	12
$V_{Ax.m}$	12	12	11	9	7	13
$V_{A.f}$	20	19	19	17	17	19
$V_{A.m}$	21	21	20	18	17	20
V_M	12	16	14	11	10	17
V_V	16	—	—	—	—	—
V_E	23	23	22	21	20	25

rank over the bootstrap results. The lower confidence intervals on matrix ranks suggest that we have high confidence that the additive genetic variance fills out at least 17 of the 20 possible dimensions in phenotype space in female flies and 18 dimensions in males.

The algorithm used to align the landmark configuration potentially introduces its own pattern of covariance into the data (Rohlf and Slice 1990; Walker 2000). This might affect our estimates of rank. Therefore, we conducted an alternative analysis based on the distances between aligned landmarks. We calculated a total of 27 distances between nearby points using the aligned landmark data. Each landmark was involved in the calculation of at least three distances. The causal matrices were then estimated for this dataset and the dimensionality calculated from the eigenvalues as with the coordinate data. The dimensionalities are very similar, although cases where the dimensionality is both higher and lower exist. When the genetic dimensionalities do differ, there is no more than one rank difference between the estimates. Note that the maximum rank is the same for both analyses, as they are both based on the same data.

The distribution of $\log_{10}$ eigenvalues in the total additive genetic variance matrix is shown in Figure 3. The agreement between the bootstrap medians and the estimates from the full analysis for coordinate data is very close up to approximately eigenvector 15. For the last eigenvectors, the full estimates exceed the bootstrap estimates by an increasing amount. The distribution of eigenvalues from the distance-based matrices is similar to that for coordinate-based analyses, although variation is slightly more concentrated in the earlier vectors.

Figure 3 shows that the logarithms of the eigenvalues decrease at a rate that is close to constant, suggesting an exponential distribution of variation across phenotype space. Regression of logrithms of the first 20 eigenvalues on vector number for the total **G**-matrix for the full coordinate analysis including vial effects was highly significant in each sex, with a slope of -0.171 ± 0.012 in females (explaining 92% of the variation) and -0.133 ± 0.005 males (explaining 97.7% of the variation). Figure 3 shows that the relationship is even more precise for females over the first 19 eigenvalues, yielding a slope of -0.151 ± 0.004 in females (explaining 98.9% of the variation). The corresponding slope for males is -0.130 ± 0.005 (explaining 97.6% of the variation). These

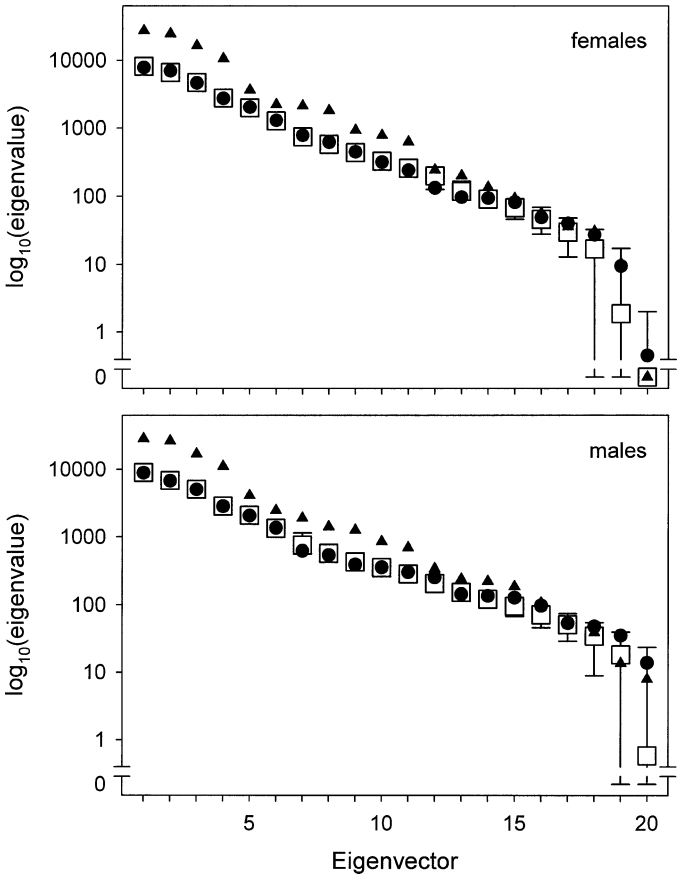

FIG. 3. Distribution of eigenvalues in the total additive genetic covariance matrices for females (above) and males (below). Solid circles are the eigenvalues from the full analysis, including vial effects. The open squares show the median values from the bootstrap estimates, while the upper and lower limits denote the 95th and fifth percentiles of the bootstrap results. Solid triangles are estimates from the distance-based analysis.

last slopes indicate that each successive eigenvalue decreases by about 29% in females and 26% in males. The rate of decrease in the distance analysis is slightly higher: -0.183 ± 0.005 for the first 19 eigenvalues in females (explaining 99.0% of the variation) and -0.175 ± 0.006 in males (explaining 98.2% of the variation). This is a decrease of about 34% per dimension. These are both remarkably shallow rates of decrease compared to other eigenvalue distributions for morphological data. For example, Kirkpatrick and Lofsvold (1992) showed distributions of eigenvalues that decline by three orders of magnitude over about four vectors, while ours declined that much over approximately 20 vectors.

DISCUSSION

The main result of our analyses is that we were able to detect additive genetic variation with high confidence in almost every dimension of form that we measured. We first discuss some other aspects of our results before returning to this fundamental result below.

A striking aspect of our results is that there is little covariance between the aligned landmark data and centroid size, indicating little allometric variation in shape. The lack of allometric shape variation in our study contrasts with strong allometries found when the environment is manipulated to create variance in size (Weber 1990). There is also strong allometry for wing shape among species in the family Drosophilidae (D. Houle and K. van der Linde, unpubl. data). Because wing size is a reasonable proxy for body size (Reeve and Robertson 1953; David et al. 1977), these results suggest that allometry in these other settings is the consequence of joint selection on size and shape, rather than of developmental constraints.

Different landmarks in our study show very different levels of phenotypic and genetic variation, with the most variable point, landmark 1, showing about four times the variation in the least variable, landmark 11 (see supplementary online material). It is possible that these patterns are biologically significant, but they are also to some extent a function of the algorithm used to superimpose the point configurations in the first step of our analyses. The algorithm we used would behave well if the variation in point locations were homogeneous and independent, an assumption that is clearly incorrect for our data. The mean landmark configuration itself influences the weight assigned to variation at each point. Thus, we have not necessarily recovered the variance-covariance pattern that best reflects the biological processes that gave rise to the variation (Rohlf and Slice 1990; Walker 2000). The relative variation of points from this and all geometric morphometric analyses must be interpreted cautiously.

Previous studies have indicated that X chromosome effects explained about 20% of the genetic variation in morphology in *D. melanogaster* (Cowley et al. 1986; Cowley and Atchley 1988), as expected from the proportion of genes found on the X chromosome. The relative variation in males and females also was accurately predicted by an assumption of perfect dosage compensation in males relative to females. Our data showed several patterns consistent with the importance of X effects, such as larger covariance between female parents and offspring than males, and between female half-sibs than for males, and larger error variances in males relative to females. Given these results, we fit a causal model with male and female additive X-linked effects assuming complete dosage compensation. We rarely detected significant X-linked additive variation (Table 2). This probably reflects insufficient power as the average size of the X effect was a substantial 13% of the total additive variance (Table 3). This is somewhat lower than the proportion of genes on the X chromosome. It is interesting to note that two QTL studies of wing shape (Zimmerman et al. 2000; Mezey et al. 2005) found a noticeable paucity of QTL on the X chromosome. This result could be explained by lower wing-shape gene density on the X chromosome and therefore, indirectly, lower power to detect QTL, since the combined effects of genes with small effects producing a QTL signal may be expected to be lower (Noor et al. 2001). However, the smaller-than-expected amount of variation attributable to the X chromosome in our study suggests that the small X-effect may be a biological phenomenon.

Our results show a consistently high heritability of landmark displacements, between 33% and 77% for all traits. All traits in both sexes showed significant additive genetic variation in either the parent-offspring or the full analysis. Con-

versely, the amount of variation that could indicate either maternal effects or dominance effects (V_M) was uniformly very low. Estimates of V_M only differed significantly from zero in two of the 25 traits.

Dimensionality of Additive Genetic Variation

We demonstrated that the rank of the combined **G**-matrices (including both autosomal and X-linked effects) was near the maximum possible value. Our best estimates of rank are 20 of 20 dimensions in both sexes, with additional variation in size that is largely independent of shape. Our bootstrap estimates of confidence limits to **G**-matrix rank are slightly lower: 17 for females and 18 for males. Thus, we are quite confident that additive genetic variation exists for almost all aspects of wing shape that we assessed. The gradual exponential decrease in eigenvalue with eigenvector number suggests that we may have run out of power to detect higher rank, rather than there being an absence of variation. These results do not support the notion that standing genetic variation constrains the ability of wing shape to evolve in this population.

Our bootstrap estimates of rank may be somewhat conservative. The simulations that support the use of bootstrapping are of the ideal case where the distribution of breeding values is Gaussian. In a natural population breeding values may well be non-Gaussian such that the sampling of particular families may have a larger impact on the rank than is true in our simulations. Bootstrap samples will on average only contain $1 - (1/e) = 0.63$ of the blocks in the original dataset, and therefore only 63% of the sire families. In a real dataset, the missing families may be the ones that fill out the variation in a particular dimension of phenotype space, reducing overall rank. The lower bootstrap estimates of eigenvalues of rank > 15 may reflect such a bias.

Our high estimate of dimensionality is consistent with results of artificial selection experiments on wing shape. Weber (1990, 1992) applied selection to seven different ratios of distances between landmarks, almost all of which were also used in the current study. The direction of selection was in each case orthogonal to the strong allometric relationships between the distances, that is, in a direction away from the most variable aspect of shape. He found a localized response in the selected region, with relatively small changes in other traits of the wing. Houle et al. (2003) also obtained large responses in a complex wing-shape index. The high rank of **G** found here suggest that these results are not due to a fortunate choice of traits, but would likely have obtained some response along most dimensions in phenotype space. These findings are particularly interesting given the very conservative nature of wing shape in the family Drosophilidae (Hansen and Houle 2003; Houle et al. 2003;). The lack of evidence for bidirectional absolute constraints instead implicates selection as the primary determinant of wing shapes across the family.

The potential for such fine-grained, high-dimensional change is only expected when many genes with distinct effects control wing development. QTL studies have supported this supposition, as many QTLs have been detected in recent analyses of wing shape (Weber et al. 1999, 2001; Zimmerman et al. 2000; Mezey et al. 2005). In each of the three populations studied more than 20 distinct QTL affecting aspects of wing shape were detected. It is important, however, to note that the large number of genes with wing-shape effects do not directly imply a low level of absolute constraint. There must not only be a large number of genes but the genes must also alter wing shape in distinct ways. Directly estimating the level of bidirectional absolute constraint is difficult with QTL approaches because reasonable power to identify QTL requires genes of large effect or a large combined effect of many genes on a localized section of a chromosome. For the purposes of estimating bidirectional constraints, the current approach is far more powerful.

The high genetic dimensionality of wing shape contrasts with the results of Kirkpatrick and Lofsvold (1992), who could only demonstrate genetic variation in a few aspects of developmental trajectories in mice, sheep, and chickens. Several key differences between the studies might explain the differences. First, each of the trajectories analyzed in the Kirkpatrick and Lofsvold study consisted of the size of a single trait at various times during growth. Such trajectories may be less free to vary. For example, the animals studied are probably not capable of becoming smaller during development. Second, Kirkpatrick and Lofsvold analyzed data from vertebrates, which may differ from *Drosophila* in the level of constraint, for example due to differences in development or to smaller population size. Finally, our analyses of simulated data suggest that the methods used by Kirkpatrick and Lofsvold (Kirkpatrick et al. 1990) can drastically underestimate rank.

We have demonstrated that fly wings have a higher number of heritable aspects of form than any previously studied set of features. This result implies that genetic variation expressed in the multiple genetic pathways involved in the development of the wing imaginal disc can cause distinct wing-shape phenotypes. An open question is whether there are any bidirectional constraints on wing shape. A more pragmatic way to phrase this question is: How many aspects of wing shape would we have to consider before we are able to find convincing evidence of constraint? Studies that analyze even more aspects of the wing, with accompanying increases in the scale of the study to increase power, could answer this question.

A related question is at what point does variation in a dimension become so small that they are effectively an absolute constraint? This question is difficult to answer because it depends on the time scale over which directional selection is applied to these least variable traits. If we expect these dimensions to be under directional selection for short periods, these may be effectively absolute constraints. Conversely, long periods of consistent directional selection may change the population mean even when the amount of additive genetic variation is never very high. Another possibility is that initially rare alleles might rise in frequency under such selection regimes, increasing variation along the selected directions. Selection on traits that are among the least variable may reveal much about the importance of genetic constraints on evolution.

An important limitation to a variation-based study of constraints is that variation can itself evolve. In a population

where the mean phenotype is at equilibrium, genetic variation can still change, decreasing due to drift and perhaps selection, and increasing due to mutation and gene flow. If the phenotype is subject to directional selection, variation might rapidly be exhausted, especially in aspects of form with little genetic variation. Ultimately, the capability of mutation to replenish variation in the direction of selection will determine whether the phenotype will be constrained. At an even longer time scale, epistatic alleles could either decrease or increase the variation available for evolution (Hansen and Houle 2003). If interacting alleles tend to mask genetic variation in some directions, genetic canalization might evolve. Conversely, decanalization is also possible.

A second limitation is that our approach cannot deal with qualitatively novel phenotypes, such as gain or loss of wing veins. A very small proportion of the individuals in our study showed such qualitatitve changes, and these were discarded from our final sample. Our results therefore apply within the phenotypic neighborhood of actual *Drosophila* wings and not to the evolution of phenotypic dimension itself.

If, despite these caveats, we accept the fundamental result of our study—that additive genetic variation exists for essentially all phenotypic dimensions—no wing shape would be beyond the reach of evolution. Thus, selection must have been the primary architect of wing shapes observed across the Drosophilidae, despite the fact that they vary little (Houle et al. 2003). To the extent that phenotypes in other organisms also show this pattern of variation, constraint-based explanations of macroevolutionary patterns would need to be reevaluated. Additional studies of the dimensionality of genetic variation are needed.

Acknowledgments

We thank J. Fry and two anonymous reviewers for their comments, J. Birdsley for managing the data collection and editing; and F. Smyth, Y. Ng, L. Carpenter, C. Emery, and W. Feng for imaging wings and editing wing images. National Institutes of Health grant RO1-GM61773-01 supported JGM during the final stages of preparing this manuscript. This research was supported by National Science Foundation grant DEB-0129219.

Literature Cited

Alberch, P. 1985. Developmental constraints: why St Bernards often have an extra digit and poodles never do. Am. Nat. 126:430–433.

Alberch, P., and E. A. Gale. 1983. A developmental analysis of an evolutionary trend: digital reduction in amphibians. Evolution 39:8–23.

Arnold, S. J. 1992. Constraints on phenotypic evolution. Am. Nat. 140:S85–S107.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9–32.

Atchley, W. R., J. J. Rutledge, and D. E. Cowley. 1981. Genetic components of size and shape. II. Multivariate covariance patterns in the rat and mouse skull. Evolution 35:1037–1055.

Bailey, D. W. 1956. A comparison of genetic and environmental principal components of morphogenesis in mice. Growth 20: 63–74.

Bier, E. 2000. Drawing lines in the *Drosophila* wing: initiation of wing vein development. Curr. Opin. Genet. Dev. 10:393–398.

Birdsall, K., E. Zimmerman, K. Teeter, and G. Gibson. 2000. Genetic variation for the positioning of wing veins in *Drosophila melanogaster*. Evol. Dev. 2:16–24.

Bradshaw, A. D. 1991. The Croonian lecture, 1991: genostasis and the limits to evolution. Philos. Trans. R. Soc. Lond. B 333: 289–305.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Univ. of Chicago Press, Chicago.

Cavicchi, S., G. Giorgi, V. Natali, and D. Guerra. 1991. Temperature-related divergence in experimental populations of *Drosophila melanogaster*. 4. Fourier and centroid analysis of wing shape and relationship between shape variation and fitness. J. Evol. Biol. 4:141–159.

Cook, O., B. Biehs, and E. Bier. 2004. *brinker* and *optomotor-blind* act coordinately to initiate development of the L5 wing vein primordium in *Drosophila*. Development 131:2113–2124.

Cowley, D. E., and W. R. Atchley. 1988. Quantitative genetics of *Drosophila melanogaster*. II. Heritabilities and genetic correlations between sexes for head and thorax traits. Genetics 119: 421–433.

Cowley, D. E., W. R. Atchley, and J. J. Rutledge. 1986. Quantitative genetics of *Drosophila melanogaster*. I. Sexual dimorphism in genetic parameters for wing traits. Genetics 114:549–566.

David, J., C. Bocquet, and M. de Scheemaeker-Louis. 1977. Genetic latitudinal adaptation of *Drosophila melanogaster*: new discriminative biometrical traits between European and equatorial African populations. Genet. Res. 30:247–255.

de Jong, G., and A. J. van Noordwijk. 1992. Acquisition and allocation of resources: genetic (co)variances, selection, and life histories. Am. Nat. 139:749–770.

Falconer, D. S., and T. F. C. Mackay. 1996. Quantitative genetics. Longman Group, Essex, U.K.

Fry, J. D. 2004. Estimation of genetic variances and covariances by restricted maximum likelihood using PROC MIXED. Pp. 11–34 *in* A. Saxton, ed. Genetic analysis of complex traits with SAS. SAS Institute, Cary, NC.

Gilchrist, A. S., R. B. R. Azevedo, L. Partridge, and P. O'Higgins. 2000. Adaptation and constraint in the evolution of *Drosophila melanogaster*. Evol. Dev. 2:114–124.

Gould, S. J., and R. C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond. B 205:581–598.

Hansen, T. F., and D. Houle. 2003. 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 Univ. Press, Oxford, U.K.

Held, L. I., Jr. 2002. Imaginal discs: the genetic and cellular logic of pattern formation. Cambridge Univ. Press, Cambridge, U.K.

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

———. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–204.

———. 2001. Characters as the units of evolutionary change. Pp. 109–140 *in* G. P. Wagner, ed. The character concept in evolutionary biology. Academic Press, New York.

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

Imasheva, A. G., O. A. Bubli, O. E. Lazebny, and L. A. Zhivotovsky. 1995. Geographic differentiation in wing shape in *Drosophila melanogaster*. Genetica 96:303–306.

Insightful Corporation. 2001. S-Plus 6.0, professional release 1. Insightful; Seattle, W.A.

Jackson, D. A. 1993. Stopping rules in principal components-analysis: a comparison of heuristic and statistical approaches. Ecology 74:2204–2214.

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

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. Genetics 124:979–993.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

———. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33: 402–416.
———. 1982. A quantitative genetic theory of life history evolution. Ecology 63:607–615.
Lawley, D. N. 1956. Tests of significance for the latent roots of covariance and correlation matrices. Biometrika 43:128–136.
Leamy, L. 1977. Genetic and environmental correlations of morphometric traits in randombred house mice. Evolution 31: 357–369.
Lu, F., and E. E. Milios. 1994. Optimal spline fitting to planar shape. Signal Processing 37:129–140.
Lunde, K., J. L. Trimble, A. Guichard, K. A. Guss, U. Nauber, and E. Bier. 2003. Activation of the *knirps* locus links patterning to morphogenesis of the second wing vein in *Drosophila*. Development 130:235–246.
Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.
Maynard Smith, J., R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. 1985. Developmental constraints and evolution. Q. Rev. Biol. 60: 265–287.
Mezey, J. G., D. Houle, and S. Nuzhdin. 2005. Naturally segregating QTL for *Drosophila* wing shape. Genetics *In press*.
Mohler, J., M. Seecoomar, S. Agarwal, E. Bier, and J. Hsai. 2000. Activation of *knot* (*kn*) specifies the 3–4 intervein region in the *Drosophila* wing. Development 127:55–63.
Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. Heredity 59:181–197.
Noor, M. A. F., A. L. Cunningham, and J. C. Larkin. 2001. Consequences of recombination rate variation on quantitative trait locus mapping studies: simulations based on the *Drosophila melanogaster* genome. Genetics 159:581–588.
Phillips, P. C., and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution 43:1209–1222.
Price, T., M. Turelli, and M. Slatkin. 1993. Peak shifts produced by correlated response to selection. Evolution 47:280–290.
Reeve, E. C., and F. W. Robertson. 1953. Studies in quantitative inheritance. II. Analysis of a strain of *Drosophila melanogaster* selected for long wings. J. Genet. 51:276–316.
Roff, D. A., and T. A. Mousseau. 1987. Quantitative genetics and fitness: lessons from *Drosophila*. Heredity 58:103–118.
———. 1999. Does natural selection alter genetic architecture? An evaluation of quantitative genetic variation among populations of *Allonemobius socius* and *A. fasciatus*. J. Evol. Biol. 12: 361–369.
Rohlf, F. J. 1998a. tpsDig: digitizing software. Ver. 1.17. Dept. of Ecology and Evolution, State Univ. of New York, Stony Brook, NY. Available via http://life.bio.sunysb.edu/morph/.
———. 1998b. tpsRegr: digitizing software. Ver. 1.17. Dept. of Ecology and Evolution, State Univ. of New York, Stony Brook, NY. Available via http://life.bio.sunysb.edu/morph/.
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.
Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol. Evol. 17:320–327.
Strang, G. 1988. Linear algebra and its applications. Saunders College Publishing, New York.
Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39:505–522.
Walker, J. A. 2000. Ability of geometric morphometric methods to estimate a known covariance matrix. Syst. Biol. 49:686–696.
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. E. 1990. Selection on wing allometry in *Drosophila melanogaster*. Genetics 126:975–989.
———. 1992. How small are the smallest selectable domains of form? Genetics 130:345–353.
Whitlock, M. C., and K. Fowler. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics 152:345–353.
Zeng, Z.-B. 1988. Long-term correlated response, interpopulation covariation, and interspecific allometry. Evolution 42:363–374.
Zimmerman, E., A. Palsson, and G. Gibson. 2000. Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. Genetics 155:671–683.

Corresponding Editor: J. Fry