Evolution, 56(10), 2002, pp. 1968–1975

PERSISTENCE OF CHANGES IN THE GENETIC COVARIANCE MATRIX AFTER A BOTTLENECK

MICHAEL C. WHITLOCK,[1,2] PATRICK C. PHILLIPS,[3,4] AND KEVIN FOWLER[5,6]
[1]*Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada*
[2]*E-mail: whitlock@zoology.ubc.ca*
[3]*Center for Ecology and Evolutionary Biology, University of Oregon, Eugene, Oregon 97403-5289*
[4]*E-mail: pphil@darkwing.uoregon.edu*
[5]*Department of Biology, University College, London, NW1 2HE, United Kingdom*
[6]*E-mail: k.fowler@ucl.ca.uk*

Abstract.—Genetic variance, phenotypic variance, and the genetic covariance matrix (**G**) can change as a result of genetic drift. These changes will persist over time to some extent and will continue if population size remains relatively small. Nine populations founded by a single pair of *Drosophila melanogaster* were measured for a series of six morphological characteristics for a large number of parent-offspring families at both the third generation after the bottlenecks and after 20 generations. From these data, the phenotypic variance, additive genetic variance, and **G** were estimated for each line at each generation. Phenotypic and genetic variances were highly correlated over time, so that the measurements made at the third generation were predictive of the state of the population 17 generations later. Genetic covariances were also somewhat stable over time; however, the **G** matrices of some lines changed significantly over the intervening generations. This change did not return the populations toward their original state before the population bottlenecks. We conclude that the genetic covariance matrix can change as a result of mild genetic drift over a short span of time.

Key words.—Genetic covariance, genetic drift, genetic variance, **G** matrix, population bottlenecks.

Received May 6, 2002. Accepted June 25, 2002.

In modern quantitative genetic theory, the additive genetic covariance matrix (**G**) plays a central role in predicting the short-term trajectory of evolution by selection or genetic drift (Lande 1979; Falconer and Mackay 1989). This **G** matrix gives a structured list of the additive genetic variances for a set of continuously distributed traits and the additive genetic covariances among each of these traits. These characteristics are needed to predict the change in the mean values of traits due to selection, drift, or migration.

For these predictions to be useful over a period of time longer than just a few generations, the **G** matrix must stay roughly constant over time (Lande 1979; Turelli 1988). Several studies have tested whether **G** is uniform in closely related taxa, that is, whether the **G** matrix has remained constant over the evolutionary time separating the two taxa (i.e., Lofsvold 1986; Arnold and Phillips 1999 and references therein). There have been relatively few experimental investigations of the constancy of **G,** however. Wilkinson et al. (1990; see also Shaw et al. 1995) found that the **G** matrix for five morphological characters in *Drosophila melanogaster* changed during 23 generations of divergent selection. Bryant and Meffert (1988) found that **G** changed in shape following population bottlenecks in the housefly *Musca domestica*, and we have shown that **G** can change drastically as a result of population bottlenecks in *D. melanogaster* (Phillips et al. 2001). Many other studies have shown changes in the additive genetic variance during selection (Wilkinson et al. 1990; Meyer and Hill 1991; Beniwal et al. 1992a,b; Shaw et al. 1995) or after inbreeding or population bottlenecks (Frankham 1980; Bryant et al. 1986; Briscoe et al. 1992; Whitlock and Fowler 1999 and references therein). These changes in **G** after inbreeding may be transient, though, because the evolutionary forces such as selection and mutation that act to determine **G** may return it to its previous shape (although see Camara et al. 2000). Thus, we need to understand more about the persistence of changes in **G.**

Certain models of evolution depend on changes in **G** and in the phenotypic covariance matrix, **P.** A model of complex trait evolution, called the variance-induced peak shift (VIPS) model (Whitlock 1995), predicts that evolution by genetic drift on a complex adaptive landscape is facilitated by drift-induced changes in the phenotypic variance. If phenotypic variance increases sufficiently for any reason, including drift induced changes in genetic variance, then the adaptive landscape may become flatter and it becomes easier for a population to evolve to higher adaptive peaks deterministically (Whitlock 1995). This process depends on the ability of **P** to change as a result of drift (demonstrated by Whitlock and Fowler 1996; Fowler and Whitlock 1999a,b and references therein) and on the short-term stability of these changes. The changed landscape must remain in its new topography long enough for selection to take the population into the area previously occupied by the domain of attraction of the alternate adaptive peak. There are no published empirical data that address the stability of the change in **G** or **P** after population bottlenecks.

To address questions about the kinds of changes in genetic variance, phenotypic variance, and **G,** we began a large experiment using a set of wing size and shape characters in *D. melanogaster*. A set of 52 inbred lines were created by single-pair population bottlenecks and expanded to larger size. At the third generation, more than 6000 families were measured, and various results have been reported previously. The changes in the phenotypic variance (Fowler and Whitlock 1999a), genetic variance (Whitlock and Fowler 1999), and fitness (Fowler and Whitlock 1999a,b) were measured for each of the lines. The phenotypic, genetic, and environmental variances were extremely variable across the bottlenecked lines,

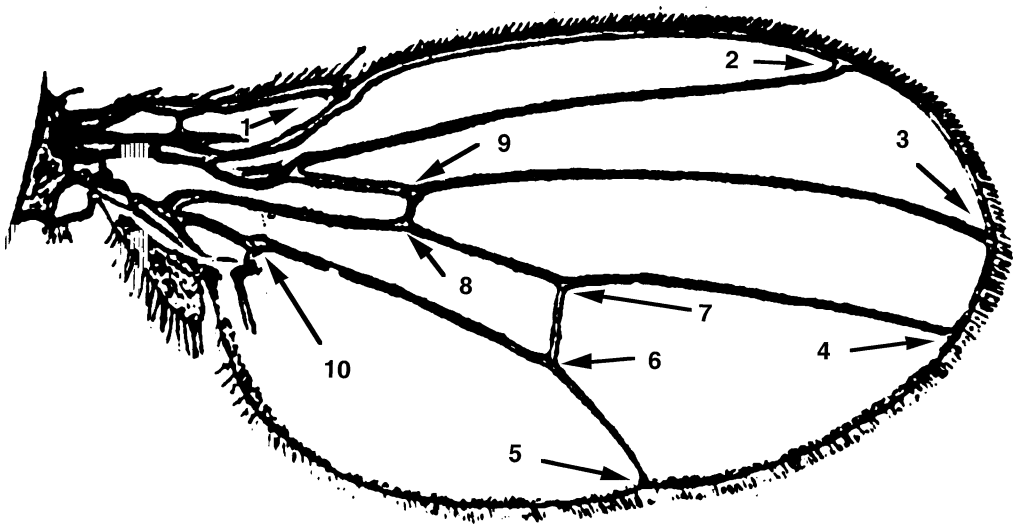

FIG. 1. The landmarks used to generate characters. Ten landmarks were measured for each wing. The measurements were consistently made from the same point of the junction of the wing veins. The characters used were wing area (the area of the polygon defined by vertices at points 1, 2, 3, 4, 5, and 10), and the angles formed by the points 5-7-4, 8-7-6, 2-9-3, 2-1-5, and 2-3-5 (with the vertex listed as the middle point).

in all cases with both increases and decreases observed. Taking the average, however, showed that the bottlenecked lines decreased in genetic variance almost exactly as predicted by classical additive theory (Wright 1951; Lande 1980; Whitlock and Fowler 1999). We further used these data to address the constancy of the **G** matrix in response to bottlenecks (Phillips et al. 2001). On average, **G** changes as predicted by theory (Lande 1979), but the **G** matrix of any given line is likely to vary in every possible way from its progenitor (Phillips et al. 2001).

A subset of these same lines was maintained for additional experiments after another 17 generations had passed. In this paper, we address the following questions: How stable are increases in genetic and phenotypic variance after population bottlenecks? How stable are the changes in **G** induced by this inbreeding? Does **G** evolve toward an equilibrium state?

MATERIALS AND METHODS

Experimental Design

The development of the bottlenecked lines has been described in detail in Fowler and Whitlock (1999a) and Whitlock and Fowler (1999) and will only briefly be recounted here. A series of randomly mated pairs from the Dahomey stock of *D. melanogaster* was used to found inbred lines, which after this one-generation bottleneck were maintained by mixing among four half-pint bottles with 18 individuals of each sex per bottle per generation for a total of 144 flies per line. Approximately 400 pairs of flies were used to create each of six control outbred populations, which were maintained with the same husbandry as the inbred lines. These lines were created in three batches, separated by approximately three months. Each of the batches had two subgroups, created on consecutive days, and each of these subgroups had a corresponding control outbred population.

At the third generation after the bottleneck (the F_3), pairs of flies from each population were randomly mated and allowed to produce offspring. All of the lines were then measured for genetic and phenotypic changes resulting from the bottleneck. After 17 more generations, in the 20th generation (the F_{20}), nine inbred lines and five of the control populations were chosen for re-examination. The nine inbred lines with characters that increased in additive genetic variance by 30% or more for any one of the characters were chosen for further analysis, and the contemporaneous controls of each of these nine were measured as well. Flies were chosen at random to produce offspring, following the same protocol as in the third generation (Whitlock and Fowler 1999). Wings from each of these parents and eight of their female offspring were mounted on microscope slides to be measured for a series of wing size and shape characters (see Fig. 1) using a computerized digitizer. On average, 60 families per line were measured, and 258 families from the control populations were also measured. In addition, enough other flies were measured to bring the total number of males and females available for estimating the phenotypic variance (V_P) to an average of 68 individuals of each sex per line, and a total of 287 males and 287 females for the control lines. The pattern of variance and covariance among these parent-offspring families was used to estimate **G** (for details, see Whitlock and Fowler 1999; Phillips et al. 2001). The outbred lines showed no heterogeneity across replicates and so were pooled into a single measure for each variance component and the **G** matrix.

Stability of Increases in V_P and V_A

Estimates of the additive genetic variance (V_A) and the phenotypic variance (V_P) for each character have been calculated previously for each of the lines in the third generation (Fowler and Whitlock 1999a; Whitlock and Fowler 1999). The new data on the 20th generation allows us to ask whether these important components have changed over time. In particular, we are interested to know whether the increases in genetic variance observed in some traits in some lines con-

tinue through subsequent generations. (This is the reason for the selection of the lines with high V_A for given characters, as outlined in the section above.) To compare across traits, each variance in the inbred lines was standardized by dividing by the corresponding variance from the outbred lines, resulting in a variance ratio.

We want to estimate the correlation of the variance ratios in the F_3 generation to those in the F_{20} generation. To do this, we have to account for the fact that the traits are not entirely independent of each other (although the mean correlation of the traits is quite small for all pairs of traits; Phillips et al. 2001). To maximize the power of this comparison, we might first observe that the variance ratios do not vary significantly among the traits (testing the variance ratios over all 52 lines in the F_3, $P > 0.5$). This implies that combining information across traits is appropriate. Thus, we use each line as a measure of the correlation between generations of the variance ratios. For each line we obtain an estimate of this correlation by finding the correlation among the values for the six traits. No statistical testing is done within a line, so the nonindependence of the six points is not relevant. The nine correlation values obtained from the set of nine lines are statistically independent, therefore we can test whether the mean correlation coefficient is different from zero using a *t*-test.

Measurement error can result in a biased estimate of a correlation. A correlation coefficient is calculated by the ratio of the covariance of two variables divided by the square root of their variances. If the true values of the variable in question are not known without error, then the estimates of the variances of the variables will be inflated by this measurement error. If the error in measurement is not correlated between the two variables, then the estimate of the covariance between the variables is unbiased. Thus, with measurement error, the denominator of estimate of the correlation coefficient is too high, while the numerator is centered on its true value. As a result, a correlation coefficient calculated from data with measurement error is biased toward zero. Fortunately, if the variance due to measurement error is known or can be estimated, the variances can be corrected and an unbiased estimate of the correlation can be produced. This is called the ''correction for attenuation,'' a standard statistical technique in the behavioral and psychological sciences (Ferguson 1981). (The calculation is very similar to the use of repeatability in quantitative genetics.)

In the current case, we are trying to calculate the correlation of two variance ratios. There is no bias in the calculation of the covariance of these ratios, but some of the variance among estimates of these ratios is due to measurement error. In this case, the measurement error is equivalent to the sampling error.

To estimate the variance due to error of the phenotypic variance ratios, we observe that the distribution of each of the characters within a line is not significantly different from a normal distribution. This implies that the variance among estimates due to sampling error should follow an *F*-distribution. (Note that this calculation requires only that the error is *F*-distributed, not that the variance ratios themselves necessarily follow an *F*-distribution. The *F*-distribution of error follows directly from the observation of normally distributed traits.)

To estimate the variance due to error of the genetic variance ratios, we estimate the sampling error from bootstrapping pseudosamples from the control data.

Stability of ***G***

Measuring differences in ***G*** *matrices*

Matrix comparisons were performed using an approach developed by Flury (1988) as modified for quantitative genetic data by Phillips and Arnold (1999). This method not only tests whether two matrices are equal to one another but also determines the ways in which they may differ, in a hierarchical series of questions. Matrices are tested for similarity in principal component structure as well as for proportionality and equality in an ordered fashion. For example, two matrices that share one principal component in common can then be tested to see if they share two, and so on. Matrices with all of their principal components in common can be tested for proportionality, and proportional matrices can then be tested for equality. The significance levels of these tests are determined by repeatedly randomizing families across populations, testing the likelihood of the tested hypothesis against the null hypothesis of no shared structure. We used 10,000 randomization runs per test, using the CPCrand program (Phillips 1998a).

There are two potential limitations in the Flury approach that may be important for its application to these data. First, tests conducted with data on few families are less likely to detect differences among matrices than are those with many families (Phillips and Arnold 1999). However, for any given comparison (e.g., F_3 vs. F_{20}), the power from line to line is approximately the same, and so relative differences among lines in the similarity of the F_3 and F_{20} retain meaning. Furthermore, tests versus the control population involve thousands of families and are therefore expected to be very powerful.

A second limitation of the Flury approach is that it tests a set of null hypotheses based on principal component structure, which is constrained to be orthogonal. Matrices may potentially share similarity in structure that is not reflected in the principal components (Houle et al. 2002). Unlike many studies of matrix stability, the populations compared here display a direct ancestor-descendent relationship that strongly motivates a null hypothesis of matrix equality, whereas mutation and drift lead to expectations of proportionality (Phillips et al. 2001). Matrices that are found to be unrelated by the Flury approach may in fact share similar nonorthogonal structures (Houle et al. 2002), but this does not negate the value of testing other degrees of similarity, especially because our focus is on divergence.

The real problem here is that strict hypothesis testing is probably not as valuable in this context as a metric that allows one to measure the degree of divergence (Marroig and Cheverud 2001). We therefore also used a second type of analysis that asks whether the lines have retained any of their covariance structure over the 17 generations between measurements. For each pair of traits in each population and each generation, the major axis of genetic variation (i.e., the first

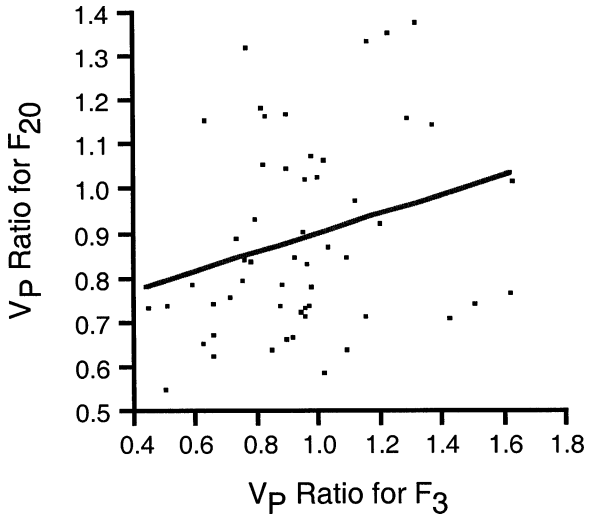

FIG. 2. The correlation between the phenotypic variance ratios at different generations. The line is the best-fit regression predicting the variance ratio in the F_{20} from that in the F_3. (Significance values cannot be directly given for this fit; see text.) A variance ratio is the variance of an inbred line divided by the same variance from the control population.

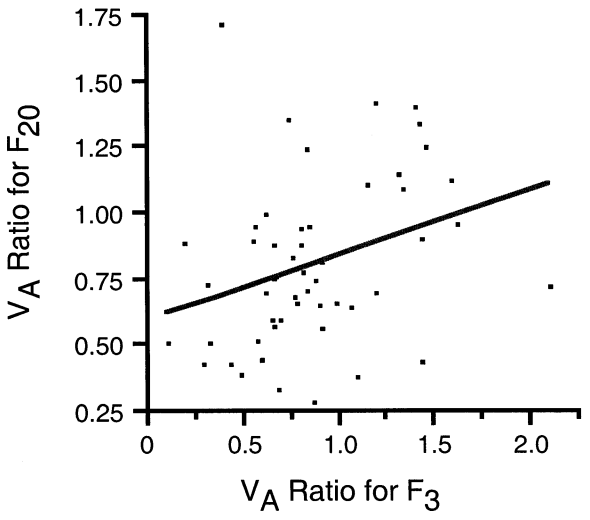

FIG. 3. The correlation between the additive genetic variance ratios at different generations. The line is the best-fit regression predicting the variance ratio in the F_{20} from that in the F_3.

principal component of the two-dimensional **G** matrix) was determined. The angle between this axis and the major axis formed by the same traits in the control lines was calculated, which we call the ''angle of deviation'' for those traits in that line (see also Phillips et al. 2001). A correlation coefficient was calculated between these angles of deviation in the two measured generations. Large positive correlations indicate that there is some stability of genetic covariance structure within lines over the generations. This approach allows for a more general measure of similarity than the CPC method and to some degree should protect against the possibility that principal components might be the wrong model for shared structure across the full matrices. Because the datapoints are not independent, normal statistical techniques are not appropriate, and a bootstrap analysis was therefore performed. The set of 15 trait-pair covariances for a line were resampled as a set within a generation, but paired at random with a randomly chosen set of estimates from the other generation. A total of 10,000 pseudo-datasets were created in this way. The *P*-value given by this technique is the proportion of pseudo-datasets that had a higher correlation in angles of deviation than the actual dataset.

Measuring the direction of change in the G matrix

If the **G** matrix of the bottlenecked lines has changed over the intervening 17 generations between the F_3 and F_{20}, we would like to know whether these changes are toward the control line or at random with respect to the ancestral state. The angles of deviation from the control data were analyzed using the bootstrap test described above to evaluate whether the average angles of deviation in the F_{20} were significantly different from that in the F_3.

RESULTS

Stability of Increases in V_P and V_A

Of the nine lines examined and six characters per line, there were 10 instances in which there was a greater than 30% increase in additive genetic variance in the F_3. Eight of the lines had one character that showed such a large increase in genetic variance, and one line had two characters that increased by that much in V_A. Of these 10 total instances of substantially increased V_A, six had V_A greater than the controls in the F_{20} sample. We compared the mean additive genetic variance ratio in generation 20 for the traits that had a high V_A in the F_3 to the mean variance ratios in the F_3. Because of the nonindependence of two of the comparisons (i.e., the pair that came from the same line), we used only one of these traits in that line at a time. The traits that had high V_A in the F_3 still had a much higher V_A in generation 20 than the average in the F_3 (mean additive genetic variance ratios 1.07 vs. 0.71, Wilcoxon sign-rank test, $P = 0.002$ or $P = 0.0001$, depending on which of the traits from the line with two substantial increases was used).

The pattern of correlation between the variance ratios for phenotypic variance in the F_3 and the variance ratios for the same characters in the F_{20} is shown in Figure 2. The phenotypic variance early after a bottleneck predicts V_P many generations later ($P < 0.05$). The raw correlation of variance ratios between generations is not extremely high ($r = 0.27$), but each of these variance ratios is estimated with a fairly large error. Correcting for attenuation (Ferguson 1981) gives an estimate of $r = 0.35$.

Similarly, the additive genetic variance for each trait is similar between the generations ($P < 0.01$). Figure 3 shows the pattern of correlation for the additive genetic variance ratios for the two generations. The correlation is not high ($r = 0.32$), but by correcting for the error variances we can estimate the correlation coefficient to be 0.63 for the additive genetic variance of the F_3 and F_{20}.

TABLE 1. Matrix comparisons using the Flury hierarchy. Each entry gives the level of the hierarchy at which the stated hypothesis could not be rejected. CPC[*x*] indicates that the hypothesis that the matrices share *x* common principal components cannot be rejected. Full CPC indicates that the hypothesis that the matrices share all six principal components cannot be rejected. Unrelated indicates that the hypothesis that any principal components are shared by the matrices can be rejected at the $\alpha = 0.05$ level. Line numbers are arbitrarily assigned for ease of reading this table.

Line number	No. of families: F_3	No. of families: F_{20}	F_3 vs. F_{20}	F_3 vs. control	F_{20} vs. control
1	74	59	equal	CPC[1]	CPC[1]
2	102	63	equal	unrelated	unrelated
3	98	56	equal	unrelated	proportional
4	90	65	proportional	unrelated	equal
5	100	58	full CPC	unrelated	full CPC
6	87	60	CPC[3]	CPC[1]	CPC[1]
7	86	55	CPC[3]	unrelated	unrelated
8	104	60	CPC[1]	CPC[1]	unrelated
9	87	64	unrelated	unrelated	unrelated

*Stability of the Outbred **G** Matrix*

The replicate control populations were not significantly variable in their **G** matrices in the F_3 (Phillips et al. 2001), nor was there evidence for significant deviations from equality in the F_{20} ($P = 0.2131$). Thus, the control populations were pooled to make a single control dataset. The residuals of midparent and offspring means from the averages of that control line were used to prevent the small variance in means among the replicate control lines from being conflated with within-line variance. The outbred control populations did not change significantly in the size or shape of their **G** matrix over the 17 generations in this study (test of equality of matrices; $P = 0.204$).

*Changes in the **G** Matrix of Bottlenecked Lines*

For six of the nine lines measured in this study, there is significant evidence of change in the **G** matrix over the course of the study (Table 1). These lines have changed in ways that vary from sharing the entire eigenstructure but differing in the magnitudes of the eigenvalues to having no eigenstructure in common. In most cases, the structure of the **G** matrix remains quite different from that of the control populations. Thus, the **G** matrix can evolve over relatively short periods of time, even when the population size is not extremely small.

For one-third of the lines, however, there is no statistically significant evidence that the **G** matrix has changed over these 17 generations. These negative results must be interpreted in view of the fact that with around 60 families per line in the F_{20} generation, we are at the limits of reliability of measuring genetic covariances (Klein 1974; Phillips 1998b). Limitations in power can lead to failure to reject the null hypothesis of similarity when it is false (Phillips and Arnold 1999). Using the most similar (Table 2) and least similar (Table 3) lines as examples, it is clear that the results for any specific comparison are likely to be strongly influenced by sampling error. For line 1, for example, the Flury test could not reject the hypothesis of equality between the F_3 and F_{20} matrices, yet it is very unlikely that these matrices are in fact equal. For instance, there are point estimates (e.g., the variance for angle

TABLE 2. Additive genetic variance-covariance matrices (**G**) for the outbred control, the bottleneck, and the 20-generation recovery lines for a replicate (line 1) that showed the highest degree of similarity across generations (see Table 1). For each element in the matrix, the estimate for the outbred control population is given on the top line, the estimate for the F_3 population shortly after the bottleneck is in the middle line, and the estimate for the F_{20} population is in the bottom line. Estimates from the control line are reprinted from Phillips et al. (2001). All variances, covariance, and standard errors have been multiplied by 10^4 for ease of viewing. Variance units are mm^4 for area and $radians^2$ for the angles. Values in parentheses give the standard error of the estimate calculated via bootstrapping across families.

Trait		Wing area	Angle 5-7-4	Angle 8-7-6	Angle 2-9-3	Angle 2-1-5	Angle 2-3-5
Wing area	C	12.10 (0.57)					
	F_3	7.31 (2.70)					
	F_{20}	7.77 (2.98)					
Angle 5-7-4	C	3.51 (0.33)	9.07 (0.36)				
	F_3	0.69 (1.70)	6.22 (1.37)				
	F_{20}	2.36 (1.66)	5.71 (1.67)				
Angle 8-7-6	C	1.50 (0.53)	−3.85 (0.45)	21.84 (1.03)			
	F_3	9.43 (3.64)	−5.86 (2.76)	45.41 (8.30)			
	F_{20}	3.19 (3.14)	−3.56 (2.09)	15.41 (4.46)			
Angle 2-9-3	C	1.03 (0.10)	0.95 (0.08)	−0.46 (0.14)	0.90 (0.04)		
	F_3	1.11 (0.49)	1.09 (0.37)	0.42 (1.02)	0.78 (0.15)		
	F_{20}	0.87 (0.48)	1.12 (0.44)	−0.10 (0.66)	0.71 (0.14)		
Angle 2-1-5	C	1.79 (0.15)	2.44 (0.14)	−1.31 (0.19)	0.80 (0.04)	1.96 (0.07)	
	F_3	0.89 (0.64)	2.25 (0.58)	−4.44 (1.65)	0.73 (0.22)	1.64 (0.44)	
	F_{20}	1.19 (0.72)	1.81 (0.66)	−2.41 (1.07)	0.84 (0.20)	1.78 (0.36)	
Angle 2-3-5	C	1.61 (0.26)	−1.48 (0.21)	2.71 (3.24)	0.09 (0.06)	0.02 (0.09)	5.79 (0.21)
	F_3	2.72 (1.34)	−0.29 (0.83)	4.78 (2.08)	0.85 (0.31)	0.72 (0.37)	4.23 (1.13)
	F_{20}	4.58 (1.84)	−1.38 (1.18)	4.65 (2.62)	0.39 (0.41)	0.43 (0.50)	7.18 (1.86)

TABLE 3. Additive genetic variance-covariance matrices (**G**) for the outbred control, the bottleneck, and the 20-generation recovery lines for a replicate (line 9) that showed the least degree of similarity across generations (see Table 1). Notations as in Table 2.

Trait		Wing area	Angle 5-7-4	Angle 8-7-6	Angle 2-9-3	Angle 2-1-5	Angle 2-3-5
Wing area	C	12.10 (0.57)					
	F_3	7.14 (1.82)					
	F_{20}	4.99 (1.16)					
Angle 5-7-4	C	3.51 (0.33)	9.07 (0.36)				
	F_3	2.79 (1.40)	11.76 (2.39)				
	F_{20}	3.64 (1.31)	10.33 (2.28)				
Angle 8-7-6	C	1.50 (0.53)	−3.85 (0.45)	21.84 (1.03)			
	F_3	2.02 (2.11)	−11.53 (3.01)	24.80 (4.59)			
	F_{20}	−4.71 (2.39)	−8.83 (2.64)	24.34 (6.19)			
Angle 2-9-3	C	1.03 (0.10)	0.95 (0.08)	−0.46 (0.14)	0.90 (0.04)		
	F_3	−0.10 (0.41)	1.29 (0.52)	−1.55 (0.78)	0.80 (0.18)		
	F_{20}	−0.16 (0.32)	0.78 (0.42)	−0.14 (0.54)	0.63 (0.13)		
Angle 2-1-5	C	1.79 (0.15)	2.44 (0.14)	−1.31 (0.19)	0.80 (0.04)	1.96 (0.07)	
	F_3	1.59 (0.65)	4.74 (1.11)	−6.03 (1.47)	0.99 (0.28)	2.78 (0.56)	
	F_{20}	0.75 (0.36)	2.23 (0.59)	−2.56 (0.68)	0.23 (0.13)	0.86 (0.23)	
Angle 2-3-5	C	1.61 (0.26)	−1.48 (0.21)	2.71 (3.24)	0.09 (0.06)	0.02 (0.09)	5.79 (0.21)
	F_3	−0.91 (0.92)	−1.078 (0.96)	1.35 (1.41)	−0.61 (0.22)	−1.54 (0.43)	3.80 (0.93)
	F_{20}	−2.47 (0.96)	−1.68 (1.28)	3.16 (2.12)	−0.97 (0.34)	−1.13 (0.41)	5.38 (0.98)

8-7-6) that differ strongly in each of the three time points sampled. Similarly, for line 9, which was found to have unrelated structure across the generations, there are a number of point estimates that can be quite similar in magnitude (e.g., angle 8-7-6). Because all of the analyses within a particular comparison group (e.g., F_3 vs. F_{20}) have approximately the same power, the overall pattern of conservation and change is best viewed by looking across all replicates. Doing this shows that the replicates vary in their degree of divergence through time, at least with respect to the common principal components model.

To examine whether the structure of the **G** matrix at the F_{20} is predicted in any way by its structure in the F_3, we ask whether the angles of deviation from the control line are significantly correlated over time. The value of this correlation is 0.58, with none of the 10,000 bootstrapped datasets reaching a value that high. Thus, we can say with confidence ($P < 0.0001$) that the covariances between characters have some constancy over generations. Combined with the results from the Flury analysis, we can see that the **G** matrix often does change over time but that this change does not erase the effects of history. Some of the changes in the **G** matrix that result from population bottlenecks persist over several generations, but the **G** matrix may also continue to change.

Given this further change in **G** over time, we should ask whether the change in **G** is such that evolution returns the **G** matrix back toward the shape of the original matrix in the control population. The randomization test shows that the pattern of change is entirely consistent with the null hypothesis of random change relative to the shape of the control outbred populations; that is, the angles of deviation from the control do not get significantly smaller over time ($P = 0.65$). In summary, **G** changes in size and shape from population bottlenecks (Phillips et al. 2001), continues to change during subsequent generations at larger population size, but not in a way that returns it to its original shape.

One caveat about these results should be emphasized. The lines chosen for further study in the F_{20} were not a random subset of all lines but represented the populations that had a character with a large increase in V_A. (The original intent of this study was to measure the persistence of large increases in V_A and V_P.) By the regression effect, the V_A estimates for these characters are probably biased upward in the F_3 (i.e., the reason they are higher than average is a combination of a truly large V_A and sampling error in the direction of an increase). This should, if anything, cause the F_{20} to appear closer than the F_3 to the control populations in **G.** The lack of this return to the control state in our data suggests that this regression effect is not great. A similar caveat applies to the results of the regression in the first section: The correlation across generations is likely slightly lower than would be expected from randomly chosen lines because of the regression effect on the characters that had large increases in V_P in the first assay.

DISCUSSION

The course of evolution is determined by a combination of the pattern of selection and the pattern of genetic variability described by the **G** matrix (Lande 1979). If **G** is constant, then the response to future selection can be predicted and the pattern of past selection can be estimated (Lande 1979; Arnold 1988). Furthermore, the **G** matrix reflects the underlying developmental and functional coupling between traits (e.g., Cheverud 1984) and the presence of genetic constraints in general (Arnold 1992). The direction of divergence of closely related species is predictable from **G** (Schluter 1996). Thus, an understanding of the stability and constancy of the **G** matrix is critical to our fuller understanding of the process of evolution. We have seen here that over several generations after population bottlenecks, the **G** matrix can change, but also some features of the **G** matrix are fairly stable over this span. We will consider these two features of the evolution of genetic covariance in turn.

We have previously demonstrated that population bottlenecks can cause severe changes in the size and shape of the **G** matrix (Phillips et al. 2001). It is possible that such changes may be only transient and that with the return of selection

as a primary force in the evolution of these populations that the **G** matrix may return to its original form. This is not the pattern observed here. In most lines **G** continues to change, albeit at a slower pace, but not in a way that is biased toward a return to its original shape. Note, however, that there was no evidence for change in the **G** matrix of the large outbred population over this time period. Most of the variance among lines in additive genetic variance and covariance will be generated by drift-induced linkage disequilibria (Avery and Hill 1977; Zeng and Cockerham 1991). Given enough time, these disequilibria will decay, revealing the true genic differences among the lines. The continuing changes observed here probably reflect both of these processes. Thus, even after the severe bottleneck, genetic drift continues to affect not only the mean phenotypes of these populations but also the pattern of genetic variability that will also affect subsequent evolution. Genetic drift can therefore alter the course of evolution in a way not typically given much attention; it is possible that the major effect that genetic drift might have on the pattern of evolution is by changing **G** rather than by changing the mean values of phenotypes directly.

We have also previously shown that population bottlenecks cause changes in genetic and phenotypic variance of morphological traits and that sometimes these variances increase (Fowler and Whitlock 1999a; Whitlock and Fowler 1999; see also Lopez-Fanjul et al. 1989). Increases in phenotypic variance have been theoretically implicated in models of evolution on complex adaptive landscapes (Whitlock 1995). The variance-induced peak shifts model argues that genetic drift is more likely to cause a shift between peaks on a complex adaptive landscape via changes in the phenotypic variance than by changes in the trait mean (Whitlock 1995). Our previous results demonstrated that most assumptions and predictions of the VIPS model were reasonable, but the VIPS model also requires that changes in phenotypic variance persist over several generations. These current data demonstrate that these changes can easily persist for at least 20 generations. Most of the lines that had traits with significant increases in V_P at the F_3 generation continued to have high V_P for those traits after 17 additional generations. The correlation in both the genetic and phenotypic variance over time was high (and in fact was probably much higher than can be shown with these noisy data; see Results). In most cases, if the phenotypic variance had changed enough to allow peak shifts, our data suggest that it may remain high long enough to allow selection to cause a peak shift.

As yet, we lack a complete theory of how genetic variance and covariance will change with genetic drift and selection. We know that the patterns of change that we will expect to see depend on genetic details such as the number and distribution of effects of the loci involved (Avery and Hill 1977; Zeng and Cockerham 1991; Whitlock 1995). We also know that there are many genetic patterns that can create identical genetic covariance between traits, but with very different evolutionary properties (Houle 1991; Gromko 1995). We will never be able to fully predict changes in **G** from theory alone, making data like those presented here that much more valuable.

Genetic and phenotypic variances and covariances are malleable under genetic drift. Population bottlenecks sometimes cause large changes in the genetic basis of traits, and these changes can continue over subsequent generations. This having been said, however, the changes induced by population bottlenecks do have some persistence. Therefore, we would expect that in the presence of genetic drift that evolution by natural selection would be neither completely predictable nor unpredictable. Over short time spans with moderate population sizes, predictions of response to selection should be reasonably accurate, but with a large degree of uncertainty. As more time passes, we would expect our predictions to become increasingly worse, as **G** continues to transform relative to its starting condition. The changes in **G** that result from genetic drift could cause substantial evolutionary divergence even under subsequent uniform selection. If **G** remains stable over long periods of time, as has not been rejected by several analyses (see Arnold and Phillips 1999), then we can infer that either genetic drift has not played a large role in the evolutionary history of the groups or that selection ultimately may operate to constrain the shape of the genetic covariance matrix.

Acknowledgments

We thank G. Geddes, J. Tu, and J. Bayle for technical assistance; the *Drosophila* lab group at UCL for many helping hands; S. Baird and R. Azevedo for help with digitizer software; S. Otto, A. Caballero, and anonymous reviewers for helpful comments on the manuscript; and the Biotechnology and Biological Sciences Research Council (UK), the National Science Foundation (US, grants BIR-9612469 and DBI-9722921), Natural Environment Research Council (UK), Natural Sciences and Engineering Research Council (Canada), and the Royal Society (UK) for financial support.

Literature Cited

Arnold, S. J. 1988. Quantitative genetics and selection in natural populations: microevolution of vertebral numbers in the garter snake *Thamnophis elegans*. Pp. 619–636 *in* B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong, eds. Proceeding of the second international conference on quantitative genetics. Sinauer Associates, Sunderland, MA.

———. 1992. Constraints on phenotypic evolution. Am. Nat. 140: S85–S107.

Arnold, S. J., and P. C. Phillips. 1999. Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*. Evolution 53: 1516–1527.

Avery, P. J., and W. G. Hill. 1977. Variability in genetic parameters among small populations. Genet. Res. 29:193–213.

Beniwal, B. K., L. M. Hastings, R. Thompson, and W. G. Hill. 1992a. Estimation of changes in genetic-parameters in selected lines of mice using REML with an animal-model. 1. Lean mass. Heredity 69:352–360.

———. 1992b. Estimation of changes in genetic-parameters in selected lines of mice using REML with an animal-model. 2. Bodyweight, body-composition and litter size. Heredity 69:361–371.

Briscoe, D. A., J. M. Malpica, A. Robertson, G. J. Smith, R. Frankham, R. G. Banks, and J. S. F. Barker. 1992. Rapid loss of genetic variation in large captive populations of *Drosophila* flies: implications for the genetic management of captive populations. Conservation Biology 6:416–425.

Bryant, E. H., and L. M. Meffert. 1988. Effect of an experimental bottleneck on morphological integration in the housefly. Evolution 42:698–707.

Bryant, E. H., S. A. McCommas, and L. M. Combs. 1986. The

effect of an experimental bottleneck upon the quantitative genetic variation in the housefly. Genetics 114:1191–1211.
Camara, M. D., C. A. Ancell, and M. Pigliucci. 2000. Induced mutations: a novel tool to study phenotypic integration and evolutionary constraints in *Arabidopsis thaliana*. Evol. Ecol. Res. 2:1009–1029.
Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110: 155–171.
Falconer, D. S., and T. F. C. Mackay. 1989. Introduction to quantitative genetics. Longman, Harlow, U.K.
Ferguson, G. A. 1981. Statistical analysis in psychology and education. McGraw-Hill, New York.
Flury, B. 1988. Common principal components and related multivariate models. Wiley, New York.
Fowler, K., and M. C. Whitlock. 1999a. The distribution of phenotypic variance with inbreeding. Evolution 53:1143–1156.
———. 1999b. The variance in inbreeding depression and the recovery of fitness in bottlenecked populations. Proc. Royal Soc. London B 266:2061–2066.
Frankham, R. 1980. The founder effect and response to artificial selection in *Drosophila*. Pp. 87–90 *in* A. Robertson, ed. Selection experiments in laboratory and domestic animals. Commonw. Agric. Bur., Slough, England.
Gromko, M. H. 1995. Unpredictability of correlated response to selection: pleiotropy and sampling interact. Evolution 49: 685–693.
Houle, D. 1991. Genetic covariance of fitness correlates: what genetic correlations are made of and why it matters. Evolution 45: 630–648.
Houle, D., J. Mezey, and P. Galpern. 2002. Interpretation of the results of common principal components analysis. Evolution 56: 433–440.
Klein, T. W. 1974. Heritability and genetic correlation: statistical power, population comparisons, and sample size. Behavior Genetics 4:171–189.
Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. Evolution 33: 402–416.
———. 1980. Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. 116:463–479.
Lofsvold, D. 1986. Quantitative genetics of morphological differentiation in *Peromyscus*. I. Tests of the homogeneity of genetic covariance structure among species and subspecies. Evolution 40:559–573.
López-Fanjul, C., J. Guerra, and A. García. 1989. Changes in the distribution of the genetic variance of a quantitative character in small populations of *Drosophila melanogaster*. Genet. Sel. Evol. 21:159–168.
Marroig, G., and J. M. Cheverud. 2001. A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys. Evolution 55:2576–2600.
Meyer, K., and W. G. Hill. 1991. Mixed model analysis of a selection experiment for food-intake in mice. Genetical Research 57:71–81.
Phillips, P. C. 1998a. CPCrand: randomization test of the CPC hierarchy. Available via darkwing.uoregon.edu/~pphil/software.html
———. 1998b. Designing experiments to maximize the power of detecting correlations. Evolution 52:251–255.
Phillips, P. C., and S. J. Arnold. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. Evolution 53:1506–1515.
Phillips, P. C., M. C. Whitlock, and K. Fowler. 2001. Inbreeding changes the shape of the genetic covariance matrix in *Drosophila melanogaster*. Genetics 158:1137–1145.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766–1774.
Shaw, F. H., R. G. Shaw, G. S. Wilkinson, and M. Turelli. 1995. Changes in genetic variances and covariances: **G** whiz! Evolution 49:1260–1267.
Turelli, M. 1988. Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution 42:1342–1347.
Whitlock, M. C. 1995. Variance-induced peak shifts. Evolution 49: 252–259.
Whitlock, M. C., and K. Fowler. 1996. The variance among populations in phenotypic variance with inbreeding. Evolution 50: 1919–1926.
———. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics 152: 345–353.
Wilkinson, G. S., K. Fowler, and L. Partridge. 1990. Resistance of genetic correlation structure to directional selection in *Drosophila melanogaster*. Evolution 44:1990–2003.
Wright, S. 1951. The genetical structure of populations. Ann. Eugen. 15:323–354.
Zeng, Z.-B., and C. C. Cockerham. 1991. Variance of neutral genetic variances within and between populations for a quantitative character. Genetics 129:535–553.

Corresponding Editor: A. Caballero