Changes in Genetic Variances and Covariances: G Whiz!

Frank H. Shaw; Ruth G. Shaw; Gerald S. Wilkinson; Michael Turelli

Evolution, Vol. 49, No. 6. (Dec., 1995), pp. 1260-1267.

Stable URL:
http://links.jstor.org/sici?sici=0014-3820%28199512%2949%3A6%3C1260%3ACIGVAC%3E2.0.CO%3B2-J

Evolution is currently published by Society for the Study of Evolution.

Evolution, 49(6), 1995, pp. 1260–1267

CHANGES IN GENETIC VARIANCES AND COVARIANCES: G WHIZ!

FRANK H. SHAW,[1] RUTH G. SHAW,[2] GERALD S. WILKINSON,[3] AND MICHAEL TURELLI[4,5]
[1]*Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, Minnesota 55455*
E-mail: fshaw@superb.ecology.umn.edu
[2]*Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, Minnesota 55108*
E-mail: shaw@superb.ecology.umn.edu
[3]*Department of Zoology, University of Maryland, College Park, Maryland 20742*
E-mail: wilkinson@zool.umd.edu
[4]*Section of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616*
E-mail: mturelli@ucdavis.edu

Key words.—Bulmer effect, constancy of **G**, genetic drift, maximum likelihood, retrospective selection analysis.

Received July 5, 1994. Accepted December 19, 1994.

This note has three objectives: first, we present a reanalysis of a large data set (Wilkinson et al. 1990), addressing the constancy of genetic variance-covariance (**G**) matrices; second, we discuss possible causes for the statistically significant changes found in **G**; and third, we discuss how such changes, or lack thereof, affect retrospective selection analyses, which attempt to reconstruct the long-term history of selection using only contemporary estimates of differences between taxa and estimates of genetic variances and covariances (Lande 1979).

A central issue in evolutionary biology is the relationship between variation within taxa and differences among taxa. From quantitative-genetics theory, we know that short-term selection response is affected by patterns of genetic variation and covariation (Falconer 1989, ch. 19). Lande (1979) extrapolated this selection theory to elucidate the selection pressures responsible for long-term changes. He showed that under certain assumptions the net selection leading to the phenotypic divergence of two groups, a and b, could be characterized retrospectively by the formula

$$\boldsymbol{\beta} = \mathbf{G}^{-1}(\bar{\mathbf{z}}_a - \bar{\mathbf{z}}_b). \quad (1)$$

Here, $\boldsymbol{\beta}$ is an $r \times 1$ vector of net selection gradients, expressing the magnitude of directional selection on each of the r traits, **G** is the $r \times r$ matrix of additive genetic variances and covariances, and $\bar{\mathbf{z}}_a$ ($\bar{\mathbf{z}}_b$) is the $r \times 1$ vector of trait means in population a (b). When estimated within a single generation, the elements of $\boldsymbol{\beta}$ express the average slope of the fitness surface with respect to change in each character, assuming, among other things, that the phenotypes have a multivariate Gaussian distribution (Lande and Arnold 1983; Mitchell-Olds and Shaw 1987). Equation (1) has been used to reconstruct the net selection gradient over many generations. This extrapolation requires several assumptions: for instance, the observed divergence must be wholly genetic, it must have been caused entirely by selection, and all of the (directly) selected traits must be included in the analysis. In addition, equation (1) is valid only if **G** is the same for both groups (e.g., populations, species, genera) and has remained constant throughout the course of their divergence. Turelli (1988) has emphasized that none of the models concerning the maintenance of heritable variation guarantee constancy of **G** and that this issue must be addressed empirically (see also Barton and Turelli 1989). Substantial efforts have recently been devoted to testing the constancy of **G** (Arnold 1981; Lofsvold 1986; Kohn and Atchley 1988; Billington et al. 1988; Wilkinson et al. 1990; Shaw and Billington 1991; Platenkamp and Shaw 1992; Brodie 1993). It nevertheless remains difficult to draw a general conclusion, because most reported studies are modest in scale and because studies differ in both experimental and statistical approaches.

To date, most studies have addressed the constancy of **G** indirectly by comparing **G** between present-day populations whose histories since divergence are not known in detail. One notable exception is the work of Bryant and his collaborators (e.g., Bryant and Meffert 1993) documenting drift-induced changes associated with population bottlenecks. To our knowledge, the most extensive experiment addressing the role of selection is that of Wilkinson et al. (1990), who compared **G** matrices for *Drosophila melanogaster* populations previously subjected to 23 generations of divergent selection on thorax length (for other examples, see Barton and Turelli 1989; Stanton and Young 1994). There are further unusual strengths of this study. It is large, involving 150–200 families for a total of 1500–1900 individuals for each of four populations, with five traits measured per individual. Moreover, the genetic design, with parents subjected to positive assortative mating and traits measured on both parents and offspring, permits relatively precise estimation of additive genetic components of variance and covariance. Finally, the design permits estimation of common environment effects that are likely to be substantial in many of the previously published studies. It thus provides estimates of additive genetic components that are free of bias due to common environment (for details of husbandry, selection, etc., see Wilkinson et al. 1990, especially p. 1994 for a discussion of maternal effects).

[5] Corresponding author

Despite the virtues of this experiment, a mistake in the original statistical analysis may undermine its impact. Let $L(\mathbf{y}_A, \hat{\mathbf{G}})$ denote the likelihood of the data, $\mathbf{y}_A$, from population A given an estimate $\hat{\mathbf{G}}$ of $\mathbf{G}$. Wilkinson et al. (1990) incorrectly tested the equality of the $\mathbf{G}$ matrices of populations A and B by comparing $L(\mathbf{y}_A, \hat{\mathbf{G}}_A)$ to $L(\mathbf{y}_A, \hat{\mathbf{G}}_B)$ (see Wilkinson et al. 1990, p. 1997 and table 3). This approach ignores the sampling variation of $\hat{\mathbf{G}}_B$. We have reanalyzed the data using the likelihood method of Shaw (1991) to check the original conclusions and to provide a context for discussing alternative mechanisms that may change $\mathbf{G}$ and their consequences for retrospective selection analyses. Our analyses yield three results: (1) as reported by Wilkinson et al. (1990), there are statistically significant differences among the $\mathbf{G}$ matrices of their four populations, (2) although selection may have played a role in producing these differences, their magnitude and direction seem consistent with the effects of genetic drift, and (3) even when estimated $\mathbf{G}$ matrices are precisely constant, retrospective selection analyses may be misleading because of unobserved transient changes in $\mathbf{G}$ produced by periods of intense selection (the "Bulmer effect"; Bulmer 1980).

Methods

Our analyses used data in which bristle numbers, denoted BB, were taken as the average of the values on both sides of the fly, with the remaining traits measured in units of 10^{-2} mm. In contrast, components reported in the Appendix of Wilkinson et al. (1990) are based on data with metric traits in ocular units (o.u.) unique to each trait (for thorax, denoted TX, an o.u. $= 4 \times 10^{-5}$ m; for wing length [WL], an o.u. $= 6.25 \times 10^{-5}$ m; and for wing width [WW] and tibia length [TB], an o.u. $= 2.5 \times 10^{-5}$ m), and with bristle number as the sum of the counts on both sides of the fly divided by 2.5. To facilitate comparisons with other experiments, the phenotypic variance-covariance matrix and correlations are given in the Appendix.

The analysis was done using Restricted Maximum Likelihood (REML; Patterson and Thompson 1971; Shaw 1987). The population comparison program pcr2 in the Quercus program package (Shaw and Shaw 1992) was adapted to take account of the assortative mating applied. This requires including extra terms in the variance-covariance matrix $\mathbf{V}$ that appears in the log likelihood:

$$-2L_1(\mathbf{y}, \boldsymbol{\Theta}) = \log(|\mathbf{V}|) + \log(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|) + (\mathbf{y} - \mathbf{X}\alpha)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\alpha). \quad (2)$$

Here, $\mathbf{y}$ is the vector of observations, assumed to follow a multivariate normal distribution, $\mathbf{X}$ is the design matrix for the fixed effects (population, sex, and generation), α is the vector of estimates of the fixed effects, and matrix transpose is indicated by prime. The vector of parameters, $\boldsymbol{\Theta}$, is partitioned into six parts corresponding to the elements of three symmetric matrices of variance and covariance components (additive genetic $\mathbf{G}$, common environment $\mathbf{C}$, and microenvironmental $\mathbf{E}$) for each of the two populations being compared. Fractions of dominance and other nonadditive components of genetic (co)variance would contribute to $\mathbf{C}$, the remainder of each to $\mathbf{E}$. The matrix $\mathbf{V}$, the variance of the multivariate normal $\mathbf{y}$, is a function of these parameters, having elements given by the expected covariance between a particular pair of traits measured in a pair of individuals whose relationship is specified. When mating is random, the covariance between traits of unrelated individuals is $\mathbf{0}$; between trait i in parent and trait j in offspring, it is $(1/2)g_{ij}$; and between trait i and trait j in a pair of full sibs, it is $(1/2)g_{ij} + c_{ij}$.

Assortative mating affects the elements of $\mathbf{V}$. To account for this, we added to each element of the $\mathbf{V}$ matrix terms appropriate for the pair of relatives and traits from the matrices below. Wilkinson et al. (1990, table 1) give complete (and compact) expressions for the entries of $\mathbf{V}$ after scaling to unit phenotypic variance for each trait. The first row and column of the matrices below correspond to the thorax trait (upon which assortative mating was performed). The second and third rows and columns might correspond to any two other traits. We here have used bristle number and wing length as an example. The terms involve the matrices of genetic and phenotypic variance-covariance components, $\mathbf{G}$ and $\mathbf{P} = \mathbf{G} + \mathbf{C} + \mathbf{E}$. In what follows, for example, G_{tw} refers to the additive genetic covariance between thorax and wing length and $P_{tb} = G_{tb} + C_{tb} + E_{tb}$ is the phenotypic covariance between thorax length and bristle number. The term ρ which appears in every element below, is the correlation between the thorax phenotypes of mates. We estimated ρ separately for each population, directly from the phenotypic correlation of thorax lengths for mating pairs. It was treated as a known parameter in our likelihood calculations. Note that, when mating is at random with respect to trait values and, hence, $\rho = 0$, all of the terms below vanish. The parent-offspring covariance matrix is not symmetric: the ijth entry pertains to the relationship between the parent's ith trait and its offspring's jth trait. The phenotypic "cross-covariance" matrices are:

$$\begin{pmatrix} \rho P_{tt} & \rho P_{tb} & \rho P_{tw} \\ \rho P_{tb} & \dfrac{\rho P_{tb}^2}{P_{tt}} & \dfrac{\rho P_{tb} P_{tw}}{P_{tt}} \\ \rho P_{tw} & \dfrac{\rho P_{tb} P_{tw}}{P_{tt}} & \dfrac{\rho P_{tw}^2}{P_{tt}} \end{pmatrix}, \quad \text{between parents,} \quad (3a)$$

$$\begin{pmatrix} \rho G_{tt} & \dfrac{\rho G_{tb}}{2} & \dfrac{\rho G_{tw}}{2} \\ \dfrac{\rho G_{tt} P_{tb}}{2P_{tt}} & \dfrac{\rho G_{tb} P_{tb}}{2P_{tt}} & \dfrac{\rho G_{tw} P_{tb}}{2P_{tt}} \\ \dfrac{\rho G_{tt} P_{tw}}{2P_{tt}} & \dfrac{\rho G_{tb} P_{tw}}{2P_{tt}} & \dfrac{\rho G_{tw} P_{tw}}{2P_{tt}} \end{pmatrix}, \quad \text{between parent and offspring,} \quad (3b)$$

and

$$\frac{1}{2P_{tt}} \begin{pmatrix} \rho G_{tt}^2 & \rho G_{tt} G_{tb} & \rho G_{tt} G_{tw} \\ \rho G_{tt} G_{tb} & \rho G_{tb}^2 & \rho G_{tw} G_{tb} \\ \rho G_{tt} G_{tw} & \rho G_{tw} G_{tb} & \rho G_{tw}^2 \end{pmatrix}, \quad \text{between offspring.} \quad (3c)$$

The program pcr2 uses the Fisher Scoring algorithm to maximize the likelihood. It can be seen from the complete development of this algorithm (Searle et al. 1992) that only

TABLE 1. Additive genetic parameters estimated for each population. Additive genetic (co)variances are given on and above the diagonal (approximate standard errors computed from the information matrix are included for descriptive purposes in parentheses), additive genetic correlations are given below (approximate standard errors computed via the delta method are included for descriptive purposes in parentheses). Estimates are for data with metric traits in units of 10^{-2} mm and with bristle number (BB) as the average of counts on both sides.

	BB	TX	WL	WW	TB
			Base		
BB	0.786 (0.092)	−0.290 (0.156)	−0.511 (0.223)	−0.512 (0.156)	−0.057 (0.103)
TX	−0.173 (0.093)	3.590 (0.601)	2.785 (0.658)	2.478 (0.474)	1.393 (0.306)
WL	−0.234 (0.103)	0.597 (0.126)	6.056 (1.109)	4.085 (0.673)	1.585 (0.431)
WW	−0.291 (0.088)	0.658 (0.112)	0.835 (0.123)	3.948 (0.537)	1.570 (0.294)
TB	−0.054 (0.098)	0.619 (0.122)	0.542 (0.129)	0.665 (0.112)	1.410 (0.237)
			Large		
BB	0.838 (0.091)	−0.111 (0.138)	−0.100 (0.193)	−0.175 (0.136)	0.067 (0.086)
TX	−0.074 (0.093)	2.670 (0.464)	3.245 (0.494)	1.952 (0.345)	1.059 (0.211)
WL	−0.041 (0.079)	0.744 (0.107)	7.130 (0.835)	3.260 (0.483)	1.312 (0.291)
WW	−0.119 (0.094)	0.746 (0.126)	0.762 (0.105)	2.566 (0.419)	0.884 (0.205)
TB	0.079 (0.101)	0.696 (0.134)	0.527 (0.109)	0.592 (0.130)	0.868 (0.168)
			Small		
BB	0.763 (0.091)	0.219 (0.175)	−0.012 (0.279)	0.327 (0.156)	0.301 (0.122)
TX	0.157 (0.127)	2.534 (0.800)	2.237 (1.075)	1.140 (0.570)	1.303 (0.448)
WL	−0.007 (0.154)	0.677 (0.283)	4.302 (1.942)	1.243 (0.949)	1.330 (0.738)
WW	0.285 (0.139)	0.545 (0.236)	0.456 (0.297)	1.726 (0.571)	0.748 (0.377)
TB	0.366 (0.157)	0.868 (0.275)	0.680 (0.320)	0.604 (0.266)	0.889 (0.349)
			Control		
BB	0.764 (0.087)	0.483 (0.150)	0.374 (0.210)	0.368 (0.128)	0.252 (0.096)
TX	0.371 (0.113)	2.217 (0.524)	2.950 (0.549)	1.420 (0.324)	1.167 (0.243)
WL	0.171 (0.096)	0.792 (0.152)	6.255 (1.020)	2.486 (0.490)	1.587 (0.360)
WW	0.285 (0.097)	0.645 (0.145)	0.672 (0.121)	2.186 (0.361)	1.028 (0.208)
TB	0.293 (0.111)	0.796 (0.174)	0.645 (0.137)	0.706 (0.144)	0.969 (0.209)

V and the first derivative matrices ($\partial \mathbf{V}/\partial\theta_i$) need adjustment to account for assortative mating in the pedigree. Just as we modified **V** by including the appropriate terms from (3), we modified ($\partial \mathbf{V}/\partial\theta_i$) by adding the first derivatives of the additional terms.

Analysis of these data requires estimation of 90 parameters in addition to ρ (five variance components and 10 covariance components for each of the three matrices, **G**, **C**, and **E**, in each of a pair of populations. Runs involving approximately 15,000 observations in 90 parameters took up to 20 CPU hours on Sun SPARCstations. Four to six scoring iterations led to convergence at the maximum of the log-likelihood function (L_1). To test the null hypotheses that the **G** matrices of two populations are the same, we required the log-likelihood L_0 of that hypothesis. This we obtained by constraining **G** to be the same in the two populations and by maximizing the log likelihood under these equality constraints. Since the **G** matrix for each population involves 15 covariance components, the constrained model has 15 fewer parameters than the unconstrained model. In the absence of additional constraints (Shaw and Geyer 1993), the likelihood-ratio statistic, $2(L_1 - L_0)$, is asymptotically distributed as χ^2, with 15 degrees of freedom. This test was used to compare the **G** of all pairs of the four populations. We have tested for differences in **G** matrices among all possible pairs of the four populations, and this raises the issue of multiple testing. Adjustment to account for multiple tests is not straightforward here. We therefore report unadjusted *P*-values. A highly conservative correction for the tests of full **G** matrices is to multiply the *P*-values by 6, the number of pairs of populations compared.

TABLE 2. Log-likelihood ratio statistics for testing equality of **G** matrices.

	Large	Small	Control
Base	20.16†	47.20**	44.76**
Large		27.12*	22.2†
Small			10.8‡

† $0.1 < P < 0.18$; ‡ $0.5 < P$; * $P < 0.05$; ** $P \ll 0.005$.

RESULTS

The comparisons revealed highly significant differences in **G** between the Base population and the population selected for small thorax ("Small") and also between the Base and Control populations (tables 1, 2). The **G** matrices of the populations selected for large and small thorax were also significantly different. In contrast, the **G** matrix of Large did not differ significantly ($P > 0.15$) from those of either the Base or Control populations, nor did the **G** matrix of Small differ significantly from that of Control.

The residuals were examined to assess agreement with the assumption of normality. Quantile-quantile plots showed the data to be very close to normal with the exception of the

bristle number trait which was considerably skewed. This was corrected by log transformation. Two representative comparisons (Base vs. Large and Base vs. Small) were run with this transformation in place, and the resultant likelihood ratios (21.8 for Base vs. Large and 48.9 for Base vs. Small) gave *P*-values slightly smaller than the analyses of untransformed data. Examination of the residuals also turned up an outlier for wing width in the Small population. Its removal from the data increased the test statistic of the Base vs. Small comparison to 57.5.

A nonparametric Rao test, proposed by White (1982), was also run on these comparisons. In contrast to the usual likelihood-ratio test, which assumes normality in the data, this test is asymptotically distribution-free. It employs a direct calculation of the variance of the gradient of the likelihood, a calculation made valid here by the large number of independent families in each population, and made possible by the use of maximum likelihood (rather than REML). The nonparametric Rao test gave test statistics of 20.66 for Base versus Large and 61.10 for Base versus Small. Thus, all these corrections to address the issue of deviations from normality yielded smaller *P*-values than the parametric tests based on untransformed data. These analyses gave us confidence that violations of the normality assumption were not appreciably affecting the tests. We here report analyses of untransformed data in order to retain the same scales for all populations.

The selected populations, as well as the Control population, tended to express lower additive genetic variance (V_A) for each trait than the Base population. In particular, V_A for thorax was substantially smaller in selected and Control populations than in the Base population, as was V_A for wing width and tibia length. Overall, decreases in V_A were as much as 39% (for tibia length in the Large population). Exceptions to the general decrease in V_A include bristle number in Large and wing length in both Large and Control populations. In general, differences in V_P tended to be in the same direction as differences in V_A. In the Small population, however, all traits except bristle number exhibited higher V_P than in the Base population. Table 3 gives estimates of the sex-specific means, heritabilities (h^2) and additive coefficient of variation (CV_A) for each trait. In each case, the increase in V_A was associated with an increase both in h^2 and CV_A, advocated by Houle (1992) as an alternative scale-independent measure of additive variation.

The pattern of genetic covariances and correlations of Large was similar to that of the Base population. There was a single difference in sign of Cov_A between bristle number and tibia length. The structure of genetic covariance differed more substantially between Base and Small, on the one hand, and between Base and Control, on the other. Note that the means for the Base and Control populations are very similar; hence, the difference in the **G** matrices cannot be attributed to a scale effect. Whereas in the Base population, Cov_A of bristle number with every other character was negative, the Cov_A involving bristle number in Small and Control were positive with a single exception in Small. Additional likelihood comparisons of the **G** matrix of Base with that of Small indicated that differences in trait covariances largely accounted for the significant difference between these matrices; collectively, the variance components were not significantly different (likelihood-ratio test statistic [lrt] = 8.7, df = 5, $P > 0.1$). Hence, the significant difference between the **G** matrices is not likely to be attributable to a scale effect. This is further supported by the fact that for all traits but bristle number, the Small population has a higher phenotypic variance than any of the other populations. Covariances involving thorax differed significantly between populations (lrt = 10.4, df = 4, $P < 0.05$), but those involving bristle number differed far more strikingly (lrt = 19.4, df = 4, $P <<< 0.005$). Evidently, the genetic relationships between traits other than those subject to direct selection can change profoundly over relatively few generations.

TABLE 3. Means (for males and females) and standardized additive genetic variances (narrow-sense heritabilities, h^2, and additive genetic coefficients of variation, CV_A) for each trait in each population. The CV_A were computed with the mean over both sexes; values computed with sex-specific means differ from those reported below by less than 9%. The metric traits are measured in mm and bristle counts (BB) are the average of counts on both sides.

		BB	TX	WL	WW	TB
Base	$\bar{x}_m$	9.61	0.93	1.59	0.89	0.65
	$\bar{x}_f$	10.03	1.08	1.86	1.02	0.71
	$\hat{h}^2$	0.56	0.36	0.38	0.54	0.41
	CV_A	9.03	1.89	1.43	2.08	1.74
Large	$\bar{x}_m$	9.45	1.00	1.64	0.94	0.68
	$\bar{x}_f$	9.65	1.15	1.91	1.07	0.74
	$\hat{h}^2$	0.58	0.30	0.52	0.37	0.30
	CV_A	9.59	1.52	1.51	1.59	1.31
Small	$\bar{x}_m$	8.68	0.83	1.48	0.81	0.60
	$\bar{x}_f$	9.26	0.96	1.74	0.93	0.66
	$\hat{h}^2$	0.56	0.20	0.16	0.20	0.17
	CV_A	9.74	1.78	1.29	1.51	1.50
Control	$\bar{x}_m$	9.21	0.93	1.58	0.90	0.64
	$\bar{x}_f$	9.82	1.06	1.84	1.02	0.69
	$\hat{h}^2$	0.55	0.24	0.40	0.40	0.30
	CV_A	9.18	1.50	1.46	1.54	1.47

In most cases, our estimates are very similar to those of Wilkinson et al. (1990), when we take into account the difference in scale. Moreover, the signs of the covariances are perfectly consistent with those of the previous analysis. The trends of differences in the estimates of particular components are also comparable, with a few exceptions. In particular, we did not find that V_A and h^2 of thorax length was greater in Small than in Base.

DISCUSSION

In contrast to Wilkinson et al. (1990), we have not found a statistically significant difference between the **G** matrix of the Base population and that of the population selected for large thorax. Our analysis does, however, agree with that of Wilkinson et al. (1990), in demonstrating statistically significant differences between the **G** matrix of the Base population and those of Small and Control. There are many cases of striking differences in genetic variances and covariances and also in their respective standardized values (fig. 1). In the comparison between the Base and Control populations, the estimated heritabilities (h^2) declined by 27%–35% for thorax length (TX), wing width (WW) and tibia length (TB). In the comparison of Small with Base, h^2 declined by more than 45% for all four length characters. Excluding correla-

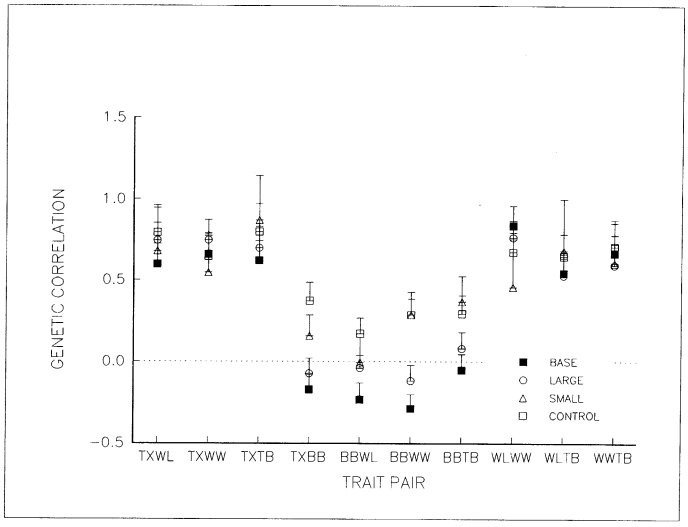

FIG. 1. Estimates of additive genetic correlations within each population. Bars indicate one standard error above the estimate.

tions involving bristle number, differences ranged as high as 45% (wing length and wing width in Small). With the exception of the remarkable constancy of h^2 for bristle number, it does not appear that either heritabilities or genetic correlations were stable over the 23 generations of this study.

Similar changes in genetic variances and covariances have occurred in several experiments (Bohren et al. 1966; Sheridan and Barker 1974). Recent analyses by Meyer and Hill (1991) and Beniwal et al. (1992a,b) demonstrated decreases in additive variances during about 20 generations of selection. The observed changes in V_A (and accompanying changes in additive genetic correlations) are too large and too protracted to be explained by inbreeding or the "Bulmer effect," which describes changes attributable to linkage disequilibrium created by selection (Bulmer 1980, ch. 9). Hence, they were attributed to changes in allele frequencies caused by selection. Most quantitative-genetic analyses predict changes in **G** matrices during selection response (Turelli 1988). These empirical studies document their magnitude. Studies with sample sizes far smaller than those of Wilkinson et al. (1990) that fail to find differences in **G** matrices across natural populations are more likely to reveal the lack of statistical power of the experimental designs (Shaw 1991) than the validity of the hypothesis that **G** matrices remain constant (cf. Arnold 1992).

Genetic Drift May Explain the Observed Changes

Although selection is the expected cause of the observed differences, the small size of the selected and Control populations makes genetic drift a plausible alternative. This is apparently supported by the highly statistically significant difference between the **G** matrices of the Base and Control populations. Yet, even in this comparison, selection is not excluded, because the populations were maintained under different culture conditions (large cage with overlapping generations for the Base vs. individual vials with discrete generations for the Control). Indeed, in *Drosophila* populations moved from vials to cages, Buzzati-Traverso (1955) observed mean wing-size changes too large to be explained by drift (Turelli et al. 1988). Thus, all of the differences we describe may result from selection.

The consequences of drift on these populations are complicated by the poolng of the replicates. During the 23 generations of selection, the census population size was 20 in each of the four replicates of the Large, Small, and Control treatments. Virgins were collected from each replicate, pooled and allowed to mate to create the populations from which the variance estimates were obtained. Thus, the expected allele frequencies in these pooled populations are the averages of the allele frequencies in the replicates. Because the number of generations of drift is not much greater than the effective population sizes, we can approximate the cumulative short-term effects of drift by assuming that the net effective size for the pooled population was roughly four times the effective size of each replicate (i.e., given that we do not expect alleles to have been fixed by drift within the replicates, we can ignore the population subdivision as a first approximation). A typical value for the ratio of effective to census population size for laboratory *Drosophila* experiments is 0.6 (Crow and Kimura 1970, table 7.6.4.2). Hence, we estimate the net effective size (denoted N_e) for the pooled Large, Small and Control populations as 48. We can approximate the expected reduction in the additive variance by

$$E(V_{A,t} \mid V_{A,0}) = V_{A,0}\left(1 - \frac{1}{2N_e}\right)^t \qquad (4)$$

(Crow and Kimura 1970, Sec. 7.5). Under these approximations, we expect that the additive variances will have been reduced by 21% on average from the values in the large Base population. Averaging the proportional changes in the variances for the five traits between the Base population and the three smaller populations, we find that these 15 variances were reduced, on average, by 23%. Such agreement with the drift expectation is not seen in the average reductions for the individual traits, which range from 0% to 45%. A more complete analysis of the drift-induced changes in genetic variances (and covariances) must consider the variance of the expected reduction of variance. In general, these variances depend on the number of loci contributing to the genetic variation of the traits, their linkage relationships and the within-locus distributions of allelic effects (Zeng and Cockerham 1991). Given our ignorance of the genetics and the complexity introduced by the pooling in the Wilkinson et al. (1990) experiment, this will not be pursued. However, our qualitative conclusion is that although selection may well have altered the **G** matrices, the observed reduction in variances seem compatible with expectations from genetic drift alone.

Possible Role of the Bulmer Effect

As emphasized by Turelli (1988), even when selection does not alter **G** appreciably by changing allele frequencies, it generally produces changes via the Bulmer effect. Assuming that many unlinked loci contribute to the variance in the character (i.e., the "Gaussian infinitesimal model"), selection on a random mating population changes the additive genetic variance approximately according to

$$\Delta \mathbf{G}_t = \frac{1}{2}\mathbf{H}_t(\Delta_s \mathbf{P}_t)\mathbf{H}_t' + \frac{1}{2}(\mathbf{G}_{LE} - \mathbf{G}_t), \quad (5)$$

where $\mathbf{H} = \mathbf{G}\mathbf{P}^{-1}$, $\mathbf{H}'$ is the transpose of $\mathbf{H}$, Δ_sP is the within-generation change in the phenotypic covariance matrix caused by selection, and $\mathbf{G}_{LE}$ is the $\mathbf{G}$ matrix that would be observed if all of the relevant loci were in linkage equilibrium (Tallis and Leppard 1988; Turelli 1988; Taylor 1993; Turelli and Barton 1994). For linked loci, the Bulmer effect is governed approximately by the harmonic-mean recombination rate. For most taxa, this mean is likely to be near 0.5 for randomly distributed loci; but it may be as small as 0.1 for organisms like *Drosophila* with few chromosomes and recombination in only one sex (Bulmer 1974; 1980, p. 160). In applications to artificial selection, $\mathbf{G}_{LE}$ is generally assumed to equal $\mathbf{G}_0$, the $\mathbf{G}$ matrix in the initial generation before selection begins (e.g., Meyer and Hill 1991). As in our analysis of genetic drift, we will assume that the Base population values estimate $\mathbf{G}_0$. For the intensity of selection practiced by Wilkinson et al. (1990), equation (5) predicts that the additive variance for thorax length after four generations is reduced by about 18% from its initial value and remains indefinitely very near that value (this model ignores allele frequency changes and assumes free recombination, which accounts for the rapid approach to equilibrium predicted by eq. [5]). However, if we use Bulmer's (1974) harmonic-mean recombination rate approximation for *Drosophila,* $r_h = 0.1$, we predict that the additive variance for thorax length will be reduced by 45% (see Bulmer 1980, eq. 9.47). Because selection was relaxed for three generations before the estimates were made, the harmonic-mean approximation predicts that the observed additive variance for thorax will be reduced by roughly 33% in the Large and Small populations relative to the initial (Base) values. There are associated reductions in the additive variances and covariances for correlated characters. In fact, the additive variances for thorax in the Large and Small populations are reduced by 26% and 29%, respectively, relative to the Base population. Although this suggests reasonable agreement with the Bulmer prediction, it must be noted that the additive variance for thorax in the unselected Control population was reduced by 38% relative to the Base. This "agreement" is surely spurious. Hence, we will not pursue a more detailed analysis of the possible relevance of the Bulmer effect to these data. Our conclusion is simply that both genetic drift and selection may plausibly explain the observed changes. Almost certainly, both processes contribute.

For organisms with more chromosomes than *Drosophila* and recombination in both sexes, harmonic-mean recombination rates are near 0.5 (Bulmer 1974). Hence, if selection were relaxed for three generations before genetic variances were estimated, only 12.5% of the Bulmer-effect reduction of variance would remain. We show below that the Bulmer effect can nonetheless seriously confound retrospective selection analyses.

Does constancy matter? The quantitative results of retrospective selection analyses depend critically on this assumption (Turelli 1988). If multivariate selection response is described exactly by the standard Gaussian equation (cf. Turelli and Barton 1994),

$$\Delta \bar{\mathbf{z}}_t = \mathbf{G}_t \boldsymbol{\beta}_t, \quad (6)$$

the net selection gradient, $\boldsymbol{\beta}_T = \Sigma_{t=0}^{T-1} \boldsymbol{\beta}_t$, is

$$\boldsymbol{\beta}_T = \bar{\mathbf{G}}^{-1}\left(\Delta_T \bar{\mathbf{z}} - \sum_{t=0}^{T-1} (\mathbf{G}_t - \bar{\mathbf{G}})\beta_t\right), \quad (7)$$

where $\bar{\mathbf{G}}$ is the average of $\mathbf{G}$ over time and $\Delta_T \bar{\mathbf{z}} = \bar{\mathbf{z}}_T - 1 - \bar{\mathbf{z}}_0$ is the cumulative change in the mean. Thus, to reconstruct accurately the net selection gradient from equation (1), the current estimate of $\mathbf{G}$ must accurately approximate the mean of $\mathbf{G}$ throughout the period of evolutionary divergence and the second term on the right-hand side of equation (7) must be smaller than the first. Turelli (1988) mistakenly implied that the second term would vanish if the variation in $\mathbf{G}$ was restricted to a multiplicative constant, so that $\mathbf{G}_t = k_t\mathbf{A}$ for some constant matrix $\mathbf{A}$. In fact, this leads to no useful simplification. Thus, even if genetic correlations were "more resistant" to change by selection than covariances (Wilkinson et al. 1990), this does not help validate retrospective selection analyses. Moreover, as demonstrated below, even if one finds identical estimates of $\mathbf{G}$ between taxa, retrospective selection analyses may still be misleading.

The Bulmer Effect May Confound Retrospective Selection Analyses Even When G Is Apparently Constant

Most estimates of $\mathbf{G}$ matrices are made in the laboratory under conditions of relaxed selection. Hence, in the context of the Bulmer-effect equation (5) for the dynamics of $\mathbf{G}$, most experiments are likely to estimate $\mathbf{G}_{LE}$ rather than the $\mathbf{G}$ matrices resulting from selection and contributing to selection response. Many episodes of selection are likely to alter $\mathbf{G}$ through genotype-environment interaction and/or changes in allele frequencies. Here we concentrate on the consequences of changes in $\mathbf{G}$ induced by the Bulmer effect that can be produced by episodes of strong, fluctuating selection of the sort documented in Darwin's finches (Gibbs and Grant 1987). We will show that retrospective selection analyses based on $\mathbf{G}_{LE}$ can lead to qualitatively incorrect conclusions concerning the direction and magnitude of selection. Allowing dominance, but no epistasis, we assume that $\mathbf{G}_{LE}$ remains constant during episodes of selection that produce a net change $\Delta_T \bar{\mathbf{z}}$ for a suite of characters. Let

$$\boldsymbol{\beta}_{retro} = \mathbf{G}_{LE}^{-1}\Delta_T \bar{\mathbf{z}} \quad (8)$$

denote the net selection gradient estimated by retrospective selection analyses, under the assumption that selection is relaxed for some generations before estimating $\mathbf{G}$. We compare this to the actual value of $\boldsymbol{\beta}_T$ obtained by taking into account the changes in $\mathbf{G}_t$ produced by the Bulmer effect (eq. 5).

For simplicity, we consider index selection on two traits (Falconer 1989, ch. 19). This involves truncation selection on a linear combination, such that only individuals with phenotypes satisfying $az_1 + bz_2 > k$ survive, for some constants *a, b,* and *k.* Let *p* denote the fraction of the population that survives. The greatest confounding occurs with selection on characters that are intrinsically uncorrelated. Let $\mathbf{E}$ denote the matrix of environmental and nonadditive genetic effects and assume that

$$\mathbf{G}_{LE} = \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} \quad \text{and} \quad \mathbf{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{9}$$

If we select for $(z_1 + z_2)$ with $p = 0.2$ for three generations, and then select for $-(z_1 + z_2)$ with $p = 0.2$ for three generations, we obtain

$$\boldsymbol{\beta}_{\text{retro}} = \begin{pmatrix} 0.18 \\ 0.18 \end{pmatrix}, \quad \text{but} \quad \boldsymbol{\beta}_T = \begin{pmatrix} -0.07 \\ -0.07 \end{pmatrix}, \tag{10}$$

so that both the magnitude and the signs of the estimated net selection intensities are incorrect. If we select for $(z_1 + z_2)$ with $p = 0.2$ for three generations, then select for $-(2z_1 + z_2)$ with $p = 0.2$ for three generations, we obtain

$$\boldsymbol{\beta}_{\text{retro}} = \begin{pmatrix} -0.09 \\ 0.96 \end{pmatrix}, \quad \text{but} \quad \boldsymbol{\beta}_T = \begin{pmatrix} -0.44 \\ 0.82 \end{pmatrix}, \tag{11}$$

so that the relative magnitudes of direct selection on the two traits are poorly approximated.

These are admittedly extreme examples chosen to illustrate the level of confounding that can be produced under the simplest genetic model that takes into account selection-induced changes in **G**. In contrast, numerical examples indicate that if selection continues in a fixed direction without reversals, $\boldsymbol{\beta}_{\text{retro}}$ and $\boldsymbol{\beta}_T$ will differ only by a constant related to the reduction in variance caused by the Bulmer effect. Similarly, weak selection over many generations or strong selection that is approximately exponential in form would produce little change in **G** from the Bulmer effect. Nevertheless, the examples above demonstrate that in addition to the other assumptions necessary to support retrospective selection analyses, one must also assume that selection neither dramatically alters phenotypic variances nor exhibits dramatic reversals of direction.

Conclusions

One motivation for estimating **G** matrices is that they will reveal the most likely paths of evolution. Recent population genetic analyses support the usual Gaussian predictions for selection response (eq. 6) and the Bulmer effect (eq. 5) over the short term, even when selection is strong and the distribution of breeding values is not quite normal (Turelli and Barton 1994). Yet predicted and realized responses to artificial selection often differ appreciably (Sheridan 1988; Hill and Caballero 1992), particularly for correlated characters (e.g., Falconer 1989, ch. 19; Gromko et al. 1991), perhaps because of the contribution of alleles of major effect or with deleterious pleiotropic effects (e.g., Mackay and Langley 1990). Moreover, as noted by Zeng (1988), **G** often says nothing about long-term evolution in a unimodal fitness landscape. As long as there is any genetic variation in the appropriate directions, the population is expected to track a moving optimum. This conclusion requires, however, that genetic variation in the appropriate directions is available, which may not be true when numerous traits are simultaneously subject to selection (Gomulkiewicz and Kirkpatrick 1992). Similarly, with multiple fitness peaks, genetic covariances can play a decisive role in determining which peak is reached (Price et al. 1993). In view of these points, we suggest that short-term predictions based on sound estimates of genetic parameters are likely to be qualitatively informative. In contrast, both the empirical results of this study and the theoretical considerations suggest that quantitative predictions for long-term selection response and retrospective analyses of selection should be interpreted with caution (Falconer 1989, ch. 12).

Acknowledgments

We thank C. Geyer for many helpful discussions, J. Coyne, T. Price, B. Walsh, and an anonymous reviewer for comments that greatly improved our presentation, and B. Svenson for producing the figure. We are grateful for computational facilities provided by the Department of Mathematics at the University of California, Riverside and the Institute for Mathematics and its Applications at the University of Minnesota. This research was supported by National Science Foundation grants DIR 9112842 and BSR 9119463.

Literature Cited

Arnold, S. J. 1981. Behavioral variation in natural populations. I. Phenotypic, genetic, and environmental correlations between chemoreceptive responses to prey in the garter snake, *Thamnophis elegans*. Evolution 35:489–509.

———. 1992. Constraints on phenotypic evolution. American Naturalist 140(Supplement):S85–S107.

Barton, N. H., and M. Turelli. 1989. Evolutionary quantitative genetics: How little do we know? Annual Review of Genetics 23:337–370.

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

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

Billington, H. L, A. M. Mortimer, and T. McNeilly. 1988. Divergence and genetic structure in adjacent grass populations. I. Quantitative genetics. Evolution 42:1267–1277.

Bohren, B. B., W. G. Hill, and A. Robertson. 1966. Some observations on asymmetrical correlated responses to selection. Genetical Research 7:44–57.

Brodie, E. D., III. 1993. Homogeneity of the genetic variance-covariance matrix for antipredator traits in two natural populations of the garter snake *Thamnophis ordinoides*. Evolution 47: 844–854.

Bryant, E. H., and L. M. Meffert. 1993. The effect of serial founder-flush cycles on quantitative genetic variation in the housefly. Heredity 70:122–129.

Bulmer, M. G. 1974. Linkage disequilibrium and genetic variability. Genetical Research 23:281–289.

———. 1980. The mathematical theory of quantitative genetics. Clarendon Press, Oxford.

Buzzati-Traverso, A. A. 1955. Evolutionary changes in components of fitness and other polygenic traits in *Drosophila melanogaster* populations. Heredity 9:153–186.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetic theory. Harper and Row, New York.

Falconer, D. S. 1989. Introduction to quantitative genetics, 3d ed. Longman, Harlow, Essex.

Gibbs, H. L., and P. R. Grant. 1987. Oscillating natural selection in Darwin's finches. Nature 327:511–513.

Gomulkiewicz, R., and M. Kirkpatrick. 1992. Quantitative genetics and the evolution of reaction norms. Evolution 46:390–411.

Gromko, M. H., A. Briot, S. C. Jensen, and H. H. Fukui. 1991. Selection on copulation duration in *Drosophila melanogaster*: predictability of direct response versus unpredictability of correlated response. Evolution 45:69–81.

Hill, W. G., and A. Caballero. 1992. Artificial selection experiments. Annual Review of Ecology and Systematics 23:287–310.
Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130:195–204.
Kohn, L. A. P., and W. R. Atchley. 1988. How similar are genetic correlation structures? Data from mice and rats. Evolution 42: 467–481.
Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402–416.
Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:110–1226.
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.
Mackay, T. F. C., and C. H. Langley. 1990. Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. Nature 348:64–66.
Meyer, K., and W. G. Hill. 1991. Mixed model analysis of a selection experiment for food intake in mice. Genetical Research 57:71–81.
Mitchell-Olds, T., and R. G. Shaw. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. Evolution 41:1149–1161.
Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554.
Platenkamp, G. A. J., and R. G. Shaw. 1992. Environmental and genetic constraints on adaptive population differentiation in *Anthoxanthum odoratum*. Evolution 46:341–352.
Price, T., M. Turelli, and M. Slatkin. 1993. Peak shifts produced by correlated response to selection. Evolution 47:280–290.
Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. Wiley, New York.
Shaw, F. H., and C. J. Geyer. 1993. Constrained variance component models. IMA Preprint Series #1189, Institute for Mathematics and Its Applications, University of Minnesota.
Shaw, R. G. 1987. Maximum-likelihood approaches applied to quantitative genetics of natural populations. Evolution 41:812–826.
———. 1991. The comparison of quantitative genetic parameters between populations. Evolution 45:143–151.
Shaw, R. G., and H. L. Billington. 1991. Comparison of variance components between two populations of *Holcus lanatus*: a reanalysis. Evolution 45:1287–1289.
Shaw, R. G., and F. H. Shaw. 1992. Quercus: programs for quantitative genetic analysis using maximum likelihood. Published electronically on the Internet, available directly from the authors or via anonymous ftp from ftp.bio.indiana.edu; directory path biology/quantgen/quercus.
Sheridan, A. K. 1988. Agreement between estimated and realised genetic parameters. Animal Breeding Abstracts 56:877–889.
Sheridan, A. K., and J. S. F. Barker. 1974. Two-trait selection and the genetic correlation. II. Changes in the genetic correlation during two-trait selection. Australian Journal of Biological Sciences 27:89–101.
Stanton, M., and H. J. Young. 1994. Selecting for floral character associations in wild radish, *Raphanus sativus* L. Journal of Evolutionary Biology 7:271–285.
Tallis, G. M., and P. Leppard. 1988. The joint effects of selection and assortative mating on multiple polygenic characters. Theoretical and Applied Genetics 75:278–281.
Taylor, N. F. 1993. Sexual selection with condition-dependent handicaps and the evolution of genetic covariances. Ph.D. diss. University of California, Davis.
Turelli, M. 1988. Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution 42:1342–1348.
Turelli, M., and N. H. Barton. 1994. Genetic and statistical analyses of strong selection on polygenic traits: What, me normal? Genetics 138:913–941.
Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rates tests for selection on quantitative characters during macroevolution and microevolution. Evolution 42:1085–1089.
White, H. 1982. Maximum likelihood estimation of misspecified models. Econometrika 50:1–25.
Wilkinson, G. S., K. Fowler, and L. Partridge. 1990. Resistance of genetic correlation structure to directional selection in *Drosophila melanogaster*. Evolution 44:1990–2003.
Zeng, Z.-B. 1988. Long-term correlated response, interpopulation covariation, and interspecific allometry. Evolution 42:363–374.
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: J. Bruce Walsh

APPENDIX

Phenotypic parameters estimated for each population, units as in table 1. Phenotypic (co)variances are given on and above the diagonal, phenotypic correlations are given below.

	Base					Large				
	BB	TX	WL	WW	TB	BB	TX	WL	WW	TB
BB	1.405	−0.081	−0.111	−0.244	0.059	1.450	0.155	−0.071	0.017	0.141
TX	−0.022	9.868	7.391	4.880	3.246	0.043	8.922	6.422	4.065	2.382
WL	−0.023	0.586	16.107	7.486	4.717	−0.016	0.582	13.629	5.965	3.128
WW	−0.076	0.576	0.691	7.276	2.838	0.005	0.519	0.616	6.881	1.938
TB	0.027	0.555	0.631	0.565	3.464	0.068	0.465	0.494	0.431	2.944
	Small					Control				
BB	1.354	0.502	0.422	0.427	0.433	1.393	0.768	0.669	0.596	0.365
TX	0.121	12.805	13.423	6.928	5.456	0.212	9.414	6.354	3.353	2.504
WL	0.069	0.716	27.416	11.628	8.861	0.144	0.525	15.590	5.614	3.856
WW	0.126	0.665	0.762	8.484	4.361	0.215	0.466	0.606	5.507	1.941
TB	0.161	0.659	0.731	0.647	5.360	0.172	0.452	0.541	0.458	3.256