The genetic structure of female life history in *D. melanogaster*: comparisons among populations

PHILIP M. SERVICE
Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA

(*Received 19 January 1999 and in revised form 16 June 1999*)

Summary

Two questions were addressed: (1) What is the genetic variance–covariance structure of a suite of four female life history traits in *D. melanogaster*? and (2) Does the genetic architecture of these traits differ among populations? Three populations of *D. melanogaster* were studied. Genetic variances and covariances were estimated by sib analysis three times for each population: immediately upon establishment of populations in the laboratory, and subsequently after approximately 6 months and 2 years of laboratory culture. Entire genetic variance–covariance matrices, as well as their individual components, were compared between populations by means of likelihood ratio tests. All traits studied were significantly heritable in at least one-half of estimates. Despite large sample sizes, additive genetic covariances were for the most part not statistically significant, and only two significant negative covariance estimates were obtained throughout the experiments. Therefore, these experiments provide little support for evolutionary life history theories that are based on negative genetic correlations among life history components. Neither do they support the idea that genetic variance for fitness components is maintained by trade-offs. Evidence suggests that the **G** matrix of one population was initially different from those of the other two populations. Those differences disappeared after 2 years of laboratory culture. At the level of individual (co)variance components, there were relatively few differences among populations, and the overall impression was that the three populations had generally similar genetic architectures for the traits studied.

1. Introduction

The additive genetic variances and covariances of a set of traits can be thought of as a 'genetic architecture' that describes the way in which a suite of trains may evolve under the influence of natural selection (Lande, 1979; Lande & Arnold, 1983). Negative genetic covariances (trade-offs) among fitness components have played an important role in the development of life history theory (Roff, 1992; Stearns, 1992), and have been proposed as a mechanism for the maintenance of genetic variation for fitness components via antagonistic pleiotropy (Rose, 1982, 1985). The existence of pervasive genetic correlation among life history traits seems to be generally accepted and many studies have reported such correlations; for example, and considering just the *Drosophila* literature, Rose & Charlesworth (1981), Giesel *et al.* (1982), Murphy *et al.* (1983), Scheiner *et al.* (1989) and Hughes (1995). Statistical tests on estimates of genetic correlations are problematic, and not all authors have been able to demonstrate that an appreciable number of estimated correlations were statistically significant (e.g. Scheiner *et al.*, 1989; Hughes, 1995). However, the underlying assumption that genetic covariance among life history traits is common appears not to have been seriously questioned. In this paper, I will argue that failure to detect significant covariance (or correlation) may not be simply an artefact of the relatively low power of many experiments. Rather, we should consider the possibility that, at least for life history traits in *Drosophila*, genetic covariation is for the most part either absent or weak. We might expect to see strong correlations only among traits that are functionally related (Roff, 1992), but even then negative correlations may be expected only when traits are under selection.

Estimates of quantitative genetic parameters (heritability, genetic correlations, etc.) are specific to the populations and environments in which they are

made. However, relatively few studies have actually addressed the question of whether there is appreciable variability in genetic architecture among populations of a species. If differences among populations are minor, then studies on single populations may accurately characterize entire species. On the other hand, if differences among populations are large, then extrapolation of single-population studies to whole species is unwarranted.

Sheridan & Barker (1974) found that the realized genetic correlation between bristle characters in *Drosophila melanogaster* was altered by artificial selection on bristle number. Wilkinson (1989) and Shaw *et al*. (1995) found significant differences between the **G** matrices of fly lines that had been subjected to divergent selection on body size in the laboratory. Five morphological traits were studied, and the observed changes in **G** were attributed primarily to changes in allele frequency under artificial selection. While those studies demonstrate that differences in genetic architecture between natural populations are possible, they do not directly address the question of whether such differences typically exist. Two comparative studies of plant populations have provided little, if any, evidence for significant differences in variance and covariance components (Shaw & Billington, 1991; Platenkamp & Shaw, 1992). Similarly, Brodie (1993) found no differences between two garter snake populations for **G** matrices involving antipredator traits. However, the authors of all those studies noted that the power to detect differences between populations was very low.

A few studies have compared genetic correlations (as distinct from covariances) between populations. Those comparisons generally involved matrix permutation tests, which have been criticized on statistical and biological grounds (Shaw, 1992). Nevertheless, there is some evidence for interpopulation variation in genetic correlation structure of morphology in amphipods (Fong, 1989), of morphology in flies (Cowley & Atchley, 1990), of flowering time and floral structures in *Mimulus* (Carr & Fenster, 1994) and of morphology in garter snakes (Arnold, 1988). Additional studies have compared genetic correlation structures at greater taxonomic distances (e.g. interspecifically), again with mixed results. Those studies are summarized by Roff (1997, chapter 3), who also discusses various methods for comparing (co)variances and correlations among populations.

In this paper, I will present comparisons of **G** matrices for a suite of four female life history traits in *D. melanogaster*. Comparisons were made between three different laboratory populations that were started from collections of wild flies at widely separated locations. Three independent estimates of **G** were obtained for each laboratory population, and nine pair-wise sets of comparisons were made between populations. I used the maximum likelihood methods developed by Shaw (1987, 1991) to estimate (co)-variance components, and to compare **G** matrices and their individual components between populations.

2. Materials and methods

(i) *Populations*

Wild *D. melanogaster* were collected from three localities: Flagstaff, Arizona; Davis, California; and Bowling Green, Ohio. The Flagstaff collection was made in the fruit and vegetable section of a supermarket; the Davis collection was made by placing banana baits in residential yards; and the Bowling Green collection was made by sweepnetting at a large outdoor fruit and vegetable market. *D. melanogaster* may be collected at all times of the year in Davis (T. Prout, personal communication), and it is reasonable to assume that there is a permanent 'natural' population in that part of California. Bowling Green has a seasonal temperate climate and if flies do overwinter in that area, conditions in the local orchards are favourable for year-to-year continuity of local populations (M. Gromko, personal communication). *D. melanogaster* may be caught in residential areas of Flagstaff in late summer (P. Service, personal observation). However, it is never abundant and it seems likely that Flagstaff is re-colonized each year by new immigrants, possibly from many sources. Thus, the Flagstaff collection is least likely to represent a population that has continuity from year to year.

Flagstaff, Davis and Bowling Green were chosen because they are reasonably distant from one another (minimally 1200 km), and because flies could be collected in those localities. Except for geographic separation and climate differences, I had no *a priori* reason to expect the populations to be genetically distinct. Comparisons among populations after introduction to the laboratory showed that they were phenotypically differentiated for early and late fecundity (data not shown). The Davis and Bowling Green populations were similar, and the Flagstaff population had markedly higher fecundity at both ages. Fecundity differences among populations declined over time in the laboratory, although they remained statistically significant at the end of the experiment. Comparisons were confounded by the fact that populations were not tested simultaneously.

As soon as possible after collection, individual wild-caught females were isolated in vials. Species identity was confirmed by examining male progeny. First-generation laboratory-reared progeny were then pooled to establish a laboratory population corresponding to each of the wild collections. The dates of collection and the numbers of wild-caught females used to start each laboratory population are given in

Table 1. *Populations and sib analyses*

	Population		
	Flagstaff	Davis	Bowling Green
Collection date (month/year)	7/90	5/91	9/91
Number of females used to start lab. population	73	111	91
Sib analysis 1			
Date begun	7/90	6/91	9/91
Number of sires	111	126	129
Number of dams	680	663	810
Sib analysis 2			
Date begun	1/91	1/92	4/92
Number of sires	97	154	145
Number of dams	644	786	844
Sib analysis 3			
Date begun	8/92	1/93	3/93
Number of sires	109	116	120
Number of dams	815	730	827

Mean number of sires per sib analysis = 123.
Mean number of dams per sib analysis = 756.

Table 1. Each laboratory population was maintained in 20 vials (27×95 mm), at a size of 1–2 thousand adults per generation, and with discrete generations of 2 weeks duration at 25 °C. Fly medium consisted of corn meal, sucrose, dextrose, agar and yeast, with propionic acid added to retard microbial growth. Single sources of corn meal and yeast were used throughout the experiment.

(ii) *Sib analyses*

Three sib analyses were performed on each population (Table 1). The first analyses used grandoffspring of wild-caught flies as sires and dams. The second sib analyses took place approximately 6 months (13 generations) after the introduction of populations to the laboratory; and the third sib analyses 18–25 months (39–54 generations) after introduction to the laboratory. Sires and dams were obtained by sampling eggs from populations and rearing flies at a density of 30 eggs per vial (approximately 5 ml medium). Sires and dams were collected as virgins. Each sire was allowed to mate with eight dams together for 2 days in vials with live yeast added to the surface of the medium. That was the only point in these experiments where medium was supplemented with live yeast. Dams were subsequently separated into individual laying tubes, and the tubes were thoroughly mixed so that dams mated to the same sire were not associated during the remainder of the experiment. After 24 h, eggs were transferred to rearing tubes with standard medium. For each dam, 24–30 eggs were transferred. Thus, the offspring of each dam were reared as a single cohort.

The following phenotypes were measured on the progeny of each dam: female development rate, early-age fecundity, late-age fecundity and adult female survivorship. Development rate is the reciprocal of the development time (days) of the first female to eclose in each full-sib cohort. The reciprocal of development time was used because larger values are associated with greater fitness. The first female was chosen because of the ease with which her eclosion time could be determined. Eclosion time was determined within 6 h. The first female to eclose was discarded. If more than one female was 'first', one was discarded – the older one if that could be determined by maturation of the cuticle.

When eclosion was complete, 8 females and 6 males from each full-sub cohort were transferred together to a fresh vial. Subsequent transfers to fresh vials were made every third day for the duration of the sib analysis. Early-age fecundity was measured as the average daily fecundity for two consecutive 24 h periods beginning approximately 6 days after eclosion. A single female from each cohort was transferred to an individual tube. Eggs were counted after 24 h, at which time the female was transferred to a fresh tube. Eggs were counted again after a second 24 h. The female was then discarded. Late-age fecundity was measured at approximately 5 weeks after eclosion, using the same methods as for early fecundity. The remaining 6 males and 6 females (after discarding 2 for fecundity measurements) were used to determine adult survivorship. Female survivorship was calculated as the proportion of the 'original' 6 females that were still alive approximately 8 weeks after eclosion. About half of all flies were dead by 8 weeks, thus maximizing the possibility of finding differences between sibships. Survivorship is strongly correlated with mean life span (P. Service, unpublished data). Life span was not determined because it would have required considerably more time and resources to continue each sib analysis until all flies were dead.

Two effects may possibly confound interpretation of the phenotypes that were measured. First, males can influence the rate at which previously unmated females lay eggs after a one-time mating (Service & Vossbrink, 1996). Because females in the present experiments were mated to their brothers, it is possible that differences among half-sib families in fecundity reflected genetic effects expressed through males rather than females. However, in contrast to the procedure used by Service and Vossbrink, females and males were continuously confined for several days or weeks before assaying fecundity. Furthermore, attempts to measure the heritability of male influence on female egg-laying rate in other populations of *D. melanogaster* have produced non-significant results (unpublished

data). Secondly, male accessory gland products have been shown to influence female survivorship (Chapman *et al*., 1995). Because mating was between full sibs, sire influences on female survivorship might reflect effects of variation in the quality or quantity of the male accessory gland fluid, if there were additive genetic variance for such effects.

(iii) *Statistical analysis*

For each dam, there was a single estimate of each phenotype – either a single offspring was measured (development rate, fecundity) or a single aggregate measure was obtained (female survivorship). As far as possible, each phenotype was measured on a different offspring (or different set of offspring). Each sib analysis was conducted in four or five blocks which were spaced about 1 week apart. Block effects (although often statistically significant) are not reported in the Section 3. Sires were nested within blocks. Each sib analysis, therefore, can be treated as a simple hierarchical design: block is a fixed effect, sire is a random effect nested within block, and for each sire there are several observations of a given phenotype (one per dam). Variance components due to block and sire effects can be estimated. However, the dam variance component is confounded with the residual (or error) variance. Therefore, phenotypic variances and covariances can be decomposed into only additive genetic and the remainder, which will be termed environmental (co)variance. This design was chosen for simplicity, ease of analysis, and because additive genetic (co)variances and correlations were the focus of this study. The numbers of sires and dams used for each sib analysis are shown in Table 1.

I used restricted maximum likelihood (REML) to estimate (co)variances, and likelihood ratio tests (LRTs) to test hypotheses concerning those parameters (Shaw, 1987, 1991). The Quercus program package (Shaw & Shaw, 1992) was used for REML analyses. All REML parameter estimates and hypothesis tests were obtained by maximizing multivariate likelihood functions. That is, all four phenotypes were analysed simultaneously. Maximum likelihood methods as implemented by Quercus provide a convenient and elegant multivariate framework for parameter estimation and hypothesis testing. In particular, it is possible to test easily the significance of covariance components, to test hypotheses about entire variance–covariance matrices (or some subset of (co)variance components), and to compare entire matrices for equality. Where possible, I also estimated parameters and conducted hypothesis tests using analysis of variance (ANOVA). In general, ANOVA is only useful for estimating (co)variances and testing whether variance components are greater than zero. Results obtained by REML and ANOVA agreed very well, and only the REML-based results are shown in the tables.

Both ANOVA and REML require normality for hypothesis testing. Normality is also necessary for variance component estimation by REML, but not by ANOVA. The effect of non-normality on REML estimation and hypothesis testing is not well known. ANOVA is generally robust to deviations from normality (Sokal & Rohlf, 1981). The UNIVARIATE procedure of SAS (SAS Institute, 1988) was used to test residuals for univariate normality. However, strictly speaking, multivariate REML requires multivariate normal distribution of observations. Residuals for each sib analysis were evaluated separately. I also examined residuals from combined sib analyses within populations, and residuals from the entire set of nine sib analyses together.

(a) *Hypothesis testing*. Only very general procedures for using LRTs for hypothesis testing in quantitative genetic experiments were outlined by Shaw (1987, 1991). A more detailed treatment seems warranted here. Briefly, to test the null hypothesis that a particular (co)variance component is equal to zero, the log likelihood is calculated when the component is unconstrained ($L_{\max}$) and when the component is constrained to zero (L_0). The quantity $2(L_{\max}-L_0)$ is asymptotically χ^2-distributed with 1 degree of freedom. When H_0 is not on the boundary, as is the case for a covariance component that can take either positive or negative values, the LRT is evaluated by consulting a χ^2-table and obtaining P values in the normal way. When H_0 is on the boundary, as is the case for a variance component that must be greater than or equal to zero, the probability associated with the LRT is obtained by halving the P value obtained from the χ^2 table (Stram & Lee, 1994).

The LRT is easily extended to null hypotheses about sets of (co)variances. To test the hypothesis that an entire **G** matrix is zero, for example, $L_{\max}$ is determined with all components unconstrained and L_0 is determined with all components constrained to zero. The number of degrees of freedom is the difference in the number of components constrained under L_0 and $L_{\max}$. When an entire **G** matrix is tested this way, H_0 is on the boundary for the variance components but not for the covariance components. There is no simple way to accommodate this complication. In the multiple (co)variance component tests reported in this paper, P values are reported as though all components were not on the boundary – a conservative procedure. If a variance component is zero, its associated covariance components must also be zero. Therefore, in order to test H_0: all $V_A = 0$, $L_{\max}$ is computed with all additive genetic covariances constrained to zero, and L_0 is computed with all additive genetic variances and covariances constrained

to zero. With four phenotypes, this test has 4 degrees of freedom. To test H_0: all $Cov_A = 0$, L_{max} is the likelihood of the unconstrained **G** matrix, and L_0 is the likelihood of the **G** matrix with all $Cov_A = 0$. With four phenotypes, this test has 6 degrees of freedom.

LRTs can also be used to compare (co)variances (or sets of (co)variances) between populations. For example, to test the null hypothesis that the **G** matrices of two populations are the same, L_{max} is determined by allowing each population to have its own **G** matrix and L_0 is determined under the constraint that the two matrices are the same. The number of degrees of freedom is the number of components constrained to be the same in both populations. Because this is a test of equality, complications do not arise from having variance components on the boundary. Quercus can compare only two populations at a time. Roff (1997) summarizes other methods for comparing **G** matrices.

The ability to use LRTs for simultaneous tests of several (co)variance components, as well as individual components, means that hypothesis testing can be hierarchical. For example, an initial step in analysis might be to evaluate the null hypothesis that $\mathbf{G} = 0$. If that hypothesis can be rejected, then we might proceed to the nested null hypothesis that all the covariances (i.e. all off-diagonal elements) are equal to zero. If the second hypothesis cannot be rejected (but the first can be), then the implication is that one or more variance components are greater than zero, and these can be tested individually. If the second hypothesis can be rejected, then the implication is, additionally, that one or more covariances is not equal to zero. Note that it is illogical to test the null hypothesis that only the variance components (diagonal elements) are equal to zero, because zero variances imply zero covariances. This hierarchical approach to testing hypotheses about (co)variance components is attractive because it may help to mitigate problems associated with multiple tests. In this paper, I have tested hypotheses about (co)variances hierarchically. However, the analyses have not been stopped upon failure to reject an inclusive null hypothesis. Such an approach might be overly conservative. For example, a single significant covariance might not be detectable in the test of H_0: all $Cov_A = 0$. However, significance of individual (co)variance components should be evaluated cautiously if the inclusive null hypothesis could not also be rejected.

Because a zero variance component implies that all associated covariance components are zero (but not vice versa), I used the following procedure to test whether individual variance components were equal to zero. L_{max} was computed under the constraint that all the associated covariances were equal to zero, but the variance component was unconstrained. L_0 was obtained by constraining the variance as well as the covariances to zero. A significant LRT implied that the variance was greater than zero. This test is analogous to the test of H_0: all $V_A = 0$.

Strictly speaking, the χ^2 approximation for LRTs is valid only when the variance–covariance matrices of the estimates are feasible (positive definite) under both L_{max} and L_0. Feasibility implies that all variances are non-negative and all correlations are in the range -1 to $+1$. Quercus provides the option of enforcing feasibility constraints on the matrices. However, the sampling distribution of the LRT is not asymptotically χ^2-distributed when feasibility is enforced this way. One solution to this problem is to use a method known as the asymptotic parametric bootstrap in order to approximate the distribution of the test (Shaw & Geyer, 1997). However, programs for general application of the asymptotic parametric bootstrap are not available. Furthermore, use of the χ^2 distribution for LRTs when feasibility constraints are imposed is conservative (Shaw & Geyer, 1997). Therefore, I have reported the results of LRTs using the χ^2 distribution, even when feasibility constraints were imposed under L_{max} and/or L_0. I also calculated LRTs using unconstrained parameter estimates. The differences between these latter LRTs and the ones using constrained estimates were negligible.

Each four-trait sib analysis yielded 20 (co)variance component estimates (10 additive genetic and 10 environmental). Considering all nine sib analyses, L_{max} involved the estimation of 72 variances and 108 covariances (and their associated correlations). No variance component estimates were negative. Of the 108 correlations, five were non-feasible: the largest was 1·48 and the smallest was $-1{\cdot}21$. All non-feasible correlations were genetic (as opposed to environmental). Feasible estimates of covariances and correlations are reported in this paper. Imposition of hypothesis constraints (L_0) never resulted in a negative variance component estimate, and frequently 'corrected' non-feasible correlations (such as when a covariance was constrained to be zero). For comparison, ANOVA-based parameter estimates yielded two non-feasible genetic correlations, both of which were also non-feasible by REML.

(b) *Heritabilities and genetic correlations*. The Quercus programs provided estimates of additive genetic and environmental (co)variances. Heritabilities and genetic correlations were calculated by the usual formulae: $h^2 = V_A/(V_A + V_E)$; and $r_A = Cov_{Aij}/(V_{Ai}\,V_{aj})^{1/2}$, where subscripts i and j refer to two traits. I have taken the approach of assuming that the heritability of a trait is significantly greater than zero if the additive genetic variance of the trait is significantly greater than zero (by LRT). Similarly, I consider a genetic correlation to be significant if its associated genetic covariance is non-zero. A difficulty

with this criterion is that correlations are ratios of covariances and variances. It is possible that a covariance estimate may be significantly different from zero by LRT but that one or both of the associated variance components will not be greater than zero. Approximate formulae for standard errors of heritabilities and genetic correlations estimated from half-sib designs are available (Falconer & Mackay, 1996; Roff, 1997). The usual formula for the standard error of a genetic correlation is particularly unsatisfactory in that it depends upon the magnitude of the correlation itself and declines to zero as the estimated correlation approaches ± 1 (Roff, 1997, p. 81 *et seq.*). Re-sampling and randomization could also be used to test the statistical significance of heritabilities and correlations. However, given that there are four traits and nine sib analyses, those methods would be prohibitively cumbersome. Cases in which a correlation is claimed to be significant (by non-zero covariance) but in which one or both of the associated variance components are not significantly greater than zero are noted in Section 3.

3. Results

(i) *Tests for normality*

The null hypothesis that the data were normally distributed could be rejected with a high level of confidence in many cases. Table 2 presents the results of tests on residuals generated by combining all nine sib analyses in a single model (with population, sib analysis, block and sire as effects). As a matter of course, the angular transformation was used for survivorship, and that trait best approximated a normal distribution. Numerous transformations were investigated for the other variables. At the level of individual sib analyses, transformations frequently improved the fit of the data to the normal distribution. However, because pairs of populations were to be analysed together, it was not feasible to transform a given variable one way for one population and another way for a second population, or to transform the variable in only one population. Furthermore, the same data were often included in multiple analyses, for example, in single-population and two-population analyses. Using one transformation for one analysis and a different one for the other would have made it difficult to compare results across analyses. Satisfactory transformations, applicable to multiple populations and multiple sib analyses within populations, could not be found for early and late fecundity or development rate.

Table 2. *Distribution of residuals for all sib analyses combined*

	Sample size	D	g_1	g_2
Early fecundity	6727	0·042	0·296	4·071
Development rate	6771	0·062	−0·003	3·065
Late fecundity	6415	0·045	0·205	2·492
Survivorship	6077	0·058	−0·145	−0·117

Approximate critical values ($P < 0{\cdot}001$): Kolmogorov-Smirnov $D = 0{\cdot}016$; skewness (g_1) = 0·100; and kurtosis (g_2) = 0·200 (Sokal & Rohlf, 1981). The angular transformation was applied to survivorship.

Table 3. *Likelihood ratio tests of H_0: all $Cov_A = 0$*

	Population		
	Flagstaff	Davis	Bowling Green
First sib analysis	**< 0·05**	> 0·50	< 0·10
Second sib analysis	< 0·10	> 0·10	> 0·10
Third sib analysis	**< 0·01**	> 0·10	**< 0·02**

Table entries are P values. d.f. = 6. Significant entries ($P < 0{\cdot}05$) are in **bold** type.

Skewness (g_1) was small but significantly non-zero for all traits except development rate (Table 2). All traits except survivorship (angular transformed) were significantly leptokurtic (positive g_2). The effects of non-normality on estimation and hypothesis testing with REML have not been investigated in detail (Roff, 1997). Therefore, the present results should be evaluated with some caution. However, close agreement in most cases between estimates from ANOVA and REML suggests that the REML estimates are reliable. Furthermore, in at least one case, LRTs based on REML were apparently conservative with non-normal data (Shaw *et al.*, 1995).

(ii) *Likelihood ratio tests on **G** matrices: single populations*

As a first step in the analysis, LRTs were performed on entire **G** matrices and on variance and covariance components as sets. For every sib analysis, H_0: $\mathbf{G} = 0$ was comfortably rejected; as was H_0: all $V_A = 0$ ($P < 0{\cdot}001$ in all cases, except one with $P < 0{\cdot}01$). However, in only three of nine sib analyses was it possible to reject H_0: all $Cov_A = 0$ (Table 3). Results of significance tests on individual additive genetic variance and covariance components may be found in the presentation on heritabilities and genetic correlations (Tables 4, 5).

(iii) *Heritabilities and genetic correlations*

Heritability estimates are presented in Table 4. All heritability estimates were in the feasible range (zero

Table 4. *Heritabilities*

	Population			
	Flagstaff	Davis	Bowling Green	Mean (SE)
Early fecundity	0·286**	0·191	0·629****	0·369 (0·133)
	0·335****	0·285****	0·174*	0·265 (0·048)*
	0·520****	0·373****	0·388****	0·427 (0·047)**
Development rate	0·347****	0·092	0·123	0·187 (0·080)
	0·246***	0·101	0·115	0·154 (0·046)*
	0·254***	0·275***	0·234****	0·254 (0·012)***
Late fecundity	0·433****	0·218*	0·190*	0·280 (0·077)*
	0·115	0·105	0·096	0·105 (0·005)***
	0·321****	0·132	0·138	0·197 (0·062)*
Survivorship	0·330**	0·504****	0·454****	0·429 (0·052)**
	0·520****	0·467****	0·651****	0·546 (0·055)***
	0·586****	0·206*	0·581****	0·458 (0·126)*

* $P < 0{\cdot}05$; ** $P < 0{\cdot}01$; *** $P < 0{\cdot}005$; **** $P < 0{\cdot}001$.
Estimates for the three sib analyses are in order (top to bottom) in each triplet. Significance levels for individual estimates those associated with LRTs of V_A. Significance levels for means are from one-tailed, one-sample *t*-tests.

Table 5. *Genetic correlations*

	Population			
	Flagstaff	Davis	Bowling Green	Mean (SE)
Early fecundity × development rate	0·215	−0·279	−0·064	−0·043 (0·143)
	−0·554	0·843*†	−0·145	0·048 (0·415)
	0·067	0·324	−0·153	0·079 (0·138)
late fecundity	0·574*	0·123	0·136	0·278 (0·148)
	−0·157	0·078	−0·120	−0·066 (0·073)
	0·501**	0·735*†	0·384	0·540 (0·103)*
survivorship	0·281	−0·342	−0·289	−0·117 (0·199)
	0·048	−0·281	−0·352	−0·195 (0·123)
	−0·127	0·315	0·189	0·126 (0·131)
Development rate × late fecundity	−0·057	−0·543	0·239	−0·120 (0·228)
	0·301	−0·500	0·678	0·160 (0·347)
	−0·811****	0·100	−0·290	−0·334 (0·264)
survivorship	0·124	0·502	0·125	0·250 (0·126)
	0·373	−0·248	0·303	0·143 (0·196)
	−0·317	0·681	0·363	0·242 (0·294)
Late fecundity × survivorship	0·751****	0·354	0·850****	0·652 (0·152)
	0·952***†	−0·024	0·875**†	0·601 (0·313)
	0·175	0·034	0·968****†	0·392 (0·291)

* $P < 0{\cdot}05$; ** $P < 0{\cdot}025$; *** $P < 0{\cdot}01$; **** $P < 0{\cdot}005$.
Estimates for the three sib analyses are in order (top to bottom) in each triplet. Significance levels for individual correlations are those associated with LRTs of Cov_A.
† Indicates that one of the variances used to calculate the covariance was not significantly greater than zero. Significance levels for means are from two-tailed, one-sample *t*-tests.

to one). The most consistently heritable trait, and the one with the highest heritability, was female survivorship. However, there were two factors which might have inflated the heritability of that trait. First, experiments were timed to reveal maximal differences in survivorship among full-sib cohorts. Secondly, survivorship is an 'aggregate' phenotype (that is, the 'mean' of several individuals). The heritability of an aggregate trait will be biased upward from the value for the corresponding individual trait because non-genetic causes of variance are underestimated. Early fecundity was also quite consistently heritable. Female development rate and late fecundity were less consistently heritable. When averaged across populations, however, all traits were significantly heritable in the first, second and third sib analyses. The approximate lower limit for statistically significant heritabilities in these experiments was 15–19%. Overall, there does

Table 6. ***G** matrices for the first sib analysis of each population*

	Early fecundity	Development rate	Late fecundity	Survivorship
Early fecundity	**68·37**	0·33	**55·53**	0·56
	23·22	−0·15	2·67	−0·46
	88·53	−0·10	5·24	−0·70
Development rate	—	**0·04**	−0·13	0·01
	—	0·01	−0·28	0·02
	—	0·03	0·17	0·01
Late fecundity	—	—	**136·71**	**2·13**
	—	—	**20·22**	0·45
	—	—	**16·83**	**0·90**
Survivorship	—	—	—	**0·06**
	—	—	—	**0·08**
	—	—	—	**0·07**

Top numbers in each group are for the Flagstaff population, middle numbers are for the Davis population, and bottom numbers are for the Bowling Green population. Individual (co)variance components that are statistically significant are in **bold** type. Approximate significance levels may be found in Tables 4 and 5. As a group, additive genetic variances were significantly different between Flagstaff and Davis, and between Flagstaff and Bowling Green ($P < 0{\cdot}05$ in both cases).

not appear to be a trend towards lower heritabilities over time in the laboratory.

Genetic correlations are shown in Table 5. Only 10 of 54 correlations were statistically significant (by test of H_0: $Cov_A = 0$). Of the 10, there were 5 in which one of the associated additive genetic covariances was not significantly greater than zero. Furthermore, for only 5 of the 10 significant correlations was it also possible to reject the more inclusive null hypothesis that all $Cov_A = 0$. Five of the significant correlations were between late fecundity and survivorship, and all those were positive. Three significant correlations (all positive) were between early and late fecundity. There was only one significant negative genetic correlation in the experiment: between development rate and late fecundity in the third sib analysis for the Flagstaff population. It is particularly interesting to note that large-magnitude correlations ($> 0{\cdot}5$, $< -0{\cdot}5$) were frequently not significantly different from zero by the test used here.

(iv) *Comparisons of **G** matrices between populations*

Pairwise comparisons between populations were performed on (1) entire **G** matrices; (2) additive genetic variances as a group; and (3) additive genetic covariances as a group. For no comparison was it possible to reject H_0: $\mathbf{G}_i = \mathbf{G}_j$, where subscripts i and j refer to populations. However, in two cases it was possible to reject H_0: all V_{Ai} = all V_{Aj}. In both cases, the sib analyses involved were ones done immediately upon introduction to the laboratory. In one comparison, it was possible to reject H_0: $Cov_{Ai} = Cov_{Aj}$. All three significant tests involved the Flagstaff population as one member of the pair. These comparisons required 27 (non-independent) statistical tests. Three significant results are more than might be expected by chance, but not markedly so. **G** matrices are shown in Tables 6–8, and statistically significant comparisons are indicated in the table footnotes.

Examination of the **G** matrices suggests several comparisons of individual (co)variances between populations that might be statistically significant. For example, considering the first sib analyses, the covariance between early and late fecundity is much higher in the Flagstaff population than in either the Davis or Bowling Green populations (Table 6). It would be possible to test all such apparently large differences between individual components. However, in order to reduce the number of tests and the chance of 'false positives', I routinely compared individual components only when a significant result was obtained for the more inclusive hypothesis. I regard such tests as planned comparisons. Thus, for the first sib analyses, individual genetic variance components were compared between the Flagstaff and Davis, and between the Flagstaff and Bowling Green populations. For the second sib analyses, individual covariance components were compared between the Flagstaff and Davis populations. I did perform a few additional *ad hoc* comparisons where they appeared warranted. No individual planned comparisons were made for the third sib analyses.

(a) *Planned comparisons*. For the first sib analyses (Table 6), comparing Flagstaff with Davis and Flagstaff with Bowling Green, the only variance component that was significantly different between populations was that for late fecundity ($P < 0{\cdot}005$ in both cases). For the second sib analyses (Table 7), comparing only covariances between Flagstaff and Davis, there were two significant differences: the covariance between early fecundity and development rate ($P < 0{\cdot}01$), and between late fecundity and

Table 7. ***G** matrices for the second sib analysis of each population*

	Early fecundity	Development rate	Late fecundity	Survivorship
Early fecundity	**52·28**	−0·68	−4·25	0·10
	39·63	**0·81**	1·40	−0·50
	16·23	−0·12	−1·61	−0·44
Development rate	—	**0·03**	0·19	0·02
	—	0·02	−0·22	−0·01
	—	0·04	0·45	0·02
Late fecundity	—	—	14·11	**1·06**
	—	—	8·25	−0·02
	—	—	11·10	**0·92**
Survivorship	—	—	—	**0·09**
	—	—	—	**0·08**
	—	—	—	**0·10**

Top numbers in each group are for the Flagstaff population, middle numbers are for the Davis population, and bottom numbers are for the Bowling Green population. Individual (co)variance components that are statistically significant are in **bold** type. Approximate significance levels may be found in Tables 4 and 5. As a group, additive genetic covariances were significantly different between Flagstaff and Davis ($P < 0{\cdot}025$).

Table 8. ***G** matrices for the third sib analysis of each population*

	Early fecundity	Development rate	Late fecundity	Survivorship
Early fecundity	**24·64**	0·05	**8·77**	−0·19
	25·04	0·31	**11·54**	0·25
	41·82	−0·18	8·10	0·35
Development rate	—	**0·02**	**−0·40**	−0·01
	—	**0·04**	0·06	0·02
	—	**0·03**	−0·18	0·02
Late fecundity	—	—	**12·44**	0·18
	—	—	9·84	0·02
	—	—	10·66	**0·90**
Survivorship	—	—	—	**0·09**
	—	—	—	**0·03**
	—	—	—	**0·08**

Top numbers in each group are for the Flagstaff population, middle numbers are for the Davis population, and bottom numbers are for the Bowling Green population. Individual (co)variance components that are statistically significant are in **bold** type. Approximate significance levels may be found in Tables 4 and 5.

survivorship ($P < 0{\cdot}05$). In all, 14 planned comparisons of individual (co)variance components were made.

(b) Ad hoc *comparisons*. In the first analysis (Table 6), the additive genetic variance for early fecundity was significantly different between the Davis and Bowling Green populations ($P < 0{\cdot}025$). In the third sib analysis (Table 8), the genetic covariance of late fecundity and survivorship was significantly different between the Davis and Bowling Green populations ($P < 0{\cdot}025$). A total of 8 *ad hoc* comparisons were made, including tests on the covariance of early and late fecundity in the first sib analyses (Table 6): despite their large magnitude, the differences between Flagstaff, on the one hand, and Davis and Bowling Green on the other, were not statistically significant ($P > 0{\cdot}05$).

4. Discussion

Two principal results emerge from these experiments. First, there is rather little evidence for genetic covariation among these life history traits and even less evidence for negative genetic correlations. Secondly, the evidence for differences in quantitative genetic architecture between populations is equivocal. There is some suggestion of differences between populations at the time of their introduction to the laboratory. However, any initial differences disappeared after 18–24 months of laboratory culture.

(i) *Genetic covariation*

It is frequently assumed that there is pervasive genetic correlation among life history traits. The existence of such correlations is presupposed by much ecological

and evolutionary life history theory: for example, r and K selection (MacArthur & Wilson, 1967), and antagonistic pleiotropy (Rose, 1982, 1985). Specifically, these theories are based on negative genetic correlations, i.e. trade-offs, between life history traits. The empirical evidence for genetic covariation (positive or negative) among life history traits in *Drosophila* is mixed (Giesel *et al.*, 1982; Scheiner *et al.*, 1989; Hughes, 1995), and in some cases few significant correlations have been found (Scheiner *et al.*, 1989; Hughes, 1995).

There are two difficulties inherent in many studies of genetic correlations. First, statistical power to detect correlations or covariances is low, and if heritabilities are modest extremely large experiments are needed to detect even moderate correlations (see, e.g. Roff, 1992, p. 58). Therefore, it is relatively easy to dismiss absence of significant results on the grounds that experiments are too small. Secondly, for many experimental designs, statistical tests of covariances (or correlations) have not been available until relatively recently. When correlations are estimated from sib designs using ANOVA, there is no simple test of significance. Using the approximate formula for the standard error of r_A that is in the literature (e.g. Falconer & Mackay, 1996, p. 316) leads, perhaps often erroneously, to the conclusion that correlations are statistically significant when they are of large absolute magnitude ($> 0{\cdot}50$). In the present study, however, only 10 of 16 correlations with an absolute magnitude $> 0{\cdot}50$ were significant by non-zero covariance. It is possible to test correlations (or covariances) using randomization or re-sampling methods (Roff, 1997). However, those methods may be cumbersome if multiple populations or several traits are being evaluated. When parent–offspring regression is used, it is possible to evaluate the significance of genetic correlations by testing cross-covariances between traits in parents and offspring (Falconer & Mackay, 1996). However, that method appears to have been little used in the *Drosophila* literature. Maximum likelihood methods do permit simple tests of genetic covariances (and, indirectly, correlations) in a broad array of experimental designs, and are less computationally demanding than randomization or re-sampling.

Some of the most convincing evidence for genetic correlations among life history traits in flies comes from correlated responses to selection in replicated experiments. For example, Rose (1984) and Luckinbill *et al.* (1984) selected for increased life span and late-age fecundity. In both cases, there was a correlated reduction in early-life fecundity. However, in a similar experiment, Partridge & Fowler (1992) did not observe reduced early fecundity, and in another experiment selection for increased life span decreased fecundity generally (Zwaan *et al.*, 1995*b*). Correlated selection responses between increased life span and late-age fecundity on the one hand, and reduced early-life fecundity on the other hand, are difficult to reconcile with the present results. For example, the genetic correlation between early and late fecundity was positive in the third sib analysis for all three populations, significantly so in two populations; and the correlation between early fecundity and survivorship was positive in two populations and negative in the other, although not significantly so in any case (Table 5). Correlated selection responses may arise for reasons other than pleiotropy, including linkage and inadvertent direct selection on the apparently correlated trait. For example, one result of selection for increased late-life fitness in flies is extension of pre-adult development time. However, it has been demonstrated that extended development time is associated with increased pre-adult survival. Therefore, slower development may enhance individual fitness in lines undergoing selection for increased life span, independently of any genetic relationship between life span and development rate (Chippindale *et al.*, 1994).

Another reason that genetic correlations estimated from breeding experiments may not agree with realized correlations observed in selection experiments is that the environments used for the two types of experiments are typically different. Lines selected for increased longevity are usually housed in population cages as adults (e.g. Rose, 1984). On the other hand, estimates of genetic parameters (e.g. this experiment) are usually made on flies housed in vials throughout their lives. Furthermore, larval densities may be much less in breeding experiments than in lines undergoing mass selection. Larval rearing density has important effects on the expression of genetic variance for life span in *D. melanogaster* (Luckinbill & Clare, 1986), and presumably on the expression of genetic covariance as well. The rearing density used in this experiment (24–30 eggs/vial) was evidently high enough to ensure significant heritabilities for most traits most of the time. However, it remains possible that the differences between estimated genetic correlations in this experiment and realized genetic correlations from selection experiments are due to artefacts of experimental procedure.

Another difficulty in comparing the results of different experiments is that phenotypes may not be defined in the same way. In the present experiments late fecundity was measured at about 5 weeks adult age. That age was chosen to be as late as practicable to ensure that a large proportion of females would still be alive and still be laying eggs. When selected lines are assayed, fecundity measurements are often continued to later ages, and selection responses may be more pronounced at later ages. However, lines selected for late-age fitness often show higher fecundity than lines selected for early-age fitness by about 2 weeks

Table 9. ***G*** *matrix and genetic correlation matrix for all populations combined* (*third sib analysis*)

	Early fecundity	Development rate	Late fecundity	Survivorship
Early fecundity	31·15***	0·05	9·62**	0·13
	0·42	—	—	—
Development rate	—	0·03***	−0·18	0·01
	0·06	0·26	—	—
Late fecundity	—	—	11·35**	0·38*
	0·51	−0·31	0·18	—
Survivorship	—	—	—	0·07***
	0·09	0·17	0·43	0·48

* $P < 0{\cdot}025$; ** $P < 0{\cdot}005$; *** $P < 0{\cdot}001$.
Numbers above the diagonal are additive genetic (co)variances; numbers below the diagonal and heritabilities and additive genetic correlations. Significance levels of heritabilities and genetic correlations may be inferred from those of corresponding (co)variances. LRTs of multiple covariance components gave the following results: H_0: $\mathbf{G} = 0$, $P < 0{\cdot}001$; H_0: all $V_A = 0$, $P < 0{\cdot}001$; H_0: all $Cov_A = 0$, $P < 0{\cdot}005$.

adult age (Rose, 1984; Partridge & Fowler, 1992), suggesting that 5 weeks should be adequately old to estimate late fecundity in a sib analysis.

Lastly, with regard to the apparent inconsistency between the present estimated genetic correlations and the results of selection experiments, it should be noted that the selection history of populations may also be important. Negative genetic correlations are to be expected when traits have been under simultaneous directional selection and alleles that affect both traits in a similar way become fixed or lost. The culture regime used in the present experiments did not select directly on life span or late-age fecundity. The correlations between those traits and early fecundity and development rate might, therefore, have been free to take on any value after many generations in the laboratory. In natural populations, both early- and life-life fitness components may be under strong selection. That would lead to the expectation of negative correlations before selection in the laboratory. However, Service & Rose (1985) have argued that in such cases the novel laboratory environment will bias correlations in a positive direction. On a similar argument, however, we would have expected to see a negative correlation arising in the present experiments between development rate and early fecundity because both were presumably under strong directional selection. Overall, that estimated correlation was very low and inconsistent in sign at the time of the third sib analyses (Table 5). A further consideration is that the laboratory environment might not have been stringent enough to produce a trade-off between development rate and early fecundity. For example, very low larval densities may suppress genetic variance for life span (Luckinbill & Clare, 1985). However, the larval densities used in these experiments (for the populations rather than the sib analyses) were similar to those used in a selection experiment that did suggest a trade-off between development rate and early fecundity (Zwaan *et al.*, 1995*a*). Finally, it is possible that additional selection in the laboratory might have been necessary to reveal negative correlations.

There is some evidence that late-age life history traits were consistently positively correlated. Five of the nine estimated genetic correlations between survivorship and late-age fecundity in this experiment were significantly positive (Table 5). Similarly, Hughes (1995) found a positive correlation between life span and late-age male mating ability. These results suggest that alleles at some loci may act to enhance late-life fitness in general, and remain segregating because directional selection is absent. However, one selection experiment did not reveal a positive correlation between life span and late-age fecundity (Zwaan *et al.*, 1995*b*).

It is likely that these experiments, particularly individual sib analyses, had relatively low statistical power to detect genetic correlations: the smallest significant correlation was 0·50. However, it is also possible that there really is little correlation among these particular traits. Multiple populations provide additional statistical power. Correlations from the three populations can be analysed by treating each population estimate as a single datum and performing a *t*-test of the hypothesis that their mean equals zero. Results of those tests are shown in the right-hand column of Table 5. Only one genetic correlation, that between early and late fecundity in the third sib analysis, was significant by this method. Given that 18 correlations were tested, a single significant result is not surprising. A second approach to the issue of statistical power is to combine data from all three populations into a single sib analysis, retaining the block structure for each population. Such an analysis was performed using the data from the third sib analysis of each population (Table 9). This is justifiable on the grounds that no planned comparisons among

populations for the third sib analysis revealed any significant differences. There was one significant *ad hoc* comparison: the genetic covariance of late fecundity and survivorship was significantly different between the Davis and Bowling Green populations ($P < 0{\cdot}025$). Combining populations into a single analysis did not appreciably alter the picture (cf. Tables 8 and 9). All genetic variances (and heritabilities) were significant. The genetic covariance between early and late fecundity was significantly positive, as it was for two of the three populations tested separately. The genetic covariance between late fecundity and survivorship was also significant overall, whereas previously it had been significant in only one of the three separate analyses. The only negative correlation was between late fecundity and development rate, and although of moderate magnitude ($-0{\cdot}31$) was not significant ($P > 0{\cdot}10$). That result is perhaps partly attributable to the modest heritabilities of the two traits involved. The remaining three estimated correlations were of relatively small magnitude ($< 0{\cdot}20$). Overall, then, the picture is still one of generally positive or non-significant genetic correlations even with very large sample sizes (345 sires and 2372 dams in this case). It seems unlikely, therefore, that the present results can be attributed simply to lack of statistical power.

The present results may not apply generally to life history characters in *D. melanogaster*. In particular, traits that may be closely related functionally, for example egg size and egg number, might *a priori* be expected to show negative genetic correlations at equilibrium (Roff, 1992). Scheiner *et al.* (1989) found evidence for trade-offs between early and late periods of fecundity schedules (a result consistent with the correlated responses observed in selection for increased longevity). However, the present results are opposite to those of Scheiner *et al.* Furthermore, Hughes (1995) did not observe a genetic trade-off between early- and late-age male mating ability in flies. One might argue that early- and late-life reproduction are not functionally constrained. But even when functional constraints are acting in a multitrait system, particular pairs of traits may exhibit positive genetic correlations (Charlesworth, 1990). Thus, if all traits in a functionally interdependent set are not investigated, important negative genetic correlations may not be detected. Furthermore, Houle (1991) argues that the sign of the genetic correlation between functionally related traits at equilibrium will depend upon the relative numbers and types of loci affecting expression of the traits, and upon the input of pleiotropic mutational variance.

In sum, there appear to be at least five (not necessarily mutually exclusive) interpretations of the present results: (1) the experiment lacks sufficient statistical power to have confidence in negative results; (2) lack of consistency between the present results and other experiments may be due to differences in design and procedure that produce different genetic architectures; (3) the laboratory environment might not have been appropriate to the expression or evolution of negative genetic correlations; (4) the life history traits studied in these experiments really are for the most part uncorrelated genetically, and that this is true for most life history traits; and (5) the four life history traits studied in this experiment are not generally representative of all fitness component sets, and the absence of negative genetic covariances cannot be used to refute arguments that some functionally related traits should show genetic trade-offs. In any case, additional and probably much larger experiments may be needed to resolve these issues. An alternative approach, investigating the pleiotropic effects of putative individual life history loci (QTLs), may also prove fruitful (Nuzhdin *et al.*, 1997).

(ii) *Multi-population genetic architecture*

There is some evidence for differences in the quantitative genetic architecture of these populations at the time that they were introduced to the laboratory. In particular, the Flagstaff population appeared to be different from the Davis and Bowling Green populations. Those results suggest that there may be differences betwen natural populations of *D. melanogaster*. However, there are at least two reasons why such an interpretation should be made cautiously. First, a necessary, but unfortunate, aspect of these experiments was that populations were not tested concurrently. Thus, differences between populations might have been due to changes in the laboratory environment over time. Arguing against this possibility is the trend toward fewer inter-population differences during the course of the experiment. Thus, there were no differences between populations for the third sib analyses (planned comparisons), even though the experiments were conducted at different times (Table 1). This trend suggests parallel adaptation to a relatively constant laboratory environment. A second reason for caution in claiming differences among populations is that Flagstaff was involved in all significant differences between populations. That may be a reflection of the likelihood that the wild population in Flagstaff does not have continuity from year to year. Thus, the collection of wild flies in Flagstaff may not have been representative of a natural population at equilibrium. A third cause of initial differences between laboratory populations might have been variation inherent in establishment of the laboratory populations from finite samples of wild flies. That, however, seems unlikely with the sample sizes used. Assuming that each wild-caught female was inseminated by a different male, initial

population sizes were approximately 150–200 (Table 1).

Initial differences between Flagstaff and the other two populations notwithstanding, the broader picture suggests that there were few differences between these three populations in genetic architecture. That result is entirely consistent with relatively high estimates of gene flow among populations of *D. melanogaster* based on studies of allozyme loci (Singh & Rhomberg, 1987*a*, *b*). Again, it is possible that the statistical power of these experiments to detect such differences between populations was low. Shaw (1991) used simulation to investigate the power of maximum likelihood methods to detect differences in V_A between populations. With 900 progeny per population (about 20% more than in these experiments) in a sib design, the power to detect differences of 2·5-fold in V_A ranged from about 50% down to about 30%, depending on the magnitude of V_E. Lower heritability reduced the power to detect differences in V_A.

Caution should always be exercised when relying on a negative result (i.e. absence of a significant difference) unless the power of the experiment is known to be high. The probable moderate power of the present experiments to detect differences between populations, together with the fact that the data were apparently not normally distributed, suggests that these experiments should be interpreted carefully. Thus, the conclusion that there is little, if any, difference in genetic architecture of these traits between populations of *D. melanogaster* should be regarded as tentative. Once again, definitive answers may come only with much larger experiments, which will be extremely difficult to carry out.

These experiments could not have been completed without the unflagging laboratory assistance of R. Leone, J. Nichols, E. Kiefer and J. Grant. I am very grateful to R. and F. Shaw for their help and advice at all stages of the maximum likelihood analyses, and without them these analyses could not have been done. I thank M. Turelli for much of the original inspiration for these experiments and for assistance with fly collection; M. Gromko for assistance with fly collection; and N. Waser and M. Price for encouraging me to persist in obtaining funding for this project. This research was supported by National Science Foundation grants BSR 90-17401 and DEB-970776, and National Institutes of Health grant R25-GM56931.

References

Arnold, S. J. (1988). Quantitative genetics and selection in natural populations: microevolution of vertebral numbers in the garter snake *Thamnophis elegans*. In *Proceedings of the Second International Conference on Quantitative Genetics* (ed. B. S. Weir, E. J. Eisen, M. M. Goodman & G. Namkoong), pp. 619–636. Sunderland, MA: Sinauer.

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.

Carr, D. E. & Fenster, C. B. (1994). Levels of genetic variation and covariation for *Mimulus* (Scrophulariaceae) floral traits. *Heredity* **72**, 606–618.

Chapman, T., Liddle, L. F., Kalb, J. M., Wolfner, M. F. & Partridge, L. (1995). Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. *Nature* **373**, 241–244.

Charlesworth, B. (1990). Optimization models, quantitative genetics and mutation. *Evolution* **44**, 520–538.

Chippindale, A. K. Hoang, D. T., Service, P. M. & Rose, M. R. (1994). The evolution of development in *Drosophila melanogaster* selected for postponed senescence. *Evolution* **48**, 1880–1899.

Cowley, D. E. & Atchley, W. R. (1990). Development and quantitative genetics of correlation structure among body parts of *Drosophila melanogaster*. *American Naturalist* **135**, 242–268.

Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th edn. Harlow, UK: Longman.

Fong, D. W. (1989). Morphological evolution of the amphipod *Gammarus minus* in caves: quantitative genetic analysis. *American Midland Naturalist* **121**, 361–378.

Giesel, J. T., Murphy, P. A. & Manlove, M. N. (1982). The influence of temperature on genetic interrelationships of life history traits in a population of *Drosophila melanogaster*: what tangled data sets we weave. *American Naturalist* **119**, 464–479.

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

Hughes, K. A. (1995). The evolutionary genetics of male life-history characters in *Drosophila melanogaster*. *Evolution* **49**, 521–537.

Lande, R. (1979). Quantitative gentic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416.

Lande, R. & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.

Luckinbill, L. S. & Clare, M. J. (1985). Selection for life span in *Drosophila melanogaster*. *Heredity* **55**, 9–18.

Luckinbill, L. S. & Clare, M. J. (1986). A density threshold for the expression of longevity in *Drosophila melanogaster*. *Heredity* **56**, 329–335.

Luckinbill, L. S., Arking, R., Clare, M. J., Cirocco, W. C. & Buck, S. A. (1984). Selection for delayed senescence in *Drosophila melanogaster*. *Evolution* **38**, 996–1003.

MacArthur, R. H. & Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press.

Murphy, P. A., Giesel, J. T. & Manlove, M. N. (1983). Temperature effects on life history variation in *Drosophila simulans*. *Evolution* **37**, 1181–1192.

Nuzhdin, S. V., Pasyukova, E. G., Dilda, C. L., Zeng, Z.-B. & Mackay, T. F. C. (1997). Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the USA* **94**, 9734–9739.

Partridge, L. & Fowler, K. (1992). Direct and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. *Evolution* **46**, 76–91.

Platenkamp, G. A. J. & Shaw, R. G. (1992). Environmental and genetic constraints on adaptive population differentiation in *Anthoxanthum odoratum*. *Evolution* **46**, 341–352.

Roff, D. A. (1992). *The Evolution of Life Histories*. New York: Chapman & Hall.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. New York: Chapman & Hall.

Rose, M. R. (1982). Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* **48**, 63–78.
Rose, M. R. (1984). Laboratory evolution of postponed senescence in *Drosophila melanogaster*. *Evolution* **38**, 1004–1010.
Rose, M. R. (1985). Life history evolution with antagonistic pleiotropy and overlapping generations. *Theoretical Population Biology* **28**, 342–358.
Rose, M. R. & Charlesworth, B. (1981). Genetics of life history in *Drosophila melanogaster*. I. Sib analysis of adult females. *Genetics* **97**, 173–186.
SAS Institute (1988). *SAS Procedures Guide*, release 6.03 edition. Cary, NC: SAS Institute.
Scheiner, S. M., Caplan, R. L. & Lyman, R. F. (1989). A search for trade-offs among life history traits in *Drosophila melanogaster*. *Evolutionary Ecology* **3**, 51–63.
Service, P. M. & Rose, M. R. (1985). Genetic covariation among life-history components: the effect of novel environments. *Evolution* **39**, 943–945.
Service, P. M. & Vossbrink, R. E. (1996). Genetic variation in 'first' male effects on egg laying and remating by female *Drosophila melanogaster*. *Behavior Genetics* **26**, 39–48.
Shaw, R. G. (1987). Maximum-likelihood approaches applied to quantitative genetics of natural populations. *Evolution* **41**, 812–826.
Shaw, R. G. (1991). The comparison of quantitative genetic parameters between populations. *Evolution* **45**, 143–151.
Shaw, R. G. (1992). Comparison of quantitative genetic parameters: reply to Cowley and Atchley. *Evolution* **46**, 1967–1969.
Shaw, R. G. & Billington, H. L. (1991). Comparison of variance components between two populations of *Holcus lanatus*: a reanalysis. *Evolution* **45**, 1287–1289.
Shaw, F. H. & Geyer, C. J. (1997). Estimation and testing in constrained covariance models. *Biometrika* **84**, 95–102.
Shaw, R. G. & Shaw, F. H. (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.
Shaw, F. H., Shaw, R. G., Wilkinson, G. S. & Turelli, M. (1995). Changes in genetic variances and covariances: G whiz! *Evolution* **49**, 1260–1267.
Sheridan, A. K. & Barker, J. S. F. (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.
Singh, R. S. & Rhomberg, L. R. (1987*a*). A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. *Genetics* **115**, 313–322.
Singh, R. S. & Rhomberg, L. R. (1987*b*). A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of genetic differentiation. *Genetics* **117**, 255–271.
Sokal, R. R. & Rohlf, F. J. (1981). *Biometry*, 2nd edn. San Francisco: W. H. Freeman.
Stearns, S. C. (1992). *The Evolution of Life Histories*. New York: Oxford University Press.
Stram, D. O. & Lee, J. W. (1994). Variance component testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177.
Wilkinson, G. S., Fowler, K. & Partridge, L. (1990). Resistance of genetic correlation structure to directional selection in *Drosophila melanogaster*. *Evolution* **44**, 1990–2003.
Zwaan, B., Bijlsma, R. & Hoekstra, R. F. (1995*a*). Artificial selection for developmental time in *Drosophila melanogaster* in relation to the evolution of aging: direct and correlated responses. *Evolution* **49**, 635–648.
Zwaan, B., Bijlsma, R. & Hoekstra, R. F. (1995*b*). Direct selection on life span in *Drosophila melanogaster*. *Evolution* **49**, 649–659.