Comparing Evolvability and Variability of Quantitative Traits

David Houle[1]

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695,
and Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637

ABSTRACT

There are two distinct reasons for making comparisons of genetic variation for quantitative characters. The first is to compare evolvabilities, or ability to respond to selection, and the second is to make inferences about the forces that maintain genetic variability. Measures of variation that are standardized by the trait mean, such as the additive genetic coefficient of variation, are appropriate for both purposes. Variation has usually been compared as narrow sense heritabilities, but this is almost always an inappropriate comparative measure of evolvability and variability. Coefficients of variation were calculated from 842 estimates of trait means, variances and heritabilities in the literature. Traits closely related to fitness have higher additive genetic and nongenetic variability by the coefficient of variation criterion than characters under weak selection. This is the reverse of the accepted conclusion based on comparisons of heritability. The low heritability of fitness components is best explained by their high residual variation. The high additive genetic and residual variability of fitness traits might be explained by the great number of genetic and environmental events they are affected by, or by a lack of stabilizing selection to reduce their phenotypic variance. Over one-third of the quantitative genetics papers reviewed did not report trait means or variances. Researchers should always report these statistics, so that measures of variation appropriate to a variety of situations may be calculated.

EVOLUTIONARY and ecological geneticists use relative measures of genetic variation to gain insight into two sorts of questions. First, we would like to be able to predict the ability of a population to respond to natural or artificial selection; this I will term "evolvability." Second, we wish to gain insight into the strength of the forces which maintain and deplete the genetic variation on which evolvability depends. For this, we need measures of the relative "variability" of traits.

At the phenotypic level, we want to know about the potential for evolution of the mean phenotype. In the short term, this depends on the additive genetic variance. An essential step in quantitative genetics is the partitioning of trait variance, V_P, into additive genetic variance, V_A, and a remainder, V_R, consisting of variance due to other genetic and environmental causes (FALCONER 1989). Narrow-sense heritability, $h^2 = V_A/V_P$, is almost always used as a summary measure of this partitioning. In part, this is because h^2 plays an important role in determining the response to selection as shown by the familiar equation

$$R = h^2 S, \qquad (1)$$

where S is the selection differential, the difference in population mean before and after selection within a

single generation, and R is the response to selection in the following generation (FALCONER 1989; TURELLI and BARTON 1990).

Heritability has been widely used to compare the genetic variation of life history traits relative to traits presumably subject to less strong selection (GUSTAFSSON 1986; ROFF and MOUSSEAU 1987; MOUSSEAU and ROFF 1987; FALCONER 1989). The robust result of such comparisons is that life history traits have lower heritabilities. The conclusion drawn from such comparisons has typically been that life history traits in some absolute sense possess less genetic variation, and will evolve less readily than other traits.

In this paper I will show that trait means, rather than V_P are appropriate for standardizing genetic variances for comparative purposes. To investigate some of the implications of mean-standardized comparative studies, I have reviewed over 200 quantitative genetic studies of animal populations. When mean-standardized measures are compared, traits closely related to fitness are more variable than morphological traits, the opposite conclusion to studies based on heritability. This has important implications for theories of the maintenance of genetic variation. Unfortunately, a very high proportion of quantitative genetic studies report h^2, but not trait means or variances. No summary measure of variation, h^2 or a mean-standardized statistic, can substitute for basic statistics about a study population.

[1] Address after March 1, 1992: Department of Biology, University of Oregon, Eugene, Oregon 97403.

EVOLVABILITY

A good measure of evolvability is the expected response to selection. Looking at Equation 1, it is natural to interpret h^2 as a sufficient descriptor of the state of a population for comparing evolvabilities. It fails to fulfill this purpose for two reasons. First, while h^2 is a dimensionless quantity R is not. Since, for comparative purposes, we are most often interested in proportional, rather than absolute change (HALDANE 1949), the potential response needs to be standardized. One straightforward way of doing so is using the population mean.

Second, the selection differential S is not solely a function of the fitness function, but also contains information about the selected population. For example, in the case of truncation selection on a normally distributed character, $S = i\sqrt{V_P}$, where i is the "intensity" of selection, determined only by the proportion of the population selected for breeding (FALCONER 1989). For a given i, the higher the standard deviation of the selected population, the more the selected individuals may differ from the original population mean. An appropriate mean-standardized predictor of evolvability under truncation selection of unknown intensity is therefore

$$R/(i\bar{X}) = V_A/(\bar{X}\sqrt{V_P}). \tag{2}$$

Here, and throughout this paper, $\bar{X}$ represents the population mean before selection. In general, for fitness function $W(X)$ and trait distribution $F(X)$,

$$S = \bar{W}^{-1} \int_{-\infty}^{\infty} (X - \bar{X})W(X)F(X) \, dX \tag{3}$$

(KIMURA and CROW 1978) where

$$\bar{W} = \int_{-\infty}^{\infty} W(X)F(X) \, dX, \tag{4}$$

the mean fitness of the population. If $W(X)$ can be approximated as a polynomial of order n, S will be a function of the first $n + 1$ moments of $F(X)$. One must specify the form that selection will take before an appropriate index of evolvability can be chosen.

In two important cases h^2 is a particularly misleading estimator of evolvability because S is a function of V_P. One such case is when X is fitness ($W(X) = X$), making

$$S = \frac{\int_{-\infty}^{\infty} (X^2 - X\bar{X})F(X) \, dX}{\int_{-\infty}^{\infty} XF(X) \, dX} \tag{5}$$

$$= \frac{V_P}{\bar{X}}. \tag{6}$$

Standardizing by the mean yields a quantity CROW (1958) called the "opportunity for selection":

$$I = \frac{V_P}{\bar{X}^2}. \tag{7}$$

I is related to the phenotypic coefficient of variation, $CV_P = 100\sqrt{V_P}/\bar{X}$. Application of Equation 1 yields a version of FISHER's "fundamental theorem of natural selection (FISHER 1930; PRICE 1972; FALCONER 1989):

$$R_W = V_A/\bar{W}. \tag{8}$$

Standardizing this response to selection by the mean yields an appropriate measure of the relative evolvability of fitness

$$I_A = \frac{V_A}{\bar{X}^2} = \left(\frac{CV_A}{100}\right)^2. \tag{9}$$

At equilibrium, the predicted gains of Equation 8 would be offset by processes, such as mutation and segregation, which restore both the original mean and genetic variance. In this case I_A is also an appropriate indicator of the variability of fitness, as large values must be the result of powerful forces maintaining genetic variance.

Equations 8 and 9 also hold for all multiplicative components of fitness (O'DONALD 1970; ARNOLD and WADE 1984), although for such characters the interpretation is complicated by the potential for correlations among them. Large values of I_A could also be the result of negative genetic correlations among traits, which attenuate the effects of selection predicted for each character separately. Negative correlations are, however, not necessarily to be expected (CHARLESWORTH 1990; HOULE 1991), and the empirical evidence for their generality is weak (HOULE 1991).

A second case where h^2 is a misleading indicator of evolvability is that of weak gaussian optimizing selection

$$W(X) = \exp\left[-\frac{(X - \theta)^2}{2\omega^2}\right], \tag{10}$$

where ω^2 is a parameter inversely proportional to the strength of selection for the optimum θ. Assuming that X is normally distributed,

$$R = (\theta - \bar{X}) \frac{V_P}{V_P + \omega^2} \frac{V_A}{V_P} \tag{11}$$

(BULMER 1980, p. 150). What data is available on the strength of stabilizing selection suggests that frequently $\omega^2 \gg V_P$ for morphological characters (TURELLI 1984), making the response to changes in θ approximately independent of V_P. Then the standardized response to selection is

$$\frac{R}{\bar{X}} \approx \frac{(\theta - \bar{X})}{\omega^2} \frac{V_A}{\bar{X}}. \tag{12}$$

A good measure of the evolvability of the population in response to arbitrary changes in θ is then $V_A/\bar{X}$. While $(\theta - \bar{X})/(V_P + \omega^2)$ is clearly important in determining the actual response to selection, its future value is not estimable and of no predictive value.

In both of these important cases, directional and weak stabilizing selection, a variance to mean ratio is the most appropriate measure of evolvability of a trait. While I_A and $V_A/\bar{X}$ are distinct quantities, they are correlated, as they share the same components. This argues that quantities related to the additive genetic coefficient of variation, $CV_A = 100\sqrt{V_A}/\bar{X}$, are useful for comparing evolvability of some traits, although certainly not all. In contrast, even in the special case of truncation selection, h^2 is not a good predictor of the response to selection (Equation 2).

VARIABILITY

The evolution of genetic variance must be the result of a balance between evolutionary processes, but in general we are quite ignorant of the relative importance of such processes, as evidenced by the debate over the maintenance of quantitative variation (BARTON and TURELLI 1989; HOULE 1989; BARTON 1990). Therefore, it is not possible to derive precisely satisfying measures of variability analogous to those for evolvabilities. Indeed, comparisons of variability data can be used to gain insight into these processes, by comparing populations in which selection, migration, and drift differ in importance.

When we casually judge a character to be phenotypically variable, we implicitly standardize by the mean. Therefore CV_A seems a reasonable dimensionless criterion for comparing genetic variabilities. CHARLESWORTH (1984, 1987) has previously used CV_A to compare genetic variabilities of Drosophila populations for this reason. Note, however, that a difference in coefficients of variation between traits may be due to an enormous diversity of factors which may be very difficult to disentangle. For example, high standardized phenotypic variability might suggest a lack of strong stabilizing selection. However, it is also possible that a highly variable character is by its nature subject to more environmental variance. A further set of potential complications applies to genetic variability, where it is possible that underlying factors which increase or maintain variance, such as mutation or balancing selection, are correspondingly strong.

An example of the sort of question that coefficients of variation (CVs) can shed light on concerns the low heritability of fitness components. Many evolutionary geneticists have assumed that this is predicted by FISHER's fundamental theorem (GUSTAFSSON 1986; ROFF and MOUSSEAU 1987; MOUSSEAU and ROFF 1987; FALCONER 1989). In the absence of perturbing forces, Equation 8 does predict that V_A, and therefore

h^2, will decline to zero (CHARLESWORTH 1987). However, fitness components do in general possess V_A (ISTOCK 1983; ROFF and MOUSSEAU 1987; MOUSSEAU and ROFF 1987), which is not too surprising given the variety of potential perturbing forces (CHARLESWORTH 1987; MAYO, BÜRGER and LEACH 1990). In the presence of additive genetic variance for fitness, the fundamental theorem makes no prediction about h^2, as FISHER and others have made clear (FISHER 1951; BARTON and TURELLI 1989; PRICE and SCHLUTER 1991). FISHER's fundamental theorem shows that the phenotypic variance of fitness itself is irrelevant to its response to selection, and thus the theorem makes no prediction concerning residual variance, V_R. Reviews of the literature have convincingly confirmed the fact that heritabilities of fitness components are lower than those of other traits (GUSTAFSSON 1986; ROFF and MOUSSEAU 1987; MOUSSEAU and ROFF 1987; FALCONER 1989), but this can be explained either by relatively low V_A for fitness components or by high V_R, or a combination of the two. Coefficients of variation provide a reasonable scale on which this question may be addressed.

The value of coefficients of variation for measuring variability depends on the degree to which they correct for whatever relationships exist between means and variances. Previous studies of phenotypic variation show that there is frequently a residual relationship between means and coefficients of variation (FISHER 1937; WRIGHT 1968; YABLOKOV 1974; ROHLF, GILMARTIN and HART 1983). Therefore, scale effects must be corrected for in analyses of coefficients of variation, as in any comparative analysis of variability. Three potential causes for such relationships have been identified. First, ROHLF, GILMARTIN and HART (1983) show that characters with small means tend to be measured with less accuracy than those with large means. This causes a negative relationship between means and CVs based on phenotypes. Second, ROHLF, GILMARTIN and HART also show that CVs of meristic traits are negatively correlated with the number of categories in the population, quite independent of any measurement errors. This predicts that both genetic and phenotypic CVs will be negatively correlated with the means of meristic traits.

The other known source of relationships between means and coefficients of variation is the relationship of the variances of parts and wholes. If the parts combine multiplicatively to produce the whole, LANDE (1977) has shown that the covariance of the parts determines the sign of the covariance between means and coefficients of variation. Positively correlated parts cause positive correlations of means and coefficients of variation. LANDE points out that since morphological traits are generally positively correlated, CVs of volumes will be larger than those of areas,

which will be larger than those of linear dimensions. Trends based on dimensionality arise for non-biological reasons, and so ought to be factored out before any analyses of variability. On the other hand, a life history trait, such as lifetime fecundity is in part due to the products of age specific mortalities, thus we can also predict that the CV of lifetime fecundity might be higher than that of fecundity over shorter periods, if the mortalities are generally positively correlated. Such relationships are biologically more interesting. If the parts combine additively to make up the whole, then the CV of the whole equals that of the parts only if the parts are perfectly correlated, and less than that of the parts otherwise (LANDE 1977; BRYANT 1986). BRYANT uses this to explain decreasing CVs through parts of ontogeny.

VARIATION DATA BASE

To reexamine evolvabilities and variabilities on the basis of mean-standardized measures, I reviewed quantitative genetic studies of animal populations. The data set was restricted to those estimates which met five criteria: (i) the characters may be assumed to be under either strong directional or weak stabilizing selection; (ii) the population studied may be assumed to be near genetic equilibrium; (iii) the population studied was outbred; (iv) the additive genetic variance was estimated using relatively unbiased methods; and (v) the variances are reported on an untransformed scale. Some of these criteria require further clarification.

First, I chose characters where there is a clear expectation of the kind of selection pressure the character is subjected to. Traits assumed to be under directional selection, which I term fitness traits, include many life history traits, and traits highly correlated with size during growth, such as weight at a particular age. Some life-history characters, such as development times of univoltine insects, were excluded from the analyses because they are more likely to be subject to optimizing selection. The traits which were assumed to suffer weak optimizing selection are primarily morphological traits which are not directly a function of growth rate. These are principally sizes of adult body parts, or meristic traits such as bristle numbers. Many of the studies included in this data set were also reviewed by ROFF and MOUSSEAU (1987) and MOUSSEAU and ROFF (1987). My definition of directionally selected and weak optimally selected traits roughly corresponds to their definitions of "life history" and "morphological" traits, with the exception of the sorts of traits noted above. Behavioral and physiological traits were excluded from this analysis because there is rarely a clear expectation about their relationship to fitness.

To meet criterion (ii), populations were presumed to be far from genetic equilibrium if they had recently been founded from inbred lines or very small numbers of individuals, had been subjected to artificial selection, or were formed by crossing or mixing individuals from more than one population. However, populations many generations beyond such events were included. Some Drosophila and Tribolium laboratory populations were included as a result of this exception. To meet criterion (iv), only studies which partitioned phenotypic variance using half-sib correlations, parent-offspring regression, or response to initial generations of selection were used. Studies using inappropriate nonrandom samples of individuals were also excluded. Maternal-offspring regressions were included despite the possibility of some bias due to maternal effects. Estimates of the same parameters by different methods were combined before analysis, as described below, to help mitigate the effects of different estimation methods.

Criterion (v) is necessary because the arguments advanced above for the use of coefficients of variation do not apply on alternative scales of measurement. For example, any transformation of fitness invalidates the simple prediction of Equation 8, since the fitness function would also need a corresponding transformation. In addition, calculating coefficients of variation on scales which potentially yield negative as well as positive means, such as log scales, renders coefficients of variation meaningless.

From 213 quantitative genetics papers reviewed, 153 included narrow-sense h^2 estimates for characters which met these criteria. Of these, 61 studies did not contain either character means or variances, or lacked both, precluding calculation of the appropriate coefficients of variation. From the 92 remaining papers, 842 acceptable sets of estimates of means and variances were extracted. A substantial number of estimates from within these papers were also excluded for the set of reasons detailed above. The data, and references to acceptable and excluded papers, are available from the author in printed form, or on disk as an ASCII, Quattro Pro or Lotus file.

Of the 201 studies cited by ROFF and MOUSSEAU (1987) and MOUSSEAU and ROFF (1987) which included "life history" or "morphological" characters, I was able to examine 147. Of these, 108 met criteria i-v, and 65 included the necessary means and variances to calculate coefficients of variation. The 27 other papers I include consist primarily of papers published between 1985 and 1990, and laboratory studies of Tribolium.

ANALYSES

Three potential measures of variability were calculated from these data: narrow sense h^2, CV_A, and the residual coefficient of variation

$$CV_R = 100 \sqrt{V_P - V_A}/\overline{X}. \qquad (13)$$

When the V_A estimate was negative, h^2 and CV_A were both set equal to zero. Analyses which included the negative estimates (with CV_A replaced by I_A) yield similar conclusions to those reported below. All three measures are far from normally distributed.

For analysis traits were classified on the basis of their relationship to body size as well as the sort of selection they are assumed to experience. Directionally selected traits which represent sizes of whole organisms or body parts during growth are designated "G" traits, while the remaining traits, primarily life history traits, are designated "L." Similarly, weakly selected traits which represent sizes of adult body parts are designated "S" traits. The remaining traits are primarily meristic in nature, and are designated "M" traits, although a few non-meristic traits which were demonstrably uncorrelated with adult size are also included in this category.

Many of the 842 sets are repeated estimates of the same parameters by different methods. Therefore median heritabilities and coefficients of variation for the same trait in the same population, sex, and environment were calculated before further analysis. This reduced the number of estimates to 400.

To correct for scale effects I use a "robust" regression algorithm which minimizes the absolute deviations from the regression line (PRESS *et al.* 1989, pp. 595–597), rather than the conventional least squares or major axis methods. The lack of normality of the data results in undue weight being given to outlier points under least squares, and renders significance tests invalid. Instead, significance tests of the robust regression were constructed from the proportion of positive slopes under bootstrap resampling, with a sample size of 200. Differences in slopes and intercepts among groups were assessed using randomization tests with a sample size of 1000. The probability values represent the proportion of randomizations where the absolute value of the parameter difference exceeds that based on the original data. Comparisonwise error rates were adjusted for multiple comparisons using the Šidák method (SOKAL and ROHLF 1981, p. 242).

COMPARISONS OF VARIATION

Evolvability comparisons: As an example of the usefulness of mean-standardized measures of evolvability, Table 1 presents variability measures for five commonly studied characters in *Drosophila melanogaster*. The last three columns present standardized evolvabilities assuming three different selection regimes: truncation, linear directional, and gaussian optimizing where the population mean is not at the optimum. In no case is heritability a good indicator of evolvability. Sternopleural bristles do generally have

high evolvabilities, as well as high h^2, but wing length, which has similar h^2, has far lower evolvability in each case. Comparison of CV_A and CV_R show that this is due to approximately fourfold lower mean-standardized variability. Similarly, fecundity generally has a high evolvability, because of its high additive genetic variance, in spite of relatively high residual variance.

Corrections for dimensionality: As noted above, the variances of traits are expected to be a function of their dimensionality. Only S and G classes contain substantial number of traits with different dimensionalities in this data set. In these classes, analysis of covariance of means and CVs on a log-log scale shows that the slopes are homogeneous, but the intercepts are significantly different among dimensionality classes for both CV_A and CV_R ($P < 0.001$). In each case, the differences among intercepts are not significantly different from the 1:2:3 relationship (linear:area:volume) expected if the linear dimensions are perfectly correlated (YABLOKOV 1974; LANDE 1977). Therefore, both CVs were divided by their dimensionalities before further analysis.

Variability comparisons: The possibility of nonindependence in this data set exists because a large proportion of the estimates come from a few species (44% from *D. melanogaster*) and a few studies (the five largest studies contribute 31% of the estimates). If species differ in the parameters being estimated, or individual studies estimate parameters with bias, the overall tests may be misleading. Therefore, analyses were carried out on medians of variables calculated within four different groupings of the data. In order of decreasing sample size these are medians of variables within (1) species, phenotype and study (the full 400 estimates); (2) species and phenotype; (3) study and class of trait; and (4) species and class of trait.

For grouping (1), there were significant regressions of h^2, CV_R, and CV_A on trait means. For the other three groupings, there are significant relationships between mean, and CV_R and CV_A. In each case, the estimated slopes are positive for h^2 and negative for CV_R. This may suggest that V_R may be disproportionately inflated at small scales due to measurement or scoring errors. CV_A has a negative relationship with mean which is most easily explained by the part to whole relationships described under VARIABILITY above.

These significant regressions require that analyses of differences among trait classes should proceed in a manner analogous to analysis of covariance. To compare the variability of two classes of traits, I first tested whether their slopes were homogeneous. I then tested whether the residuals from a joint regression differed in location. This approach yields ambiguous results if the slopes differ and the regression lines cross within the range of the data, or at higher means. For h^2 and

TABLE 1

Evolvabilities of representative traits in *D. melanogaster*

					Evolvability		
Trait	N^a	h^2	CV_R	CV_A	Truncation $V_A/(\bar{X}\sqrt{V_P})$	Linear $I_A \times 10^2$	Shift optimum $V_A/\bar{X}$
Sternopleural bristles	21	0.44	7.97	8.39	0.061	0.70	0.13
Wing length	31	0.36	2.09	1.56	0.009	0.02	0.05
Fecundity	12	0.06	39.02	11.90	0.035	1.42	5.38
Longevity	7	0.11	27.73	9.89	0.033	0.98	0.19
Development time	2	0.28	4.48	2.47	0.012	0.06	0.12

^{a} Number of studies median estimates calculated from.

CV_R there is a strong expectation that they will be influenced by measurement error at small scales, so it might be that the class with the more positive slope would have larger parameter values in the absence of error. However, since these data include traits measured on a wide variety of scales, and measurement error is not the only reason for a relationship between means and the other parameters, I prefer to regard such results as inconclusive.

As an example, the data for grouping 2, medians within phenotype within species, are graphed in Figure 1. Since the traits measured spanned so many orders of magnitude in their means and coefficients of variation, I log transformed these variables in all groupings of the data before analysis to increase the influence of points near the center of the distribution and aid the presentation. Similar results were obtained from analyses of untransformed data.

The analysis of grouping 2 is presented in some detail in Tables 2–4. In this grouping, for all three parameters, none of the individual trait class regressions are significantly different from 0 (results not shown). The analyses of h^2 in Table 2 show that none of the tests for homogeneity of slopes are significant when adjusted for the fact that six comparisons are made. Comparisons of the residuals from common regressions confirms results previously found for h^2 data (ROFF and MOUSSEAU 1987; MOUSSEAU and ROFF 1987). S traits have significantly higher h^2 than G or L traits, and M traits higher h^2 than L traits.

Table 3 presents analyses of CV_R, and here opposite results are obtained. As may also be seen in Figure 1b, L traits have the highest CV_R, while S traits perhaps the lowest. Comparison of S and M traits yields an ambiguous result. The results for CV_A, presented in Figure 1c and Table 4 are similar to those for CV_R. S traits have significantly lower CV_A than each of the other three classes.

A summary of the results obtained from data of all groupings is presented in Table 5. These results are similar to those for grouping 2, although the groupings with smaller sample sizes allow fewer opportuni-
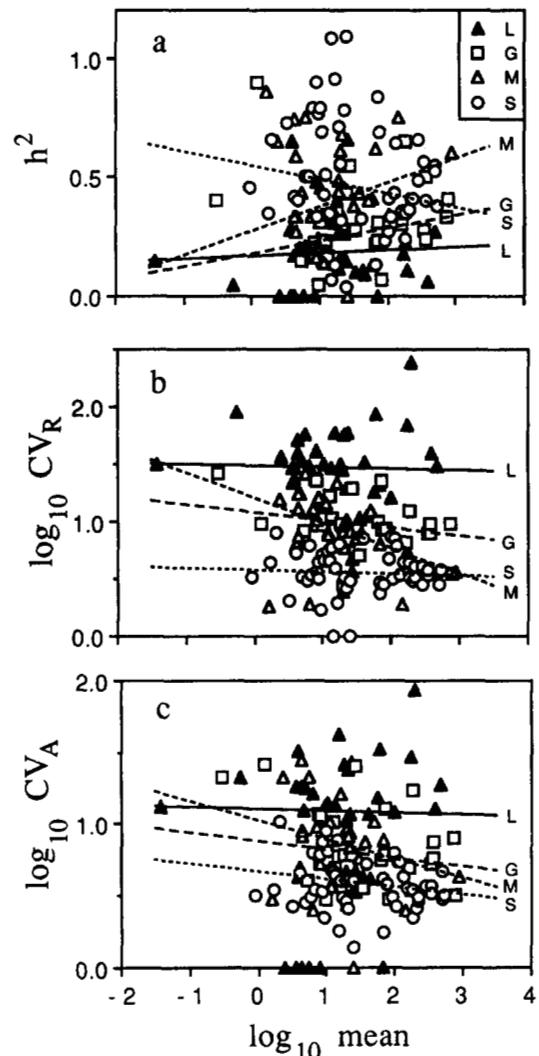

FIGURE 1.—Plot of h^2, CV_R and CV_A vs. trait means for each of the four trait classes. Points are medians for phenotypes within species. The lines are the robust regressions within trait classes.

ties for discrimination among classes. In almost all cases the above conclusions that trait classes with high h^2 are the trait classes with the lowest CVs are confirmed. When using CVs as an indicator of variability, there is certainly no evidence that traits which are

TABLE 2

Comparisons of h^2 among trait classes for data grouping 2

Class of trait	M	G	L
S			
Slopes differ?	†	NS	NS
S residual	0.073	0.095**	0.159**
Column residual	0.010	−0.098	−0.125
M			
Slopes differ?		NS	NS
M residual		0.099†	0.165**
Column residual		−0.020	−0.056
G			
Slopes differ?			NS
G residual			0.080†
Column residual			−0.009

The first row in each cell reflects the probability that slopes do not differ between groups. The numbers in the second and third rows are means of residuals from a common regression on the row and column class data. The upper value is for the row class, the lower for the column class. Significant differences among residuals are indicated in row 2.

† $P < 0.05$, uncorrected for simultaneous tests; * $P < 0.05$, corrected for simultaneous tests; ** $P < 0.01$, corrected for simultaneous tests; NS, not significant.

TABLE 3

Comparisons of CV_R among trait classes for data grouping 2

Class of trait	M	G	L
S			
Slopes differ?	*	NS	NS
S residual	−0.248**	−0.262**	−0.597**
Column residual	0.467	0.809	1.166
M			
Slopes differ?		NS	NS
M residual		−0.196†	−0.627**
Column residual		0.144	0.539
G			
Slopes differ?			NS
G residual			−0.390**
Column residual			0.364

See Table 2 for explanation.

TABLE 4

Comparisons of CV_A among trait classes for data grouping 2

Class of trait	M	G	L
S			
Slopes differ?	NS	NS	NS
S residual	−0.089**	−0.134**	−0.164**
Column residual	0.382	0.517	0.601
M			
Slopes differ?		NS	NS
M residual		−0.114	−0.303
Column residual		0.087	0.012
G			
Slopes differ?			NS
G residual			−0.258
Column residual			−0.124

See Table 2 for explanation.

TABLE 5

Summary of trait class comparisons for data groupings 1–4

		Parameter		
Grouping	N	h^2	CV_R	CV_A
1	400	L G M?S	S?M G L	S?M G L
2	148	L G M S	S?M G L	S M G L
3	102	L G M S	S M G L	S M G L
4	50	L G M S	S M G L	S L M G

Trait classes are listed in order from smallest to largest. Horizontal lines connect classes which are not significantly different at $P < 0.05$ (adjusted for multiple comparisons) by the analysis of covariance-randomization tests described in the text. Order of classes within homogeneous sets does not necessarily reflect relative ranking. A question mark between classes represents an ambiguous result where classes have significantly different slopes, and the regression lines cross within or above the range of the data. N is the sample size of each data grouping.

TABLE 6

Spearman rank-correlations among variability parameters

Param- eter	CV_R	CV_A	I_A	$V_A/\bar{X}$
h^2	−0.5266****	0.0637	0.0650	0.1879***
CV_R		0.7217****	0.7167****	0.3030****
CV_A			0.9980****	0.5754****
I_A				0.5793****

*** $P < 0.001$; **** $P < 0.0001$.

likely to be under weak stabilizing selection are more variable than traits more closely related to fitness.

The position of G traits in these regressions seems to reflect the fact that CVs decrease as animals approach adulthood (BRYANT 1986). As shown in Figure 1, G tends to diverge from L and approach S and M at large scales. Pairwise comparisons never reveal significant slope differences involving G and L, but this may simply be due to insufficient sample size. This reveals a strong limitation of these analyses, in that a tremendous variety of traits, measured on different scales are being compared. With these data it is not possible to correct for such factors as age or experimental error directly. No more than a very broad suggestion of the patterns of variability should be gleaned from these analyses.

Table 6 presents the rank correlations between the various measures of evolvability and variability, calculated for data grouping 1. Heritability is not significantly correlated with CV_A or I_A, although it is weakly correlated with $V_A/\bar{X}$. Thus h^2 is not a suitable surrogate for other measures of genetic variability. As expected, CV_A, I_A, and $V_A/\bar{X}$ are all moderately to highly correlated, and h^2 is significantly negatively correlated with CV_R. CV_R is also highly positively correlated with CV_A and I_A, which suggests parallels be-

tween the processes shaping environmental and genetic variances.

DISCUSSION

There are two distinct kinds of questions that require comparing standardized estimates of genetic variation: first how fast will a character respond to a given selective pressure; and second how much variability is maintained in relation to the evolutionary forces which act on it. In both these cases, evolvability and variability, I suggest that the genetic coefficient of variation CV_A provides a great deal of relevant information. However, while it is straightforward to define measures of evolvability appropriate to any particular circumstance, it is not possible to define a single measure appropriate for all circumstances. For characters subject to strong stabilizing selection (ω^2 of order V_P or less), V_P does not drop neatly out of the predicted response to selection. Even the ranking of selection responses changes with the fitness function. Comparing evolvabilities of traits likely to experience very different selection regimes is pointless. For variability it is not possible to be so precise about what measure is most appropriate, although it seems clear that coefficients of variation are useful, at least on a heuristic level.

My review of coefficients of variation of strongly and weakly selected characters suggests that fitness traits have higher standardized genetic variability. Thus, the low h^2 of fitness is best explained by proportionately extremely large V_R values. In retrospect, this is not surprising, as, for example, there will always be very large CV_P for fitness in cases where juvenile mortality is high. This suggests one explanation for the large residual variability of fitness components. Fitness traits tend to integrate variability over the lifetime of the organism, and result from all the selective forces acting on other characters (PRICE and SCHLUTER 1991). Hence it is likely that a larger number of loci, a wider range of environmental variables, and more interactions among them will affect fitness traits relative to morphological traits. This suggestion is consistent with CABANA and KRAMER's (1991) review of I calculated from horizontal life-history studies. They show that random age-specific survival alone can account for much of the rise in I with age.

Another factor which may contribute to the higher CV_R of fitness components is higher nonadditive genetic variance. Some simple models of balancing selection predict little or no additive variance for fitness, but plenty of dominance or epistatic variance. However, in the best studied cases, nonadditive genetic variance never accounts for a large proportion of phenotypic variance (MATHER and JINKS 1982; MUKAI 1988). Therefore, it seems more likely that the majority of the higher CV_R values of fitness traits is due

to environmental causes or genotype-environment interactions.

A very different potential explanation for the high CV_R of fitness components is the selection pressure on V_P itself. Stabilizing selection favors genotypes which minimize phenotypic variance of traits, but phenotypic variance of traits under directional selection may be favored, neutral, or disfavored for a given trait mean (LANDE 1980). The fitness of alleles which affect only the variance of a trait is approximately determined by the second derivative of the fitness function. If this derivative is near zero, as with linear fitness functions, then phenotypic variance will be a neutral character. Thus, the variance of fitness components may on the average experience weaker selection than that of other traits. If the fitness function is concave upward, an accelerating function of trait value (e.g., SCHAFFER and SCHAFFER 1977), variance may actually be favored.

The differences found in CV_A are perhaps of more fundamental interest. One potential explanation is that the genetic variance in composite characters such as fitness will be the result of variation in a large proportion of the genome. Under most models of the maintenance of genetic variance, the genetic variance in a trait will tend to be correlated with the number of loci which affect it (HOULE 1991). Increased genetic variance of fitness components can be explained simply if the number of loci affecting fitness is large enough to compensate for the small variance in fitness that each locus is likely to contribute. One testable prediction of this hypothesis is that the increase in CV_A due to new mutation should be larger for fitness components than for morphological characters. Alternatively, loci with effects on fitness components may be particularly likely to have protected polymorphisms. This may occur through antagonistic pleiotropy (ROSE 1982, 1985), or genotype-environment interactions (MUKAI 1988; GILLESPIE and TURELLI 1989). There is no direct evidence for such protected polymorphisms, however.

Heritability has long assumed an inordinately central role in experimental quantitative genetics. A measure of this emphasis is that of the quantitative genetic studies reviewed for this paper which otherwise met my criteria for inclusion, only a handful did not report h^2 values, while almost 40% failed to report either trait means or variances. This has obscured much useful information. While I make a case that coefficients of variation convey important information about relative variabilities and evolvabilities, no set of summary ratios can be more useful than the estimated moments of the phenotype or genotype distribution. Authors should always report these with the results of quantitative genetic experiments, so that readers have the important fundamental statistics before them.

A similar argument suggests that means and variances of fitness components should be reported on an untransformed scale when natural selection is of interest. For example, the natural scale for measuring fecundity is that of zygotes produced, and non-linear transformations can only make it more difficult to predict relative fitnesses. While it may be convenient to, for example, transform to normality so that F tests are unbiased, the resulting statistics may be of little use for predicting response to selection or assessing variability. Now that resampling tests of statistical significance, like jackknifing and bootstrapping, are available, much of the statistical rationale for transformations to normality is gone.

While heritability may be of little use for comparative purposes, it is necessary for making predictions of absolute responses to selection. In many organisms it is possible to estimate selection differentials for particular life stages (ARNOLD and WADE 1984), in which case knowledge of h^2 allows the response to selection to be predicted directly. Because of its resemblance to F ratios, h^2 is an indicator of the effort needed to demonstrate that a trait possesses statistically significant V_A, and the confidence one can place in such an estimate. On the other hand, a consequence of this is that minimizing environmental variance is a common experimental goal, thereby systematically inflating h^2. Coefficients of variation are less susceptible to this source of experimental basis. The fact that h^2 values are frequently similar for different morphological traits may be due either to common evolutionary processes which lead V_A and V_R to be similar, or to a choice of experimental methods which lead to the same results.

A related effect of an emphasis on estimating h^2 is that experiments are often designed to decrease the variance of h^2, although many of the questions a study seeks to address do not depend directly on h^2. For example, it is frequently a primary goal to determine whether a trait possesses significant V_A. When parent-offspring regression or artificial selection is used, there is no conflict between the goals of estimating V_A and h^2. However, when sib data are used, the two goals conflict if the total number of individuals which can be measured limits the size of the experiment. An optimal design for estimating V_A maximizes the number of families in an experiment (SOKAL and ROHLF 1981), while the optimal design for h^2 involves balancing family size and family number (ROBERTSON 1959).

Finally, these results point to a fundamental experimental problem in evolutionary quantitative genetics. While it is now a truism that most traits possess additive genetic variance, it is still difficult to obtain precise estimates of genetic variance for traits with low h^2. Thus, for fitness components it is impractical to make estimates of evolvabilities or variabilities which meaningfully restrict the values of potential causal parameters. Simple point estimates of moments of fitness traits are unlikely to shed much further light on the maintenance of genetic variance in fitness. On the other hand, artificial selection experiments designed to preclude the operation of natural selection might more readily indicate the relative evolvability and variability of fitness components.

LITERATURE CITED

ARNOLD, S. J., and M. J. WADE, 1984 On the measurement of natural and sexual selection: theory. Evolution 38: 709–719.

BARTON, N. H., 1990 Pleiotropic models of quantitative variation. Genetics 124: 773–782.

BARTON, N. H., and M. TURELLI, 1989 Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23: 337–370.

BRYANT, E. H., 1986 On the use of logarithms to accommodate scale. Syst. Zool. 35: 552–559.

BULMER, M. G., 1980 The Mathematical Theory of Quantitative Genetics. Oxford University Press, Oxford.

CABANA, G., and D. L. KRAMER, 1991 Random offspring mortality and variation in parental fitness. Evolution 45: 228–234.

CHARLESWORTH, B., 1984 The evolutionary genetics of life histories, pp. 117–133 in Evolutionary Ecology, edited by B. SHORROCKS. Blackwell, Oxford.

CHARLESWORTH, B., 1987 The heritability of fitness, pp. 21–40 in Sexual Selection: Testing the Alternatives, edited by J. BRADBURY and M. B. ANDERSSON. John Wiley & Sons, London.

CHARLESWORTH, B., 1990 Optimization models, quantitative genetics, and mutation. Evolution 44: 520–538.

CROW, J. F., 1958 Some possibilities for measuring selection intensities in man. Hum. Biol. 30: 1–13.

FALCONER, D. S., 1989 Introduction to Quantitative Genetics, Ed. 3. Longman Scientific & Technical, Harlow.

FISHER, R. A., 1930 The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

FISHER, R. A., 1937 The relation between variability and abundance shown by the measurements of the eggs of British nesting birds. Proc. R. Soc. B 122: 1–26.

FISHER, R. A., 1951 Limits to intensive production in animals. Br. Agric. Bull. 4: 217–218.

GILLESPIE, J. H., and M. TURELLI, 1989 Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121: 129–138.

GUSTAFSSON, L., 1986 Lifetime reproductive success and heritability: empirical support for Fisher's fundamental theorem. Am. Nat. 128: 761–764.

HALDANE, J. B. S., 1949 Suggestions as to quantitative measurement of rates of evolution. Evolution 3: 51–56.

HOULE, D., 1989 The maintenance of polygenic variation in finite populations. Evolution 43: 1767–1780.

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

ISTOCK, C. A., 1983 The extent and consequences of heritable variation for fitness characters, pp. 61–96 in Population Biology: Retrospect and Prospect, edited by C. R. KING and P. S. DAWSON. Columbia University Press, New York.

KIMURA, M., AND J. F. CROW, 1978   Effect of overall phenotypic selection on genetic change at individual loci. Proc. Natl. Acad. Sci. USA **75:** 6168–6171.

LANDE, R., 1977   On comparing coefficients of variation. Syst. Zool. **26:** 214–217.

LANDE, R., 1980   Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. **116:** 463–479.

MATHER, K., and J. L. JINKS, 1982   *Biometrical Genetics*, Ed. 3. Chapman & Hall, London.

MAYO, O., R. BÜRGER and C. R. LEACH, 1990   The heritability of fitness: some single gene models. Theor. Appl. Genet. **79:** 278–284.

MOUSSEAU, T. A., and D. A. ROFF, 1987   Natural selection and the heritability of fitness components. Heredity **59:** 181–197.

MUKAI, T., 1988   Genotype-environment interaction in relation to the maintenance of genetic variability in populations of *Drosophila melanogaster*, pp. 21–31 in *Proceedings of the Second International Conference on Quantitative Genetics*, edited by B. S. WEIR, E. J. EISEN, M. J. GOODMAN and G. NAMKOONG. Sinauer, Sunderland, Mass.

O'DONALD, P., 1970   Change of fitness by selection on a quantitative character. Theor. Popul. Biol. **1:** 219–232.

PRESS, W. H., B. P. FLANNERY, A. A. TEUKOLSKY and W. A. VETTERLING, 1989   *Numerical Recipes in Pascal.* Cambridge University, Cambridge, U.K.

PRICE, G. R., 1972   Fisher's "fundamental theorem" made clear. Ann. Hum. Genet. **36:** 129–140.

PRICE, T., and D. SCHLUTER, 1991   On the low heritability of life history traits. Evolution **45:** 853–861.

ROBERTSON, A., 1959   Experimental design in the evaluation of genetic parameters. Biometrics **15:** 219–226.

ROFF, D. A., and T. A. MOUSSEAU, 1987   Quantitative genetics and fitness: lessons from *Drosophila.* Heredity **58:** 103–118.

ROHLF, F. J., A. J. GILMARTIN and G. HART, 1983   The Kluge-Kerfoot phenomenon—a statistical artifact. Evolution **37:** 180–202.

ROSE, M., 1982   Antagonistic pleiotropy, dominance, and genetic variation. Heredity **48:** 63–78.

ROSE, M. R., 1985   Life history evolution with antagonistic pleiotropy and overlapping generations. Theor. Popul. Biol. **28:** 342–358.

SCHAFFER, W. M., and M. V. SCHAFFER, 1977   The adaptive significance of variation in reproductive habit in the Agavaceae, pp. 261–276 in *Evolutionary Ecology*, edited by B. STONEHOUSE and C. M. PERRINS. Macmillan, London.

SOKAL, R. R., and F. J. ROHLF, 1981   *Biometry*, Ed. 2. W. H. FREEMAN, San Francisco.

TURELLI, M., 1984   Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theor. Popul. Biol. **25:** 138–193.

TURELLI, M., and N. H. BARTON, 1990   Dynamics of polygenic characters under selection. Theor. Popul. Biol. **38:** 1–57.

WRIGHT, S., 1968   *Evolution and the Genetics of Populations. Vol. 1. Genetics and Biometrical Foundations.* University of Chicago Press, Chicago.

YABLOKOV, A. V., 1974   *Variability of Mammals.* Amerind Publishing, New Delhi.