Genetica **102/103:** 241–253, 1998.

How should we explain variation in the genetic variance of traits?

David Houle
Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, Ontario, Canada M5S 3G5 (Phone: (416) 978-1014; Fax: (416) 978-8532) E-mail: dhoule@zoo.toronto.edu)

Key words: genetic variance, mutation, selection, canalization

Abstract

Recent work has called attention to large differences among traits in the amount of standardized genetic variance they possess. There are four general factors which could play a role in causing this variation: mutation, elimination of deleterious variation, selection of favorable alleles, and balancing selection. Three factors could directly influence the mutational variability of traits: canalization, the mutational target size, and the timing of trait expression. Here I carry out simple tests of the importance of some of these factors using data from *Drosophila melanogaster*. I compiled information from the literature on the mutational and standing genetic variances in outbred populations, inferred the relative mutational target size of each trait, its a timing of expression, and used models of life history to calculate fitness sensitivities for each trait. Mutational variation seems to play an important role, as it is highly correlated with standing variance. The target size hypothesis was supported by a significant correlation between mutational variance and inferred target size. There was also a significant relationship between the timing of trait expression and mutational variance. These hypotheses are confounded by a correlation between timing and target size. The elimination and canalization hypotheses were not supported by these data, suggesting that they play a quantitatively less important role in determining overall variances. Additional information concerning the pleiotropic consequences of mutations would help to validate the fitness sensitivities used to test the elimination and canalization hypotheses.

Introduction

Recent reviews demonstrate that quantitative traits can be extremely different in the amount of variation they display in outbred, unselected populations (Roff & Mousseau, 1987; Mousseau & Roff, 1987; Houle, 1992). Most strikingly, on a mean standardized scale, life history traits on average have an order of magnitude more additive genetic variance than do morphological traits (Houle, 1992). On the more traditional heritability scale, a substantially lower proportion of the phenotypic variance in life history traits is additive genetic (Roff & Mousseau, 1987; Mousseau & Roff, 1987). Taken together, these observations suggest that both genetic and non-genetic variances vary between traits, and that there is a relationship between genetic and non-genetic variance. To me the most fundamental question these data raise is, Why do traits differ in their levels of genetic variance?

This question may have answers at two levels. First, traits may differ in the generation of new variation, a property which Wagner, Booth and Bagheri-Chaichian (1997) define as the *variability* of a trait. This property is determined by the functional architecture of the trait (Houle, 1991), that is length and nature of the genetic pathways from genotype to phenotype, the mutational properties of those loci, and their interactions with other loci in creating the phenotype. Second, natural selection and other population genetic processes will influence the fate of this newly arising variation, leading to either elimination of variant alleles, fixation of new alleles, or balanced polymorphisms.

I have elaborated this basic dichotomy into a catalogue of seven simple hypotheses that can account for variation in trait genetic variances, listed in Table 1.

Table 1. Hypotheses that can explain differences in genetic variance between traits

	Hypothesis	Explanation	Predictions
Hypotheses concerning the maintenance of genetic variance	Mutation	Mutation helps determine standing genetic variance.	CV_M positively correlated with CV_A.
	Elimination	Deleterious alleles influencing more strongly selected traits are more rapidly eliminated.	V_A/V_M negatively correlated with fitness sensitivity.
	Selective sweep	Advantageous alleles on their way to fixation increase genetic variance.	No clear prediction. May be more likely if elimination rejected.
	Balance	Traits differ in likelihood of balanced polymorphism.	No clear prediction. May be more likely if mutation or elimination rejected.
Hypotheses concerning mutational variance	Canalization	Genotypes that affect the genetic variance of other loci are selected.	CV_I negatively correlated with fitness sensitivity?
	Mutational target	Traits with complex functional architecture are larger targets for mutation.	CV_I positively correlated with CV_M. CV_M and CV_I positively correlated with mutational target size.
	Timing	Traits expressed later in life inherit variation from earlier stages.	CV_M and CV_I correlated with timing of expression.

The first four concern the maintenance of genetic variance per se. The elimination hypothesis is based on mutation-selection balance, where strong selection will eliminate deleterious alleles more rapidly than weak selection. The selective sweep hypothesis is based on the idea that traits differ in the degree to which they are at genetic equilibrium. When a trait is not at equilibrium, advantageous alleles on their way to fixation can temporarily create much larger amounts of genetic variance than deleterious alleles maintained by mutation-selection balance. If a steady supply of mildly advantageous mutants sweep through the population, this can lead to a stable, high level of genetic variance (Kimura & Ohta, 1971; Bürger, 1993). The balance hypothesis holds that traits differ in their propensity to support balanced polymorphisms. For example, Rose (1982) proposed that traits which are strongly selected will often be involved in trade-offs with other traits, leading to an enhanced probability of balanced polymorphisms, a conclusion opposite to that predicted by the elimination hypothesis.

The mutation hypothesis simply asserts that levels of variation are related to the amount of mutational variance a trait experiences. This may obviously be the case under mutation-selection balance if mutation and selection do not have precisely complementary effects. Less obviously, a large supply of new variation may also lead to more rapid selective sweeps. Even if balancing selection is an important factor preserving variation, mutation may still play a causal role if the conditions favoring such polymorphisms change frequently, leading to the loss of old polymorphisms and the gain of new ones.

The final three hypotheses concern processes that can determine the levels of mutational variance. The canalization hypothesis proposes that selection on trait variation may also lead to the fixation of modifiers that alter the effects of alleles at other loci, leading to canalization or decanalization of traits (Waddington, 1957; Lande, 1980; Wagner, Booth & Bagheri-Chaichian, 1997). The mutational target hypothesis notes that each gene which is capable of affecting trait expression will be the target of independent mutational events. Thus, traits with complex functional architectures are influenced by many loci, and hence affected by mutational damage in a relatively large proportion of the genome (Houle, Morikawa & Lynch, 1996). The timing hypothesis proposes that the expression of a trait later in life will display more variance than the same trait early in life, due to the cumulative effects of the same allelic variants. The mutational target and timing hypotheses for variation in variance are related by the assumption that variance in a trait is compounded from variance in the more-or-less independent component processes that influence trait expression. They are also to some extent confounded with each other, as

one of the factors that must influence mutational target size is the timing of expression of the loci involved. A gene expressed only early in life can still affect traits manifested only late in life, for example by damaging the overall health of the organism.

Both of these variance-compounding processes could in theory be balanced by the evolution of canalization. In this context, canalization would take the form of non-independence in trait expression that would damp out the effects of component or earlier events. A well-known example is compensatory growth in mammals, where variance in growth rate early in life tends to be compensated for by later variation in growth as maturity approaches (Riska, Atchley & Rutledge, 1984). On the other hand, other processes also lead to the evolution of target size, such as gene duplication and divergence. Mutational target size may also be increased through the evolution of condition dependence (Rowe & Houle, 1996).

It is important to note that these seven hypotheses for variation in genetic variances are not exclusive. All can be correct to some degree, although the elimination, selective sweep and balance hypotheses cannot simultaneously apply to the same loci. Although it is possible that all the hypotheses are correct, three have been advocated as primary by different research groups in recent years. The elimination hypothesis is the basis for the traditional explanation of the low heritability of life history traits (Falconer, 1981; Roff & Mousseau, 1987; Mousseau & Roff, 1987). The idea is that all else being equal, stronger selection will lead to lower levels of genetic variance than weak selection.

I recently showed that the additive genetic coefficients of variation of life-history traits are much larger than those of morphological traits, inconsistent with the simplest versions of the elimination hypothesis (Houle, 1992). This led me to propose the mutational target hypothesis as the primary explanation for the levels of variation. Consistent with this, the coefficient of mutational variance of life history traits is significantly larger than for morphological traits (Houle, Morikawa & Lynch, 1996).

Finally, the canalization hypothesis has recently been championed by Stearns and Kawecki (1994; Stearns, Kaiser & Kawecki, 1995), on the basis of their studies of the effects of single P-element insertions in *Drosophila melanogaster*. They showed a negative correlation between the fitness sensitivity of a series of life-history traits and the genetic variance generated by insertions. The same relationship was also observed between sensitivity and environmental variance. Stearns, Kaiser and Kawecki (1995) noted that their results were also consistent with the timing hypothesis.

In this paper, I will use data from *Drosophila melanogaster* to test predictions of five of the hypotheses listed in Table 1. *D. melanogaster* is the best choice for these comparisons because three types of data on trait variances and variabilities are available for a number of traits: 1) the amount of variation produced by a transposable element insertion (V_I); 2) the rate that trait variance increases as a result of spontaneous mutations (V_M); and 3) the standing additive genetic variance in outbred populations (V_A). In order to compare traits measured in different units, I chose to place each of these estimates on a mean standardized scale as a coefficient of variation (CV), as the strength of selection against variant genotypes is often determined relative to the mean. Also of interest is the ratio V_A/V_M, which is the number of generations of mutation required to supply the observed trait variances. Under mutation-selection balance, this corresponds to the average persistence time of a cohort of mutations introduced in the same generation (Crow, 1979; Houle, Morikawa & Lynch, 1996). The results of these tests support the mutation, mutational target, and timing hypotheses, and suggest that neither the canalization nor elimination hypotheses have quantitatively dominant effects on trait mutational or standing variances.

The data

Tables 2 and 3 show the data used for each of the 10 traits for which at least some data on mutational variances are available. A number of these traits require definition. Fitness is an index based on several generations of competition between homozygous wild-type flies and heterozygotes for a marked, recessive lethal bearing chromosome (Sved & Ayala, 1970; Houle et al., 1992). It summarizes variation in most aspects of life history under competitive conditions in the laboratory. Viability is the probability that a fly survives from egg to adulthood. Productivity is the number of adult flies produced by a mated female, and thus measures viability times fecundity. Size was measured as dry weight in the study used to estimate CV_I (Stearns, Kaiser & Kawecki, 1995), but as wing length by the studies used to estimate CV_M and CV_A (see Houle, Morikawa & Lynch, 1996 for references).

Table 2. Trait variation, mutational target size, and timing of expression

Trait	CV_M	CV_I	CV_A	V_A/V_M	Time	Target size
Abdominal bristles	0.24	2.04	6.11	646.01	1	1
Sternopleural bristles	0.39	3.01	7.38	367.97	1	1
Adult size[a]	0.14	1.92	1.54	127.91	1	2
Development time	0.43	1.20	2.47	33.72	1	2
Viability	1.57	38.50	10.40	43.75	1	3
Early fecundity	1.22	19.74	8.81	52.12	2	4
Late fecundity	2.56	50.46	28.79	126.23	3	4
Longevity	1.35	14.44	9.06	45.22	3	4
Productivity	2.24	–	–	–	2	4
Fitness	4.15	–	–	–	2	5

[a] Variation in adult size was assessed from wing dimensions, except for CV_I, where dry weight was used.

Table 3. Fitness sensitivities, resulting from 1% and 10% changes for each trait

Trait	1%		10%	
	no tradeoff	Roff	no tradeoff	Roff
Abdominal bristles	0.03	0.03	2.94	2.94
Sternopleural bristles	0.04	0.04	4.19	4.19
Wing length – growth	2.96	2.98	27.10	27.27
Wing length – development time	–	0.04	–	4.66
Development time	1.00	0.05	9.52	4.49
Early fecundity	0.96	1.13	9.63	11.18
Late fecundity	0.04	0.00	0.37	0.03
Longevity	0.00	0.00	0.03	0.00
Viability	1.00	1.13	10.00	11.21
productivity, fitness				

Variation and variability

The variation data are summarized in Table 2. Three types of data on trait variation will be used in this paper. First, data on mutational coefficients of variation (expressed as a percentage of the mean, $CV_M = 100\sqrt{V_M/X}$) and the corresponding standing additive genetic coefficients of variation (CV_A) are taken from Houle, Morikawa and Lynch (1996). In that paper, we lumped all data on fecundities into single CV_M and CV_A values. To examine the relationship between selection and trait variances, it may be more appropriate to examine age-specific fecundities. Following Stearns and Kawecki (1994), I split the fecundity schedules into days 1–14 of adult life as early fecundity, and days 15–40 as late fecundity. Coefficients of variation were calculated by taking median values estimated during the appropriate period. Data on lifetime fecundity was dropped from the analyses.

In Houle, Morikawa and Lynch (1996), we did not include estimates of CV_M for development time. These data are available based on two analyses of an experiment by Mukai and associates where mutations accumulated in the near absence of natural selection over 60 generations (Mukai & Yamazaki, 1971; Yoshimaru & Mukai, 1985). Mukai and Yamazaki (1971) give the accumulated mutational variance in development time at two generation times, but give the mean development time only as a deviation from the control mean. Yoshimaru and Mukai (1985) report means and variances on a transformed scale y = ln(1/d), where d is development time standardized to a control mean of 1, and y the transformed value. These values were used to estimate the mean and variance on the original scale through the following Taylor expansions:

$$\overline{d} \approx \frac{1 + \frac{V_y}{2} + \frac{V_y^2}{8}}{\exp(\overline{y})} \qquad V_d \approx \frac{V_y + 7\frac{V_y^2}{4}}{\exp(\overline{y})^2}$$

To estimate V_M, the estimates of the variance were regressed on generation time, with the line constrained to pass through the origin.

Data on the genetic variance generated by P-element insertions (V_I) were obtained from four sources (Eanes et al., 1988; Mackay, Lyman & Jackson, 1992; Stearns, Kaiser & Kawecki, 1995; Lyman et al., 1996). If insertions cause a change in the mean of the trait, that is positive and negative effects do not balance each other, then these estimates of variation are biased downwards, as it is the control mean that is the appropriate point about which moments should be calculated. Unfortunately, Stearns, Kaiser and Kawecki (1995) give no data on genotypes that contain no new insertions, which would be the appropriate control. Lyman et al. (1996) found variable effects of insertion on trait means, which they attributed to a marker gene inserted along with the P-element in their study. Therefore neither the Stearns, Kaiser and Kawecki nor the Lyman et al., data can be corrected to the variance around the control mean. Accordingly, I have not corrected any estimates for this source of bias. The coefficient of variation for weight was divided by 3 to correct for the fact that weight is proportional to volume (Lande, 1977; Houle, 1992) and allow comparisons with the CV_M data available for wing length.

Two studies give useable estimates of CV_I values for egg-to-adult viability (Eanes et al., 1988; Lyman et al., 1996). Eanes et al. (1988) reported data on egg-to-adult viability for males carrying varying numbers of unselected P-inserts on the X-chromosome. Therefore, the among-line genetic variance includes a contribution from the variance in insertion number, as well as the desired variance of the effect of insertions. In this data set, the mean number of insertions was 3.26, with a variance of 8.52. The among-line variance, V_l, was corrected to correspond to the variance due to a single insertion, V_a, as

$$V_a = \frac{V_l - 8.52a^2}{3.26},$$

where *a* is the average homozygous effect of an insertion, about 1% for this data set. The data of Lyman et al. (1996) make it clear that the X chromosome contributes substantially less variance per insertion than the two autosomes. To get a single combined estimate of CV_I for viability, Eanes et al.'s estimate was combined with the data of Lyman et al. (1996) for the X-chromosome. The autosomal and X-chromosome estimates were averaged by weighting the two autosomes twice as heavily as that for the X, to compensate for the difference in size between chromosomes.

Mackay, Lyman and Jackson (1992) report data on V_M per insertion due to homozygous P-element insertions on the third chromosome for sternopleural and abdominal bristles. These were averaged with the estimates from Lyman et al. (1996). Mackay, Lyman and Jackson also measured the relative viability of these lines, but unfortunately there was a non-linear relationship between viability and insertion number, which makes it difficult to extrapolate their results to a comparable estimate of the variance in effect of a single insertion.

Target size and timing of expression

Also shown in Table 2 are indices of mutational target size and timing of expression. For timing, traits are readily classified into those determined at eclosion (development time, bristle traits, viability, wing length), early in adulthood (early fecundity, productivity, fitness), and late in adulthood (longevity and late fecundity).

Inferring the mutational target size of the functional architecture underlying traits is more problematic as we have little direct information to go on. The ranking of traits shown in Table 2 was derived from the following arguments. Bristle numbers in natural populations are influenced primarily by known neurogenic loci (Lai et al., 1994; Long et al., 1995). In addition, they are affected by alleles with large effects on growth (i.e., *bobbed*). While this may be hundreds or possibly thousands of loci, it is nevertheless probably a subset of loci active during the larval phase. These traits are assigned the lowest index of target size, 1. Body size and development time should be more complex, because all loci that potentially have effects on larval growth rate and the timing of pupation will influence both traits, and growth rate must be influenced by many aspects of metabolism, development, and behavior. In addition, development time must be influenced by additional genes determining the timing of eclosion. On the other hand, wing length is also influenced by genes involved in the development of the wing per se. Therefore, these traits are both assigned rank 2.

Direct evidence from saturation mapping of portions of the *D. melanogaster* genome suggest that approximately 5,000 genes are capable of mutating to recessive lethal alleles (Judd, Shen & Kaufmann, 1972), so this sets a lower limit to the target size of viability. The mutational correlation between devel-

opment time and viability is not significantly different than 1 (Mukai & Yamazaki, 1971; Yoshimaru & Mukai, 1985), arguing that they share a large proportion of loci with small effects on each. Any locus that influences growth rate is likely to have an effect on viability, as well, because this must affect competitive ability and development time, as noted above. On the other hand, there are likely to be loci influencing viability that are not involved in growth rate. This suggests that viability is both genetically complex and more complex than development time and size.

A large number of studies make clear that fecundity is strongly dependent on growth rate and size in *D. melanogaster* (Roff, 1981; Zwaan, Bijlsma & Hoekstra, 1995). Similarly, it seems very likely that many loci with effects on viability have pleiotropic effects on condition, and therefore on fecundity as well. In addition, there are a large number of loci uniquely involved in oogenesis, and probably some additional functional architecture unique to more general aspects of adult metabolism and physiology. This argues for a higher target size ranking for fecundity than viability. It has been argued that more loci may be involved in determining fecundity late in adult life that early in life. This seems unlikely, as it is difficult to imagine what genetic pathways could be involved in egg production late in life than are not involved early in life. This is supported by the high positive mutational correlation between fecundity early and late in life (Houle et al., 1994). I therefore ranked fecundity at both times as equally complex. By these arguments, productivity, the product of female fecundity times the viability of her offspring, should be ranked the same as fecundity per se.

A similar argument to that for fecundity argues that adult lifespan should be influenced by genes active in the pre-adult period affecting growth and viability. In addition, there are certainly additional processes involved in adult metabolism and maintenance, as noted above. Some of these may not affect fecundity. The pathways involved in oogenesis that affect fecundity probably do not all affect longevity. The balance of these factors is unclear, so I ranked longevity as equal in target size to fecundity and productivity.

Finally, every locus in the genome must be capable of influencing fitness, so fitness is given the top complexity rank of 5.

Fitness sensitivities

I calculated fitness sensitivities (Δw) for each of the traits with mutational data, using a similar approach to Stearns and Kawecki (1994). Stearns and Kawecki intentionally utilized a life history model that did not incorporate fitness component tradeoffs created by pleiotropic effects, which I call the 'no tradeoffs' model. With this model, Stearns and Kawecki showed that their ranking of the fitness sensitivities was robust to reasonable alterations in juvenile and adult mortality, and to the definition of fitness as R_0 or r. On the basis of this, they implied that their fitness sensitivities would hold up to other sorts of changes in assumptions. I also calculated fitness sensitivities with Roff's (1981) model of *D. melanogaster* life history. This differs from the 'no tradeoffs' model principally in assuming that development time and adult size are functionally related. I also used a model based on laboratory culture of *D. melanogaster* (Houle and Rowe, unpublished) that yielded very similar results to the Roff model.

Life history models

Fitness in the Stearns and Kawecki model was calculated as

$$R_0 = e^{-m_j \alpha} cw \int_2^{\omega} e^{-m_a x} F(x) dx \,,$$

where m_j is the daily mortality during the pre-adult period, m_a is the daily mortality rate during the adult period, α is the development time from egg to adulthood, ω is the maximum adult life span, w is weight, c is a constant determining the proportionality between weight and lifetime fecundity, and $F(x)$ is proportional to the fecundity on day x of adult lifespan. The daily fecundity function is

$$F(x) = [1 - e^{-b(x-2)}]e^{-a x}$$

(MacMillan, Fitz-Earle & Robson, 1970; Roff, 1981), where a and b are constants determining the shape of the fecundity schedule. These were assumed to have the values a = 0.12 and b = 0.45. Since fecundity is linear with the product cw, the numerical values of these variables do not affect fitness sensitivities. To obtain the fitness sensitivity of wing length, weight was assumed to be proportional to the cube of length. To calculate sensitivities, the baseline values α = 10, ω = 40, m_j = 0.1, and m_a = 0.2 were assumed.

The Roff model has two major differences from the Stearns and Kawecki model. First, r is used as the measure of fitness, obtained from the relation

$$e^{-m_j \alpha} cw \int_2^{\omega} e^{-r(x+\alpha)} e^{-m_a x} F(x) dx = 1 .$$

Second, size is assumed to depend on development time as $w = g(\alpha - p)$. This creates a trade-off between development rate and size. The value of *p* was assumed to be 1.59, and *g* given the baseline value of 0.3125. This tradeoff creates an intermediate optimum development time, which was found numerically to be approximately 8.02 days for the baseline parameter values. In order to yield comparable sensitivities to the 'no trade-offs' model, the population growth rate was calculated over a period of 14.44 days, the generation time (Stearns, 1992) calculated using the baseline parameters.

Calculating fitness sensitivities

A sensitivity is defined as the percentage alteration in fitness resulting from a given proportional change in a trait. If fitness is a linear function of trait value, then relative sensitivities are not affected by the size of the change in the trait. This assumption is not seriously violated for the five traits used by Stearns and Kawecki in the 'no tradeoffs' model; however, it does not apply to bristle numbers or development time in the Roff model. Therefore, in addition to the sensitivities to a 10% change given in Stearns and Kawecki, I also calculated sensitivities to 1% change in trait values.

The fitness sensitivities for the life history traits are shown in Table 3. In each case, these were calculated by changing the trait value in the direction most detrimental to fitness. For development time, only lengthening development time had a negative impact on fitness for the 'no tradeoffs' model, while in the Roff model shortening development time was most costly. Longevity was decreased by changing the value of ω from 40 to 36 days, which was the approach used by Stearns and Kawecki (T. J. Kawecki, pers comm.). This mimics a 10% change in lifespan under ideal laboratory conditions where extrinsic mortality is low.

For early and late fecundity, the sensitivities calculated by Stearns and Kawecki (1994) were incorrect. Instead of reducing fecundity throughout the defined period by 10%, they assumed that fecundity was reduced by 10% from days 4-14 for early fecundity, and by 10% from days 30-40 for late fecundity (T. J. Kawecki, pers comm.). These assumptions lead to substantial underestimates of the sensitivities for these traits, because the trait is not reduced by a full 10%, and especially given that the fitness sensitivities of the earliest days are the largest.

In addition to these traits intimately involved in life history, I also estimated the fitness sensitivities of abdominal and sternopleural bristle numbers from the results of Nuzhdin, Fry and MacKay (1995). These authors assumed that the fitness function of both traits was Gaussian ($w(x) = \exp[-(x-\theta)^2/(2Vs)]$, where θ is the trait optimum, and V_s is the strength of selection). Nuzhdin, Fry and MacKay estimated that V_s = 51 for sternopleural bristle numbers, and V_s = 77 for abdominal bristles. I assumed that the trait optimum, θ, was equal to the control population mean; 15.35 for sternopleural, and 15.4 for abdominal bristles. Two other recent studies using the same fitness model gave estimates of V_s which flank these: García-Dorado and González (1996) estimated that V_s = 124 for abdominal bristle number, and Mackay, Lyman and Hill (1995) estimated V_s = 24 for abdominal bristles and V_s = 6 for sternopleurals. These latter estimates are likely too low, as the rate of divergence of inbred lines was used to derive the estimates, and the lines diverged in a non-linear fashion not fully explicable by selection, mutation and drift alone (Mackay, Lyman & Hill, 1995). The results of my analysis are insensitive to the actual values of V_s in this range (results not shown).

Examining the sensitivities

The two life history models generally gave comparable results with the important exception of size and development time, which were assumed to be involved in a tradeoff in the Roff model. In the 'no tradeoff' model, the fitness sensitivity of development time is entirely due to its high correlation with total juvenile mortality. If juvenile mortality is assumed to be unrelated to development time, the sensitivity of fitness to changes in development time drops to 0. Both models allow size to be altered by, in effect, reducing growth rate. In this case, there is no tradeoff and size has the highest sensitivity of any trait in the model. The Roff model, in linking development time and size, allows another route to altering size through changes in development time. With the tradeoff, any change in development time is partially compensated for: shorter development lowers mortality and decreases generation time; longer development increases size and fecundity. The result of this simple and well-justified (Roff, 1981; Zwaan, Bijlsma & Hoekstra, 1995) change to the no-tradeoffs

Table 4. Spearman rank correlations of variation and sensitivities

	Target size	CV_M	CV_I	CV_A	V_A/V_M	$1\%\Delta w_{n.t.}$	$1\%\Delta w_{Roff}$
Timing	0.82^{**}	0.68^{*}	0.62	0.66^{a}	−0.16	−0.34	−0.38
Target size		0.85^{**}	0.67^{a}	0.67^{a}	−0.52	0.13	0.10
CV_M			0.83^{**}	0.93^{***}	−0.55	0.06	−0.09
CV_I				0.95^{***}	−0.10	−0.22	−0.29
CV_A					−0.24	−0.35	−0.48
V_A/V_M						−0.35	−0.14
$1\%\ \Delta w_{n.t.}$							0.86^{**}

^{a} $0.05 < P < 0.10$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

model is the conversion of high sensitivity traits into low sensitivity ones.

The issue of what life-history model is most appropriate is complex. To test the elimination hypothesis, we should use the model that best reflects the total average fitness impact of alleles influencing the trait, as this controls their rate of elimination from the population. For traits, that have a small direct impact on fitness, this can lead to large discrepancies between sensitivities calculated with different assumptions.

For the canalization hypothesis, the most relevant model is one that reflects direct fitness effects of the trait, as well as pleiotropic effects that are functionally inseparable from the trait. This is because one of the possible outcomes of canalization is that some pleiotropic effects are minimized, while others are left intact. An example of a pleiotropic effect that seems necessary is in the 'no tradeoffs' model, where Stearns and Kawecki (1994) assumed that a change in development time affects total juvenile mortality by changing the time a fly is at risk of death before becoming an adult. It is reasonable that this pleiotropic effect of development time on juvenile mortality is a necessary one, particularly under natural conditions where there are many sources of extrinsic mortality that are beyond the evolutionary control of the organism. Therefore, the fact that fitness sensitivity of development time in their model depends entirely on this pleiotropic effect is reasonable, even though development time per se is unselected in the models with R_0 as the fitness criterion. I have chosen to explore the Roff model in addition to the 'no tradeoffs' model because I think it is also likely that the pleiotropic effect of development time on size is a necessary one, because there is ample evidence for a trade-off between development time and size. It is also likely that a change in growth rate also has a pleiotropic or plastic effect on development time, but this effect is not included in the Roff model.

In addition, I assume that deleterious changes in size are most likely to occur through changes in growth rate rather than development time per se.

Testing the hypotheses

Table 4 gives the Spearman rank correlation coefficients between each of the predictor and dependent variables in this study. Because small mutational effects are probably typical of mutations that contribute substantially to genetic variance in natural populations, the fitness sensitivities used to calculate the correlations shown utilized the 1% values given in Table 3. The generally high positive correlations among the coefficients of variation must be interpreted cautiously as they are potentially inflated by autocorrelation, because they are standardized by similar mean values. On the other hand, if variances do scale with the trait mean, then standardization by mean is appropriate, because it eliminates autocorrelation based on the measurement scale.

The strongest correlation is between CV_A and CV_M (r_s = 0.95, P = 0.0009, 8 df), which tends to confirm the hypothesis that mutation is an important factor determining standing variance. This relationship is shown in the upper panel of Figure 1.

The lower panel of Figure 1 shows the similarly strong relationship between CV_M and CV_I (r_s = 0.83, P = 0.01, 8 df) . With just the CV_I values from Stearns, Kaiser and Kawecki (1995), the correlation is similar, but not significant (r_s = 0.80, P = 0.10, 5 df). This suggests that CV_M and CV_I may be measuring the same thing. Stearns and Kawecki (1994; Stearns, Kaiser & Kawecki, 1995) interpreted CV_I as being influenced solely by canalization, but the correlation with CV_M suggests that this is not correct. Because mutations of all kinds can occur in non-functional regions, or

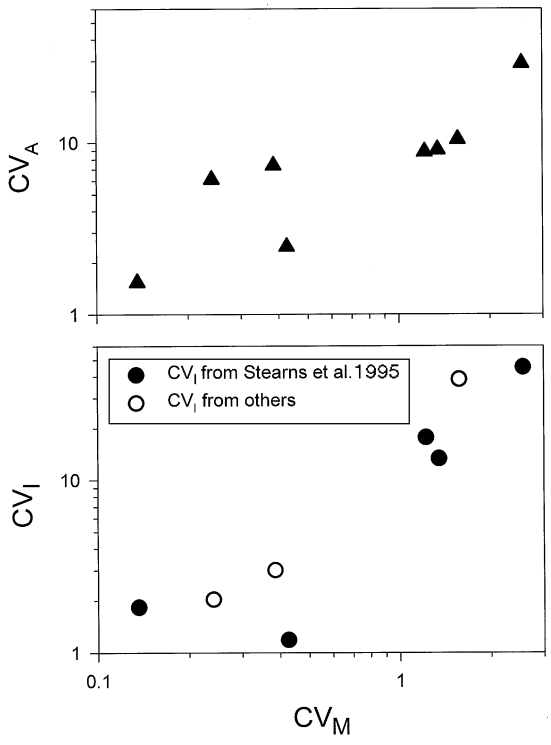

Figure 1. Relationship of CV_M to CV_I and CV_A.

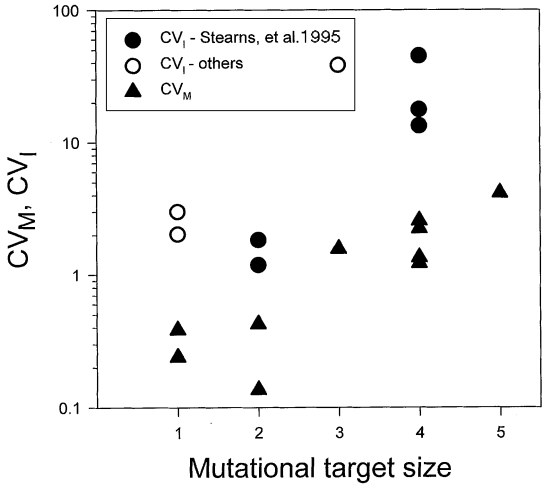

Figure 2. Relationship between CV_M, CV_I, and mutational target size index.

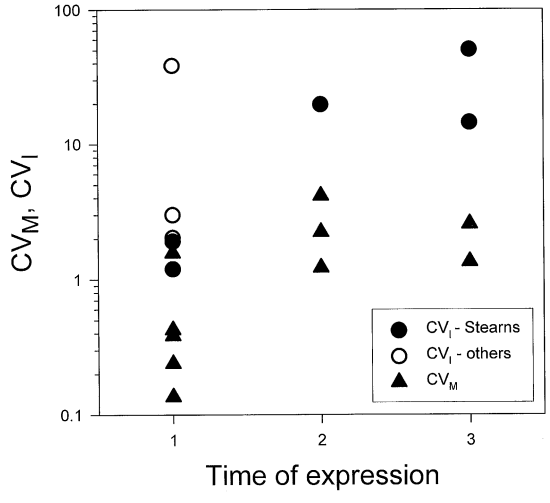

Figure 3. Relationship between CV_M, CV_I, and timing of trait expression.

regions that are functional but do not affect the trait in question, then CV_M and CV_I both estimate the effects of random genetic perturbations on the genome, which must consist of the probability that a mutation hits the mutational target for the trait, multiplied by the average effect of a hit on the trait.

My rankings of target size and timing of expression are related, as shown by the significant positive correlation between them (r_s = 0.82, P = 0.004). Both time of expression and target size are significantly positively correlated with CV_M and CV_I (Table 4). The relationships between CV_M, CV_I, and mutational target size are shown in Figure 2. The relationships between CV_M, CV_I, and timing of trait expression are shown in Figure 3. The correlations with the target size index support the mutational target hypothesis, while those with time of expression support the timing hypothesis. The strong correlation between target size and timing makes it difficult to discriminate the two, although the higher and more significant correlations with target size may suggest that the mutational target hypothesis has slightly stronger support. Examination of Figure 3 suggests that the correlations with timing are due to the differences between characters expressed at eclosion and those expressed later in life.

Despite the differences between the 'no tradeoffs' and Roff life history models, the sensitivities they generate are highly correlated with each other (r_s = 0.81, P $<$ 0.001). There are no significant correlations between sensitivities for either model and any of the measures of trait variance. Figure 4 shows the relationship of all the measures of trait variation to fitness sensitivities to 1% changes in trait value, calculated with the Roff model. Longevity and late fecundity, which have sensitivities very near 0, were arbitrarily assigned fitness sensitivities of 0.001 for presentation.

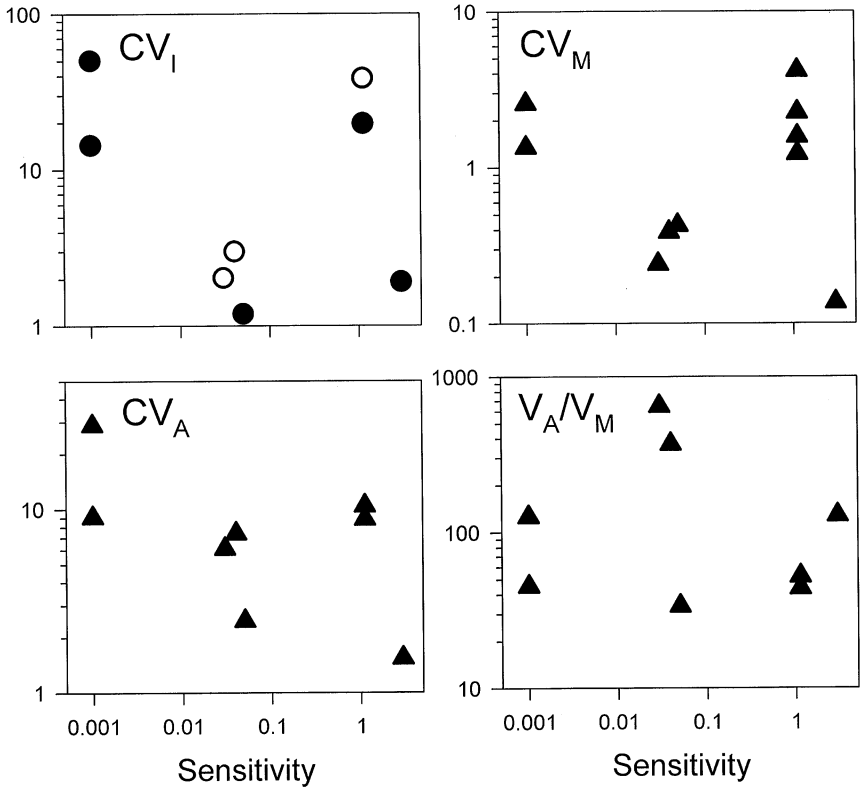

Figure 4. Relationship between sensitivities of fitness to 1% change in trait values in the Roff model, and measures of trait variation. Symbols as in Figures 2, 3.

In addition to these sensitivities, a large number of other fitness sensitivity models were also generated as described above, and by varying the baseline parameters (cf. Stearns & Kawecki, 1994). None of these models resulted in significant correlations with any of the dependent variation variables.

A simple expectation of the canalization hypothesis is that traits with higher sensitivities for fitness will be more canalized (Stearns & Kawecki, 1994; Stearns, Kaiser & Kawecki, 1995) . Although, as noted above, CV_I does not provide clear-cut information about the average effects of loci required to test the canalization hypothesis directly, canalization could still play a role in determining both CV_I and CV_M . However, the lack of relationships between fitness sensitivities and CV_I and CV_M suggest that canalization is not a dominant force in controlling trait variability.

A simple expectation of the elimination hypothesis is that variation created by spontaneous mutation will be eliminated more quickly in traits that are under stronger selection. The lack of a significant relationship between V_A/V_M and sensitivity therefore casts doubt on the elimination hypothesis. The two alternatives for the maintenance of variation, the selective sweep and the balance hypotheses, do not yield clear-cut predictions that can be tested with these data.

Tests of multiple hypotheses

Because none of the seven hypotheses in Table 1 are mutually exclusive, it is possible that consideration of several predictors simultaneously would reveal evidence for more than one hypothesis. To test for such effects, I carried out stepwise multiple regression with both entry and removal thresholds set at P = 0.15 (SAS Institute, 1990). All variables except the indices of target size and timing were log-transformed before analysis.

For the two mutational coefficients of variation, the relevant predictor variables are target size, timing, and sensitivities. When CV_M is the dependent variable, target size is the only variable to enter the model, at $P = 0.011$. When timing is omitted from the predictors, the slope for the sensitivities is negative, consis-

tent with the canalization hypothesis, although neither the 'no-tradeoff' nor the Roff sensitivities explain significant variance (P = 0.50 and 0.37, respectively). When CV_I is the dependent variable, only target size enters the model, at $P = 0.002$. In the full models, the slopes for sensitivities are negative, although they explain very little variance (P = 0.42 and 0.81).

When CV_A is the dependent variable, then CV_M can be added to the list of predictor variables. The best model includes CV_M, at P = 0.007, but no other predictors. In the full models, the slopes for sensitivities are negative, consistent with the elimination hypothesis, although they explain very little variance (P = 0.49 and 0.35). In summary, there is very little evidence that multiple predictors explain more of the variance in CV_M or CV_A.

Discussion

The results provide evidence in support of the mutational target or timing hypotheses for the explanation of variation in trait genetic variances. They suggest that the canalization and elimination hypotheses are less important. The chief evidence supporting the mutational target hypothesis is the high correlation between my measure of rank mutational target size and mutational variance.

Confirmatory evidence includes the correlations between mutational, insertional, and standing genetic variances. However, these latter correlations may have been influenced by the fact that standardizing all three variances by trait means can introduce positive covariances among them. The unstandardized variances are certainly autocorrelated due to measurement scale, so comparing variances as a dimensionless quantity is certainly worthwhile. However, any particular standardization technique is not guaranteed to remove all autocorrelation.

The timing hypothesis is also lent credence by these same correlations of mutational and standing coefficients of variation, as the timing of trait expression is correlated with my measure of mutational target size. Although the overall pattern of trait variances seems to be explained more satisfactorily by trait target size (compare, for example, Figures 2, 3), the timing hypothesis seems compelling when applied to the much larger mutational, insertional, and standing genetic variance of late fecundity compared to early fecundity. It is difficult to imagine that different genetic pathways are involved in these two traits, but easy to imagine how the average effect of alleles could be larger for traits expressed later in life. For example, if variation in fecundity is influenced by rates of gain and loss of some limiting resource, then under many circumstances variation will increase over time (Houle, unpublished).

If this is correct, this would mean that the fitness sensitivities for late fecundity are misleading with respect to the elimination hypothesis, as they do not reflect pleiotropic effects on early fecundity. The alternative would be to assume that early and late fecundity have the same fitness sensitivities. Similarly, the true fitness sensitivity of longevity is probably dominated by the pleiotropic effects of mutations on other processes early in life. For longevity, it is difficult to specify what those processes are, so an alternative fitness sensitivity is not easy to derive. This points to a more general problem with fitness sensitivities. The actual fitness sensitivities relevant for the elimination hypothesis are based on all of the pleiotropic effects of alleles, and not just their effects on the trait under study. While it is clear that there is massive pleiotropy among mutations affecting life history traits (Mukai & Yamazaki, 1971; Yoshimaru & Mukai, 1985; Lynch, 1985; Houle et al., 1994), the nature of that pleiotropy is not always clear. The life history models I used include many obviously appropriate pleiotropic interactions (for example between development time and viability, size and fecundity, and, for the Roff model, size with development time), but there are many other possible pleiotropic effects that could be included. For example, mutational effects have positively correlated effects on many life history traits, probably through their effects on overall health of the organism (Houle et al., 1994). In addition, it has been suggested that at equilibrium growth rate and viability, and longevity and fecundity are subject to tradeoffs.

A potential general consequence of large amounts of pleiotropy is that the direct effects of mutations on any trait of interest may be relatively independent of their effects on fitness. This is supported by the rather weak and variable relationship between the effects of P-element insertions on viability and bristle number (Mackay, Lyman & Jackson, 1992; Lyman et al., 1996). Therefore the fitness impacts of alleles affecting different phenotypic traits may be more similar than suggested by the role of the traits studied. If this is so, then the strong relationship between CV_A and CV_M, and a weak one between sensitivity and V_A/V_M, is what one would expect to see under mutation-selection

balance. Further work on fitness sensitivities in relation to pleiotropic effects would clearly be desirable.

On the other hand, the failure to confirm this prediction of the elimination hypothesis may also favor the alternative hypotheses for the maintenance of genetic variation: selective sweep and balancing selection. Both processes could render levels of genetic variance unpredictable, as a single balanced polymorphism, or a single allele sweeping to fixation, can generate very large amounts of genetic variance.

The canalization hypothesis received little support in this study. There are a large number of differences in data and assumptions between this study, which rejects the influence of canalization, and those of Stearns (Stearns & Kawecki, 1994; Stearns, Kaiser & Kawecki, 1995) who strongly favored it. First, Stearns, Kaiser and Kawecki assumed that insertional variance reflects the variance in effect of loci that affect the trait. This disregards the probable role of mutational target size in influencing the effects of insertions. Second, Stearns, Kaiser and Kawecki made what I believe to be inappropriate assumptions about mutational effects on fecundity, leading to underestimates of the fitness sensitivities for early and late fecundity. Third, additional data on the effects of insertions already in the literature includes two bristle traits with modest sensitivities and low insertional variance, and one trait, egg-to-adult viability, with high sensitivity and high insertional variance, in contradiction to the predictions to the canalization hypothesis. Fourth, their statistical treatment gives a misleading impression of the strength of the correlations between sensitivities and insertional and environmental variances in their data, as they treated traits as fixed effects rather than random ones, as would be appropriate given their hypothesis. The result is pseudo-replication due to the treatment of replicate experiments as independent with respect to their hypothesis. Finally, the conclusion by Stearns, Kaiser and Kawecki that their ranking of fitness sensitivities is robust depends on the pattern of pleiotropy they assumed. One well-supported change to their assumed pattern of pleiotropy, introducing a tradeoff between size and development time, changes development time from a trait with high sensitivity to one with modest sensitivity.

Although the simplest prediction of the canalization hypothesis was not confirmed in this study, theoretical work suggests that this is a naive prediction. Broadly speaking, canalization occurs whenever selection on variances leads to the evolution of modifiers of that variance. This may lead either to decreases or increases in trait variances, so it is not clear that there is a simple expectation to test. For example, Lande (1980) shows that for fitness functions that are concave upward, alleles that increase the variance of a trait are favored, conditional on the trait mean. In addition, if mutations are biased in direction towards lower fitness, then either canalization or decanalization may result, depending on whether eliminating the alleles or obscuring their effects yields the highest fitness. Even under stabilizing selection which is, a concave downward fitness function, Wagner, Booth and Bagheri-Chaichian (1997) have shown that canalization will be strongest at intermediate strengths of stabilizing selection, rather than with the strongest selection, if mutation-selection balance is responsible for the variance. Given this complicated pattern of expectations, we cannot yet predict the overall patterns we would expect to see if canalization were important during evolution.

Despite this ignorance, I predict that canalization rarely plays a dominant role in determining mutational or standing variances. Modifier selection is in general fairly weak, and there are a large number of other processes that are obviously directly relevant to mutational and standing variance. For example, a major factor in determining mutational target size must be the process of gene duplication and divergence. While selection on trait variance may play some role in this process, it is far more likely to be the direct result of the effects of the duplications, for example in determining the level of expression of a gene product, or in allowing the evolution of new functions. At a very different evolutionary level, Rowe and Houle (1996) have shown that direct selection for increases in the size of a costly trait is expected to lead to an increase in the genetic variance of the trait without modifier evolution. In addition, Wagner, Booth and Bagheri-Chaichian (1997) emphasize that modifier alleles themselves will usually have direct fitness effects. I expect that such direct selection will often overwhelm the results of variance modifier selection.

Finally, it is important to note that the data in this study are often imprecisely estimated, and the results of many different studies have had to be combined to detect patterns. Ideally, mutational, insertional, and standing genetic variances should be estimated by the same methods in the same population. Fitness sensitivities should be estimated separately for direct and pleiotropic effects. Further work on whole-organism measures of variation and covariation would be very helpful in constructing better tests of these hypotheses. Regardless of the flaws in this limited exercise, I

hope that it is clear that comparisons of this nature are potentially very informative.

Acknowledgements

I thank Stephen C. Stearns, Tadeusz J. Kawecki, Locke Rowe, Alexei Kondrashov, Günter Wagner and a reviewer for their comments on previous versions of this manuscript.

References

Bürger, R., 1993. Predictions of the dynamics of a polygenic character under directional selection. J. Theor. Biol. 162: 487–513.

Crow, J.F., 1979. Minor viability mutants in *Drosophila*. Genetics 92: s165–s172.

Eanes, W.., C. Wesley, J. Hey, D. Houle & J.W. Ajioka, 1988. The fitness consequences of *P* element insertion in *Drosophila melanogaster*. Genet. Res. 52: 17–26.

Falconer, D.S., 1981. Introduction to Quantitative Genetics, 2nd edn. Longman, London.

García-Dorado, A. & J.A. González, 1996. Stabilizing selection detected for bristle number in *Drosophila melanogaster*. Evolution 50: 1573–1578.

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

Houle, D., 1992. Comparing evolvability and variability of quantitative traits. Genetics 130: 195–204.

Houle, D., D.K. Hoffmaster, S. Assimacopoulos & B. Charlesworth, 1992. The genomic mutation rate for fitness in *Drosophila*. Nature 359: 58–60.

Houle, D., K.A. Hughes, D.K. Hoffmaster, J. Ihara, S. Assimacopoulos, D. Canada & B. Charlesworth, 1994. The effects of spontaneous mutation on quantitative traits. I. Variance and covariance of life history traits. Genetics 138: 773–785.

Houle, D., B. Morikawa & M. Lynch, 1996. Comparing mutational variabilities. Genetics 143: 1467–1483.

Judd, B.H., M.W. Shen & T.C. Kaufman, 1972. The anatomy and function of a segment of the *X* chromosome of *Drosophila melanogaster*. Genetics 71: 139–156.

Kimura, M. & T. Ohta, 1971. Theoretical aspects of population genetics. Princeton University Press, Princeton.

Lai, C., R.F. Lyman, A.F. Long, C.H. Langley & T.F.C. Mackay, 1994. Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster*. Science 266: 1697–1702.

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

Lande, R., 1980. Genetic variation and phenotypic evolution during allopatric speciation. Amer. Natur. 116: 463–479.

Long, A.D., S.L. Mullaney, L.A. Reid, J.D. Fry, C.H. Langley & T.F.C. Mackay, 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 139: 1273–1291.

Lyman, R.F., F. Lawrence, S.V. Nuzhdin & T.F.C. Mackay, 1996. Effects of single *P*-element insertions on bristle number and viability in *Drosophila melanogaster*. Genetics 143: 277–292.

Lynch, M., 1985. Spontaneous mutations for life-history characters in an obligate parthenogen. Evolution 39: 804–818.

Mackay, T.F.C., R.F. Lyman & M.S. Jackson, 1992. Effects of *P* element insertions on quantitative traits in *Drosophila melanogaster*. Genetics 130: 315–332.

Mackay, T.F.C., R.F. Lyman & W.G. Hill, 1995. Polygenic mutation in *Drosophila melanogaster*: non-linear divergence among unselected strains. Genetics 139: 849–859.

MacMillan, M.F., F. Fitz-Earle & D.S. Robson, 1970. Quantitative genetics of fertility. I. Lifetime egg production of *D. melanogaster* - theoretical. Genetics 65: 349–353.

Mousseau, T.A. & D.A. Roff, 1987. Natural selection and the heritability of fitness components. Heredity 59: 181–197.

Mukai, T. & T. Yamazaki, 1971. The genetic structure of natural populations of *Drosophila melanogaster*. X. Developmental time and viability. Genetics 69: 385–398.

Nuzhdin, S.V., J.D. Fry & T.F.C. Mackay, 1995. Polygenic mutation in *Drosophila melanogaster*: the causal relationship of bristle number to fitness. Genetics 139: 861–872.

Riska, B., W.R. Atchley & J.J. Rutledge, 1984. A genetic analysis of targeted growth in mice. Genetics 107: 79–101.

Roff, D., 1981. On being the right size. Amer. Natur. 118: 405–422.

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

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

Rowe, L. & D. Houle, 1996. The lek paradox and the capture of genetic variance by condition dependent traits. Proc. Roy. Soc. London, Ser. B 263: 1415–1421.

SAS Institute, I., 1990. SAS/STAT User's Guide, Version 6, 4th edn. SAS Institute, Cary, NC.

Stearns, S.C., 1992. The Evolution of Life Histories. Oxford, Oxford.

Stearns, S.C. & T.J. Kawecki, 1994. Fitness sensitivity and the canalization of life-history traits. Evolution 48: 1438–1450.

Stearns, S.C., M. Kaiser & T.J. Kawecki, 1995. The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*. J. Evol. Biol. 8: 539–557.

Sved, J.A. & F.J. Ayala, 1970. A population cage test for heterosis in *Drosophila pseudoobscura*. Genetics 66: 97–113.

Waddington, C.H., 1957. The Strategy of the Genes. MacMillan Co., New York.

Wagner, G., G. Booth & H. Bagheri-Chaichian, 1997. A population genetic theory of canalization. Evolution 51: 329–347.

Yoshimaru, H. & T. Mukai, 1985. Relationships between the polygenes affecting the rate of development and viability in *Drosophila melanogaster*. Jap. J. Genet. 60: 307–334.

Zwaan, B., R. Bijlsma & R.F. Hoekstra, 1995. Artificial selection for developmental time *Drosophila melanogaster* in relation to the evolution of aging: direct and correlated responses. Evolution 49: 635–648.