Pleiotropic Models of Quantitative Variation

N. H. Barton

Department of Genetics and Biometry, University College London, London NW1 2HE, England

Manuscript received May 23, 1989
Accepted for publication November 21, 1989

ABSTRACT

It is widely held that each gene typically affects many characters, and that each character is affected by many genes. Moreover, strong stabilizing selection cannot act on an indefinitely large number of independent traits. This makes it likely that heritable variation in any one trait is maintained as a side effect of polymorphisms which have nothing to do with selection on that trait. This paper examines the idea that variation is maintained as the pleiotropic side effect of either deleterious mutation, or balancing selection. If mutation is responsible, it must produce alleles which are only mildly deleterious ($s \approx 10^{-3}$), but nevertheless have significant effects on the trait. Balancing selection can readily maintain high heritabilities; however, selection must be spread over many weakly selected polymorphisms if large responses to artificial selection are to be possible. In both classes of pleiotropic model, extreme phenotypes are less fit, giving the appearance of stabilizing selection on the trait. However, it is shown that this effect is weak (of the same order as the selection on each gene): the strong stabilizing selection which is often observed is likely to be caused by correlations with a limited number of directly selected traits. Possible experiments for distinguishing the alternatives are discussed.

THE main application of quantitative genetics to artificial and natural populations has been to use the pattern of genetic variances and covariances to predict the response of the mean phenotype to selection. This prediction only requires the assumption that the joint distribution of phenotypes and breeding values is Gaussian, and so is fairly robust to the underlying genetics. However, to understand morphological evolution in the longer term, we must find out what maintains genetic variation, and how variation will change under selection. This has proved difficult and contentious. The central difficulty is that, although the methods of Mendelian genetics cannot be used directly to study alleles with small effects, the evolution of the phenotypic variance does depend on the numbers and distribution of effects of individual genes (BARTON and TURELLI 1987; TURELLI and BARTON 1989).

There is a further problem, in that the basic observations conflict: we see both abundant polygenic variation, and strong stabilizing selection that should rapidly eliminate that variation. Genetic variation is reflected directly in correlations between relatives, and indirectly in sustained responses to directional selection that take the phenotype well beyond its original range. Evidence for stabilizing selection comes from the reduced fitness of extreme phenotypes [reviewed by CHARLESWORTH, LANDE and SLATKIN (1982) and ENDLER (1986)], and perhaps more convincingly, from the constancy of form over geological time (CHARLESWORTH, LANDE and SLATKIN 1982; MAYNARD SMITH 1983).

Following ROBERTSON (1967), we can contrast two kinds of explanation for these apparently contradictory observations. The simplest possibility invokes direct selection on the character of interest: there might be a balance between mutation and stabilizing selection (LANDE 1975); selection on the character might induce overdominance on the underlying loci (GILLESPIE and TURELLI 1989); or frequency-dependence might maintain a diversity of phenotypes (ROUGHGARDEN 1972; SLATKIN 1979). Such direct explanations have received most attention, because they are mathematically tractable, and because they offer the possibility that variation might be understood in terms of measurable parameters. However, even without going into genetic details, the sheer number of quantitative characters and gene loci makes direct explanations seem implausible. Suppose we accept the view that quantitative variation can be understood character by character. Can Gaussian stabilizing selection of strength $V_s \approx 20V_g$ [the typical value suggested by TURELLI (1984)] act independently on a large number (m) of characters? (Here, the fitness of an individual with phenotype z is $w(z) = \exp(-z^2/2V_s)$; the optimum is arbitrarily set at zero). Genetic variation around the optimum reduces fitness by a factor $\sqrt{V_s/(V_s + V_g)} \approx \exp(-V_g/2V_s)$ for each character, and so a naive load argument suggests a net reduction by $\exp(-mV_g/2V_s)$. This sets a limit of at most 100 independent characters.

Such arguments have been criticized because they assume that fitnesses are constant and combine multiplicatively (SVED, REED and BODMER 1967; EWENS

1979). The genetic variance in fitness leads to a more robust (albeit weaker) limit. Assuming Gaussian breeding values for the trait, the variance of squared deviations, var(z^2), would be $2V_g^2$. Since fitness declines with $z^2/2V_s$, the total genetic variance in fitness would be $\approx m(V_g/V_s)^2/2$. (I assume that selection on each character is weak enough that $\exp(-z^2/2V_S) \approx 1 - z^2/2V_S$.) This formula differs from that given by CHARLESWORTH (1987), who calculated the (much smaller) *additive* component under a two-allele model; it is consistent with TACHIDA and COCKERHAM'S (1988) results, in the limit of large numbers of loci, and with truly independent characters.

Though we know very little about the genetic variance in relative fitness (as opposed to its components), it is hard to believe that it is much greater than ≈ 0.25. It must lie between the additive genetic variance, and the total variance. Fisher's fundamental theorem implies that the former will be small (CHARLESWORTH 1987), while the total variance averages ≈ 1 across the range of species surveyed by CLUTTON-BROCK (1988). This admittedly rough figure of ≈ 0.25 would suggest a limit of at most 200 or so independent characters.

Genetic data set other kinds of constraints. Comparison of mutation rates for quantitative characters ($\Sigma\mu \approx 0.01$ or more) (TURELLI 1984) with the net mutation rate to deleterious alleles ($\Sigma\mu \approx 0.25$ for egg-to-adult viability in *Drosophila*) (SIMMONS and CROW 1977; CROW and SIMMONS 1983) suggest that a significant fraction of genes ($0.01/0.25 \approx 1/25$) may affect each trait, so that each gene must affect many traits. The argument is somewhat weakened if the mutation rate to *all* genes reducing fitness is larger than the rate quoted here from measurements of viability (CHARLESWORTH and CHARLESWORTH 1987); however, $\Sigma\mu$ cannot be very much larger without producing an excessive mutation load. (In the spirit of this paper, I am assuming here that most mutations affecting quantitative traits are in fact deleterious: there could of course be any number of genes with little effect on fitness, and with specific effects on metric characters.) Differences between species or selected lines in any one trait can involve a large number ($\approx 10^2$) of genes (WRIGHT 1968; BARTON and CHARLESWORTH 1984), again suggesting widespread pleiotropy.

Conversely, major mutants commonly have pleiotropic effects on a significant fraction of morphological traits (DOBZHANSKY and HOLZ 1943; WRIGHT 1968). Evidence here is not as clear as one would wish: it is not obvious that minor mutants will have similarly wide pleiotropic effects, and there has been no systematic survey of the effects of spontaneous mutants on multivariate morphology. An interesting counterexample is provided by abdominal and sternopleural bristle number in *Drosophila melanogaster*. The weak genetic correlation between these characters (SHERIDAN and BARKER 1974; DAVIES and WORKMAN 1971) does not in itself imply that there is no correlation in the effects of individual genes. However, DAVIES (1971) located the differences between selected lines to chromosome regions that primarily affected only one or other character. The nature and extent of pleiotropy is an important open question. At present, the strongest evidence of its general importance lies in the complex biochemical and developmental processes through which genotype affects the arbitrary metric characters which we choose to measure (WRIGHT 1968).

Widespread pleiotropy does not necessarily imply that variation in one trait will be affected by selection on other traits: in LANDE's (1980) model of Gaussian allelic effects, a locus can have an arbitrary combination of effects on many traits. However, only a limited number of alleles can segregate at each locus, and so it is unreasonable to suppose that selection on one locus could give independent responses in any of an arbitrarily large number of phenotypic directions (TURELLI 1985). Moreover, if variation at a locus primarily affects just the net activity of some enzyme, its effects will be constrained to a single degree of freedom (*cf.* WAGNER 1989; KEIGHTLEY and KACSER 1987). Patterns of expression in different tissues or at different times might vary, but effects will still be constrained to a few degrees of freedom at each locus, however many alleles segregate.

One could reconcile these observations with "direct" explanations by supposing that there is in fact variation for only a few independently evolving principal components, such as size: variation in measured characters might simply reflect selection on a much smaller set of traits (*cf.* KIRKPATRICK and LOFSVOLD 1989). This possibility cannot be excluded. However, the fact that selection has produced a profusion of delicate adaptations, each involving many traits, and the fact that artificial selection can sculpt almost all aspects of a plant or animal, has led to the view that evolution can proceed along any of a large number of phenotypic degrees of freedom (DARWIN, 1859, Ch. 6; FISHER 1930, pp. 38–41; CHARLESWORTH, LANDE and SLATKIN 1982). It therefore seems more likely that quantitative variation is based on the overlapping effects of many genes on many characters, and that any one character will be strongly influenced by selection on others (*cf.* CHARLESWORTH 1984).

If pleiotropy is indeed widespread, then the variation of any one character may be a mere side-effect of polymorphisms maintained for quite other reasons. This is likely to be the case even if one considers *sets* of characters, as in the multivariate analyses developed by LANDE and ARNOLD (1983). One can still measure

only a tiny fraction of the traits which affect fitness, and statistical analysis of more than a dozen or so characters requires inordinate effort (MITCHELL-OLDS and SHAW 1987). Polygenic variation could be analysed by postulating that fitness effects are mediated by large numbers of hypothetical characters. However, it seems more promising to consider just the net effects of alleles on fitness and on the character of interest, as suggested by HILL and KEIGHTLEY (1988): the effects of alleles on hidden metric characters, and the consequences of those characters for fitness, are summarized in the net effect on fitness. This distinction between direct and pleiotropic selection is to some extent semantic, since it depends on whether it is more helpful to suppose that the effects of genes on fitness act through a set of intermediate "traits." For example, GILLESPIE and TURELLI (1989) show that if allelic effects fluctuate at random, stabilizing selection on a character can induce overdominance, and can maintain variation. However, if one extends this argument to stabilizing selection on many characters, influenced by many genes, then the model converges to one of pleiotropic overdominance (GILLESPIE 1984). Though the overdominance at each locus is caused here by stabilizing selection on quantitative characters, the outcome would be the same with any kind of overdominance of the same strength.

We now need consider only two broad classes of pleiotropic models: variation may be maintained by deleterious mutations, or by balancing selection. It might seem that because we cannot measure selection on individual polygenes (or even count these polygenes), investigation of pleiotropic variation would be hopeless. However, observations of high heritabilities, strong stabilizing selection, and large responses to selection do impose surprisingly strong constraints on both classes of pleiotropic models.

MUTATION/SELECTION BALANCE

Suppose that mutations at any of the $2n$ genes in a diploid individual can affect a neutral character, z. Mutations occur at a rate μ per locus, and each multiplies the fitness of heterozygotes by a factor $(1 - s)$. We assume that $\mu \ll s \ll 1$, so that any particular allele is rare, and the number of deleterious alleles per individual, k, follows a Poisson distribution with mean $\bar{k} = 2n\mu/s$. The mean fitness of the population is $\exp(-2n\mu)$.

Allele i has effect α_i on the character of interest; α_i is drawn from a symmetric distribution with variance α^2 and fourth moment $3\kappa\alpha^4$. κ is a measure of the kurtosis of the distribution of mutant effects: if all effects are equal in absolute value, $\kappa = 1/3$, while if they are drawn from a normal distribution, $\kappa = 1$. For simplicity, I consider only variation in mutant effects, α. Variation in s would also affect the results (HILL and KEIGHTLEY 1988). KEIGHTLEY and HILL (1983, 1988, 1989) have made extensive studies of this kind of model: they assume that allelic effects are drawn from a reflected gamma distribution of the form $|\alpha_i|^{\beta-1}\exp(-|\alpha_i|/\omega)$. In the present notation, this gives $\alpha^2 = \omega^2\beta(\beta + 1)$, $\kappa = (\beta+2)(\beta+3)/[3\beta(\beta+1)]$. Note that one has some choice over the variance α^2, the kurtosis, κ, and the number of loci, n. One could either consider all loci, in which case n and κ would be high, and α^2 low, or only those loci with significant effects, in which case n and κ would be lower, and α^2 higher. The overall constraint is that the increase in variation due to mutation is $V_m = 2n\mu\alpha^2$ per zygote per generation.

Maintenance of variation: The equilibrium genetic variance is $\bar{k}\alpha^2 = (2n\mu\alpha^2/s)$. The gametic mutation rate to alleles which reduce viability when homozygous has been estimated for the second chromosome of *D. melanogaster* (SIMMONS and CROW 1977; CROW and SIMMONS 1983; CHARLESWORTH 1987). Extrapolating to the whole genome, $n\mu = 0.013$ for lethals, and $\approx 0.13 - 0.43$ for detrimentals. The effect on fitness in the heterozygous state has been inferred from allele frequencies in nature, on the assumption of negligible local inbreeding. Estimates for recessive lethals and detrimentals are similar, with $s \approx 2\%$ (CROW and SIMMONS 1983; p. 27). Thus, substantial heritabilities ($V_g \approx V_e$) are expected if the variance of effects, α^2, across *all* deleterious alleles, is $\approx 0.03V_e$. At first sight, such a value is not implausible: random mutations are expected to have a wide range of pleiotropic effects (WRIGHT 1968), and responses to artificial selection may involve recessive lethals or detrimentals (*e.g.* YOO 1980). Support comes from RUSSELL, SPRAGUE and PENNY'S (1963) observations of spontaneous mutations in maize. RUSSELL *et al.* measured a mutation rate to alleles affecting the trait of $n\mu = 6 \times 10^{-2}$; these alleles had a variance of effects of at least $\alpha^2 = 0.03V_e$, and perhaps considerably more (LANDE 1975; TURELLI 1984). These values would account for substantial genetic variability. However, evidence on both the variance of effects, and the mutation rate, of alleles affecting quantitative traits is flimsy (TURELLI 1984). A much stronger argument can be made from the rate at which mutation introduces genetic variance: since $V_m = 2n\mu\alpha^2$, $V_g = V_m/s$. A range of experiments has shown that $V_m \approx 10^{-3}V_e$ (LANDE 1975; HILL 1982; LYNCH 1988). With selection against deleterious alleles of a few percent, this mutational input could not maintain significant variability.

In reality, the selection coefficient will vary between alleles: the genetic variance is determined by the average of α^2/s, weighted by mutation rate. One would expect mutants with a large effect on the character to have large effects on fitness: if the relation is strong, so that s increases faster than α^2, then most variation

will be contributed by alleles of small effect, whereas if it is weak, major mutants will be more important. Thus, it is not clear how variation between alleles would affect the genetic variance. It is also unclear whether it is safe to extrapolate from the Drosophila data on alleles with noticeable homozygous effects: we have no direct way of investigating alleles with undetectable effects on homozygotes. However, the relatively good estimates of $V_m = 10^{-3}V_e$ show that if *any* kind of mutation/selection balance (direct or pleiotropic) is to maintain quantitative variation, then alleles with effects on the trait must be only mildly deleterious ($s \approx V_m/V_e = 10^{-3}$). This is quite possible, but would imply that the mutations revealed by studies on Drosophila viability are largely distinct from those responsible for variation in other metric characters. Moreover, the extra influx of very mildly deleterious alleles cannot be great: the measured mutation rate to alleles which reduce Drosophila viability is already large enough to cause a substantial load (CROW and SIMMONS 1983).

Apparent stabilizing selection: Since individuals with extreme values of the trait will tend to carry more deleterious alleles, we expect a negative correlation between fitness and deviation from the mean. This will give the false appearance of stabilizing selection on the character itself. The regression of relative fitness on squared deviation, γ, is a practical measure of the strength of stabilizing selection (LANDE and ARNOLD 1983). An alternative, used more often in theoretical work, is to write fitness as $W = \exp(-(z - z_{opt})^2/2V_s)$; now, γ corresponds to $1/2V_s$, if one makes the approximation that the load is not large ($\bar{W} \approx 1$). With pleiotropic mutations, the regression gives:

$$V_s = \alpha^2(3\kappa + 4n\mu/s)/2s \qquad (1a)$$

This can be rewritten in a dimensionless form, in terms of the apparent genetic load associated with the character–that is, the difference between the fitness of individuals with the optimal breeding value, and the mean fitness:

$$\frac{V_g}{2V_s} = \frac{2n\mu}{(3\kappa + 4n\mu/s)} \qquad (1b)$$

If an individual typically carries rather few deleterious alleles that can affect the character ($\bar{k} = 2n\mu/s \ll 1$), the apparent load is $\approx 2n\mu/3\kappa$, whereas if this number is large, it is $\approx s/2$. There is a simple explanation for these alternative limits. If individuals carry at most one deleterious allele, then individuals extreme for one trait will be extreme for all traits, and so the whole genetic load ($\approx 2n\mu$) will appear to be associated with each trait. On the other hand, if $\bar{k}$ is large, deviations in any one trait will be a poor predictor of the number of deleterious alleles: the apparent load will be smaller by a factor $\approx 1/\bar{k}$.

A further prediction of the pleiotropic model is that individuals extreme for one trait will also tend to be extreme for other, uncorrelated, traits. Consider two traits, z_1 and z_2, with mean zero, which are affected by the same set of loci; each allele has random effects α_{1i}, α_{2i} on the two traits. The genetic covariance between them will be $2n\mu\mathrm{cov}(\alpha_1, \alpha_2)/s$, and so if the allelic effects are uncorrelated, the traits will be uncorrelated. However, one can easily show that the genetic covariance between the *squares* of the traits, standardized relative to their genetic variances, is just $\mathrm{cov}(z_1^2, z_2^2)/[\mathrm{var}(z_1)\mathrm{var}(z_2)] = 1/\bar{k}$.

Note that TURELLI (1985, Eq. 11b) argued that selection on one trait could not give the appearance of selection on another unless the traits were correlated. However, this argument assumed a normal distribution of phenotypes. The selection is here induced because the Poisson distribution of deleterious alleles is not quite normal. More generally, covariances between deviations ($\mathrm{cov}(z_1^2, z_2^2)$) involve fourth moments, which may be significant when the distribution of allelic effects is leptokurtic, as will be the case if the relevant loci are close to fixation (BARTON and TURELLI 1987).

For a given value of V_s, this model predicts a lower genetic variance than does TURELLI's (1984) "rare alleles" approximation, in which selection acts directly on the character ($V_g = 4n\mu V_s/(3\kappa + 4n\mu/s) < 4n\mu V_s$, since $\kappa > 1/3$). The different predictions of the two models illustrate the important point that one cannot calculate the equilibrium variance from a purely phenotypic model which balances the erosion of variation by the observed selection against the introduction of variation by mutation: the same values of V_s and $V_m = 2n\mu\alpha^2$ give different predictions, depending on the underlying genetics. With pleiotropy, selection differentials on the character alone do not suffice to predict the equilibrium variance.

How much stabilizing selection will in fact be induced by the pleiotropic effects of deleterious mutations? The Drosophila data discussed above suggest that the total number of deleterious alleles per diploid genome is large ($\bar{k} = 2n\mu/s \approx 30$). Observations of stabilizing selection in natural and artificial populations give typical loads of $V_g/2V_s \approx 2.5\%$ (LANDE 1975; TURELLI 1984), which is comparable with estimates of $s/2 \approx 1\%$ from Drosophila. $V_g/2V_s$ might be expected to be greater than estimates of $s/2$, for two reasons. The apparent load should be given by an average of s which is weighted towards alleles with large effects on the character, and will therefore be larger than the average of s for spontaneous mutations. Also, the Drosophila estimates are derived from the persistence of recessive lethals and detrimentals in nature (CROW and SIMMONS 1983, p. 27) Since these

are a selected sample, they will have milder average effects on heterozygotes than spontaneous mutations.

This argument assumes that variation (V_g) is maintained by a mutation/selection balance. However, we have seen that the low value of V_m makes this implausible: it also makes it unlikely that deleterious mutations contribute significantly to measures of stabilizing selection. This can be seen by assuming a Gaussian distribution of breeding values (so that κ is small). The part of the regression of squared deviations on the character which is attributable to deleterious mutations is then $V_g/2V_s = V_m/2V_g$, which is negligibly small. Pleiotropic mutation is made even less likely as an explanation for variation and apparent stabilizing selection by ENDLER's (1986) survey of natural populations: he finds that selection may be remarkably strong, so that V_s is often much less than the figure of $20V_e$ used here.

The response to selection: Ignoring for the moment the difficulties in accepting mutation-selection balance as the main cause of polygenic variation, we can ask whether a model of pleiotropic mutation/selection balance would be compatible with sustained responses to directional selection. If $\bar{k}$ is large, the distribution of breeding values will be approximately Gaussian, and so the initial response will be the product of the heritability and the selection differential. The selection coefficient at a locus produced by directional selection of intensity i is approximately $i\alpha/\sqrt{V} = i\alpha/\sqrt{V_g/h^2}$, where V is the phenotypic variance, and $h^2 = V_g/V$ (KIMURA and CROW 1978). Substituting for V_g gives $i\sqrt{h^2 s/2n\mu}$; if artificial selection is intense, and heritability high ($i = 1$, $h^2 = 0.5$), this will be much larger than natural selection against the deleterious allele ($i\sqrt{h^2} > \sqrt{2n\mu s}$). Thus, directional selection is likely to overwhelm intrinsic effects, and fix deleterious alleles. The maximum response will be ($n\Gamma\alpha$), where $\Gamma\alpha$ is half the mean effect of "+" alleles; for a normal distribution of α, Γ is $1/\sqrt{8\pi} = 0.20$. Relative to the initial genetic standard deviation, the maximum response is $\Gamma n/\sqrt{\bar{k}}$.. This could be large, and so is consistent with observations. At the selection limit, half the loci will be fixed for deleterious alleles. Deleterious alleles are of course likely to be partially recessive: rather than being fixed, they may be brought to high frequency, in a balance between artificial and natural selection. There would be a substantial fitness loss, since entirely homozygous flies have very low fitness (SIMMONS and CROW 1977; CROW and SIMMONS 1983).

A further prediction is that as rare advantageous alleles increase in frequency, the genetic variance should increase; it should eventually fall as advantageous alleles are fixed. Such changes in variance are rarely seen under directional selection. However, this is not decisive evidence against this model, or others in which variation is based on rare alleles. The variance will only change slowly if the number of loci is large (BULMER 1980), and furthermore, the distribution of allele frequencies in the base and the selected populations will be distorted by drift. KEIGHTLEY and HILL (1989) show that these effects can easily mask the expected rise and fall in variance.

HILL and KEIGHTLEY (1988) have examined the response to selection due to new mutations with deleterious effects, taking into account the effects of drift. They find that, as here, strong directional selection will overwhelm the effects of pleiotropy: rapid responses are possible, despite the associated loss of fitness. Because they include recurrent mutation, there is no selection limit: in the model set out here, the limit would be reached when all favorable alleles in the base population have been fixed. In practice, one would expect a continued response as new mutations occurred; one would also expect that if the original variation had not been entirely exhausted, the trait would return toward its original value.

BALANCING SELECTION

Suppose now that quantitative variation is maintained as a side effect of balanced polymorphisms at n loci; for simplicity, suppose that the effects of all the alleles on the trait are additive. The polymorphisms might be maintained either by overdominance or by frequency-dependent selection. However, since the relation between fitness and genotype varies considerably between models with frequency-dependence, only overdominance will be analysed in detail. Furthermore, since overdominance cannot easily maintain more than a few alleles at each locus (LEWONTIN, GINZBURG and TULJAPURKAR 1978), two alleles per locus are assumed. Only results on the limits to selection are sensitive to the form of the polymorphism. Overdominance has been considered as a mechanism for maintaining quantitative variability by WRIGHT (1935a), ROBERTSON (1956), BULMER (1973), and GILLESPIE (1984). All these authors showed that overdominance could account for high heritabilities despite stabilizing selection directly on the character. Moreover, GILLESPIE (1984, Eq. 4) showed (using a multiallelic model) that overdominance introduced phenotypic variation in a manner analogous to mutation: the ratio between the selective advantage of homozygotes and the number of alleles plays the same role as V_m. ROBERTSON's (1956) treatment produced all the results needed for the biallelic case: in this section, the aim is to present these in the same framework as for mutation/selection balance.

With random mating, the three genotypes at locus i have frequencies q_i^2: $2p_iq_i$: p_i^2, their relative fitnesses are $1 - s_i$: 1:1 $- t_i$ ($s_i, t_i \ll 1$), so that selection maintains a polymorphism with stability $S_i = s_it_i/(s_i + t_i)$ and

equilibrium $p_{0i} = S_i/t_i$. (S_i is the rate of return toward the equilibrium: $dp_i/dt \approx -S_i(p_i - p_{0i})$ when $p = p_0$. It is also the segregation load associated with the polymorphism. The contributions of the genotypes to the trait are $-\alpha_i$: 0: $+\alpha_i$. As before, the allelic effect α_i varies randomly across loci, and is drawn from a symmetric distribution with variance α^2 and fourth moment $3\kappa\alpha^4$. α_i^2 is assumed to be independent of the equilibrium frequency, but may be correlated with the strength of selection, as measured by the segregation load S_i.

Maintenance of variation: The equilibrium genetic variance is $2n\overline{p_0q_0\alpha^2}$, where the overbar denotes a mean across loci. Clearly, a great deal of variability can be maintained by this mechanism. This is true even if only a small fraction of molecular polymorphisms are maintained by balancing selection: 100 or so loci with equal allele frequencies and variance of effects $\alpha^2 = 0.02V_e$ would suffice. The result would not be greatly altered by allowing more alleles per locus, since the average heterozygosity ($\overline{2p_0q_0}$) cannot be greater than 1.

Apparent stabilizing selection: Individuals with extreme phenotypes will tend to be more homozygous, and so will tend to have lower fitness. The covariance between relative fitness and the squared deviation of the trait is $\mathrm{cov}(W/\bar{W}, z^2) = -2n\overline{\alpha^2 S p_0 q_0}$. The variance of z^2 is $n(2n-3)m_2^2 + nm_4$, where m_2 and m_4 are the second and fourth central moments of effects of individual loci. Using the assumption that allelic effects are independent of the equilibrium frequency, this reduces to:

$$\mathrm{var}(z^2) = 4\alpha^4 n(2n-3)(\overline{p_0q_0})^2 + 3\kappa n\overline{\alpha^4 p_0q_0[2 - 15p_0q_0(p_0^2 + q_0^2]}. \quad (2a)$$

This unpleasant expression differs from that derived by ROBERTSON (1956), because he assumed a normal distribution of breeding values, and because he calculated the curvature of the relation between fitness and the trait, rather than the regression of relative fitness on squared deviation. The expression can be simplified by noting that the second term will be negligible if the number of loci is large relative to the kurtosis of allelic effects. The apparent genetic load is then:

$$V_g/2V_s = -V_g\mathrm{cov}(W/\bar{W}, z^2)/[2var(z^2)] = \bar{S}/[2(1 + C^2)] \quad (2b)$$

where $\bar{S}$ is the average segregation load per locus, weighted by p_0q_0, and C is the coefficient of variation of p_0q_0 across loci. This is the same result as ROBERTSON (1956). It is remarkably similar to that for mutation/selection balance: the apparent load is at most half the segregation load at a single locus, $\bar{S}/2$. It may be somewhat reduced by variation across loci, but since overdominance cannot easily maintain extreme polymorphisms, this reduction will be small, Thus, if pleiotropic overdominance is to account for observed stabilizing selection, $\bar{S}$ must be large: ≈5% for equal allele frequencies, and more if frequencies differ. Because the overall variance in fitness is limited, this is only plausible if the number of polymorphisms involved is small, and creates serious difficulties for the model.

The response to selection: Is the observed response to selection consistent with pleiotropic overdominance? Note that although the long-term response may be largely due to new mutations (HILL 1982), the large initial responses which are often seen over the first ten or twenty generations do constrain variation in the base population. When the favoured allele is close to fixation, the selection coefficient due to natural selection is $-S_i/p_0$, and the coefficient due to artificial selection is $i\alpha\sqrt{V}$. (Here p_0 is the frequency of the favoured allele in the absence of artificial selection.) Directional selection will overwhelm balancing selection if its effect on a locus is greater than the average strength of balancing selection at extreme frequencies: $i\alpha\sqrt{V} > = \bar{S}/p_0$. With equal allele frequencies, this leads to the condition $i\sqrt{h^2} > \bar{S}\sqrt{2n}$. Since there cannot be a very large number of polymorphisms under strong balancing selection (note that the last term is proportional to the coefficient of variation in fitness caused by balancing selection), directional selection is likely to dominate. The maximum response to selection, assuming equal allele frequencies, is $n\alpha/2$; relative to the initial genetic standard deviation, this is $\sqrt{n/2}$, so that hundreds of polymorphisms are needed to allow responses of several genetic standard deviations. ROBERTSON (1956) showed that the loss of fitness due to small deviations in the mean $\Delta\sqrt{V}$, is $\bar{S}\Delta^2/2h^2$. This may not be excessive for small changes. However, at the selection limit, fitness could be reduced by $\approx n\bar{\bar{S}}/2$, where $\bar{\bar{S}}$ is the unweighted average of S_i. This would be large if there were enough strongly selected polymorphisms to account for polygenic variation and substantial responses to selection.

Frequency-dependent selection: Variability in haploids or predominantly self-fertilizing species cannot be maintained by overdominance. The model discussed above is restricted in other ways. Overdominance can maintain rare alleles only if it is highly symmetric (LEWONTIN, GINZBURG and TULJAPURKAR 1978). In contrast, if frequency-dependence maintains the polymorphisms, alleles can be kept at low frequencies, allowing a much larger response to selection.

A difficulty with frequency-dependence is that, in the simplest models, all genotypes will have equal fitnesses at equilibrium: no stabilizing selection can be

induced on the trait. Stabilizing selection will appear only if fitness effects are nonadditive. Suppose that the kth allele at a locus has frequency p_k, and adds effect α_k to the trait. At equilibrium, let the fitness of genotype kl be w_{kl}; the marginal fitnesses of all alleles must be the same: $\Sigma w_{kl}p_l = \bar{w}$ for all k. As before, the effects of alleles on the trait are drawn from a symmetric distribution with kurtosis κ, and are independent of effects on fitness. The regression of fitness on squared deviations gives an apparent load of $V_g/2V_s = \delta\omega/3[(\kappa - 1) + (\kappa + 1)/H]$. Here, H is the homozygosity, Σp_k^2, and $\delta\omega$ is the average fitness excess of heterozygotes ($\delta\omega = \Sigma(1 - w_{kk}/\bar{w})p_k^2$). Thus, extreme phenotypes will only be less fit if there is some element of overdominance: heterozygotes must be, on average, fitter than homozygotes. Whilst frequency-dependence can maintain polymorphisms despite underdominance, the polymorphism is more likely to be stable with some overdominance. Stabilizing selection is therefore expected as a side-effect of polymorphisms maintained by frequency-dependence. However, it will be weak, since the apparent load is roughly the product of the fitness excess of heterozygotes, and the homozygosity ($\delta\omega \approx (1 - w_{kk}/\bar{w})H$). By the same reasoning as for overdominance, $\delta\omega$ cannot be large without producing an excessive genetic variance in fitness. Although frequency-dependence could maintain substantial quantitative variability, and could allow a greater response to selection than with overdominance, it cannot easily account for observations of stabilizing selection.

DISCUSSION

Neither mutation/selection balance nor balancing selection alone can easily account for both the high heritabilities and the strong stabilizing selection which are commonly observed. Mutation produces polygenic variation so slowly that the net selection against the alleles involved must be weak if they are to reach significant frequencies ($s \approx V_m/V_e \approx 10^{-3}$): there must be a class of alleles with significant effects on the character, but very little effect on fitness. If (following TURELLI 1985) one imagines that all fitness effects are mediated by a number of additive traits, each under stabilizing selection such that $V_s = 20V_e$, this requires that the *total* variance of effects of each allele, summed over all traits, be very small: $\Sigma\alpha^2 = 2V_sV_m/V_e \approx 0.04$. This paper gives the obvious generalization of TURELLI's (1985) argument, by accounting for all pleiotropic effects on fitness, however mediated. It shows that mutation-selection balance is an unlikely cause of quantitative variation, even if one considers rather few traits: one must suppose that there is an abundant supply of mutations that affect metric characters, but which hardly reduce fitness ($s \approx 10^{-3}$). Moreover, mutation at a rate as low as $V_m \approx 10^{-3}V_e$ cannot induce a significant amount of apparent stabilizing selection.

Balancing selection could maintain considerable amounts of quantitative variation. However, in order to account for observed selection responses, it must be spread over many weakly selected loci. If this is so, then it cannot create the appearance of stabilizing selection on neutral traits. Measurements of strong stabilizing selection are therefore likely to reflect selection caused by those traits, or by a limited number of correlated traits.

These pleiotropic models have been presented as a general alternative to the popular models in which variation in a trait is explained solely by the effects on fitness of that trait. Both are caricatures of reality; the main aim of this paper is to focus attention on an alternative class of explanations. In the introduction, it was argued that direct models cannot explain morphological variation in *all* of an indefinite number of traits. However, *some* traits are known to directly cause large fitness differences (*e.g.*, bill shape in the Galapagos finches, male secondary sexual characters) (ENDLER 1986). This is trivially true for life history traits which are components of fitness. It is quite reasonable to suppose that stabilizing selection of strength $V_s \approx 20V_e$ acts on a few tens of independent character combinations. This would still be compatible with observations of strong stabilizing selection on any arbitrarily chosen character: one would generally expect some correlation with some directly selected trait.

An organism can be described by an essentially infinite number of characters, and many evolutionists believe that changes can occur in any of a very large number of degrees of freedom. The extreme view would be that each segregating allele has a unique vector of effects on the organism: a population could then evolve along a number of dimensions given by the number of alternative alleles (minus the number of loci). Even if each locus was constrained to have effects along a single direction (WAGNER 1989), several thousand degrees of freedom would still be available. Strong stabilizing selection cannot act on all of these variables: for the majority, genetic variance must be explained by pleiotropic effects of the kind described here. This is not to propose a neutral theory of morphological variation: the argument is that while every trait may affect fitness, its variability is determined primarily by the net selection on the genes involved. The prediction is that selection on a large number of traits could be factored into strong stabilizing selection on only a few principal components (their number being limited by the overall load); and that on the remaining components, one would see at most weak stabilizing selection, of the same order as the selection on the underlying polymorphisms. [The strength of stabilising selection on each component is

given by the eigenvalues of the matrix of selection gradients on variance components, γ (LANDE and ARNOLD 1983); these are the elements of Λ in Eq. 9 of PHILLIPS and ARNOLD (1989).]

How might one distinguish the many possible explanations for quantitative variation? There are two important issues. First, what is the distribution of effects of individual loci? This question cuts across the distinction between direct and pleiotropic models, and is independent of the mechanism maintaining variation. Among direct models, the Gaussian approximation (LANDE 1975) involves many alleles of small effect, whereas the "house of cards" approximation (TURELLI 1984) leads to a highly leptokurtic distribution of effects. Among the pleiotropic models, overdominance would involve a moderate number of loci with common alleles, while mutation/selection balance relies on a large number of loci, each near fixation. It might seem that since we can hardly count or map the polygenes, we cannot expect to measure the distribution of their effects. However, indirect approaches offer some comfort. TURELLI (1984) argued that because the production of genetic variability, relative to the standing genetic variance, is much larger than per-locus mutation rates ($V_m/V_g \gg \mu$), mutant alleles must have effects which are large relative to the standing variance, and hence must be rare. Another argument is that if loci in the base population are close to fixation, one would expect a transient increase in variance under directional selection, and a large variance in response. While drift and linkage disequilibrium can obscure these patterns (KEIGHTLEY and HILL 1989), they might be observed in large populations, if care was taken to eliminate disequilibrium before measurements were taken. TACHIDA and COCKERHAM (1988) have suggested another possibility: extending work by WRIGHT (1985b), they show that under stabilizing selection, the variance in fitness will be largely additive if alleles are rare, but largely epistatic (additive by additive) if alleles are common. It is extremely hard to measure genetic variation in fitness (CHARLESWORTH 1987), and rash to suppose that much fitness variation can be attributed to stabilizing selection on additive characters at equilibrium. However, TACHIDA and COCKERHAM's suggestion could be applied more rigorously to a biometric analysis of the *squares* of deviations of the character, rather than to fitness. This is essentially an analysis of the higher moments of the distribution of breeding values, originally suggested by FISHER, IMMER and TEDIN (1932) as a way of investigating dominance.

The second issue is whether fitness differences can be attributed to the character under study, to characters correlated with it, or to pleiotropic effects of the underlying genes. A systematic survey of the effects of spontaneous mutations on fitness and on quantitative characters in Drosophila would show whether the belief in widespread pleiotropy was justified, and would establish the contribution of deleterious mutations to quantitative variation. Experimental manipulations or mechanistic arguments may demonstrate direct effects (*e.g.*, ANDERSSON 1982; GRANT 1986), though it is not easy to show quantitatively that all apparent selection can be accounted for by direct selection.

In one of the few attempts to investigate selection via pleiotropic effects, LINNEY, BARNES and KEARSEY (1971) examined stabilizing selection on bristle number. Selection clearly does not act on bristle number itself, since it occurs at the larval stage (KEARSEY and BARNES 1970). By showing that a population consisting of a variety of homozygous flies showed the same reduction in variance as the outbred base population, they eliminated any role for overdominance. However, the alternatives of frequency-dependence or deleterious mutation remain open.

ROBERTSON (1967) suggested another approach, using a kind of genetic manipulation. Suppose that when selection for (say) increased bristle number is relaxed, the population returns towards its original value. This could be because too many bristles are inherently bad, or because alleles which increase bristle number are deleterious for other reasons. If one places a chromosome from the selected line into a homozygous background that reduces bristle number, the mean could be brought to its original value. If selection acts directly on bristles, one would see no further change, whereas if it acts on pleiotropic effects, one would see a further decrease below the initial value. SPIERS (1974) reported little change in lines manipulated in this way, supporting a pleiotropic interpretation. However, these lines were maintained at low density; since stabilising selection might only be significant under crowded conditions, he repeated the experiment at high density. Two of the four manipulated lines showed no return to the original mean, whilst two showed a $\approx$50% return. However, SPIERS (1974) found no consistent relation between bristle number and survival or competitive ability: he concluded that the relaxations were best explained by linkage to deleterious alleles, rather than by stabilising selection on bristle number or characters correlated with it.

If one accepts that variation in most quantitative traits is to be explained by the pleiotropic effects on fitness of the underlying genes, then the problem is to understand the nature and causes of fitness variation. This is an important but difficult problem, on which we have little reliable information. The possibility that most additive variance in fitness is simply due to mutation cannot be rejected (CHARLESWORTH 1987). If that is so, the slow generation of quantitative

variation by mutation requires that the selection against the underlying polygenes be correspondingly weak. On the other hand, if substantial fitness variation is due to balanced polymorphisms (as is suggested by MUKAI and NAGANO's (1983) data on viability variation in southern populations of *Drosophila melanogaster*), then these polymorphisms could maintain substantial heritabilities. The models discussed here do not of themselves distinguish between these possibilities: however, they do suggest many straightforward predictions which can be used to test the alternatives.

Thanks to JERRY COYNE, BILL HILL, LINDA PARTRIDGE, MICHAEL TURELLI, and two anonymous reviewers for their critical comments. This work was supported by grants from the National Science Foundation (BSR-8866548) the Science and Engineering Research Council (GR/E/08507), and by the Institute of Theoretical Dynamics, University of California, Davis.

LITERATURE CITED

ANDERSSON, M. B., 1982 Female choice selects for extreme tail length in a widowbird. Nature **199:** 818–820.

BARTON, N. H., and B. CHARLESWORTH, 1984 Genetic revolutions, founder effects and speciation. Annu. Rev. Ecol. Syst. **15:** 133–164.

BARTON, N. H., AND M. TURELLI, 1987 Adaptive landscapes, genetic distance, and the evolution of quantitative characters. Genet. Res. **49:** 157–173.

BULMER, M. G., 1973 The maintenance of the genetic variability of quantitative characters by heterozygote advantage. Genet. Res. **22:** 9–12.

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

CHARLESWORTH, B., 1984 The cost of phenotypic evolution. Paleobiology **10:** 319–327.

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

CHARLESWORTH, B., R. LANDE and M. SLATKIN, 1982 A neo-Darwinian commentary on macroevolution. Evolution **36:** 474–498.

CHARLESWORTH, D., AND B. CHARLESWORTH, 1987 Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. **18:** 237–268.

CLUTTON-BROCK, T. H., 1988. Reproductive success, pp. 472–485 in *Reproductive Success*, edited by T. H. CLUTTON-BROCK. University of Chicago Press, Chicago.

CROW, J. F., and M. J. SIMMONS, 1983 The mutation load in *Drosophila*, pp. 2–35 in *The Genetics and Biology of Drosophila*, Vol. 3c, edited by M. ASHBURNER, H. L. CARSON, and J. N. THOMPSON. Academic Press, New York.

DARWIN, C., 1859 *On the Origin of Species by Means of Natural Selection.* John Murray, London.

DAVIES, R. W., 1971 The genetic relationship of two quantitative characters in *Drosophila melanogaster*. II. Location of the effects. Genetics **69:** 363–375.

DAVIES, R. W., and P. C. WORKMAN, 1971 The genetic relationship of two quantitative characters in *Drosophila melanogaster*. I. Responses to selection and whole chromosome analysis. Genetics **69:** 353–361.

DOBZHANSKY, T., AND A. M. HOLZ, 1943 A re-examination of the problem of manifold effects of genes in *Drosophila melanogaster*. Genetics **28:** 295–303.

ENDLER, J. A., 1986 *Natural Selection in the Wild.* Princeton University Press, Princeton, N.J.

EWENS, W. J., 1979 *Mathematical Population Genetics.* Springer-Verlag, Berlin.

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

FISHER, R. A., F. R. IMMER and O. TEDIN, 1932 The genetical interpretation of statistics of the third degree in the study of quantitative characters. Genetics **17:** 107–124.

GILLESPIE, J. H., 1984 Pleiotropic overdominance and the maintenance of genetic variation in polygenic characters. Genetics **107:** 321–330.

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

GRANT, P. R., 1986 *Ecology and Evolution in Darwin's Finches.* Princeton University Press, Princeton, N.J.

HILL, W. G., 1982 Predictions of response to artificial selection from new mutations. Genet. Res. **40:** 255–278.

HILL, W. G., AND P. D. KEIGHTLEY, 1988 Interrelations of mutation, population size, artificial and natural selection, pp. 57–70 in *Proceedings of the Second International Conference on Quantitative Genetics*, edited by B. S. WEIR, E. J. EISEN, M. M. GOODMAN and G. NAMKOONG. Sinauer Press, Sunderland, Mass.

KEARSEY, M. J., and B. W. BARNES, 1970 Variation for metrical characters in *Drosophila* populations. II. Natural selection. Heredity **25:** 11–21.

KEIGHTLEY, P. D., and W. G. HILL, 1983 Effects of linkage on response to directional selection from new mutations. Genet. Res. **42:** 193–206.

KEIGHTLEY, P. D., and W. G. HILL, 1988 Quantitative genetic variability maintained by mutation-selection balance in finite populations. Genet. Res. **52:** 33–43.

KEIGHTLEY, P. D., and W. G. HILL, 1989 Quantitative genetic variability maintained by mutation-selection balance: sampling variation and response to subsequent directional selection. Genet. Res. **54:** 45–57.

KEIGHTLEY, P. D., and H. KACSER, 1987 Dominance, pleiotropy and metabolic structure. Genetics **117:** 331–341.

KIMURA, M., 1965 A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proc. Natl. Acad. Sci. USA **54:** 731–736.

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.

KIRKPATRICK, M., and D. LOFSVOLD, 1989 The evolution of complex quantitative characters. Genome (in press).

LANDE, R., 1975 The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. **26:** 221–235.

LANDE, R., 1980 The genetic covariance between characters maintained by pleiotropic mutation. Genetics **94:** 203–215.

LANDE, R., and S. J. ARNOLD, 1983 The measurement of selection on correlated characters. Evolution **37:** 1210–1226.

LEWONTIN, R. C., L. R. GINZBURG and S. D. TULJAPURKAR, 1978 Heterosis as an explanation for large amounts of genic polymorphism. Genetics **88:** 149–170.

LINNEY, R., B. W. BARNES and M. J. KEARSEY, 1971 Variation for metrical characters in *Drosophila* populations. III. The nature of selection. Heredity **27:** 163–174.

LYNCH, M., 1988 The rate of polygenic mutation. Genet. Res. **51:** 137–148.

MAYNARD SMITH, J., 1983 The genetics of stasis and punctuation. Annu. Rev. Genet. **17:** 11–25.

MITCHELL-OLDS, T., and R. SHAW, 1987 Regression analysis of natural selection: statistical inference and biological interpretation. Evolution **41:** 1149–1161.

MUKAI, T., and S. NAGANO, 1983 The genetic structure of natural populations of *Drosophila melanogaster*. XVI. Excess additive genetic variance of viability. Genetics **105:** 115–134.

PHILLIPS, P. C., and S. J. ARNOLD, 1989 Visualizing multivariate selection. Evolution **43:** 1209–1222.

ROBERTSON, A., 1956 The effect of selection against extreme deviants based on deviation or on homozygosis. J. Genet. **54:** 236–248.

ROBERTSON, A., 1967 The nature of quantitative genetic variation, pp. 265–280, in *Heritage from Mendel*, edited by R. A. BRINK. University of Wisconsin Press, Madison.

ROUGHGARDEN, J., 1972 Evolution of niche width. Am. Nat. **106:** 683–718.

RUSSELL, W. A., G. F. SPRAGUE and L. H. PENNY, 1963 Mutations affecting quantitative characters in long-time inbred lines of maize. Crop Sci. **3:** 175–178.

SHERIDAN, A. K., and J. S. F. BARKER, 1974 Two-trait selection and the genetic correlation. II. Changes in the genetic correlation during two-trait selection. Aust. J. Biol. Sci. **27:** 89–101.

SIMMONS, M., and J. F. CROW, 1977 Mutations affecting fitness in Drosophila populations. Annu. Rev. Genet. **11:** 49–78.

SLATKIN, M., 1979 Frequency- and density-dependent selection on a quantitative character. Genetics **93:** 755–771.

SPIERS, J. G. C., 1974 The effects of larval competition on a quantitative character in *Drosophila melanogaster*. Ph.D. thesis, University of Edinburgh.

SVED, J. A., T. E. REED and W. F. BODMER, 1967 The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55:** 469–481.

TACHIDA, M., and C. C. COCKERHAM, 1988 Variance components of fitness under stabilizing selection. Genet. Res. **51:** 47–53.

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., 1985 Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics. **111:** 165–195.

TURELLI, M., and N. H. BARTON, 1989 The dynamics of polygenic characters under selection. Theor. Popul. Biol. (in press).

WAGNER, G. P., 1989 Multivariate mutation-selection balance with constrained pleiotropic effects. Genetics **122:** 223–234.

WRIGHT, S., 1935a Evolution in populations in approximate equilibrium. J. Genet. **30:** 257–266.

WRIGHT, S., 1935b The analysis of variance and the correlation between relatives with respect to deviations from an optimum. J. Genet. **30:** 243–256.

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

YOO, B. H., 1980 Long-term selection for a quantitative character in large replicate populations of *Drosophila melanogaster*. II. Lethals and visible mutants with large effects. Genet. Res. **35:** 19–31.

Corresponding editor: M. TURELLI