# 11

# THE NEUTRAL DIVERGENCE OF QUANTITATIVE TRAITS

*There are some enterprises in which a careful disorderliness is the true method.*
— Herman Melville, *Moby Dick*

*Walsh updates 3 Nov 2011 to v. 27 September 2008, see head of Tex Code for details*

In the preceding chapter, we learned how the opposing forces of random genetic drift and mutation lead to an equilibrium level of within-population genetic variance. In contrast, the phenotypic variance among isolated populations may continue to increase nearly indefinitely for neutral characters, as isolated demes or species recurrently acquire and become fixed for independent mutations. Here, we explore neutral factors that can drive the evolutionary dynamics of the among-population variance. As in Chapter 10, we will start with the situation in which the time span is short enough that most of the change in population-mean phenotypes is driven by drift acting on existing variation rather than by new alleles introduced by mutation. We then explore the consequences of longer-term divergence, with mutation playing an increasingly dominant role, showing that eventually the rate of divergence for neutral characters may become essentially independent of local effective population sizes.

We conclude by using this theory to first develop statistical tests of whether an observed pattern of phenotypic divergence is consistent with model of strict neutral drift and mutation. We then review the application of these tests to both standard morphological traits and to divergence in the patterns of gene regulation across species. Although few quantitative traits may actually evolve in a purely neutral fashion, a more compelling case for selection can always be made if the hypothesis of neutrality can be formally rejected. For example, an observed divergence of isolated lines that is significantly less than the neutral expectation provides evidence of stabilizing selection, whereas the reverse supports a role for diversifying selection. In addition, as populations become diminishingly small in size, drift begins to overwhelm selection, promoting nearly neutral patterns of evolution. Recall that Chapters 7-9 considered the complementary topic of tests for departures from neutrality at specific *loci*, as opposed to specific *traits*, which is our focus here.

## SHORT-TERM DIVERGENCE

We start with the special case in which all gene action is additive and random genetic drift is the only evolutionary force. Most of the predictions of this model can be expressed in terms of two observable quantities: the additive genetic variance in the base population $\sigma_A^2(0)$, and the effective population sizes $N_e$ of the isolated

lineages. The expected among-population genetic variance, $\sigma_B^2(t)$, under neutrality is obtained by noting that the mean genotypic value at a diallelic locus $i$ is $2a_ip_i$ (there being two genes per locus, each with additive effect $a_i$ with probability $p_i$, and effect 0 with probability $1-p_i$). The variance among populations for this locus is (from the definition of the variance) the expected value of the square of additive effects minus the square of their expected values, or $E\{[2a_ip_i(t)]^2\}-\{E[2a_ip_i(t)]\}^2$. This simplifies to $4a_i^2\sigma_{p_i}^2(t)$, where $\sigma_{p_i}^2(t)$ is the expected among-population variance in allele frequency. Summing over all loci, assuming negligible gametic-phase disequilibrium, and substituting from Equation 2.12a gives

$$\sigma_B^2(t) = 4\sum_{i=1}^{n} a_i^2 p_i(0)[1-p_i(0)]\left\{\frac{1}{N_{fo}} + \left[1-\left(1-\frac{1}{2N_e}\right)^t\right]\right\} \tag{11.1a}$$

$$= \left(\frac{1}{N_{fo}} + 2f_t\right)\sigma_A^2(0) \tag{11.1b}$$

where $N_{fo}$ is the effective number of founders per line, the inbreeding coefficient $f_t$ follows from Equation 2.4c, and the time index is defined such that $t=0$ denotes the final generation of the base population and $t=1$ denotes the founding generation for the isolated lines. Equation 11.1 shows that, under the assumptions of this ideal model, the expected variance among genotypic means of isolated populations increases linearly with the inbreeding coefficient, asymptotically approaching a limit (as $f_t \to 1$) that is very close to twice the additive genetic variance in the base population (Wright 1951). Equation 11.1 describes how any initial variation is partitioned by drift during the random (and differential) fixation of these initial alleles in the diverging populations. Under the assumption of additivity, Equation 11.1b holds regardless of the number of alleles at the underlying loci.

Ignoring the generally minor contribution ($N_{fo}^{-1}$) from the baseline founder effect, this limiting result may be obtained in a simpler manner. Because the probability of fixation of a neutral allele is equal to its initial frequency, when the process of random drift is completed, a proportion $p_i(0)$ of the populations will have genotypic value $2a_i$, while the remaining proportion, $1-p_i(0)$, will have genotypic value 0. The mean genotypic value is therefore $2a_ip_i(0)$ and the mean squared value is $(2a_i)^2p_i(0)$, which yields the among-population variance $4a_i^2p_i(0)[1-p_i(0)] = 2\sigma_{A_i}^2(0)$.

The expression for $\sigma_B^2(t)$ give by Equation 11.1 only considers the true genetic divergence among lines (the **evolutionary variance**), which can in principle be obtained by an analysis of variance of phenotypic variation within and among lines. If, however, one simply focuses on the raw variance of the observed means, additional sources of variation, associated with finite sample sizes, will contribute to the observed divergence (Hill 1972; Lynch 1988). For example, when the mean phenotype of each line is determined using $n$ progeny from $N/2$ matings (involving $N/2$ males and females, for a total parental sample size of $N$), there can be three additional sources of variance to add to Equation 11.1:

i) The segregational variance $(1-f_{t-1})\sigma_A^2(0)/(Nn)$ of the mean offspring value about the mean breeding value of their parents resulting from the sampling of $Nn/2$ individuals. This follows as the segregational variance equals half the total variance (Chapters 15 and 22);

ii) The sampling variance $\sigma^2_{E_m}/(N/2)$ associated with maternal effects resulting from the sampling of $N/2$ mothers;

iii) The residual variance $\sigma^2_{E_s}/(Nn/2)$ associated with special environmental effects averaged over the entire progeny pool.

Finally, the among-line variances in consecutive generations will be correlated as a consequence of shared ancestry,

$$\sigma_B(t,t') = \left(\frac{1}{N_{fo}} + 2f_t\right)\sigma^2_A(0) \qquad \text{for } 0 < t < t' \tag{11.2}$$

Equation 11.2 assumes no transmission of maternal effects across generations, which if present would further inflate this covariance.

A few words should also be said about the potential importance of nonadditive gene action. From Table 10.2, it can be seen that in the presence of dominance, the among-population variance (in the absence of any new mutation) eventually asymptotes at $\sigma^2_B = 2\sigma^2_A + 2\sigma_{ADI} + \sigma^2_{DI}$. Thus, dominance can magnify or reduce the among-population variance depending upon the magnitudes of $\sigma^2_{DI}$ and $\sigma_{ADI}$ and on the sign of the latter. In addition, the asymptotic contribution from epistatic interactions involving additive effects is equal to $2^n\sigma^2_{A^n}$ for $n$-locus epistasis, i.e., $4\sigma^2_{AA}$ for additive × additive epistasis, and $8\sigma^2_{AAA}$ for additive × additive × additive epistasis. Thus, epistasis involving large numbers of loci can, in principle, greatly magnify the among-population variance, even if it appears to be of relatively minor importance within populations.

## Sampling Error

We now consider the sampling properties of the among-population genetic variance by reference to a particular experimental design, again assuming a character with a strictly additive basis (Hill 1972; Lynch 1988). Starting from a base population with additive genetic variance $\sigma^2_A(0)$, $L$ replicate lines are isolated and subsequently maintained each generation with $N/2$ random monogamous matings. Due to the fact that only a finite number of lines is studied, the among-population variance that actually develops in any particular experiment, $\widehat{\sigma}^2_B(t)$, will deviate from the expectation $\sigma^2_B(t)$ given by Equation 11.1b. Moreover, due to finite sample sizes within populations, the among-population variance estimated by the investigator, $\mathrm{Var}(B,t)$, will further deviate from $\widehat{\sigma}^2_B(t)$. This first source of variation, $\sigma^2[\widehat{\sigma}^2_B(t) - \sigma^2_B(t)]$, is a function of population-genetic structure and, for a fixed system of mating, is largely beyond the control of the investigator. The second source of variation, the **sampling variance**, $\sigma^2[\mathrm{Var}(B,t) - \widehat{\sigma}^2_B(t)]$, arises in estimating $\widehat{\sigma}^2_B(t)$ from the among-line sample variance $\mathrm{Var}(B,t)$. Its contribution can be minimized by the use of large sample sizes.

Since our concern here is variation in divergence do to genetic changes generated by random drift, we focus on the situation in which the among-line divergence has been measured in such a way as to eliminate nongenetic causes (such as environmental trends). Suppose that the same experiment has been repeated many different times, on each occasion starting with $L$ lines from the same base population. Due to the variation in the drift process and the finite number of observed

lines, each set of experimental lines will develop its own temporal pattern of realized among-population variance. The expected variation in the realized variance among these hypothetical replicate experiments provides a measure of confidence that one can have in the results of any single experiment. Letting $\widehat{\sigma}_B^2(t)$ be the realized among-population variance at generation $t$ for a particular experiment, the expected variance of this quantity among replicate experiments is

$$\sigma^2[\widehat{\sigma}_B^2(t)] \simeq \frac{4\sigma_A^4(0)}{L-1}\left[\frac{1}{2N_{fo}^2} + 2\left(1+\frac{1}{N_{fo}}\right)f_t^2 + \sigma_f^2(t)\right] \tag{11.3}$$

Although, in practice, one generally performs such a divergence experiment only once, the utility of Equation 11.3 is that it is entirely expressed in terms of observable parameters, so that some idea of the reliability of estimates of $\sigma_B^2(t)$ can be determined in advance. In most situations, the terms in Equation 11.3 involving the founder number ($N_{fo}$) will be of second or third order and can be ignored.

The variance $\sigma_f^2(t)$ in the amount of actual inbreeding between individuals in the population requires additional comments. This has been examined in detail in Lynch (1988), drawing heavily from the results of Weir et al. (1980) and Cockerham and Weir (1983). For freely recombining loci, $\sigma_f^2$ is zero when the pedigree structure is fixed, e.g., obligate selfing, full-sib mating, the maximum avoidance systems of Wright (1921), and the circular systems of Kimura and Crow (1963); and even with fairly tightly linked loci, $\sigma_f^2(t)$ is generally negligible in any generation under selfing or full-sib mating. However, under most natural mating schemes, some individuals mate by chance with closer relatives than do others. This results in variation in $f$ among members of the same population, which because of sampling, accumulates as among-population variance in $f$. The theoretical value of $\sigma_f^2(t)$ under different systems of mating is of special interest because empirical studies usually do not record the essential pedigree information for its computation. For larger population sizes, even with unlinked loci, if the sexes are separate and matings are monogamous, its squared coefficient of variation $[\mathrm{CV}(f_t)]^2 = \sigma_f^2(t)/f_t^2$ can attain values of 0.1 to 1.0 in the first two to four generations of isolation, which is enough to contribute significantly to $\sigma^2[\widehat{\sigma}_B^2(t)]$. However, after six or so generations have passed, $\sigma_f^2(t)$ can be safely ignored regardless of the population size, even with tightly linked loci.

Ignoring the initial founder effect, these results indicate that the coefficient of variation of the among-population variance is $\sqrt{2\{1+[\mathrm{CV}(f_t)]^2\}/(L-1)}$, which is generally on the order of $\sqrt{2/L}$, although in some cases being as high as $2/\sqrt{L}$. Thus, studies of phenotypic divergence *need to have very large number of replicates to be statistically reliable*. For example, if it is desirable to reduce the standard error of the among-line variance to 10% of the expectation under the null hypothesis of neutrality and additivity, a minimum of 200 lines should be studied.

One can assess the fit of the additive theory to actual data under two different settings. In the first, we have a single estimate of the among-line variance and we compare this result to the value expected from theory (see Example 11.2). In the second setting, we have a series of among-line estimates at different time points, allowing us to consider the temporal pattern of increase in $\sigma_B^2$, which as noted above, should eventually reach a constant as $f \to 1$. When such a temporal sequence of $\mathrm{Var}(B,t)$ is available, these may be regressed on $f_t$. Under the null hypothesis of neutral additive genes, from Equation 11.1b the expected slope of such a regression

is $2\sigma_A^2(0)$. However, because of shared ancestry, consecutive estimates of mean phenotypes obtained from the same lines are nonindependent (Equation 11.2), violating a fundamental assumption of ordinary least-squares (OLS) regression analysis, and generalized least squares (GLS) must be used instead (Chapter 16, LW Chapter 8). For example, once the lines have become completely inbred (and ignoring mutation), all future values of $\widehat{\sigma}_B^2(t)$ must be fixed, and therefore should not be given equal weight in the regression analysis. The expected covariance of $\widehat{\sigma}_B^2$ between generations with inbreeding levels $f_t$ and $f_{t'}$ is

$$\sigma[\widehat{\sigma}_B^2(t), \widehat{\sigma}_B^2(t')] \simeq \frac{4\sigma_A^4(0)}{L-1}\left[\frac{1}{2N_{fo}^2} + 2\left(1+\frac{1}{N_{fo}}\right) f_t f_{t'} + \lambda_1^{t'-t}\sigma_f^2(t)\right] \quad \text{for } t < t' \quad (11.4)$$

where $\lambda_1 = 1 - 1/(2N)$. Lynch (1988) provides approximate expressions for the standard errors of the slope and intercept that account for the intrinsic correlations in the data, assuming measurements of $\mathrm{Var}(B,t)$ in progressive generations. Chapter 16 also considers the same problem, but in the context of response in a selection experiment and frames the solution in a GLS framework. The variance of the regression coefficient increases with the duration of the experiment, but is essentially constant after the fourth generation of inbreeding. At that point, the standard error ranges from approximately $4\sigma_A^2(0)/\sqrt{L}$ under obligate self-fertilization to $3\sigma_A^2(0)/\sqrt{L}$ with larger $N_e$, implying coefficients of variation in the range of $1.5/(f\sqrt{L})$ to $2/f\sqrt{L}$. For large $f$, these are not greatly different from the sampling variances of single-point estimates noted above.

### Confidence Intervals on a Sample Variance

Given the critical role played by the sample variance in empirical tests of the additive-drift model, we digress here to briefly consider a few statistical issues related to estimating a variance from a sample. Provided individual observations used to estimate a sample variance are uncorrelated with $y_i \sim N(\mu, \sigma^2)$, then (LW Equation A5.14c) for a sample of size $n$ we have for $\mathrm{Var} = \sum(y_i - \overline{y})^2/(n-1)$ that

$$(n-1)\mathrm{Var} \sim \sigma^2 \chi_{n-1}^2 \quad (11.5a)$$

As a result, confidence intervals for the true variance $\sigma^2$ based on the observed sample variance Var follow from critical values for a $\chi^2$ distribution. Letting $X_{p,n}$ satisfy $\Pr(\chi_n^2 \le X_{p,n}) = p$, then

$$\Pr(X_{\alpha/2,n} \le \chi_n^2 \le X_{1-\alpha/2,n}) = 1-\alpha \quad (11.5b)$$

From Equation 11.5a, substituting $(n-1)\mathrm{Var}/\sigma^2$ for $\chi_{n-1}^2$, we have

$$\Pr\left(X_{\alpha/2,n-1} \le \frac{(n-1)\mathrm{Var}}{\sigma^2} \le X_{1-\alpha/2,n-1}\right) \quad (11.5c)$$

$$= \Pr\left(\frac{1}{X_{\alpha/2,n-1}} \ge \frac{\sigma^2}{(n-1)\mathrm{Var}} \ge \frac{1}{X_{1-\alpha/2,n-1}}\right) = 1-\alpha \quad (11.5d)$$

giving

$$\Pr\left[\left(\frac{n-1}{X_{1-\alpha/2,n-1}}\right)\mathrm{Var} \le \sigma^2 \le \left(\frac{n-1}{X_{\alpha/2,n-1}}\right)\mathrm{Var}\right] = 1-\alpha \tag{11.6}$$

This motivates a $(1-\alpha)100\%$ confidence interval for the true variance $\sigma^2$ given the observed sample variance Var. As shown in Figure 11.1, confidence intervals for $\sigma^2$ are asymmetrical about Var, and tend to be quite large even for modest sample sizes.

**— Insert Figure 11.1 Here–**

We can also use Equation 11.5c to access the significance of an observed sample variance given some assumed value $\sigma_0^2$. Rearranging Equation 11.5c gives

$$\Pr\left[\left(\frac{\sigma_0^2}{n-1}\right)X_{\alpha/2,n-1} \le \mathrm{Var} \le \left(\frac{\sigma_0^2}{n-1}\right)X_{1-\alpha/2,n-1}\right] = 1-\alpha \tag{11.7}$$

An observed sample variance outside of this interval is said significantly different at level $\alpha$ from that expected under the null. Figure 11.1 plots these critical values (scaled by $\sigma_0^2$) as a function of sample size. Equation 11.7 gives critical values for a two-sided test. Values for one-sided tests easily follow by replacing $\alpha/2$ by $\alpha$ in the suitable upper or lower critical value.

If the true variance is really $\sigma_1^2 \neq \sigma_0^2$, then the **power** (LW Appendix 5) for this parameter value is just the probablity that a sample variance falls outside of the interval given by Equation 11.7, which is a function of the sample size $n$ and the assigned significance $\alpha$ for the test. Letting $\beta$ denote the probability of a type II error (failing to declare a test significant when the null is false), we can obtain this from Equation 11.7 by noting that now $[(n-1)/\sigma_1^2]\,\mathrm{Var} \sim \chi^2_{n-1}$. Multiplying all terms of Equation 11.7 by $(n-1)/\sigma_1^2$ gives the probablity $\beta$ of a sample variance failing to be declared significant as

$$\begin{aligned}\beta &= \Pr\left[\left(\frac{\sigma_0^2}{n-1}\right)\left(\frac{n-1}{\sigma_1^2}\right)X_{\alpha/2,n-1} \le \chi^2_{n-1} \le \left(\frac{\sigma_0^2}{n-1}\right)\left(\frac{n-1}{\sigma_1^2}\right)X_{1-\alpha/2,n-1}\right] \\ &= \Pr\left[\left(\frac{\sigma_0^2}{\sigma_1^2}\right)X_{\alpha/2,n-1} \le \chi^2_{n-1} \le \left(\frac{\sigma_0^2}{\sigma_1^2}\right)X_{1-\alpha/2,n-1}\right] \end{aligned} \tag{11.8a}$$

Hence, the power $1-\beta$ is

$$\Pr\left[\chi^2_{n-1} \le \left(\frac{\sigma_0^2}{\sigma_1^2}\right)X_{\alpha/2,n-1}\right] + \Pr\left[\chi^2_{n-1} \ge \left(\frac{\sigma_0^2}{\sigma_1^2}\right)X_{1-\alpha/2,n-1}\right] \tag{11.8b}$$

---

**Example 11.1.** Consider a sample variance estimated from $n = 10$ observations (e.g., the between-group variance estimated from the means of ten replicate lines). Since $\Pr(\chi^2_9 \le 2.700) = 0.025$ and $\Pr(\chi^2_9 \le 19.023) = 0.975$, Equation 11.6 gives the 95% confidence interval ($\alpha = 0.05$) on the true variance $\sigma^2$ as between (9/19.023)Var and (9/2.7)Var, or 0.473·Var to 3.333·Var, for an uncertainty in $\sigma^2$ spanning almost a full order of magnitude.

What observed values of the sample variance are unlikely given an assumed variance of $\sigma_0^2$? From Equation 11.7, the upper and lower critical values (for a two-sided test with $\alpha = 0.05$) are $(2.700/9)\sigma_0^2 = 0.3 \cdot \sigma_0^2$ and $(19.023/9)\sigma_0^2 = 2.11 \cdot \sigma_0^2$. Finally, what is the power of this design (again taking $\alpha = 0.05$) when $\sigma_1^2 = \sigma_0^2/2$? Equation 11.8b gives the power as

$$\Pr\left(\chi_9^2 \leq \frac{2.700}{2}\right) + \Pr\left(\chi_9^2 \geq \frac{19.023}{2}\right) = 0.39$$

and hence a type II error rate of 61% when the true variance is half the assumed variance. A similar calculation assuming $\sigma_1^2 = 2\sigma_0^2$ gives a power of 0.20, or a type II error rate of 80%. Useful **R** commands for these calculations are **`pchisq(x,n)`**, which returns $\Pr(\chi_n^2 \leq x)$, and hence **`1- pchisq(x,n)`** returns $\Pr(\chi_n^2 \geq x)$, while **`qchisq(p,n)`** returns $X_{p,n}$.

---

## Empirical Observations

As an example of the application of the preceding results, consider the results from a large drift experiment with laboratory cultures of the flour beetle *Tribolium castaneum* (Rich et al. 1984). The authors followed twelve replicate populations at four population sizes (1:1 sex ratio, random mating) over 20 consecutive generations. Each generation, the mean pupal weight (in $\mu$g) of each population was obtained from a bulk sample of 100 random individuals. The additive genetic variance was estimated to be 460 in the base population. The observed $\mathrm{Var}(B,t)$ are plotted as a function of $f_t$ in Figure 11.2, along with the expected divergence $2\sigma_A^2(0)f_t = 920f_t$ (solid lines). The dashed lines, obtained by using Equation 11.3 for the expected variance and substituting this into Equation 11.7 (using $\alpha = 0.05$ and $n = 12$), give the limits of the among-population variance beyond which there is less than a 5% chance for the realization of the drift process under the null to generate these values. Since these bounds ignore measurement error, they may be regarded as conservative confidence limits (as they assumed a smaller variance and hence are too narrow). Nevertheless, almost all of the observations, with the exception of the clusters of the late generations at $N = 10$ and 20, lie within these limits. The least-squares regressions of the data are given by the dotted lines. The slope of each regression is less than the expected 920, but all are within two standard errors of the expectation. The observed patterns are fairly consistent with a hypothesis of random drift of neutral additive genes. The observed declines in $\mathrm{Var}(B,t)$ late in the experiment at the two smallest population sizes may have simply arisen by chance and remained there due to intergenerational correlations (Equation 11.2).

**–Insert Figure 11.2 Here–**

The results of some other short-term divergence experiments given in Figure 10.3 show no evidence for nonlinear increases in the among-population variance with inbreeding. Eisen and Hanrahan (1974) have argued that the divergence of

growth and reproductive rates in inbred lines of mice is more rapid than can be accounted for by the additive genetic variance in the base population, and Bryant et al. (1986) suggested the same for morphological traits in bottlenecked housefly lines. The implication of these authors is that some nonadditive variance is converted by inbreeding into $\sigma_A^2$ (Chapter 10), leading to a faster between-line divergence. In neither case was it verified that the departures from expectations were significant, but this lack of significance is tempered by the low power of their designs.

### Lande's $F$ Test

Is an observed divergence over a modest amount of time significantly different than expected by drift? For the case in which one has only a single estimate of the among-population divergence, Lande (1977) suggested the statistic

$$F = \frac{\mathrm{Var}(B,t)}{t \cdot \mathrm{Var}(A,0)/N_e} \tag{11.9a}$$

as a test for neutrality. As noted by Lande, under approximate assumptions, this follows an $F$ distribution, which we can show as follows. Assuming the trait is normally distributed, the sample mean $\overline{y}_i \sim N[\mu(0), \sigma_B^2(t)]$, where we have assumed that sampling variance terms are small enought to be ignored. From Equation 11.5a,

$$\mathrm{Var}(B,t) = \frac{1}{L-1}\sum_{i=1}^{L}(\overline{y}_i - \overline{y}_\cdot)^2 \sim \frac{\sigma_B^2}{L-1}\chi_{L-1}^2 \tag{11.9b}$$

Ignoring the (usually) small founder effect, Equation 11.1b gives

$$\sigma_B^2(t) = 2f_t\sigma_A^2(0) = 2\left[1 - \left(1 - \frac{1}{2N_e)}\right)^t\right]\sigma_A^2(0) \simeq t\sigma_A^2(0)/N_e \quad \text{for } t \ll N_e \tag{11.9c}$$

and hence

$$\mathrm{Var}(B,t) \sim \frac{t\sigma_A^2(0)/N_e}{L-1}\chi_{n-1}^2 \tag{11.9d}$$

Assuming that $\mathrm{Var}(A,0)$ is a good estimate of $\sigma_A^2(0)$, substitution into Equation 11.9a gives

$$F \simeq \frac{\chi_{L-1}^2}{L-1} \sim F_{L-1,\infty}$$

The last step follows from the definition of an $F$ distribution (LW Appendix 5). Hence Lande's $F$ statistic follows an $F$ distribution with $L-1$ numerator and infinite denominator degrees of freedom. More generally, since $\sigma_A^2(0)$ is estimated by $\mathrm{Var}(A,0)$, the denominator degrees of freedom are those associated with this estimate.

A couple of approximations were required to reach this point. One check of their validity is that if $x \sim \sigma^2\chi_n^2/n$, then $\sigma^2(x) = 2\sigma^4/n$ (LW Equation A5.15b). Hence, the numerator should have a variance approximately equal to $2[2f_t\sigma_A^2(0)]^2/(L-1)$. Ignoring the added contribution from sampling error, this can be seen to be approximately true for large $N_{fo}$ by reference to Equation 11.3. However, with selfing and

full-sib mating, the expected variance is about twice and 1.5 times too high respectively. Thus, Lande's approach should be restricted to lines with at least moderate effective size. Moreover, as we will see below, all of the proceeding formulae for $\sigma_B^2$ become questionable for $t > N_e$, because they ignore the contribution from new mutations. Hence, Lande's $F$ test is best thought of as one for short-term divergence, such as would be seen in a laboratory experiment or at most a modest amount of time in a set of natural populations.

---

**Example 11.2.** Lande (1977) used Equation 11.9a to evaluate the results of a 12-year divergence experiment involving five populations of *Drosophila pseudoobscura* (Anderson 1973). Two of the populations had been maintained at 25°C, two at 27°C, and one at 16°C. They were then raised in two common environments (16 and 25°C) and measured for wing length. Estimates of the additive genetic variance for these two environments were 0.88 and 0.77, while the among-population variances were approximately 6.62 and 4.37 respectively. An approximate upper bound for the number of generations of divergence is $t = 150$, whereas the effective population size probably always exceeded $N_e = 1000$. The use of these extreme bounds gives conservative estimates of $F$, making it more difficult to demonstrate diversifying selection on wing length. Even so, the ratios of observed to expected among-population variance are 50 and 38, both of which are highly significant (comparing these with the critical $F$ ratio with four degrees of freedom in the numerator, and infinite degrees of freedom in the denominator). Thus, the hypothesis that the observed line divergence is solely attributable to random genetic drift can be rejected confidently. More likely, the different thermal conditions resulted in selection for different wing lengths.

---

## LONG-TERM DIVERGENCE

Our previous results were simply concerned with how any initial variation is partitioned among lines during drift/inbreeding. While this is occurring, new variation is constantly being generated by mutation, further driving divergence (Haldane 1949). Polygenic mutation was first incorporated into the theory of population divergence by Dempster (appendix in Bailey 1959) and was subsequently studied by Lande (1976), Chakraborty and Nei (1982), and Lynch and Hill (1986). Again focusing on a character with a purely additive genetic basis, starting with an ancestral-population genetic variance of $\sigma_A^2(0)$, and assuming the infinite-alleles model, the expected variance of genotypic means for replicate populations isolated $t$ generations in the past is

$$\sigma_B^2(t) = 2\sigma_m^2 t + 2[\sigma_A^2(0) - 2N_e\sigma_m^2][1 - e^{-t/(2N_e)}] \tag{11.10}$$

where $\sigma_m^2$ is the per-generation mutational rate of input of genetic variance, as described in Chapter 10. This expression shows that as $t$ becomes large, the expected rate of increase of the among-population variance for a neutral quantitative trait becomes a constant $2\sigma_m^2$ per generation. The same formulation applies to the among-species genetic covariance for a pair of traits, if the mutational rate of production

of covariance between the traits is substituted for $\sigma_m^2$ (Lande 1979).

Thus, under the infinite-alleles model, the asymptotic divergence rate is independent of the population size, just as it is in the neutral theory of molecular evolution (Chapter 6; Kimura 1983). Although the expected number of new mutations entering a population each generation is $2Nu$ per locus, the probability of fixation of a new mutation is its initial frequency $1/(2N)$, so the expected number of mutations fixed per locus per population per generation is simply $u$. For a set of $L$ populations, with each fixed mutation causing an increase in expected among-population variance of $\sim E[(2a)^2]$, and $n$ loci contributing, the asymptotic divergence rate is $nuLE[(2a)^2/L] = 2\sigma_m^2$.

Under the assumptions of the infinite-alleles model, the asymptotic divergence rate of $2\sigma_m^2$ is a fairly general result. It is independent of the degree of dominance of new mutations, of the linkage relationships of the constituent loci, and of the mating system (Lynch and Hill 1986). This is because both dominance and gametic-phase disequilibrium are transient properties of alleles en route to loss or fixation, and not cumulative phenomena, and because the probability of fixation of a new neutral mutation is equal to its initial frequency regardless of the breeding system.

How long should populations be isolated before one should start to worry about the contribution of new mutations to their divergence? From Equation 11.10, it can be seen that this depends on the initial level of genetic variance and on the effective sizes of the derived isolates. In Figure 11.3 it is assumed that the initial base population is in drift-mutation equilibrium, so that $\sigma_A^2(0) = 2N_e\sigma_m^2$, and that the isolated lineages have rapidly attained the same effective sizes ($N_e$). Under these circumstances, by the time $N_e$ generations have elapsed, polygenic mutation subsequent to the isolation event has caused about 20% of the divergence, whereas for $t > 3N_e$ generations, the majority of the divergence is due to new mutations.

**–Insert Figure 11.3 Here–**

As emphasized in the preceding chapter, alternatives exist to the infinite-alleles model, raising questions about the appropriate structure of a neutral null model. For example, Cockerham and Tachida's (1987) model, which assumes a finite number of alleles with each new mutational effect being independent of the prior allelic state (the house-of-cards model), yields an equilibrium among-population variance

$$\sigma_B^2 = 2[1 - E(H)]\sigma_A^2(\infty) \tag{11.11}$$

where from Chapter 10, $E(H)$ is the expected heterozygosity per locus, and $\sigma_A^2(\infty) = 2nE(a^2)$ is the expected additive genetic variance in a population of infinite size. Note that under this model, not only does the among-population variance not build up indefinitely, but as $4N_eu \to \infty$, driving the heterozygosity to 1.0, the among-population component of variance asymptotically approaches zero. This is because under the house-of-cards model, replicate populations that are each effectively infinite in size will individually harbor the same alleles with the same frequency spectrum defined by the mutational interconversion rates.

If nothing else, these dichotomous results indicate that although neutral models are essential to demonstrating the necessity of invoking natural selection to explain

an observed pattern of divergence, the actual *construction* of the null model depends on unresolved biological issues. Using the Zeng-Cockerham (1993) bridge model examined in Chapter 10 (wherein the effect of a mutant allele is given by $a_m = \tau a_o + a$, where $a_o$ is the effect in the ancestor and $a$ a random deviation), the equilibrium among-population variance becomes

$$\sigma_B^2 = \frac{4E(a^2)}{(1-\tau)^2[1+4N_e u(1-\tau)]} \tag{11.12}$$

where $\tau = 1$ under the Lynch-Hill model (infinite alleles) and 0 under the Cockerham-Tachida model. For $\tau < 1$, the approach to the equilibrium level of divergence is defined by the mutation rate ($u$), assuming an identical $N_e$ in the base and descendant populations,

$$\sigma_B^2(t) = [1-(1-u)^{2t}]\sigma_B^2 \tag{11.13}$$

and hence quite slow (approximately $2u$ per generation).

Finally, we note that the expression for the variance of the among-population variance (i.e., the variance of $\sigma_B^2(t)$ among replicate experiments with mutational input) is algebraically complex, and has only been worked out for the infinite-alleles model (Lynch and Hill 1986). However, if it is assumed that the number of loci is large and the distribution of mutational effects is normal with mean zero, the variance of the realized among-population variance approaches $2(2\sigma_m^2 t)^2/L$ for large $t$. This is simply twice the square of the expected among-population variance. Thus, for large $t$, the coefficient of variation of a realized among-population variance based on $L$ lines is expected to be on the order of $\sqrt{2/L}$, so as we have noted before, unless $L$ is quite large, estimates of $\sigma_B^2$ can deviate quite far from the expectation.

## Effectively Neutral Divergence and the Estimation of Rates of Mutational Variance

As discussed in detail in LW Chapter 12, the theoretical expectations of the neutral model provide the basis for estimating the rates of polygenic mutation. Starting from an inbred base population, experimental lines with known times of divergence can be used to estimate the amount of polygenic mutation that is necessary to account for the distribution of the resultant mean phenotypes. In one of the earliest endeavors of this sort, Russell et al. (1963) started with several lines of maize that had been maintained by prolonged self-fertilization. They then performed a dichotomous branching experiment for five generations in which each plant was self-fertilized to produce two new daughter sublines. Seed was saved from each generation, so that at the end of the experiment members of all generations could be assayed simultaneously in a common environment, and then sib analysis was used to estimate the additive genetic variance for the *total* population each generation. Assuming the within-population variance to be in drift-mutation equilibrium, this type of population expansion should give rise to an average rate of increase in the total genetic variance of $2\sigma_m^2$/generation. In accordance with this prediction, the regressions of the genetic variance on time were positive for every character investigated (Figure 11.4). The rate of polygenic mutation for each of the traits is thus estimated by one-half the slope.

Results from many other experiments of this sort were reviewed in LW Chapter 12. Although a number of additional results have emerged since then, most of these are confined to a small number of model systems, and the conclusions reached in our earlier review remain unaltered. Here, we simply give a brief update, providing references only to post-1998 papers. Most estimates are framed in terms of the mutational heritability, $h_m^2 = \sigma_m^2/\sigma_e^2$. Estimates of $h_m^2$ for a diversity of morphological, physiological, and life-history traits in *D. melanogaster* are consistently in the range of 0.001 to 0.005. Mutational heritabilities for body size and life-history traits in nematodes fall in the range of 0.001 to 0.008 (Vassilieva et al. 2000; Baer et al. 2006; Ostrow et al. 2007), and the same is true for life-history traits in the microcrustacean *Daphnia pulex* and in the grape phylloxera *Daktulosphaira vitifoliae* (Downie 2003). Thus, essentially all studies with invertebrates imply $0.001 < h_m^2 < 0.01$ for complex traits.

Although the numbers of studies are still rather limited, estimates of $h_m^2$ for some land plants and vertebrates appear to be several-fold higher than those noted above. Mutational heritabilities for growth and reproductive traits in *Arabidopsis thaliana* are in the range of 0.001 to 0.008 (Schultz et al. 1999; Shaw et al. 2000; Chang and Shaw 2003; Kavanaugh and Shaw 2005), but the average $h_m^2$ for maize, from the study of Russell et al. (1963), is 0.0092. In addition, mutational heritabilities for morphological and reproductive traits in mice fall in the range of 0.003 to 0.023 (Casellas and Medrano 2008). Thus, there is at least a rough indication that mutational heritabilities are increased in organisms with longer life spans, which might in principle be a consequence of elevated rates of mutation per generation (Chapter 4).

Finally, it should be emphasized that in all mutation-accumulation experiments, fitness declines in the vast majority of lines, indicating that mutations are on average deleterious, although the fraction of mutations that are beneficial remains unclear (Shaw et al. 2002; Keightley and Lynch 2003; Charlesworth and Eyre-Walker 2007; Eyre-Walker and Keightley 2007; Dickinson 2008; Hall et al. 2008). Equally importantly, for characters that influence fitness only indirectly (e.g., morphology), the fraction with negative pleiotropic effects on fitness remains unclear. Hence, estimates of $h_m^2$ from mutation-accumulation experiments with their very small effective population sizes may overestimate, perhaps significantly, the actual *usable* amount of $h_m^2$ for most populations. Further, if one imagines that deleterious pleiotropic effects are often small, as $N_e$ increases, the fraction of all new mutations which are effectively neutral decreases, so that the effective value of $h_m^2$ is likely a decreasing function of $N_e$. What is unclear is whether this plateaus fairly quickly or continues to decrease over a large range of $N_e$. Resolving these issues are critical to any attempts to utilize estimates of mutational heritability to infer long-term mechanisms of evolution, as illustrated in the following section.

**–Insert Figure 11.4 Here–**

## TESTING THE NULL HYPOTHESIS OF NEUTRAL PHENOTYPIC DIVERGENCE

One of the enduring problems in evolutionary biology is the struggle to demonstrate that various aspects of biodiversity are products of diversifying selection. It is one thing to concoct plausible adaptive scenarios to explain patterns of morphological, physiological, or behavioral divergence, but quite another to formally demonstrate that an observed level of divergence cannot simply be explained by a null model of random genetic drift. The preceding theory suggests a number of ways in which this might be done. Lande's $F$ test (Equation 11.9) is one such approach, wherein one compares the amount of divergence across a set of lines/populations with the expected between-population variance ($\sigma_B^2$). As mentioned, this test is best applied over short time scales ($t << N_e$), such as might occur in a selection experiment or over a short to modest amount of time in nature.

Here we focus on tests potentially over much longer time scales and ask whether an observed amount of total divergence $d = |\,\mu(t) - \mu(0)\,|$ within a single lineage is excessively large (or small) relative to drift. As with comparisons based on the amount of divergence over a set of lines or populations, the expected total divergence within a single lineage is also a function of the between-population variance $\sigma_B^2$. Tests for unusual amounts of divergence (large or small) are framed by asking what critical values for an effective population size $N_e$ or mutation variance $\sigma_m^2$ are consistent with the amount of divergence and whether these values are biological reasonable. If they are, the hypothesis of drift alone accounting for the pattern is not rejected. It should be stressed that considerable selection could have shaped the observed pattern, and yet we can still fail to reject the drift model.

While these tests are widely used, they have several important caveats. For any analysis of this sort to be meaningful, one must be confident that the magnitude of population divergence is genetic, and not inflated by environmental effects on phenotypes. This is clearly problematical when populations cannot be assayed in a "common-garden environment", such as using data entirely from the fossil record. In addition, the divergence of means is best estimated by a formal analysis of variance (e.g., Björklund 1991) so as to eliminate the inflation of the divergence by sampling error of the within-population means. A further complication is that expressions for critical effective population sizes or mutation rates ignore the sampling error of all other terms. Finally, as we have discussed, the infinite-alleles model for neutral trait evolution (where $\sigma_B^2$ is an ever-increasing function of time) may be too extreme, and the usable amount of $\sigma_m^2$ and $h_m^2$ likely decreases with increasing $N_e$. Combining all these considerations, it should be clear that the following tests for neutrality cannot be regarded as very rigorous in a statistical sense, and are best employed as diagnostic guides for future study.

### Lande's Brownian Motion Model of Neutral Trait Evolution

The basic structure of tests for neutral trait divergence is that $\mu_t \sim (\mu_0, \sigma_B^2[t])$, namely the mean at time $t$ has as expected value equal to the initial mean $\mu_0$ and has variance $\sigma_B^2(t)$. To proceed further, we need additional assumptions about the actual *distribution* from which the means are sampled, which is generally assumed to be Gaussian (normal). Support for this assumption traces back to Lande (1976), who framed mean divergence in terms of a Brownian motion process (Appendix 1). Under the simplest Brownian motion model, if the current value of a random variable

(such as a population mean) is $x$, the change in value over a small time interval is zero with constant variance $b$. Under this model, the distribution of values at time $t$ is normal with mean $x_0$ (the initial value) and variance $\sigma_t^2 = bt$ (Appendix 1). Assuming a strictly additive model with no environmental trends, Lande noted that if we sample $N_e$ individuals, their mean breeding value (i.e., the trait mean) would have a sampling variance each generation of $\sigma_A^2/N_e$, which is used for $b$. Hence, at generation $t$, the distribution of phenotypic means is approximately normal with expected mean $\mu_0$ (the initial mean) and variance

$$\sigma_t^2 = t\sigma_A^2/N_e \tag{11.14a}$$

This assumes a constant additive variance as well as a constant effective population size during the period of divergence being considered. Since drift can also change $\sigma_A^2$, the assumption of a constant $\sigma_A^2$ is reasonable only for $t \ll N_e$, unless the initial variance is close to its mutation-drift equilibrium value. More generally, if the additive variance and $N_e$ are both changing each generation, then under the Brownian motion model (taking the first time point at $i = 0$ and the last at $i = t-1$ bringing us up to generation $t$),

$$\sigma_t^2 = \sum_{i=0}^{t-1} \sigma_A^2(i)/N_e(i) \tag{11.14b}$$

For example, assuming a constant effective population size, the additive genetic variance at time $t$ under drift and mutation is given by Equation 10.19b. Substituting this into Equation 11.14b gives

$$\begin{aligned}\sigma_t^2 &= \frac{1}{N_e}\sum_{i=0}^{t-1}\left[2N_e\sigma_m^2 + [\sigma_A^2(0) - 2N_e\sigma_m^2]\exp(-i/2N_e)\right] \\ &= 2\sigma_m^2 t + [\sigma_A^2(0) - 2N_e\sigma_m^2]\left(\frac{1}{N_e}\sum_{0=1}^{t-1}\exp(-i/2N_e)\right)\end{aligned} \tag{11.14c}$$

which recovers our previous expression (Equation 11.10) for $\sigma_B^2$ under drift and mutation by noting that

$$\frac{1}{N_e}\sum_{i=0}^{t-1}\exp(-i/2N_e) \simeq 2(1 - \exp[-t/2N_e)] \tag{11.15}$$

This useful identity follows by recalling that the partial sum of a geometric series is

$$\sum_{i=0}^{k-1} x^i = \frac{1-x^k}{1-x}.$$

Taking $x = \exp(-1/2N_e)$ and noting from a first-order Taylor series that

$$1 - \exp(-1/2N_e) \simeq 1 - \left(1 - \frac{1}{2N_e}\right) = \frac{1}{2N_e}$$

returns Equation 11.15.

Thus, as expected, the variance $\sigma_t^2$ of the Brownian motion process just corresponds to the between-group drift variance $\sigma_B^2(t)$. The notion that (under a pure drift model) the additive variance changes until it reaches a drift-mutation equilibrium value has resulted in different parameterizations of $\sigma_t^2$ in tests of drift (Lande 1976, Turelli et al. 1988). Lande assumed a constant variance, $\sigma_t^2 = t\sigma_A^2/N_e$, but since part of his concern was evolution in the fossil record, he replaced $\sigma_A^2$ by $h^2\sigma_z^2$, giving

$$\sigma_t^2 = h^2\sigma_z^2 t/N_e \tag{11.16a}$$

His logic was that $\sigma_z^2$ could be estimated directly from a sample in the fossil record, while $h^2$ values for many morphological traits fall within a relatively narrow window. Hence, either a representative value for $h^2$ could be used, or different values tried to examine the robustness of any conclusions. This results in a test based on joint considerations of $N_e$ and $h^2$. Conversely, Turelli et al. (1988) note that if the population has been at its current size sufficiently long enough so that additive variance is at its mutation-drift equilibrium value, then (assuming the infinite-alleles model) $\sigma_A^2 = 2N_e\sigma_m^2$, giving

$$\sigma_t^2 = 2tN_e\sigma_m^2/N_e = 2t\sigma_m^2, \tag{11.16b}$$

Under this setting, $N_e$ does not appear and tests are based on whether the required values of $\sigma_m^2$ to be consistent with drift are plausible.

**Tests Based on the Brownian Motion Model**

Under the Brownian motion model, $\mu_t \sim N(\mu_0, \sigma_t^2)$, leading to tests of either excessive (or too little) divergence based on simple Normal theory. Suppose an absolute divergence of $d = |\,\mu(t) - \mu(0)\,|$ is observed. The probability of this under drift alone is given by

$$\Pr\left(|\,\mu(t) - \mu(0)\,| \le d\right) = \Pr\left(\frac{|\,\mu(t) - \mu(0)\,|}{\sigma_t} \le \frac{d}{\sqrt{\sigma_t}}\right) = \Pr\left(|\,U\,| \le \frac{d}{\sqrt{\sigma_t}}\right) \tag{11.17}$$

where $U$ is a unit normal random variable. Lande's (1976) original test was based on the constant variance assumption, $\sigma_t^2 = h^2\sigma_z^2 t/N_e$. Recalling that $\Pr(|\,U\,| \le 1.96) = 0.95$, Lande's critical effective population size below which there is a $< 5\%$ probability of a deviation as large as $d$ satisfies

$$1.96 = \frac{d}{\sqrt{th^2\sigma_z^2/N_e}}, \quad \text{implying} \quad (1.96)^2 th^2\sigma_z^2 = N_e d^2 \tag{11.18a}$$

Equation 11.18a allows one to determine critical values for either divergence time $t$, heritability $h^2$, or $N_e$ that are consistent with drift. For example, solving for $N_e$ gives

$$\widehat{N_e} = \frac{t \cdot h^2 \cdot 1.96^2}{d_*^2} = 3.84 \cdot \frac{t\,h^2}{d_*^2} \tag{11.18b}$$

where $d_* = d/\sigma_z$ is the divergence scaled in phenotypic standard deviations. Drift with $N_e > \widehat{N_e}$ is unlikley to generate the observed amount of divergence. For a more general test of significance level $\alpha$, one replaces 1.96 by $z_{1-\alpha/2}$. Likewise, if one is

comparing the means of two species that had a common ancestor $\tau$ generations ago, then $t = 2\tau$ and $d$ is the absolute difference between their means.

---

**Example 11.3.** Reyment (1982) observed a change of $1.49\sigma_z$ over roughly 5 x $10^5$ generations in the size of a Cretaceous foraminifer. Taking a typical heritability value of 0.3, Equation 11.18b gives the largest population size consistent with this amount of divergence as

$$\widehat{N_e} = 3.84 \cdot \frac{t\,h^2}{d_*^2} = 3.84 \cdot \frac{5 \times 10^5 \cdot 0.3}{1.49^2} \simeq 260,000$$

However, paleontological data suggests that the effective population size was greater than $10^6$, suggesting that drift could not account for such a rapid divergence. Assuming $h^2$ values of 0.5, 0.7, and 1 yields critical $N_e$ values of 433,000; 607,000; and 867,000, so that only for assumed $h^2$ values close to one does the critical maximal size under drift approach the assumed size of $N_e > 10^6$.

---

As noted by Turelli et al. (1988), the population size test given by Equation 11.18b is really *two-sided*. Lande's original test examines whether $N_e$ may be too large to account for the observed divergence (as might occur if directional selection was changing the mean). However, one can also inquire as to whether the *stability* of population means is too great to be compatible with neutrality (too little divergence). For a two-tailed test for neutrality with a 5% overall significance level, we use a 2.5% probability cutoff for the observed divergence being too small to be consistent with the model and a 2.5% cutoff for excessively high divergence. Since $\Pr(|U| \geq 2.24) = 0.025$, the critical maximum population size in a test that evolution has been too fast for drift is

$$\widehat{N_e}(fast) \leq \frac{t \cdot h^2 \cdot 2.24^2}{d_*^2} = 5.02 \cdot \frac{t\,h^2}{d_*^2} \tag{11.19a}$$

Since populations with smaller $N_e$ should show more drift (and divergence), Equation 11.19a gives the largest value of $N_e$ consisted with drift generating the observed divergence. If our assumed $N_e$ exceeds $\widehat{N_e}(fast)$, we reject the hypothesis that drift can account for this fast a divergence (using the values in Example 11.3 returns $N_e \simeq 340,000$). Likewise, since $\Pr(\,|\,U\,|\, < 0.03) = 0.025$, the critical minimal population size in a test that evolution has been too slow (support for stabilizing selection) is

$$\widehat{N_e}(slow) \geq \frac{t \cdot h^2 \cdot 0.03^2}{d_*^2} = 0.0009 \cdot \frac{t\,h^2}{d_*^2} \tag{11.19b}$$

If our assumed $N_e$ is *less* than $\widehat{N_e}(slow)$, we reject the hypothesis that drift can account for this slow a divergence (applying Equation 11.19b with the values from Example 11.3 gives a critical minimal $N_e$ of 61). More generally, for a two-sided test at overall significance level $\alpha$ ($\alpha/2$ for too much divergnece and $\alpha/2$ for too little), 2.24 and 0.03 are replaced by $z_{1-\alpha/4}$ and $z_{(1+\alpha/2)/2}$.

The second important modification offered by Turelli et al. (1988) is that the equilibrium additive variance $2N_e\sigma_m^2$ is a function of $N_e$, allowing us to alternatively express critical values (for too much or too little divergence), not in terms of the effective population size but rather in terms of the mutational heritability $\sigma_m^2$. Now $\sigma_t^2 = 2t\sigma_m^2$ and Equation 11.18a becomes $(1.96)^2 t(2N_e\sigma_m^2) = N_e d^2$, or (for too much divergence using the two-sized correction, 2.24 for 1.96), the smallest value for $\sigma_m^2$ below which drift is unlikely (at $\alpha = 0.05$) to account for the divergence is

$$\sigma_m^2(fast) = \frac{d^2}{(2.24)^2 2t} = 0.10\frac{d^2}{t}$$

Thus, if estimates of $\sigma_m^2$ and $t$ are available, one can test the neutral hypothesis without an estimate of $N_e$. To compare this with Equation 11.19a (which uses the scaled divergence $d^*$), we express this in terms of the scaled mutational variance. Since $\sigma_e^2 = (1-h^2)\sigma_z^2$,

$$\sigma_{m*}^2 = \frac{\sigma_m^2}{\sigma_z^2} = (1-h^2)(\sigma_m^2/\sigma_e^2) = (1-h^2)h_m^2 \tag{11.20}$$

Taking a generous range of heritability values (0.25 - 0.75) and recalling the empirical range for the mutational heritability $h_m$ we typically expect $\sigma_{m*}^2$ to fall within the range of $10^{-2}$ to $10^{-4}$.

The hypothesis of drift is rejected if the mutational variance is too small to account for the observed divergence

$$\sigma_{m*}^2(fast) \leq 0.10 \cdot \frac{d_*^2}{t} \tag{11.21a}$$

Just as a smaller $N_e$ allows for more divergence (and hence we set a critical upper value in Equation 11.18a), so does a larger mutational variance $\sigma_m^2$, and we set a critical lower value. Above that value drift could account for the observed divergence.

---

**Example 11.4.** Let's return to Reyment's foraminifer data from Example 11.3. Using the original Lande model, we rejected the hypothesis that drift could have accounted for the divergence. Applying Equation 11.21a, the hypothesis of drift accounting for excessive divergence is not rejected when

$$\sigma_{m*}^2 < 0.10 \cdot \frac{1.49^2}{5 \times 10^5} = 4.4 \times 10^{-7}$$

Turelli et al. note that this is several orders of magnitude lower than typical values of the scaled mutation variance, and thus this pattern of divergence is not too excessive for drift.

---

Conversely, the divergence is too slow to be accounted for by drift if the assumed variance is greater than the critical value

$$\sigma_{m*}^2(slow) \geq \frac{d_*^2}{2t \cdot 0.031^2} = 520.29 \cdot \frac{d_*^2}{t} \tag{11.21b}$$

One important caveat for such tests of stabilizing selection is that laboratory–based estimates of $\sigma^2_{m*}$ are measured under very small effective population sizes and hence many to most deleterious pleiotropic effects are likely effectively neutral in these settings (Chapter 6). As the population size increases, the actual usable amount of polygenic mutational variance in a trait is likely considerably less than the laboratory estimates, perhaps by orders of magnitude. Hence, observations of stabilizing selection based on the perceived polygenic mutation rate being to high to generate so little divergence may be very biased and this test should be used with caution.

Which of the two variance assumptions (Equation 11.16a or 11.16b) should one use in a test? In our view, the constant variance assumption (Equation 11.16a) is less problematic, as the usable amount of $\sigma^2_m$ and $h^2_m$ may *decrease* with $N_e$. In such cases, $2N_e\sigma^2_m/N_e$ may *not* be a constant over $N_e$, greatly complicating tests based on critical mutational variances. Conversely, most trait heritabilities typically fall within a modest window of values, and one can vary the assumed value of $h^2$ to examine its consequences.

Finally, it is worth noting that the methodology outlined above has primarily been used to test the neutral hypothesis with two-point analyses (i.e., with phenotypic measures at two points in time in a vertical lineage, or from two extant species derived from a recent common ancestor). A temporal series of data provides a more powerful means of analysis, as it then becomes possible to look for statistical trends in mean phenotypes or for correlations in rates of change in adjacent intervals, neither of which are expected in a strictly neutral model (at least under the infinite-alleles model) (Charlesworth 1984; Bookstein 1988; Estes and Arnold 2006). For example, Bookstein (1989), using results from the theory of random walks, notes that instead of considering the starting and ending points, if one instead considers the largest (absolute) scaled deviation $D_*$ *anywhere* in the time series, one can obtain tighter confidence intervals. One potential pitfall this with approach is that if one has a rather sparse time series to consider, the critical values suggested by Bookstein should instead be replaced by the order statistics for these values (which can easily be generated via simulation). For example, with a time series of 5 points, the largest deviation are expected to be less than a time series (over the same period) that has (say) 500 sampled time points. Just as with a two-point divergence analysis, environmental trends severely compromise this approach.

**– Mike: Want to add a few sentences about testing this within a phylogeny? –**

### Divergence in Morphological Traits

Numerous attempts have been made to apply the above procedures, or simple variants of them, to data from the fossil record to test the hypothesis that levels of morphological divergence over geological time scales have been driven by directional selection. For example, in the first of such studies, Lande (1976) showed that change in tooth-size dimensions over a 42 million year period in early horse evolution are consistent with the hypothesis of random genetic drift if the heritabilities of the traits had been near 0.5 and the long-term effective population size was smaller than 60,000 or so individuals. Given the generally high levels of heritability observed for mammalian morphological traits (Lynch and Walsh 1998), an assumption of $h^2 = 0.5$ is

not unreasonable, and the argument that the long-term $N_e$ in such lineages could be smaller than the critical value $N_e^* = 60,000$ is also plausible (Chapter 4), which would imply that drift could have acted alone to cause the observed changes. Analyses of tooth morphometrics in two additional lineages of extinct mammals (condylarths and oreodonts) suggest critical effective sizes of 80,000 to 120,000 below which the observed changes would be compatible with a neutral hypothesis. Thus, only if the effective sizes of these ancient mammalian taxa were actually in excess of $10^5$, a matter that remains unclear, would the observed changes require some mechanism of directional selection.

Several other studies of this nature have been applied to aspects of mammalian skull evolution. For example, by taking the upper and lower limits to mutational heritability, $\sigma_m^2/\sigma_e^2$, to be $10^{-2}$ and $10^{-4}$, Lynch (1990) found that the rates of evolution of cranial morphology in a wide array of placental mammalian lineages are one to two orders of magnitude below the minimum neutral rate, and Lemos et al. (2001) observed a similar pattern in marsupials. The only exception to this general trend concerns the races of modern man, which appear to have diverged at a rate slightly above the minimum neutral expectation (Lynch 1990; Ackermann and Cheverud 2004; Roseman 2004). Although they leave many questions unanswered, these kinds of results put in perspective previous arguments that rates of morphological evolution are exceptionally high in mammals, and especially so in the great apes (e.g., Cherry et al. 1982; Wyles et al. 1983; Van Valen 1985). Clearly, the predominant mode of evolution in mammalian skeletal morphology has been one of stabilizing selection, not of strong diversifying selection. Similarly, Spicer (1993) found widespread evidence of stabilizing selection on a variety of morphological traits in *Drosophila*, but some caution is in order here as the tests were based on critical mutation variances (Equation 11.21b). As mentioned, this approach likely generates many spurious calls of too little divergence, and hence spurious calls of stabilizing selection.

### Divergence in Levels of Gene Expression

These types of comparative morphological studies can now be extended to molecular-level traits (Fay and Wittkopp 2007). For example, modern genomics tools (such as **microarray analyses** and **RNA-Seq**) allow us to measure the level of expression for essentially the full repertoire of an individual's genes. The amount of mRNA present (either measured by the intensity of hybridization against probes for a gene or directly from the amount present in massive sequencing of an RNA pool) is a typically quantitative trait, showing both genetic and environmental variation, here in the amount of transcript present. Thus, with appropriate controls, it is possible to isolate the genetic component of gene-expression variance among individuals and sometimes between closely related species.

The general conclusion from such work is that stabilizing selection plays a prominent role in reducing levels of genetic variance in gene expression below the neutral expectation – both within and among species, levels of variation are much lower than expected based on presumed levels of mutational variance. For example, using lines of the nematode *C. elegans* from a long-term (280 generations) mutation-accumulation experiment, Denver et al. (2005) estimated $\sigma_m^2$ for several thousand

genes. By comparing levels of variation among a global collection of natural isolates, they found that ratios of standing levels of genetic variance to $\sigma_m^2$ were generally no greater than a few hundred. Given that this ratio provides an estimate of $4N_e$ under the assumption of neutrality in a selfing organism (as opposed to $2N_e$ in an outcrosser), these observations provide a firm rejection of the hypothesis that gene expression levels evolve in a neutral fashion. Rifkin et al. (2005) were able to estimate mutational heritabilities for mutation-accumulation lines of *D. melanogaster* by factoring out the variance at the individual fly level to obtain an estimate of $\sigma_e^2$. They found a median $h_m^2 \simeq 2.4 \times 10^{-5}$ across all genes, and showed that although interspecific variance in the expression of a gene was correlated with its mutational variance (in qualitative accordance with the neutral theory), the absolute level of divergence was too low to be compatible with neutrality (consistent with the results from Denver et al. 2005).

This conclusion of strong stabilizing selection on gene expression appears to extend to mammals, despite the fact that the efficiency of selection would be expected to be reduced as a consequence of low effective population sizes (Chapter 4). For example, Lemos et al. (2005) found that levels of gene-expression variance among intraspecific strains of mice average about two orders of magnitude below the minimum neutral expectation, whereas those between mouse species and between human and chimpanzee are eight to ten orders of magnitude too low for neutrality. Evaluating primates more broadly (human, chimpanzee, orangutan, and rhesus macaque), Gilad et al. (2006) found that the among-species variance in expression of most genes did not increase with divergence time, contrary to the neutral expectation; this study was particularly nicely designed in that it employed only DNAs for which the sequences were identical across all four species.

Quite contrary to the preceding interpretation, Khaitovich et al. (2004, 2005) have argued that gene expression in the great apes evolves in a largely neutral fashion. However, their arguments are based on observations that are only loosely connected with neutral expectations: a positive correlation between levels of within- and among-species variation for the expression of different genes; and a linear increase in among-species expression divergence with time. Because the genetic components of within- and between-species variance are both driven by mutation, they are indeed expected to be correlated under the neutral model. However, because gene expression is a function of both the genetic and environmental background of an individual, unless the latter is factored out in a quantitative-genetic analysis, such measures provide uncertain information on the more relevant levels of genetic divergence. Genes whose expression is strongly influenced by the environment may naturally exhibit higher levels of variation both within and among samples. In addition, unless the actual rate of divergence is consistent with the rate of polygenic mutation, linear patterns of evolutionary diversification need not imply neutrality, and may instead be a consequence of random fluctuating selection. A further complication is that Khaitovich et al. use human probes to measure differences in expression among their species, but sequence divergence between the probes and target sites generate reduced levels of hybridization, resulting in an increase in expression divergence over time. Broadley et al. (2008) report a similar linear divergence of expression variance with time in a series of 14 taxa in the Brassicaceae, but again the probe was based on a single species (*Arabidopsis*). Thus, the conclusion that primate gene expression

is evolving in a neutral fashion appears to be questionable, and has in fact been essentially retracted by in a more recent analysis (Chaix et al. 2008), which suggests an elevation in rate specific to the human lineage.

---

**Example 11.5.** As developed in Appendix 1, the Ornstein-Uhlenbeck process provides a model of Brownian motion drift coupled with a resorting force back to some optimal value $\theta$, as might be expected with drift and stabilizing selection. Bedford and Hartl (2009) used such a process to fit the pattern of expression divergence within a clade of seven species of *Drosophila*. Under the Ornstein-Uhlenbeck (OU) model, the expected change in the mean value of a process at value $x$ is $a(\theta - x)$, so that if $x < \theta$, it increases, while for $x > \theta$ it decreases. The parameter $a$ which measures the strength of the restoring force is also a measure of the strength of stabilizing selection. As with Brownian motion, the value of the process at time $t$ is normally distributed (Equation A1.33b), but now with mean and variance

$$\mu_t = x_o \exp(-at) + \theta[1 - \exp(-at)], \quad \sigma_t^2 = \frac{b}{2a}[1 - \exp(-2at)]$$

Thus for large $t$ the mean value approaches the optimal value $\theta$ while the divergence variance saturates at a value $b/(2a)$, where $b = \sigma_A^2/N_e$ under the constant variance model, giving an asymptotic variance of $\sigma_A^2/(2N_e a)$. Bedford and Hartl found that, in accordance with the OU model, the divergence variance does not linearly increase with time, but rather quickly approaches an asymptotic value. They also introduced a maximum likelihood estimator for $a$ (and hence the strength of stabilizing selection) using divergence data.

---

Taken together, these results suggest, perhaps not surprisingly, that at both the phenotypic and gene-regulatory levels, mammalian evolution is primarily characterized by periods of stabilizing selection, although relatively brief episodes of directional selection cannot be ruled out. However, it must also be emphasized that the interpretation of conservative rates of evolution is far from clearcut. In principle, evolutionary divergence rates that are below the expectation of the Lynch-Hill model may be a consequence of the general opposition of selection to all allelic changes associated with the trait, but there might also simply be a fraction of mutations that is truly neutral and another that has strong negative pleiotropic effects on fitness. In that case, an observed level of divergence could actually be entirely based on neutral mutations, but with the appropriate measure of mutational variance being lower than the actual value observed in mutation-accumulation experiments (where even highly deleterious mutations can accumulate). Alternatively, if the Cockerham-Tachida model is a more appropriate model, then one would expect cumulative levels of divergence to plateau in time rather than to increase indefinitely, not because of direct selective constraints but because of limited availability of alternative allelic states.

## Population Subdivision for Quantitative Traits

An alternative approach to testing the neutral hypothesis of divergence focuses on subpopulations of the same species, isolated by semipermeable migration barriers, as discussed in Chapter 2. By obtaining allele-frequency estimates for a diversity of neutral molecular markers from multiple subpopulations, one can partition the allelic diversity for the entire metapopulation (measured as heterozygosity under the assumption of panmixia) into its within- and among-subpopulation components (Cockerham 1973; Nei 1987; Weir 1996). The fraction of diversity associated with subpopulation divergence is generally called $F_{ST}$ (or sometimes $G_{ST}$) in deference to Wright (1951), who first suggested this measurement of population subdivision.

Now consider a quantitative trait with a purely additive-genetic basis, and let $Q_{ST}$ denote the level of population subdivision for allele frequencies at the loci underlying the trait. Letting the genetic variance for the trait in the entire metapopulation under the assumption of panmixia be $\sigma_G^2$, then from the theory developed earlier in this chapter as well as in Chapter 10, the within- and among-subpopulation components of variance can be represented as $\sigma_{GW}^2 = (1 - Q_{ST})\sigma_G^2$ and $\sigma_{GB}^2 = 2Q_{ST}\sigma_G^2$, respectively. It follows that

$$Q_{ST} = \frac{\sigma_{GB}^2}{\sigma_{GB}^2 + 2\sigma_{GW}^2} \tag{11.22}$$

While the term $Q_{ST}$ is do to Spitze (1993), Prout and Barker (1989) and Lande (1992) also proposed this approach. This is a very general result, applying to a wide range of population structures and migration patterns provided the character does indeed have an entirely additive genetic basis (Whitlock 1999). Leinonen et al. (2006) proposed the related measure $P_{ST}$ based on phenotype measures of divergence, but considerable caution is required when using this in place of $Q_{ST}$ (Brommer 2011).

Equation 11.22 provides a potential empirical method for testing the hypothesis of neutral divergence among subpopulation means. If isolates of sufficient numbers of families from multiple subpopulations can be grown in a common environment, then appropriate statistical methods (Lynch and Walsh 1998) can be used to estimate $\sigma_{GW}^2$ and $\sigma_{GB}^2$. The resultant estimate of $Q_{ST}$ can then be compared to a parallel measure of subdivision ($F_{ST}$) derived from putatively neutral markers. Under the assumption of neutrality $Q_{ST}$ should not be significantly different from $F_{ST}$. However, $Q_{ST} > F_{ST}$ is expected if subpopulation differentiation has been primarily driven by adaptive divergence, whereas the opposite pattern is expected if the mean phenotypes of all or most subpopulations are kept relatively uniform by stabilizing selection for the same optima. One of the first formal tests of $Q_{ST} = F_{ST}$ was proposed by O'Hara and Merilä (2005), while Whitlock (2008) and Whitlock and Guillaume (2009) provide a broad overview of the use of comparisons of $Q_{ST}$ and $F_{ST}$ in tests for selection, including an evaluation of the form of the distributions of both statistics. Holand et al. (2011) provides a recent example of these tests in action.

Because of the requirement for assays in a common-garden arena, joint studies of $Q_{ST}$ and $F_{ST}$ are not common. However, the majority of results, over a diverse assemblage of animals and land plants, support a hypothesis of $Q_{ST} \simeq F_{ST}$ or $Q_{ST} > F_{ST}$ (Leinonen et al. 2008). As the latter results are qualitatively consistent with adaptive differentiation, they are clearly at variance with the observations on longer-term divergence noted above.

One major caveat with respect to this strategy for testing for neutral divergence

is that, even under neutrality, the expected value of $Q_{ST}$ will not necessarily equal $F_{ST}$ if the trait of interest is influenced by nonadditive genetic effects. This is because, as outlined in Chapter 10 and above, with nonadditive gene action, the within- and among-subpopulation components of genetic variation for neutral characters under short-term divergence are no longer equal to $\sigma^2_{GW} = (1-f)\sigma^2_G$ and $\sigma^2_{GB} = 2f\sigma^2_G$ (where $f$ is the parameter estimated by $F_{ST}$), but instead are influenced by a number of higher-order terms. In general, because the within-population genetic variance declines less rapidly with inbreeding under nonadditivity (and sometimes even increases; Chapter 10), $Q_{ST}$ as defined by Equation 11.22 will tend to be smaller than $F_{ST}$ under neutrality, although exceptions do exist (Goudet and Büchi 2006; Goudet and Martin 2007; López-Fanjul et al. 2003, 2006, 2007). By encouraging the false impression of stabilizing selection, this general behavior makes conclusions regarding adaptive divergence based on elevated $Q_{ST}$ conservative, while rendering observations of $Q_{ST} < F_{ST}$ ambiguous.

### QTL Analysis of Divergent Lines

The preceding approaches rely on large samples of multiple populations or at two or more intervals. However, there is a situation in which one might test for adaptive divergence with just a single cross between two isolated lineages. By examining a battery of polymorphic markers segregating in the $F_2$ generation of such a cross, one may search for QTLs associated with phenotypic measures of various traits (Lynch and Walsh 1998). Under the neutral hypothesis, the relative abundances of "plus" and "minus" marker alleles associated with small vs. large phenotypes are expected to be randomly distributed should not differ significantly from a 1:1 ratio.

This general strategy will be biased if the parental lines are intentionally selected to have extreme phenotypes, as the high line would then naturally be expected to be enriched with "plus" alleles. However, Orr (1998) suggested a way around this problem. If enough QTLs have been identified so that their distribution of effects can be approximated, given the level of phenotypic divergence between the lines, one can computationally evaluate the probability that the observed number of "plus" alleles in the high line could have arisen by chance. Although Orr (1998) gives a few examples of the application of this method to some artificial systems, it has not been extensively applied, the main limitation being the development of a fairly accurate estimate of the distribution of QTL effects, which in turn requires a survey of high-density markers in a substantial number of $F_2$ individuals. However, the recent success of high-density marker genome wide association studies (GWAS) for mapping (especially in humans) may reawaken interest in Orr's method.

**Literature Cited**

Ackermann, R. R., and J. M. Cheverud. 2004. Detecting genetic drift versus selection in human evolution. *Proc. Natl. Acad. Sci. USA* 101: 17946–17051. [11]

Anderson, W. W. 1973. Genetic divergence in body size among experimental populations of *Drosophila pseudoobscura* kept at different temperatures. *Evolution* 27: 278–284. [11]

Baer, C. F., N. Phillips, D. Ostrow, A. Avalos, D. Blanton, A. Boggs, T. Keller, L. Levy, and E. Mezerhane. 2006. Cumulative effects of spontaneous mutations for fitness in *Caenorhabditis:* role of genotype, environment and stress. *Genetics* 174: 1387–1395. [11]

Bailey, D. W. 1959. Rates of subline divergence in highly inbred lines of mice. *J. Heredity* 50: 26-30. [11]

Benford, T., and D. L. Hartl. 2009. Optimization of gene expression by natural selection. *PNAS* 106: 1133–1138. [11]

Björklund, M. 1991. Sexual dimorphisms and mating system in the grackles (*Quiscalus* spp.: Icterinae). *Evolution* 45: 608–621. [11]

Bookstein, F. L. 1988. Random walk and the biometrics of morphological characters. *Evol. Biol.* 23: 369–398. [11]

Broadley, M. R., P. J. White, J. P. Hammond, N. S. Graham, H. C. Bowen, Z. F. Emmerson, R. G. Fray, P. P. M. Iannetta, J. W. McNicol, and S. T. May. 2008. Evidence of neutral transcriptome evolution in plants. *New Phytologist* 180: 587–593. [11]

Brommer, J. E. 2011. Whither $P_{ST}$? The approximation fo $Q_{ST}$ by $P_{ST}$ in evolutionary and conservation biology. *J. Evol. Biol.* 24: 1160–1168. [11]

Bryant, E. H., L. M. Combs, and S. A. McCommas. 1986. Morphometric differentiation among experimental lines of the housefly in relation to a bottleneck. *Genetics* 114: 1213–1223. [11]

Casellas, J., and J. F. Medrano. 2008. Within-generation mutation variance for litter size in inbred mice. *Genetics* 179: 2147–2155. [11]

Chaix, R., M. Somel, D. P. Kreil, P. Khaitovich, and G. Lunter. 2008. Evolution of primate gene expression: drift and corrective sweeps? *Genetics* 180: 1379–1389. [11] ******* Not currently cited**

Chakraborty, R., and M. Nei. 1982. Differentiation of quantitative characters between populations or species. I. Mutation and random genetic drift. *Genet. Res.* 39: 303–314. [11]

Chang, S. M., and R. G. Shaw. 2003. The contribution of spontaneous mutation to variation in environmental response in *Arabidopsis thaliana:* responses to nutrients. *Evolution* 57: 984–994. [11]

Charlesworth, B. 1984. Some quantitative methods for studying evolutionary patterns in single characters. *Paleobiology* 10: 308–318. [11]

Charlesworth, J., and A. Eyre-Walker. 2007. The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc. Natl. Acad. Sci. USA* 104: 16992–16997. [11]

Cherry, L. M., S. M. Case, J. G. Kunkel, and A. C. Wilson. 1979. Comparisons of frogs, humans, and chimpanzees. *Science* 204: 435. [11]

Cockerham, C. C. 1973. Analyses of gene frequencies. *Genetics* 74: 679-700. [11]

Cockerham, C. C., and H. Tachida. 1987. Evolution and maintenance of quantitative genetic variation by mutations. *Proc. Natl. Acad. Sci. USA* 84: 6205–6209. [11]

Cockerham, C. C., and B. S. Weir. 1983. Variance of actual inbreeding. *Theor. Pop. Biol.* 23: 85–109. [11]

Denver, D. R., K. Morris, J. T. Streelman, S. K. Kim, M. Lynch, and W. K. Thomas. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans.* *Nature Genetics* 37: 544–548. [11]

Dickinson, W. J. 2008. Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in *Saccharomyces cerevisiae.* *Genetics* 178: 1571–1578. [11]

Downie, D. A. 2003. Effects of short-term spontaneous mutation accumulation for life history traits in grape phylloxera, *Daktulosphaira vitifoliae.* *Genetica* 119: 237–251. [11]

Eisen, E. J., and J. P. Hanrahan. 1974. Genetic drift and inbreeding depression measured from control populations of mice. *Can. J. Genet. Cytol.* 16: 91–104. [11]

Estes, S., and S. J. Arnold. 2006. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Amer. Natur.* 169: 227–244. [11]

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. *Nature Rev. Genet.* 8: 610–618. [11]

Fay, J. C., and P. J. Wittkopp. 2007. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100: 191–199. [11]

Gilad, Y., A. Oshlack, G. K. Smyth, T. P. Speed, and K. P. White. 2006. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440: 242–245. [11]

Goudet, J., and L. Büchi. 2006. The effects of dominance, regular inbreeding and sampling design on $Q_{ST}$, an estimator of population differentiation for quantitative traits. *Genetics* 172: 1337-1347. [11]

Goudet, J., and G. Martin. 2007. Under neutrality, $Q_{ST} \leq F_{ST}$ when there is dominance in an island model. *Genetics* 176: 1371–1374. [11]

Haldane, J. B. S. 1949. Suggestions as to quantitative measurement of rates of evolution. *Evolution* 3: 51–56. [11]

Hall, D. W., R. Mahmoudizad, A. W. Hurd, and S. B. Joseph. 2008. Spontaneous mutations in diploid *Saccharomyces cerevisiae:* another thousand cell generations. *Genet. Res.* 90: 229–241. [11]

Hill, W. G. 1972. Estimation of genetic change. I. General theory and design of control populations. *Anim. Breed. Abstr.* 40: 1–15. [11]

Holand, A. M., H. Jensen, J. Tufto, and R. Moe. 2011. Does selection or genetic drift explain geographic differentitation of morphological characters in house sparrows? *Genet. Res. Camb.* 93: 367–379. [11]

Kavanaugh, C. M., and R. G. Shaw. 2005. The contribution of spontaneous mutation to variation in environmental responses of *Arabidopsis thaliana:* responses to light. *Evolution* 59: 266–275. [11]

Keightley, P. D., and M. Lynch. 2003. Toward a realistic model of mutations affecting fitness. *Evolution* 57: 683–685. [11]

Khaitovich, P., S. Pääbo, and G. Weiss. 2005. Toward a neutral evolutionary model of gene expression. *Genetics* 170: 929–939. [11]

Khaitovich, P., G. Weiss, M. Lachmann, I. Hellmann, W. Enard, B. Muetzel, U. Wirkner, W. Ansorge, and S. Pääbo. 2004. A neutral model of transcriptome evolution. *PLoS Biol.* 2: 682–689. [11]

Kimura, M. 1983. *The neutral theory of molecular evolution.* Cambridge University Press, Cambridge, UK. [11]

Kimura, M., and J. F. Crow. 1963a. On the maximum avoidance of inbreeding. *Genet. Res.* 4: 399–415. [11]

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30: 314–334. [11]

Lande, R. 1977. Statistical tests for natural selection on quantitative characters. *Evolution* 31: 442–446. [11]

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* 33: 402–416. [11]

Lande, R. 1992. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution* 46: 381–389. [11]

Leinonen, T., Cano, J. M., H. Makinen, and J. Merilä. 2006. Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of threespine sticklebacks. —si J. Evol. Biol. 19: 1803–1812. [11]

Leinonen, T., R. B. O'Hara, J. M. Cano, and J. Merilä. 2008. Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *J. Evol. Biol.* 21: 1-17. [11]

Lemos, B., C. D. Meiklejohn, M. Cáceres, and D. L. Hartl. 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* 59: 126-137. [11]

López-Fanjul, A. Fernández, and M. A. Toro. 2003. The effect of neutral nonadditive gene action on the quantitative index of population divergence. *Genetics* 164: 1627-1633. [11]

López-Fanjul, A. Fernández, and M. A. Toro. 2006. The effect of genetic drift on the variance/covariance components generated by multilocus additive × additive epistatic systems. *J. Theor. Biol.* 239: 161-171. [11]

López-Fanjul, A. Fernández, and M. A. Toro. 2007. The effect of dominance on the use of the $Q_{ST} - F_{ST}$ contrast to detect natural selection on quantitative traits. *Genetics* 176: 725-727. [11]

Lynch, M. 1988. Design and analysis of experiments on random drift and inbreeding depression. *Genetics* 120: 791–807. [11]

Lynch, M. 1990. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Amer. Natur.* 136: 727–741. [11]

Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40: 915–935. [11]

Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits.* Sinauer Assocs., Inc., Sunderland, MA. [11]

Nei, M. 1987. *Molecular evolutionary genetics.* Columbia Univ. Press, New York, NY. [11]

O'Hara, R. B., and J. Merilä. 2005. Bias and precision in $Q_{ST}$ estimates: Problems and some solutions. *Genetics* 171: 1331–1339. [11]

Orr, H. A. 1998. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 149: 2099-2104. [11]

Ostrow, D., N. Phillips, A. Avalos, D. Blanton, A. Boggs, T. Keller, L. Levy, J. Rosenbloom, and C. F. Baer. 2007. Mutational bias for body size in rhabditid nematodes. *Genetics* 176: 1653–1661. [11]

Prout, T., and J. S. F. Barker. 1989. Ecological apects of the heritability of body size in *Drosophila buzzatii*. *Genetics* 123: 803-813. [11]

Rich, S. S., A. E. Bell, D. A. Miles, and S. P. Wilson. 1984. An experimental study of genetic drift for two quantitative traits in *Tribolium. J. Heredity* 75: 191-195. [11]

Rifkin, S. A., D. Houle, J. Kim, and K. P. White. 2005. A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438: 220–223. [11]

Roseman, C. C. 2004. Detecting interregionally diversifying natural selection on modern human cranial form by using matched molecular and morphometric data. *Proc. Natl. Acad. Sci. USA* 101: 12824–12829. [11]

Russell, W. A,, G. F. Sprague, and H. L. Penny. 1963. Mutations affecting quantitative characters in long-time inbred lines of maize. *Crop Sci.* 3: 175–178. [11]

Schultz, S. T., M. Lynch, and J. H. Willis. 1999. Spontaneous deleterious mutation in *Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA* 96: 11393–11398. [11]

Shaw, F. H., C. J. Geyer, and R. G. Shaw. 2002. A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana. Evolution* 56: 453–463. [11]

Shaw, R. G., D. L. Byers, and E. Darmo. 2000. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana. Genetics* 155: 369–738. [11]

Spicer, G. S. 1993. Morphological evolution of the *Drosophila virilis* species group as assesed by rate tests for natural selection on quantitative characters. *Evolution* 47: 1240–1254. [11]

Spitze, K. 1993. Population structure in *Daphnia obtusa:* quantitative genetic and allozymic variation. *Genetics* 135: 367-374. [11]

Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* 42: 1085–1089. [11]

Van Valen, L. 1985. Why and how do mammals evolve unusually rapidly? *Evol. Theory* 7: 127–132. [11]

Vassilieva, L. L., A. M. Hook, and M. Lynch. 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans. Evolution* 54: 1234–1246. [11]

Weir, B. S. 1996. *Genetic data analysis II.* Sinauer Assocs., Inc., Sunderland, MA. [11]

Weir, B. S., P. J. Avery, and W. G. Hill. 1980. Effect of mating structure on variation in inbreeding. *Theor. Popul. Biol.* 18: 396–429. [11]

Whitlock, M. C. 1999. Neutral additive genetic variance in a metapopulation. *Genet. Res.* 74: 215-221. [11]

Whitlock, M. C. 2008. Evolutionary inference from $Q_{ST}$. *Mol. Ecol.* 17: 1885-1896. [11]

Whitlock, M. C., and F. Guillaume. 2009. Testing for spatially divergence seelction: Comparing $Q_{ST}$ to $F_{ST}$. *Genetics* 183: 1055-1063. [11]

Wright, S. 1921. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* 6: 124–143. [11]

Wright, S. 1951. The genetical structure of populations. *Ann. Eugenics* 15: 323–354. [11]

Wyles, J. S., J. G. Kunkel, and A. C. Wilson. 1983. Birds, behavior, and anatomical evolution. *Proc. Natl. Acad. Sci. USA* 80: 4394-4397. [11]

**Figure 11.1.** Confidence limits (A) and critical values (B) for $\sigma^2$ estimated from a sample of $n$ observations. **A (Top):** Upper and lower values for the 95% confidence interval in $\sigma^2$ based on an observed sample variance Var. For example, for $n = 10$, the 95% confidence interval for $\sigma^2$ is 0.44·Var to 3.33·Var. **B (Top):** Upper and lower 5% critical values for an observed sample variance given an assumed variance $\sigma^2$. For example, for $n = 10$, 95% of the values of Var are expected to fall within the interval $0.30{\cdot}\sigma^2$ to $2.11{\cdot}\sigma^2$.

**Figure 11.2.** Observed and expected levels of the among-population variance for pupal weight in a divergence experiment with the flour beetle *Tribolium.* The lines are described in the text. Data from Rich et al. (1984).

**Figure 11.3.** The expected fraction of among-population variance attributable to mutations arising subsequent to the isolation event. It is assumed that the base population is in drift-mutation equilibrium, $\sigma_A^2(0) = 2N_e\sigma_m^2$, with the same effective size as the daughter species, so that from Equation 11.10, the divergence due to base-population variance is $2\sigma_A^2(0)[1-e^{-t/(2N_e)}]$. To obtain the actual number of generations of divergence for any population size, multiply the horizontal axis by $N_e$.

**Figure 11.4.** The increase in additive genetic variance (within- plus among-population components) in an expanding set of lines of corn. From Russell et al. (1963).