Lecture 02:
Tests based on between-line divergence

UNE course:
The search for selection
3 -- 7 Feb 2020
Bruce Walsh (University of Arizona)
jbwalsh@email.arizona.edu

Outline

- Quantifying divergence
- Short-term divergence (drift)
  - Lande's CV test
- Long-term divergence (mutation)
  - Lande's Brownian-motion model (N_e-based)
  - MDE version (Var(m)-based)
    - Ornstein-Uhlenbeck process
- Time-series based tests

Between-line divergence

- Due to both partitioning (i.e., fixation of alternative alleles) of the initial additive variance {short time scales} and the accumulation of new mutations {longer time scales}, two initially identical populations (i.e., the same mean), will diverge over time

- A number of tests are based on whether an observed amount of divergence is too fast (directional selection), or too slow (stabilizing selection), relative to the expectation under drift

The between-population variance due to the initial additive variance will approach 2Var(A) over time

More precisely,

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

For t > N_e, divergence via mutation becomes important

Estimating the Among-group Variance

With L replicate populations, a common estimate in the literature for $\sigma_B^2(t)$ is

$$V_B(t) = \frac{1}{L-1} \sum_{i=1}^{L} [\bar{z}_i(t) - \bar{z}.(t)]^2 \tag{12.8a}$$

the sample variance among the sample means, $\bar{z}_1, \cdots, \bar{z}_L$, of the replicate population. When just two populations are being considered (as in some of the tests developed below), their squared difference

$$d^2(t) = [\bar{z}_1(t) - \bar{z}_2(t)]^2 \tag{12.8b}$$

is often used. This is easily related to Equation 12.8a by noting for $L = 2$ that

$$V_B = \frac{1}{2-1} \sum_{i=1}^{2} \left(\bar{z}_i - \frac{\bar{z}_1 + \bar{z}_2}{2}\right)^2 = \frac{(\bar{z}_1 - \bar{z}_2)^2}{4} + \frac{(\bar{z}_2 - \bar{z}_1)^2}{4} = \frac{d^2}{2} \tag{12.8c}$$

These expressions for V_B overestimate the true among-line variance σ_B^2, as the sample means are measured with error. In particular, $\bar{z}_i = \mu_i + e_i$, so that

$$\sigma^2(\bar{z}_i) = \sigma^2(\mu_i) + \sigma^2(e_i) = \sigma_B^2 + \frac{\sigma_z^2}{n} \tag{12.8d}$$

where σ_z^2 is the trait variance and n is the sample sized used to estimate μ_i. When $\sigma_B^2 = 2f_t\sigma_A^2 \gg \sigma_z^2/n$ (which is equivalent to $2f_th^2 \gg 1/n$), the difference between $\sigma^2(\bar{z}_i)$ and σ_B^2 is small.

As suggested by a number of authors (Lynch 1988a, 1990; Turelli et al. 1988; Björklund 1991; Savalli 1993), a simple way to avoid this issue is to estimate the among-group variance from a standard one-way ANOVA, with

$$V_B(t) = \frac{MS_B - MS_W}{n_0} \tag{12.8e}$$

Key point: use ANOVA to estimate between-population variance

Russ Lande

Test for short-term divergence

- Lande proposed a test for whether an observed amount of divergence (between-population variance) was sufficiently different from that expected under drift (assuming N_e is roughly known).

- Computes a test whose statistic is roughly F-distributed.

- Could also use this approach to look to too little divergence (stabilizing selection), but not a lot of power

Lande's Constant Variance Test, F_{CV}

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

$$F_{CV} = \frac{V_B(t)}{t \cdot V_A(0)/N_e} \qquad (12.9a)$$

$$F_{CV} \sim \frac{\chi^2_{L-1}}{L-1} \sim F_{L-1,\infty}$$

This assumes $V_A(0)$ know without error. If estimated, then the denominator degrees of freedom is that of the estimate

Example 12.2. Lande (1977b) used Equation 12.9a 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, respectively, 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_{CV}, making it more difficult to demonstrate diversifying selection on wing length. The resulting values (for the two environments) were

$$F_{CV,1} = \frac{6.62}{150 \cdot 0.88/1000} = 50.15, \quad \text{and} \quad F_{CV,2} = \frac{4.37}{150 \cdot 0.77/1000} = 37.84$$

Given that $\Pr(F_{4,\infty} > 4.6) = 0.001$, both values are highly significant. 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.

Longer-term divergence tests

- The previous test assumes that the divergence is simply due to drift of the initial variance, and that mutation has yet to become important (i.e., $t << N_e$)

- Over longer time scales, much of the divergence is generated by new mutations, and hence tests are based upon the mutation variance (or mutation heritability), rather than N_e.

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

$$\sigma_B^2(t) = 2\sigma_m^2 t + 2\left(\sigma_A^2(0) - 2N_e\sigma_m^2\right)\left(1 - e^{-t/(2N_e)}\right) \tag{12.10}$$

where σ_m^2 is the per-generation mutational rate of input of genetic variance, as described in Chapter 11. 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 σ_m^2 (Lande 1979a).

Under the incremental mutation model, divergence is a linear function of the mutational variance, and independent of the effective population size
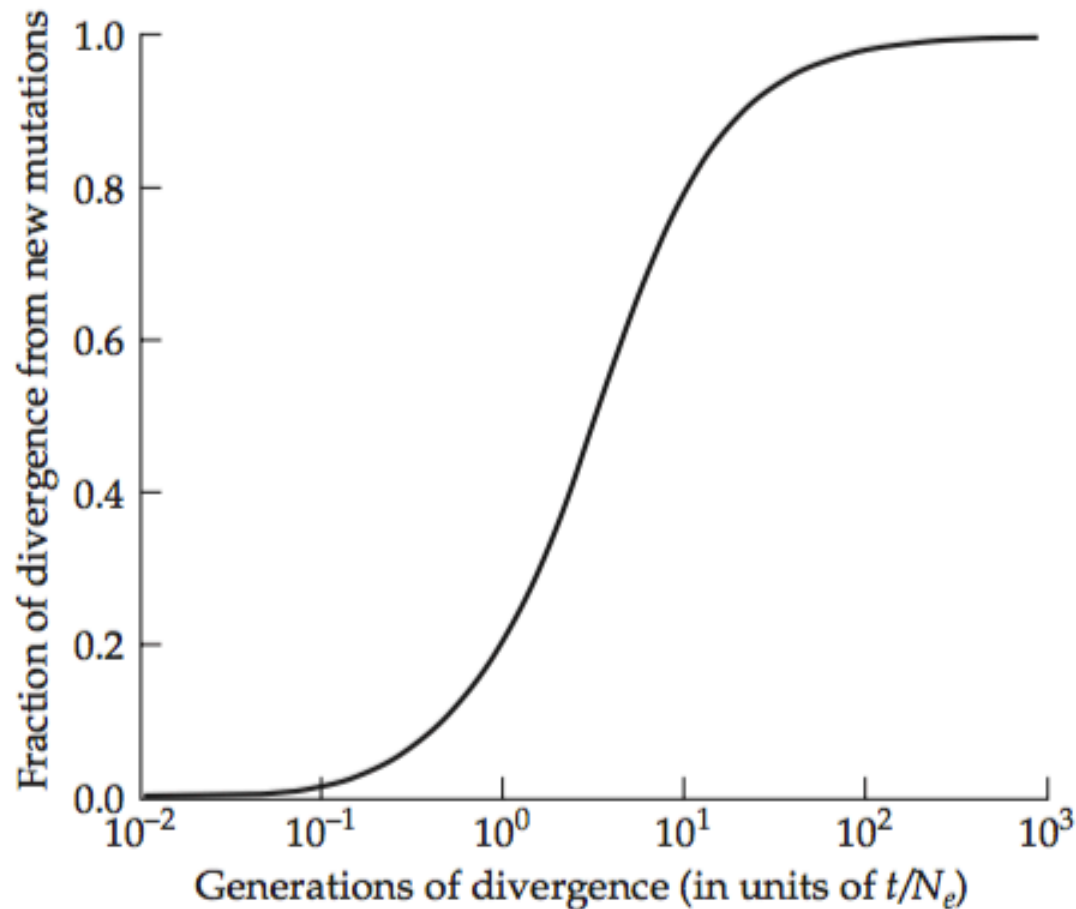
Figure 12.3 The expected fraction of neutral among-population variance attributable to mutations arising subsequent to the isolation event. It is assumed that the base population is in drift-mutation equilibrium, $\sigma_A^2(0) = 2N_e\sigma_m^2$, with the same effective size as the daughter species, so that from Equation 12.1b, the divergence due to base-population variance is $4N_e\sigma_m^2[1 - e^{-t/(2N_e)}]$. To obtain the actual number of generations of divergence for any population size, the horizontal axis is multiplied by N_e.

Divergence via mutation

- Under the incremental model variance increases without limit

- Under either the HOC or regression models, the divergence eventually reaches an upper limit

$$\tilde{\sigma}_B^2 = \frac{4E(a^2)}{(1-\tau)^2[1+4N_e u(1-\tau)]} \qquad (12.12)$$

For $\tau < 1$, the temporal 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,

$$\sigma_B^2(t) = [1 - (1-u)^{2t}]\,\tilde{\sigma}_B^2 \qquad (12.13)$$

and hence is quite slow (approximately $2u$ per generation).
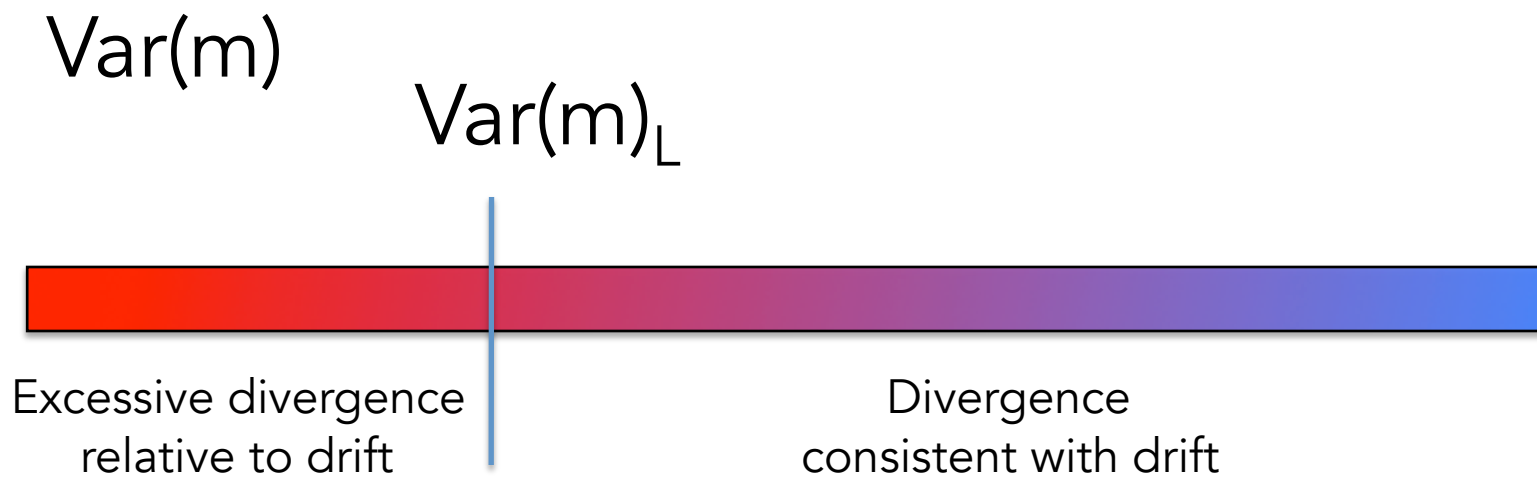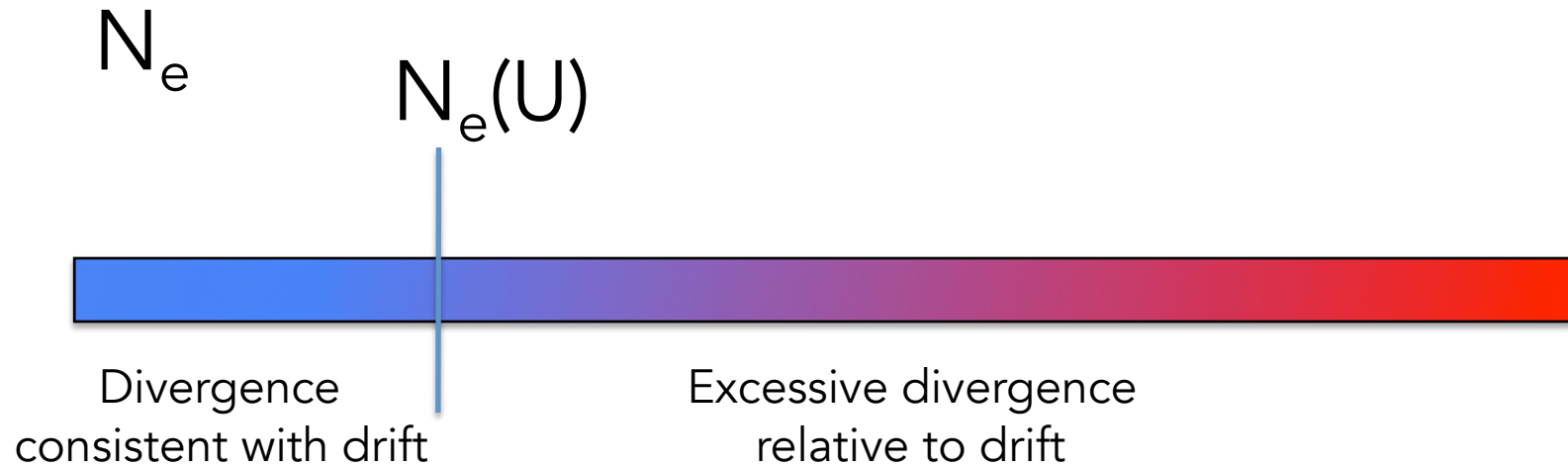
The effective mutational variance

- Most estimates of the mutational variance or mutation heritability are obtained in very small populations
  - Under this setting, the effects of selection against mutations is largely ignored (as $4N_e|s| < 1$)

- As N_e increases, the fraction of mutations that are effectively neutral likely decreases, so that for large values of N_e, the effective mutational -variance heritability may be orders of magnitude less than small-population estimates

Rate-based tests

- Either small values of N_e or large values of Var(m) lead to divergence. Hence, rate tests find either the
  - Largest value of N_e that is consistent with drift, $N_e(U)$.
    - $N_e > N_e(U)$ rejects drift
  - Smallest value of Var(m) consistent with drift, Var (m) $_L$
    - Var(m) < Var (m) $_L$ rejects drift

Directional selection (excessive divergence)

N_e

$N_e(U)$

Divergence consistent with drift

Excessive divergence relative to drift

$Var(m)$

$Var(m)_L$

Excessive divergence relative to drift

Divergence consistent with drift

Rate-based tests (cont.)

- Under test of stabilizing selection, look for too little divergence relative to drift
    - Smallest value of N_e that is consistent with drift, $N_e(L)$.
        - $N_e < N_e(L)$ rejects drift
    - Largest value of var(m) consistent with drift, $Var(m)_U$
        - $Var(m) > Var(m)_U$ rejects drift

Stabilizing selection (too little divergence)

N_e

$N_e(U)$

Too little divergence
relative to drift

Lack of Divergence
consistent with drift

Var(m)

$Var(m)_L$

Lack of divergence
consistent with drift

Too little divergence
relative to drift
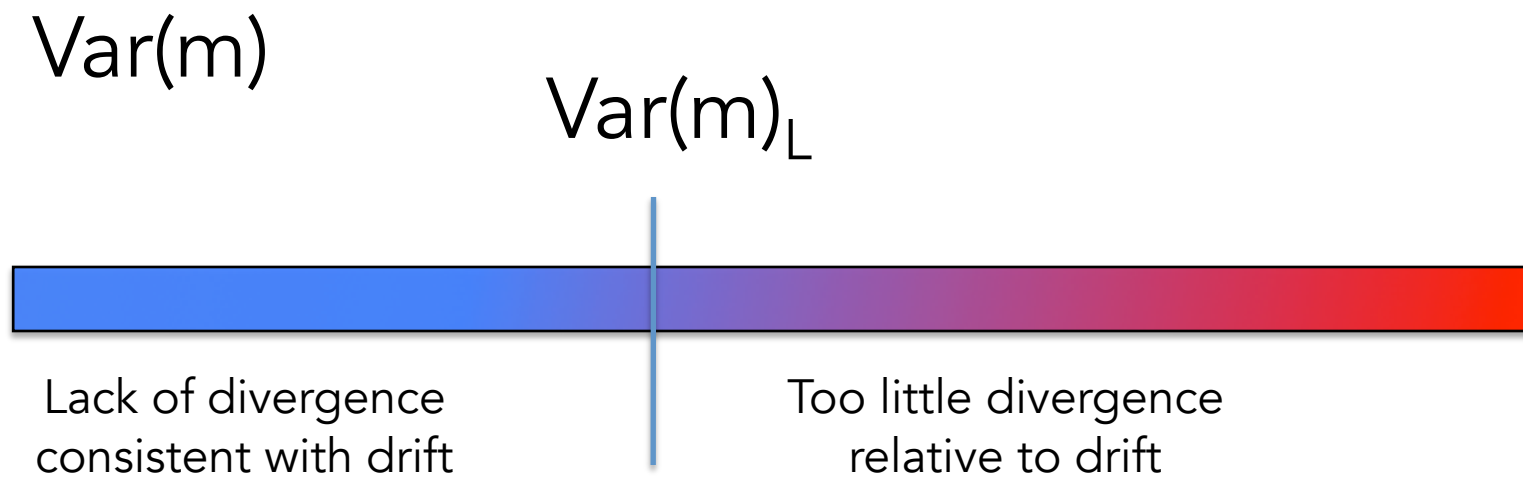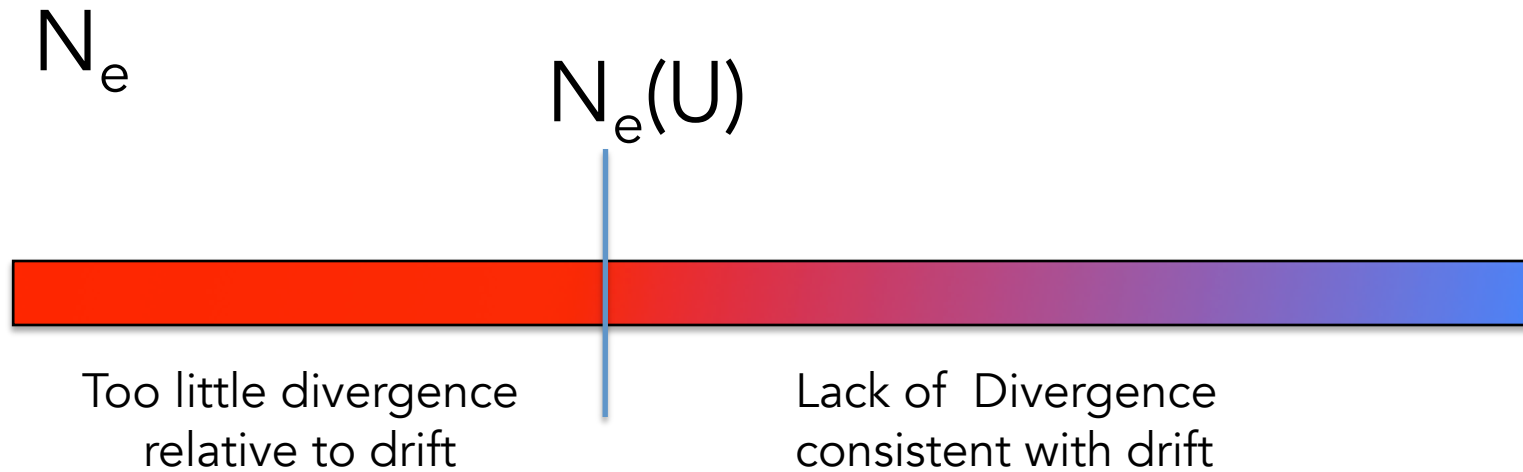
Lande's Brownian-motion model

- Under the Brownian-motion model, the mean at time t follows a normal distribution with a variance of
  - t Var(A)/N_e = t h^2 Var(z)/N_e

Conversely, Turelli et al. (1988) noted that if the population has been at its current size sufficiently long enough that the additive genetic variance is at its mutation-drift equilibrium value, then (assuming the infinite-alleles model) $\sigma_A^2 = 2N_e\sigma_m^2$, yielding

$$\sigma_t^2 = 2tN_e\sigma_m^2/N_e = 2t\sigma_m^2, \qquad (12.16b)$$

Tests Based on the Brownian Motion Model

Under the Brownian motion model, the mean phenotype $\mu_t \sim N(\mu_0, \sigma_t^2)$, providing the basis for tests of either too much, or too little, divergence based on simple normal theory. Suppose an absolute divergence of $d = |\mu(t) - \mu(0)|$ is observed, where $\mu(t)$ and $\mu(0)$ are the means from two samples from the same population taken t generations apart. The probability of this level of divergence under drift alone is given by

$$\Pr\left(|\mu(t) - \mu(0)| \le d\right) = \Pr\left(\frac{|\mu(t) - \mu(0)|}{\sigma_t} \le \frac{d}{\sigma_t}\right) = \Pr\left(|U| \le \frac{d}{\sigma_t}\right) \qquad (12.17)$$

where U is a unit normal random variable. Lande's (1976) original test (distinct from his 1977 F_{CV} test; Equation 12.9a) was based on the constant variance assumption, $\sigma_t^2 = th^2\sigma_z^2/N_e$. Recalling that $\Pr(|U| \le 1.96) = 0.95$, Lande's critical effective population size below which there is a $< 5\%$ probability of an absolute deviation as large as d satisfies

$$1.96 = \frac{d}{\sqrt{th^2\sigma_z^2/N_e}}, \quad \text{implying} \quad (1.96)^2 th^2\sigma_z^2 = N_e d^2 \qquad (12.18a)$$

Equation 12.18a 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 the upper bound, $N_{e,u}$ on the effective population size that is compatible with drift yields

$$N_{e,u} = \frac{t \cdot h^2 \cdot 1.96^2}{d_*^2} = 3.84 \cdot \frac{t\, h^2}{d_*^2} \qquad (12.18b)$$
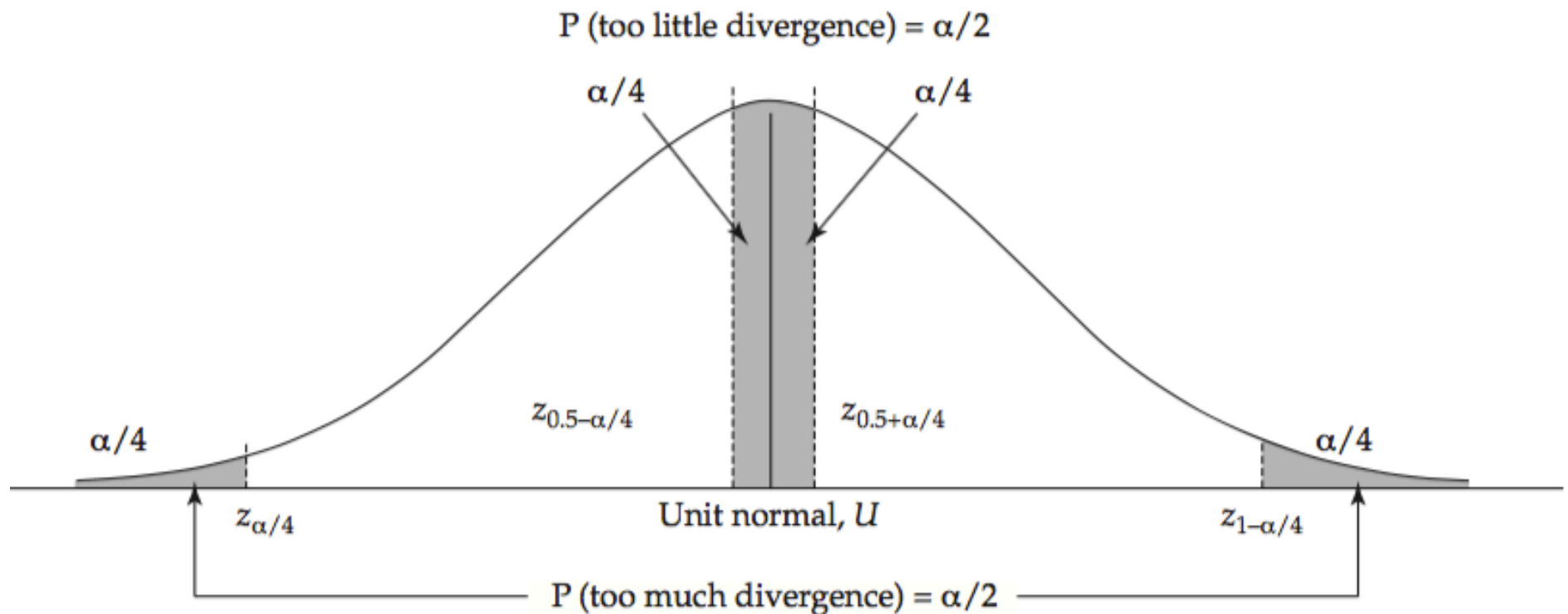
Figure 12.5 Critical values for an α-level test of a departure from drift having either too little, or too much, absolute divergence. Too much absolute divergence occurs when the unit-normal scaled test score is either in the lower $\alpha/4$ or upper $\alpha/4$ tail (for a total probability of $\alpha/2$). Too little absolute divergence occurs when the unit–normal scaled test score is too close to zero, namely, a region of probability $\alpha/4$ below zero and a region of probability $\alpha/4$ above zero (for a total probability of $\alpha/2$). Here, z_p satisfies $\Pr(U \leq z_p) = p$, where U is a unit–normal random variable. See the text for further details.

$$N_{e,u} \leq \frac{t \cdot h^2 \cdot 2.24^2}{d_*^2} = 5.02 \cdot \frac{t\,h^2}{d_*^2} \qquad (12.19a)$$

Because populations with smaller N_e should show more drift (and divergence), Equation 12.19a gives the largest value of N_e that is consistent with drift generating the observed amount of divergence. If the assumed N_e exceeds $N_{e,u}$, we reject the hypothesis that drift can account for this fast a divergence. Likewise, because $\Pr(|U| < 0.03) = 0.025$, the critical lower-bound population size $N_{e,l}$ in a test that evolution has been too slow (support for stabilizing selection) is

$$N_{e,l} \geq \frac{t \cdot h^2 \cdot 0.03^2}{d_*^2} = 0.0009 \cdot \frac{t\,h^2}{d_*^2} \qquad (12.19b)$$

If our assumed N_e is *less* than $N_{e,l}$, we reject the hypothesis that drift can account for this slow a divergence.

Example 12.3. Reyment (1982) observed a change of $1.49\sigma_z$ in the size of a Cretaceous foraminifer over roughly $5 \cdot 10^5$ generations. Taking a typical heritability value of 0.3, Equation 12.18b (i.e., assuming a one-sided test, namely, a test only for excessive divergence) gives the upper bound, $N_{e,u}$, on population size consistent with this amount of divergence as

$$N_{e,u} = 3.84 \cdot \frac{t\,h^2}{d_*^2} = 3.84 \cdot \frac{5 \cdot 10^5 \cdot 0.3}{1.49^2} \simeq 260,000$$

However, paleontological data suggest that the census population size $N \gg 10^6$, implying that drift was unlikely to account for such a rapid divergence (of course, caveats from Chapter 3 apply in that usually $N_e \ll N$). Assuming h^2 values of 0.5, 0.7, and 1.0 yields $N_{e,u}$ values of 433,000, 607,000, and 867,000 respectively, so that only for assumed h^2 values close to one does the critical N_e under drift approach the assumed size of $N_e > 10^6$.

Using the two-sided test (strict departure from drift, either too little or too much absolute divergence), the value of the 3.84 used above is replaced by 5.02 (Equation 12.19a), resulting in an ~31% (5.02/3.84 = 1.307) increase in $N_{e,u}$ value, with critical values of ~ 340,000, 566,000, 794,000, and 1,133,000 for h^2 values of 0.3, 0.5, 0.7, and 1.0, respectively. Similarly, the lower critical $N_{e,l}$ (the size below which the lack of divergence is too improbable under drift) is 61 (using Equation 12.19b with $h^2 = 0.3$).

The structure of the tests given by Equations 12.17 through 12.19 depends on N_e and h^2. A second approach is to instead base tests on the mutational variance, σ_m^2, alone. The idea is that if N_e has been roughly constant for a sufficient amount of time, then the additive-genetic variance for a neutral trait approaches its mutation-drift equilibrium value, $2N_e\sigma_m^2$ (Equation 11.20c). Under this condition, Equation 12.16b shows that the among-group variance becomes $\sigma_B^2 = 2t\sigma_m^2$, giving the **MDE** (mutation-drift equilibrium) version of Lande's F test (Equation 12.9a) as

$$F_{MDE} = \frac{V_B(t)}{2t\sigma_m^2} \qquad (12.20a)$$

We can also arrive at this test by substituting $2N_e\sigma_m^2$ for $V_A(0)$ in Equation 12.9a. As above, V_B is best estimated from the among-group variance in a one-way ANOVA (Equation 12.9f).

$$\Pr\left[\left(\frac{L-1}{X_{1-\alpha/2,L-1}}\right)V_B(t) \leq 2t\sigma_m^2 \leq \left(\frac{L-1}{X_{\alpha/2,L-1}}\right)V_B(t)\right] = 1-\alpha$$

$$\Pr\left[\left(\frac{L-1}{X_{1-\alpha/2,L-1}}\right)\frac{V_B(t)}{2t\sigma_e^2} \leq h_m^2 \leq \left(\frac{L-1}{X_{\alpha/2,L-1}}\right)\frac{V_B(t)}{2t\sigma_e^2}\right] = 1-\alpha$$

A slightly different formulation of this test is based in terms of the observed rate of divergence (Lynch 1990). Letting $\Delta = (V_B/t)/\sigma_e^2$ be the estimated rate of divergence scaled in units of the environmental variance, Equation 12.20c becomes

$$\text{Pr}\left[\left(\frac{(L-1)/2}{X_{1-\alpha/2,L-1}}\right)\Delta \leq h_m^2 \leq \left(\frac{(L-1)/2}{X_{\alpha/2,L-1}}\right)\Delta\right] = 1 - \alpha \qquad (12.20\text{d})$$

yielding the upper and lower bounds on the mutational heritability h_m^2 consistent with drift. For $\alpha = 0.05$ and $L = 2$, Equation 12.20b becomes

$$\text{Pr}\left(0.10 \cdot \Delta \leq h_m^2 \leq 509 \cdot \Delta\right) = 0.95 \qquad (12.20\text{e})$$

Thus, the hypothesis of drift is rejected (at $\alpha = 0.05$) if the mutational heritability is too small to account for the observed divergence rate, namely

$$h_m^2 < 0.10 \cdot \Delta \simeq 0.10 \cdot \frac{d_*^2}{t} \qquad (12.21\text{a})$$

Conversely, the divergence is too slow to be accounted for by drift if the assumed mutational heritability is too high to account for the observed divergence rate, or when

$$h_m^2 \geq 509 \cdot \Delta \simeq 509 \cdot \frac{d_*^2}{t} \qquad (12.21\text{b})$$

Example 12.4. We now return to Reyment's foraminifer data from Example 12.3. Using the original Lande model (Equation 12.18b), we rejected the hypothesis that drift could have accounted for the divergence. Applying Equation 12.21a, the hypothesis of drift accounting for excessive divergence is not rejected when

$$h_m^2 > 0.10 \cdot \frac{d_*^2}{t} = 0.10 \cdot \frac{1.49^2}{5 \cdot 10^5} = 4.4 \cdot 10^{-7}$$

This critical value of the mutational heritability is several orders of magnitude lower than typical values of this parameter, implying that this pattern of divergence is not too excessive for drift.

Thus, we reach two very different conclusions depending on whether the constant variance (Example 12.3) or equilibrium variance (Example 12.4) assumption is used. Which is the better choice? In our view, the constant variance assumption (Equation 12.16a) is less problematic, as the usable amount of σ_m^2 and h_m^2 may *decrease* with increasing N_e. In such cases, $2N_e\sigma_m^2/N_e$ may *not* be a constant over N_e, 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 in Equation 12.16a to examine its consequences.

Ornstein-Uhlenbeck Models

Under the OU model, the expected change in the mean value of a process at a value of x is $a(\theta - x)$ with $a > 0$, so that if $x < \theta$, it increases, while it decreases for $x > \theta$. The parameter a, which measures the strength of the restoring force, is a measure of the strength of stabilizing selection. Under the standard model of Gaussian stabilizing selection (Example 5.6; Equation 16.17), where ω^2 measures the strength of selection (smaller ω^2 implies stronger selection), Example A1.13 shows that

$$a = \frac{\sigma_A^2}{\sigma_z^2 + \omega^2} \tag{12.22a}$$

As with Brownian motion, the value of the process at time t is normally distributed (Equation A1.33b), but now with mean and variance

$$\mu_t = \theta + [x_o - \theta]e^{-at} \tag{12.22b}$$

$$\sigma_t^2 = \frac{b}{2a}[1 - e^{-2at}] \tag{12.22c}$$

where $b = \sigma_A^2/N_e$ under the constant-variance model. For large t, the mean value approaches the optimal value (θ), while the divergence variance approaches an asymptotic value of

$$\frac{b}{2a} = \frac{\sigma_z^2 + \omega^2}{2N_e} \tag{12.22d}$$

Summary

- Under the Brown-motion model (the incremental mutation model), under neutrality, means linearly diverge without limit from each other

- Under the HOC or regression mutational models, eventually the between-population variance reaches a limit

- A limiting divergence also occurs under stabilizing selection (OU model)

Times Series tests

- The above tests are based on external comparisons: one has two time points and a guess for t, and either Ne, or h^2_m or Var(m)

- A second type of data is a stratophenetic series, a time series of means for a population over some part of the fossil record.

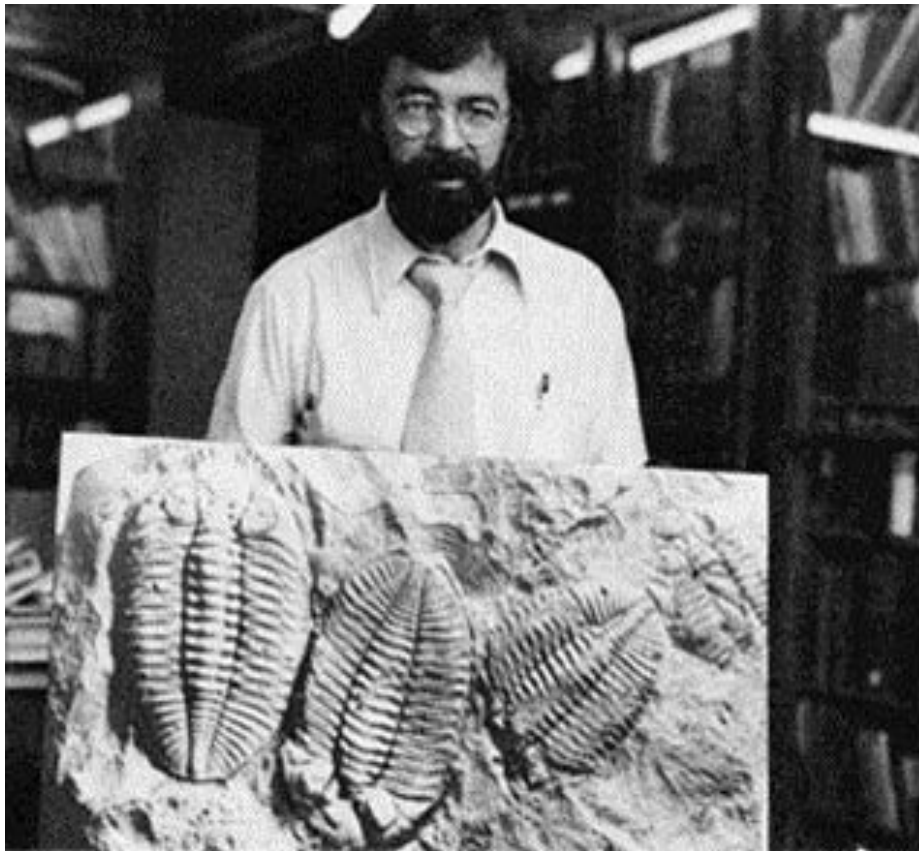 – Time series tests are based on internal characteristics of just this series alone.
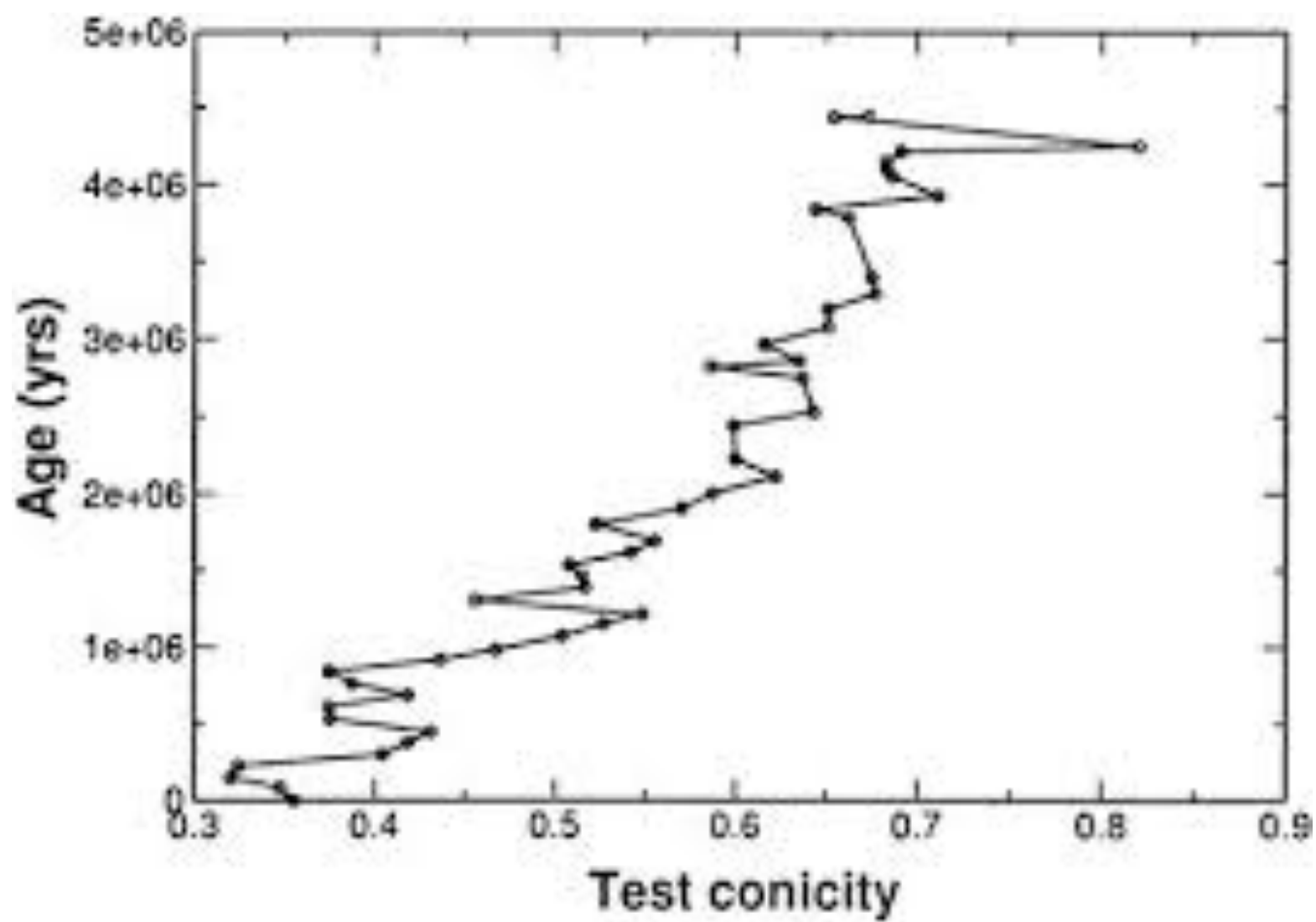
Dave Raup
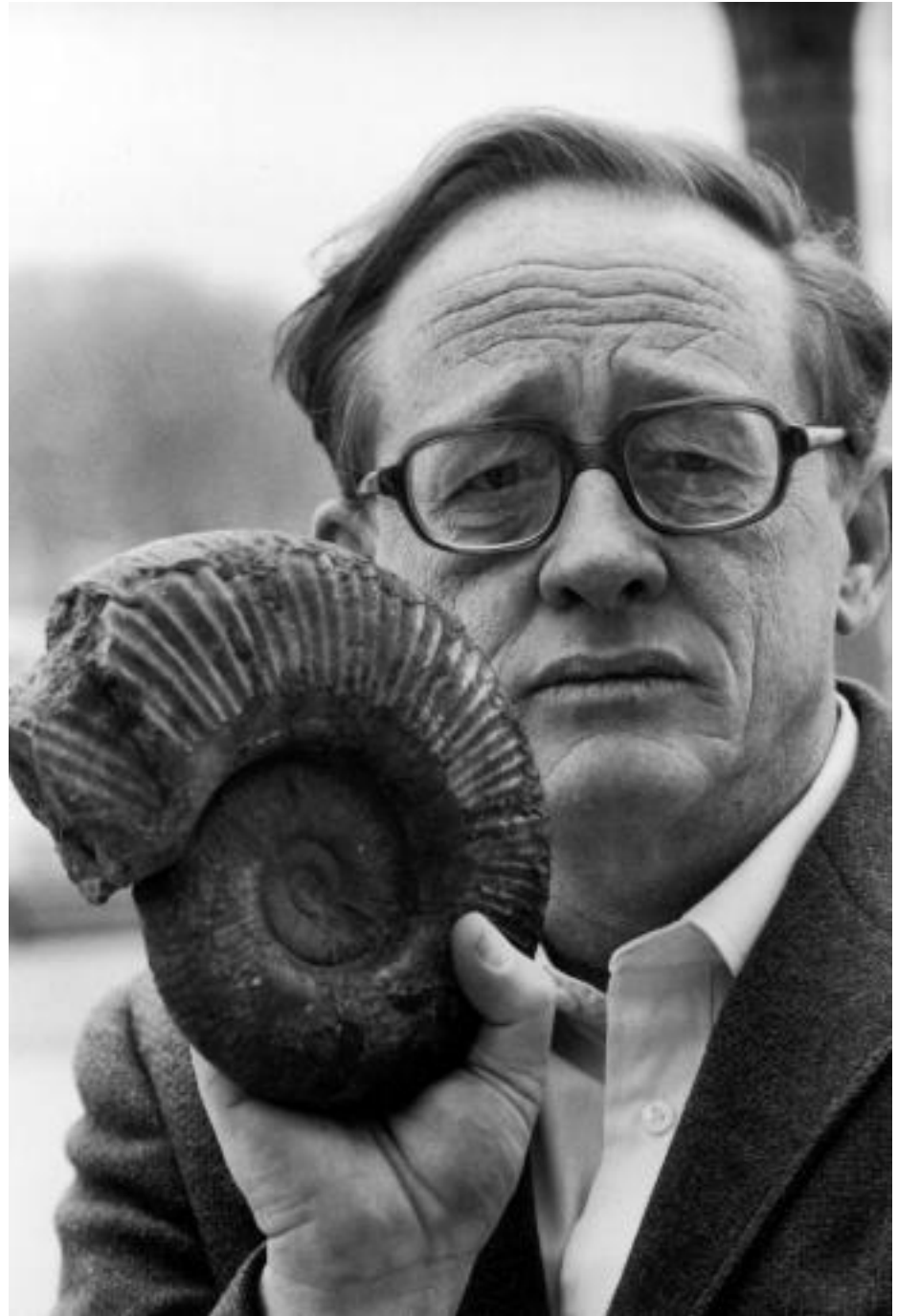
Niles Eldredge

Stephen Jay Gould

- Raup proposed the use of statistical tests of the pattern of divergence in fossil sequences to test the Eldredge-Gould suggestion that punctuated equilibrium (excess periods of stasis, followed by rapid change) is common in the fossil record
  - Eldredge and Gould suggested that species-level selection might generate such a pattern
  - Raup framed the null model as a random walk

There are two important caveats with random-walk models. First, any observed phenotypic trend could be entirely environmental, with changes in the mean being independent of any underlying genetic change. Second, as highlighted by Raup (1977), a pattern indistinguishable from a random walk can mask significant underlying selection, such as short, episodic bursts of directional selection in shifting directions or stabilizing selection with drift occurring in the optimal value. These are examples of **hierarchical models** of random change, wherein selection is driving the generational change, but the focus of selection (either directional or stabilizing) is randomly changing, generating an random walk.

Departments from a symmetric random walk

- Runs test: Each step equally likely (under drift) to be increasing or decreasing
  - Too few runs (changes in sign): directional trend
  - Too many runs: stabilizing selection
- Bookstein's largest scaled excursion
- Hurst exponents
- Hunt's model-weighting method

.

of the largest scaled excursion,

$$\gamma = \frac{\max_k |S_k|}{\hat{\sigma}\sqrt{n}} \qquad (12.23b)$$

Namely, the largest absolute value of the walk ($\max_k |S_k|$) over any of the sampled times, expressed in terms of the expected standard error of the walk value at the final sample time ($\sigma\sqrt{n}$). For $\gamma > 1$, critical values (the upper p in the tail of the null distribution) are very closely given by the corresponding $p/4$ critical values for a unit normal. For example, the upper 5% tail corresponds to $\gamma = 2.25$, consistent with the value of 2.24 for a normal with $p/4 = 0.0125$. The upper 1% and 0.1% upper tail probabilities correspond to γ values of 2.8 and 3.5, respectively. Series with values exceeding these critical values are said to be **improbably directional**, consistent with directional selection (or an environmental trend). Conversely, a series that does not vary enough is said to be **improbably constrained**, consistent with some sort of stabilizing selection or other cause of stasis. The lower 5%, 1%, and 0.1% values correspond to γ values of 0.62, 0.49, and 0.41, respectively. Failure to reject the null of a random walk still allows for considerable selection, either due to a lack or power or randomness in the direction of selection over time. Multivariate random-walk tests are discussed by Bookstein (2013).

Another widely used test for departures against the null of a symmetric random walk (closely related to Bookstein's approach), called **scaled range analysis,** is based on **Hurst exponents** (Hurst 1951). The idea behind this approach is that the absolute difference of a symmetric random walk $|x_t - x_0|$ scales as $\sigma\sqrt{t}$ (which can be estimated from Equation 12.23a). Defining the **standardized range,** $R(\tau)$, for a time interval (τ) as

$$R(\tau) = \frac{|x_\tau - x_0|}{\sigma} \qquad (12.24a)$$

one then regresses $R(\tau)$ on ever-increasing values of τ, fitting the log-log regression

$$\ln[R(\tau)] = H \ln(\tau) + \epsilon \qquad (12.24b)$$

where the slope (H) is the Hurst exponent (i.e., $R \propto \tau^H$). Under a symmetric random walk with uncorrelated increments, absolute trait divergence is expected to scale with the square root of time, giving $H = 0.5$. As increments become more positively correlated, H increases to 1.0 (**directional persistence**), consistent with directional selection. As adjacent increments become increasingly negatively correlated, H decreases to zero (**anti-persistence**), consistent with stabilizing selection or some other form of stasis. Roopnarine (2001) discussed permutation tests for the significance of $H \neq 0.5$. Gingerich's (1993) *LRI* (log rate versus log interval) method is a version of this test, where the slope (G) of his *LRI* regression is simply $G = H - 1$ (Roopnarine et al. 1999).

While straightforward and widely applied in the early literature, these methods typically have low power, meaning that the null hypothesis of a symmetric random walk is hard to reject (Roopnarine et al. 1999; Roopnarine 2001; Sheets and Mitchell 2001). This is especially the case with stratophenetic series, with their usual incompleteness and sporadic coverage due to the vagaries of the fossilization process. Further, as noted by Sheets and Mitchell (2001), there is an asymmetry of detection in that stabilizing selection is easier to detect than directional selection. They showed that the Hurst exponent (and, by extension, the *LRI* method) has the highest power to detect stabilizing selection, followed by Bookstein's approach, and then the runs test. Conversely, for detecting directional selection, the runs test is often the most powerful, followed by the Hurst exponent, and then Bookstein's approach.

Gene Hunt

Hunt's method

- Low power for tests of departures from symmetric random walks creates a "tyranny of the null hypothesis," potentially overinflating the role of drift

- Hunt initially considered three basic models:

  - a symmetric random walk (with an incremental mean value of zero)

  - a directional (or generalized) random walk (mean increment different from 0)

  - stasis.

Example 12.5. Anderson et al. (2000) proposed that a series of candidate models can be compared via their **Akaike weights**, an approach used by Hunt (2007) to assess the relative fit of a series of candidate models for long-term evolution (Figure 12.6). Suppose one has a series of models that were fit using maximum likelihood. If these models are not nested (so that likelihood-ratio tests for comparing fit are not available; LW Appendix 4), then comparison statistics, such as the **Akaike information criterion** (**AIC**; Akaike 1974) can be used to rank them. AIC rewards goodness of fit (higher log-likelihood, L), while penalizing for the number of parameters (k) with smaller AIC values implying a better model. With $n < 40k$ observations, Anderson et al. suggested that a corrected version of AIC should be used,

$$AIC_c = -2\ln(L) + 2k + \frac{2k(k+1)}{n-k-1} \qquad (12.25a)$$

which differs from the standard AIC measure in the addition of the last term. Suppose that one has a set of m candidate models and computes the support for model i relative to the best fitting of all m models,

$$\Delta_i = AIC_{c,i} - \min(AIC_c) \qquad (12.25b)$$

The resulting Akaike weights for each of the candidate models are given by

$$w_i = \frac{e^{-(\Delta_i/2)}}{\sum_{j=1}^{m} e^{-(\Delta_i/2)}} \qquad (12.25c)$$

Table 12.1 Summary of the 251 fossil sequences examined by Hunt (2007), each fit using three models of divergence: random walk, directional selection, and stasis. Counts given under the Trait and Fossil group categories are the numbers of times a model had the highest Akaike weight (Example 12.5) for a fossil sequence. For example, 13 of the 251 sequences (0.052) had directional selection as the model with the highest support, while 5 of 114 (0.044) size-related traits had directional selection as the most-supported model. Values under the Median column correspond to the median fraction of support over all sequences for a given model. For example, half of all sequences had a support for directional change model of 0.06 or less, while 95% of all sequences have a fractional support for directional selection in the 0.04 to 0.08 range. The fossil groups are planktonic and benthic microfossils (Plank and Benth) and macrofossils (Macro).

Model	Median, 95% CI	Trait				Fossil group		
		All	Size	Shape	Other	Plank	Benth	Macro
Directional	0.06 (0.04, 0.08)	13	5	4	4	5	3	5
Random	0.47 (0.39, 0.56)	123	67	43	13	24	57	42
Stasis	0.34 (0.20, 0.50)	115	42	68	5	12	37	66
		251	114	115	22	41	97	113
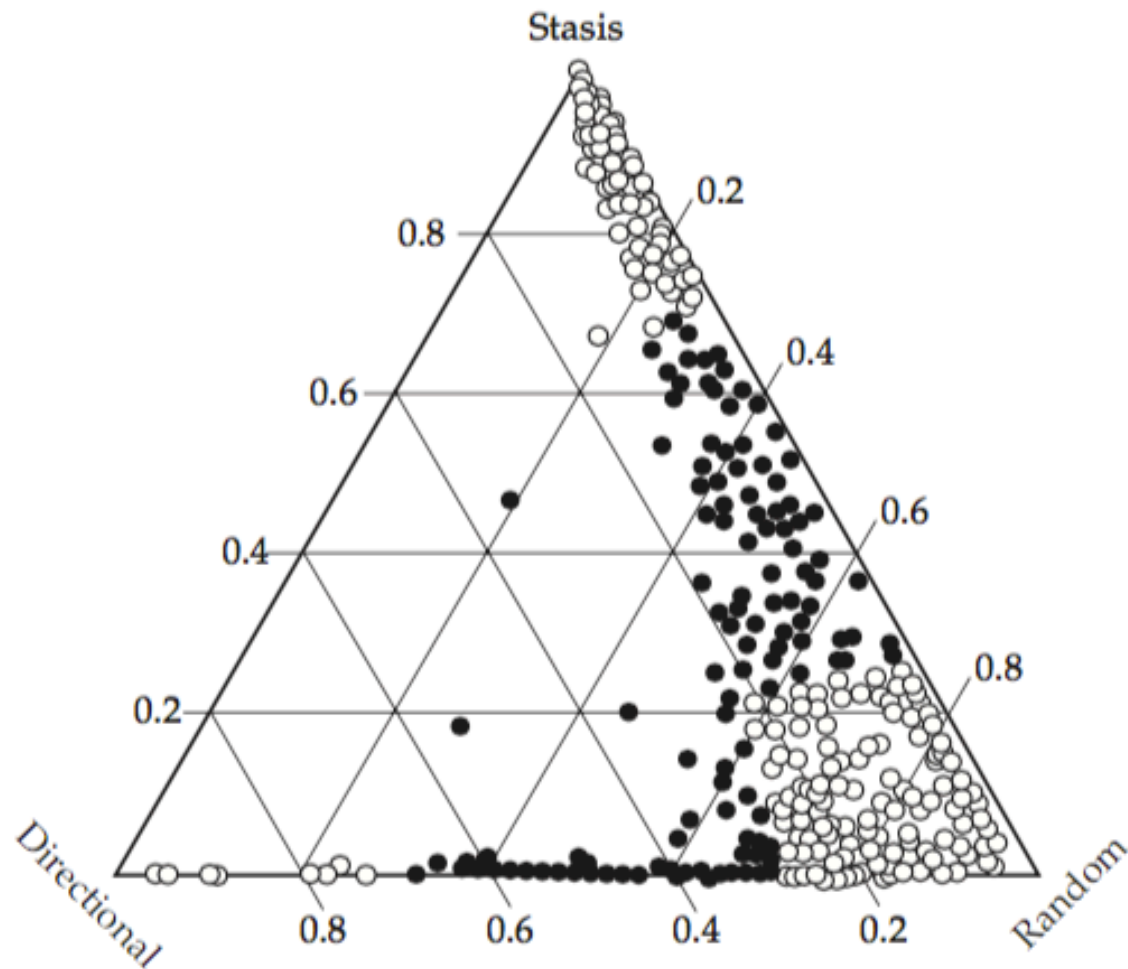
Figure 12.6 A De Finetti diagram of the support for the random walk, directional walk, and stasis models. Each point corresponds to the coordinates of the Akaike weights for these three models (which sum to one) for a single stratophenetic series. Points near vertices corresponds to almost 100% support for a particular model, hence the lables at the vertices. Points along an edge of the triangle indicate very little support for the model perpendicular to that edge. Unfilled points indicate strong support (weight for most supported model at least 2.7 times the weight of any other model). (After Hopkins and Lidgard 2012.)