Ann. Rev. Ecol. Syst. 1988. 19:445–71

PHYLOGENIES AND QUANTITATIVE CHARACTERS

Joseph Felsenstein

Department of Genetics, University of Washington, Seattle, Washington 98195

Dedicated to the memory of Sewall Wright (1889–1988), who did not doubt for a minute the relevance of his work to systematics.

INTRODUCTION

Systematists and evolutionary geneticists don't often talk to each other, and they routinely disparage each other's work as being of little relevance to evolution. Systematists sometimes invoke the punctuationist argument that most evolutionary change does not occur by individual selection and hence that within-population phenomena are largely irrelevant to evolution. They sometimes make the pattern-cladist contention that evolutionary processes cannot be inferred from any observation about pattern, nor can any knowledge of evolutionary processes inform inferences about pattern. More frequently systematists feel, with considerable justification, that evolutionary geneticists have simply failed to concern themselves with morphological characters and data on differences between species. Evolutionary geneticists in turn dismiss the idea that studies comparing species anciently diverged, using morphological characters far removed from the level of the gene and using nonquantitative methods, can either be sound in their inferences of pattern or can shed much light on evolutionary processes.

Changes in the methods of collecting and analyzing data are invalidating some of these views. Molecular data, such as nucleic acid sequences, give us observations at the gene level that are comparable across many species, and consideration of these data has drawn population geneticists across the species boundary. The availability of microcomputers and digitizers is leading to more quantitation of morphological data by systematists, and the field of

0066-4162/88/1120-0445$02.00

morphometrics, needed to derive meaningful characters from this flood of digitized coordinates, is expanding rapidly. My argument is that the methods used to study the evolution of quantitative characters within populations can profitably be used on a phylogenetic scale to illumine the connection between pattern and process. Futuyma (29) has reached much the same conclusion; so have Atchley (2), Shaffer (51), and of course Lande (38, 39). The moment seems ripe to consider the matter.

INFERRING TREES GIVEN COVARIANCES

Two Results From Quantitative Genetics

Two major results from quantitative genetics are directly relevant: the theory of selection response and the theory of genetic drift. Although I frequently use equations, many of them matrix equations, to discuss this theory, the reader uncomfortable with this can skip the equations and look for the verbal summaries that accompany them.

The results will be be stated in terms of variances, covariances, and covariance matrices. The covariance of two random variables is the expected value of the product of the deviations of both from their expectations:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]. \qquad 1.$$

The covariance is the numerator of the correlation between the variables (its denominator is the product of the standard deviations of the two variables, $\sigma_X\sigma_Y$). A covariance matrix is the generalization of a variance to many dimensions. It contains as its elements the covariances between all pairs of variables, and on the diagonal, the variances of those variables.

RESPONSE TO SELECTION The classic formula for predicting the response to selection (which arises from work by R. A. Fisher, Sewall Wright, and Jay Lush) is $R = h^2S$, where R is the response to one generation of artificial selection, h^2 is the heritability of the character, and S is the selection differential, which is the difference between the mean of the selected parents and the mean of the population from which they were chosen (17). The heritability is the fraction of variance in the character that is additive genetic variance. When it is low, selection response is reduced because some of the superior parents we choose owe their desirable phenotypes to nontransmissible environmental factors or to interactions between combinations of genes that will be disrupted by Mendelian segregation.

For natural selection Lande (39) has recast this classical formula in terms of Sewall Wright's (61, 65) concept of a fitness surface. It is well approximated by

$$\Delta z = V_A \nabla \ln \bar{W}, \quad 2.$$

where Δz is the change in the phenotype, V_A the additive genetic variance, and $\nabla \ln \bar{W}$ is the slope of the log mean fitness. This tells us that a population will move uphill on the fitness surface at a rate proportional to the additive genetic variance and to the slope of the log fitness. Lande derived a multivariate version of this as well, in which each of these quantities is replaced by a vector or matrix:

$$\mathbf{\Delta z} = \mathbf{G} \nabla \ln \bar{W}, \quad 3.$$

so that $\mathbf{\Delta z}$ is the vector of phenotype changes, $\mathbf{G}$ the additive genetic covariance matrix, and $\nabla \ln \bar{W}$ the vector of slopes of log mean fitness in each variable (which is a generalization of the slope to multiple dimensions). In multiple dimensions the population may not move directly uphill on the log fitness surface, as the $\mathbf{G}$ term tends to make it move more in those directions that have more additive genetic variance. There has long been an analogous formula for artificial selection in the theory of index selection (see 17, Ch. 13, or the more extended treatment in 5, Ch. 11). Lande & Arnold's discussion of fitness regressions (40) uses viability and reproduction data to estimate the fitness surface and so enables prediction of phenotypic changes using Formula 2.

GENETIC DRIFT Wright (64) considered genetic drift which is acting on the loci controlling a quantitative character. If different lines diverge so that genes in the same line have coefficient of kinship f relative to genes in different lines, the between-lines variance component for the character that was generated by this genetic drift would be $2fV_A$, where V_A was the additive genetic variance in the character. The extension of this formula to multiple correlated characters is straightforward. The matrix of covariances between changes in different characters is given by the same formula, with V_A simply replaced by the additive genetic covariance matrix $\mathbf{G}$.

When the kinship f is generated by genetic drift, we get $f = [1 - (1-1/(2N_e))]^t$, or approximately $f = t/(2N_e)$ if the divergence time t is small compared to the effective population size N_e.

Brownian Motion and Covariances

THE BROWNIAN MOTION MODEL The context in which these results can be applied is the Brownian motion model of evolutionary change. This was first developed by Edwards & Cavalli-Sforza (15) in their pathbreaking work on phylogenies from gene frequencies (the paper that introduced both parsimony and maximum likelihood approaches to estimating phylogenies). They

approximated the genetic drift of gene frequencies by assuming instead that each coordinate (each allele frequency, in their case) carried out a Brownian motion on an infinite scale. Maximum likelihood for the model they posed has been further developed and made workable by them, by Elizabeth Thompson, and by me (6, 16, 19, 22, 58). Figure 1 depicts a realization of this model.

Edwards & Cavalli-Sforza's assumption of Brownian motion is an approximation, a relatively good one for gene frequencies as long as they have not changed much. What interests us here is whether we could use the same model for change in quantitative characters. Is it reasonable to assume that each character is independently wandering, according to a Brownian motion, on an infinite scale? If so, then we would have a mathematically tractable probabilistic model for the evolution of quantitative traits and could use it to infer phylogenies.

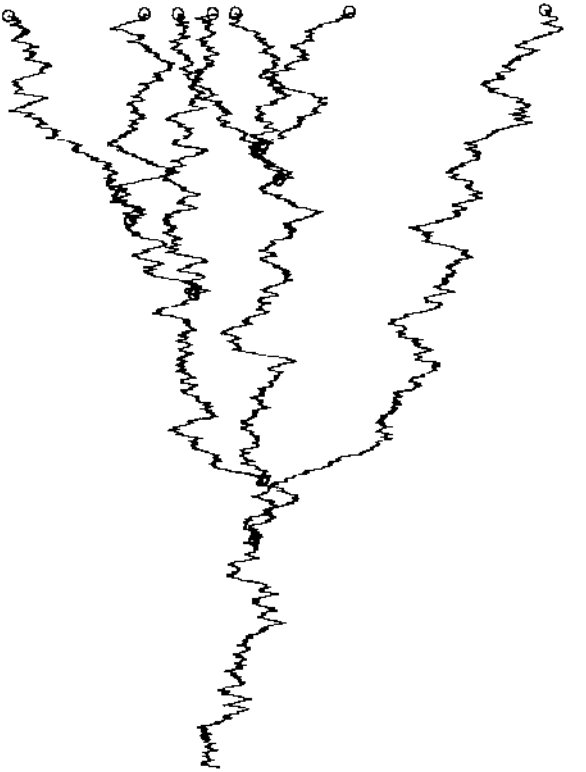

Figure 1 Outcome of a simulation of species changing their horizontal position by Brownian motion and speciating at random times. The vertical dimension is time. Tips and forks are indicated by circles. More typical simulation outcomes show more tangled branches of the tree.

The answer, of course, is that we cannot. Even if each character can be regarded as carrying out a Brownian motion on a scale, different characters will not be changing independently of each other or at the same rate. To see why and to see what possibilities there are of making the characters fit the original Brownian motion model, we must consider what evolutionary forces could cause a character to change in this way. The two candidates for such forces are genetic drift and natural selection. We can use the formulae given above to discover what covariances and correlations would be expected to exist between changes in different characters, and what relationships would exist between the rates of change in different characters.

CHANGE AND COVARIANCE When a character undergoes Brownian motion on a scale, expected displacement after an amount of time t is equally often positive and negative and has an expectation of zero. The actual displacement is drawn from a normal distribution (normal, because Brownian motion is the sum of a very large number of small independent displacements). This normal distribution has mean zero and a variance proportional to time. Note that the width of the distribution will be proportional to its standard deviation, which will increase as the square root of time. That square root dependence may seem not to be in accord with intuition; it comes about because change in different time periods is independent, and there are considerable amounts of reversal and wandering over areas previously traversed.

When there are two or more characters, each one has expectation of change zero, but they may have different variances of net displacement (in this sense different rates of change) and they may change in a correlated manner. The natural measure of their dependence is the covariance between their net displacements. The correlation between displacements can be computed from this covariance. In speaking of the covariances or the variances, we are of course looking at the characters prospectively and delineating the distribution of changes that may happen; a posteriori the net change after time t simply has some actual value in each character, and it makes no sense then to speak of its distribution unless we have many replicates of the same process.

Covariances of Change and Evolutionary Processes

COVARIANCES DUE TO GENETIC DRIFT When a single locus changes by genetic drift, it does not change on an infinite scale as does a Brownian motion, but the changes in successive generations are independent, which gives the two processes a close similarity. As the locus drifts closer to fixation, the amount of change per unit time gets progressively less, and this causes departure from the Brownian motion approximation. When we consider a quantitative character controlled additively by many loci, each locus

undergoing genetic drift, then the net change in that character may be approximated by Brownian motion. Even though some loci may be approaching fixation after a period, others may have new genetic variation newly arising as the result of mutation.

A number of recent studies have considered the theory of genetic drift and mutation in quantitative characters. Chakraborty & Nei (8) investigated a character with multiple alleles arranged along a discrete scale, each capable of mutating to the next allele. Smouse et al (54) and Chakraborty (7, 9) discussed measures of divergence between populations in quantitative characters in the presence of genetic drift. Another relevant group of papers is that of Rogers & Harpending (48) and Lewontin (41, 42) and the subsequent discussion of the differences between them and me (27, 43, 49) on how informative differences between populations in quantitative characters will be as indicators of phylogeny, relative to single loci.

Lynch & Hill (44) presented the theory of mutation and drift of a quantitative character. They found a particularly simple pattern. Like other authors, they found that the variance of the change in the character per generation is $V_A/(N_e)$ and the loss in within-population, genetic variance per generation is $V_A/(2N_e)$. However, they made an interesting observation. If mutation adds variance V_M to the genetic variance V_A every generation, when the loss and gain of variance balance each other and determine V_A, the equilibrium level of additive genetic variance is $V_A = 2N_eV_M$, so that the variance in the change of the character is expected to be $2V_M$ per generation. This is simply a quantitative characters version of the molecular clock, as it predicts the same rate of change in different lineages, even though they may have different effective population sizes.

When multiple characters are considered, the results are analogous. The covariance matrix of changes in characters is, as we have seen above, given by $\mathbf{G}/N_e$, where $\mathbf{G}$ is the matrix of additive genetic covariances between characters. $\mathbf{G}$ in turn is the result of a balance between mutation and genetic drift. If the input of mutational variance is given by the matrix $\mathbf{M}$, then we expect $\mathbf{G} = 2N_e\mathbf{M}$, and the covariance matrix of character change is then simply $2\mathbf{M}$. The pattern of changes of populations is as shown in Figure 2.

The important pattern emerging from all this is that there are expected to be differences between characters in rates of change and correlations between change in different characters. Therefore, it would not be justified to treat the characters as if they were evolving independently and at equal rates. However, as the covariance of changes is directly proportional to the within-population additive genetic covariances, we could in principle use quantitative genetic experiments to measure those and use them to transform the characters to new scales that made the characters independent and equivalent. This is not practical with many organisms.

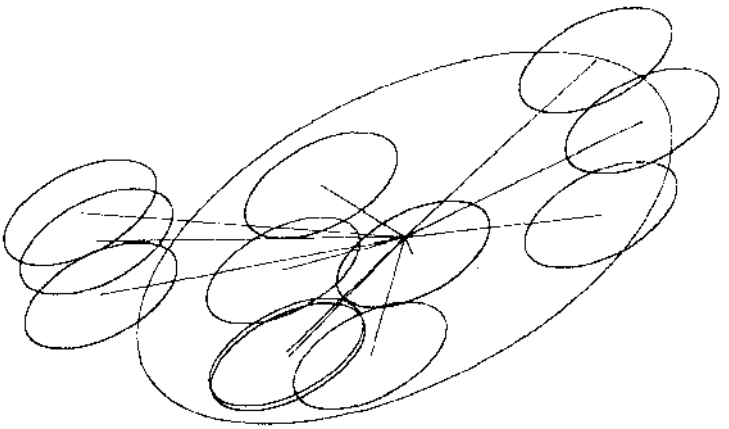

Figure 2 Outcome of a simulation of divergence of 12 populations by genetic drift, when there are two correlated characters (horizontal and vertical dimensions). The small ellipses show the pattern of the within-population additive genetic covariances, the large ellipse shows the pattern of the between-population covariation.

COVARIANCES DUE TO NATURAL SELECTION Genetic drift has only a limited plausibility as the source of observed differences between species. Particularly when the changes are observable and large, it is hard to imagine that they are not visible to natural selection and affected by it. One approach to seeing what are the sources of variance and covariance in change is to start from the selection response equations given above. A character may wander on a quantitative scale as a result of changes in the direction and intensity of selection. The formula $R = h^2S$, valid in each generation, shows that the relationship between the variance in the the net reponse R and the variance of the selection differential is $\mathrm{Var}(R) = h^4\ \mathrm{Var}\ (S)$. The variance of response is directly proportional to the variance of the selection differential, and to the square of the heritability.

For multiple characters, there is a corresponding equation. If **G** is the additive genetic covariances, **P** the phenotypic covariances, and **S** the covariance matrix of selection differentials, the covariance matrix **V** of changes can be shown to be

$$\mathbf{V} = \mathbf{G}\,\mathbf{P}^{-1}\,\mathbf{S}\,\mathbf{P}^{-1}\,\mathbf{G}, \qquad 4.$$

where $\mathbf{P}^{-1}$ is the inverse of the matrix of covariances of phenotypes, so that the covariances of selection response involve genetic covariances, phenotypic covariances (inversely), and covariances in the selection response. This can be related to Lande's form of the selection response equations by noting that

the middle terms of the right-hand-side of Equation 4, $\mathbf{P}^{-1}\,\mathbf{S}\,\mathbf{P}^{-1}$, can be rewritten as $\mathbf{C}$, the covariance of the slopes of log $\bar{W}$, so that

$$\mathbf{V} = \mathbf{G}\,\mathbf{C}\,\mathbf{G}. \qquad 5.$$

What this means is that if we see covariance between changes in different characters, it can have either of two different sources (or both). If the characters are genetically correlated, so that the genetic variations that affect one also affect the other, then we expect that natural selection acting on one of them also causes a correlated response in the other. This is reflected by the two terms $\mathbf{G}$ in Equation 5. This source of correlated change has been discussed by Shaffer (51), who gives references to many earlier comments on this point.

The term $\mathbf{C}$ reflects the other possible source of correlation, which should be called "selective covariance" in analogy to Stebbins' term "selective correlation" (57a:31). Selective covariance is the covariance in the distribution of traits, owing to covariance of the changes in these traits brought about by a correlation of their selection pressures. For example, a period of cold temperatures may select a species of mammal to be darker in coloration, have relatively shorter limbs, and be larger (respectively, Gloger's, Allen's, and Bergmann's Rules). Each of these has selective advantage in the altered environment. When one of these characters changes, the others are likely also to change, and in a predictable direction. There is no implication that there is any genetic correlation between these characters: the genes for coat color may be entirely different from the genes for relative limb length. In this example, the correlations in changes in these characters are due to the $\mathbf{C}$ term, not the $\mathbf{G}$ terms, of Equation 5. Zeng (66) has discussed these issues in a mathematical treatment of correlated response to long-term natural selection, coming to an equation almost identical to Equation 5, and conclusions that are completely compatible with the present argument.

The result is fairly depressing. We wish to have a way to transform the characters so that we have removed their covariances and placed them on scales on which all will evolve according to identical but independent Brownian motions. Variation in the selection differential can indeed produce change that is indistinguishable from Brownian motion. But the covariances depend on two things, the additive genetic covariances and the selective covariances; while we might conceivably collect genetic data relevant to the first, it is hard to see how we are to know the second.

The relevance of quantitative genetic parameter estimates to assessing between-species variation has been stressed by Atchley (2), Cheverud (11, 12), and Schluter (49a). Cheverud has argued, on what I think are insufficient grounds, that the genetic covariation will be molded by natural selection until

those dimensions in which variation is most tolerated by selection are those in which most genetic variance occurs. Atchley has argued (2) that those directions in which there is the highest heritability should be considered more important for systematic purposes. Being the easiest directions in which the population can respond to selection pressure, they might more properly be de-emphasized, just as we tend to discount size as too easy to change. Schluter (49a) has used estimates of genetic covariances to transform the characters prior to estimating a phylogeny. Neither of these arguments takes into account the selective covariance and its implications for assessing the phylogenetic meaning of multivariate data.

Before more looking closely at that point, it is worth noting that the model of natural selection used in these equations is implausible. In this model the selection pressure varies independently of the phenotype of the population. This could happen only if in each generation selection were simply directional, with fitness rising steadily as one proceeded in one direction along the phenotype scale. Variation in selection would involve the selection sometimes favoring higher, sometimes lower phenotype.

AN ADAPTIVE PEAK MODEL A more realistic model would involve the population attempting to climb an adaptive peak whose location wandered with time. This cannot be handled with the same equations, because with it selection differential covaries with the population mean and selection is more likely to be downward, the higher the mean population phenotype. If we assume that the adaptive peak has the shape of a normal distribution (a mathematically convenient assumption) and that its location changes through time by wandering according to a Brownian motion, mathematical solutions are readily obtained. If the amount of genetic variance and covariance remains roughly constant through time, selection will cause the population to move in each generation toward the peak of the adaptive surface. But since the peak continually moves, the population can never quite catch it.

The exact motions of the population mean can be calculated once we know three things: the shape of the peak, the covariance matrix of its Brownian motion in the phenotype space, and the population's genetic covariance matrix. I hope to present the details of the theory elsewhere, but the most important result is fairly easy to see. The population mean will differ from the position of the adaptive peak, coming to vary about it according to a normal distribution whose shape stabilizes and ultimately does not depend on elapsed time. Thus the adaptive peak, as it wanders through the phenotype space, tows the population with it. In the long run, the population more or less goes where the peak goes, and the departure of the population from the peak becomes smaller and smaller in comparison with the total distance the peak has moved.

The result is dramatic. If a branch of a phylogeny is long enough for this peak movement to dominate the process (that is, long compared to the time it would take the population to climb a substantial fraction of the way up the peak if the peak held still), then the main factor affecting the population's movements is the covariance matrix of the peak's movements. This result is even more depressing than the one mentioned above. In this model the genetic covariances have little influence, as long as a reasonable amount of genetic variation is available in all directions in the phenotype space. If we want to correct for the covariances of the characters, what we really need to know is the covariance matrix of peak movements. This cannot be obtained from quantitative genetic experimentation. It requires directly observing the process of phenotypic change in paleontological time, or inferring it from the differences between species.

INFERRING COVARIANCES GIVEN TREES

The problem of inferring the covariances and correlations between characters is a central problem in the comparative method. The usual way the problem arises is that we wish to know whether (say) brain size and body size are changing in a correlated fashion in the evolution of mammals. We ignore here the related problem of correlating changes in a character with observations on an environmental variable, although that involves somewhat similar logic. For the problem of relating change in different characters, four general approaches have been practiced.

The Nonphylogenetic Approach

The simplest way to make traditional verbal comparative methods quantitative seems to be to collect data on a number of species and to use standard regression and correlation statistics to test whether there is a significant relationship between brain size and body size. This has been a fairly common approach, but it is not justifiable statistically. The statistical models employed assume that the (x,y) data pairs are drawn independently from the same distribution, and they simply aren't. The species come from the tips of a phylogeny, which means that they occur in clusters. In any statistical test, the number of degrees of freedom used is in effect too large. A discussion of this in more detail, including references to some data examples will be found in my own paper on phylogenies and the comparative method (25).

The Analysis of Variance Approach

Harvey & Mace (33) realized that the nonindependence of the data points is a problem, and they have tried to correct for it by using groups from the classification system and analysis of variance. Considering the classification

system as a hierarchical design, they used analysis of variance to find out the level at which one had accounted for all the statistically detectable variance in each character. Thus, if most of the variance was between families, but within orders, and there was no additional effect of order or any higher level detectable, they would use families as the units instead of species and hope that this removed most of the effects of the phylogeny.

This is no doubt a better practice than simply treating species as independent, but it still has a number of problems. One is ignoring any phylogenetic clustering that occurs within families and also any that occurs between families, as these are not truly independent in their evolution but are themselves clustered. One is assuming that a given taxon (such as family) has the same phylogenetic meaning throughout one's data, when in fact some families may be considerably older and more diverse than others. Finally and most seriously, one is assuming that the groups in the classification system are all monophyletic, which is not true in most contemporary classifications, whether or not it ought to be.

Cheverud et al (13) have described a multivariate method that uses the levels in the classification system in a more complex way. It is hard for me to see what are the assumptions of this method, as they were not explicitly stated or investigated by Cheverud, but as no other details of an estimate of the phylogeny are employed, the method would appear to bear some relationship to the analysis of variance method.

The Parsimony Approach

Ridley (47) and Huey & Bennett (34) have employed approaches that take the structure of the phylogeny more completely into account. Using a phylogeny for the group and parsimony or similar methods, they have reconstructed the states or quantitative values of the characters at the interior nodes of the tree. It can then be seen whether changes in the two characters along interior branches of the tree are correlated, by considering those changes as data and treating the correlation statistically.

If we could reconstruct interior states in a tree without error, this approach would be promising, although it would still have one problem. Different branches of the tree are of different length so that characters would be expected to change more in one branch than in another. Thus, although the changes in different branches would be independent, they would not be identically distributed; identical distribution is also required for use of standard regressions and correlations. This difficulty could be circumvented by ignoring the magnitude of the changes and retaining only their signs, so that one simply asked whether positive change in one character was accompanied by positive change in the other, using a nonparametric sign test.

But the central problem with this method is that the reconstructions of

interior states are not without error, even if the phylogeny is known. Those interior states are not data, they are inferences from data. If two sister species are in the same derived state, parsimony reconstructs their common ancestor as having been in that state. It need not have been: that state could have arisen as a parallel change in both lineages. If two characters both show derived states in the same pair of sister species, they appear to have changed in a perfectly correlated manner. If instead one character had a parallelism, the correlation may have been less perfect than appears. There has been no attempt to check on the effects of these errors of reconstruction on parsimony approaches to comparative methods. These methods are probably more accurate than the other approaches described so far, but no one yet knows how accurate they are.

The Contrasts Approach

My own suggestion (25) for dealing with this problem involves assuming that the characters evolved by Brownian motion. The null hypothesis is that each character has independently evolved by Brownian motion, perhaps with different rates, but uncorrelated with each other. Given these assumptions, we can use the details of the phylogeny to define a series of contrasts which should be statistically independent. Thus, if the phylogeny shows us that two duck species are sister species, and so are two goose species, the difference in phenotype between the two ducks should be statistically independent of the difference between the two geese. This is expected because the evolutionary events within the duck lineage are distinct from, and independent of, those in the goose lineage. Thus, it is possible, if we have knowledge of the phylogeny, to extract from it contrasts between species that are statistically independent.

If we have not only a tree topology but also branch lengths that define expected amounts of change, we can go farther. The Brownian motion is expected to accumulate twice as much variance in a branch of length 2 as in a branch of length 1. Thus, we can use the branch lengths in the tree to compute the variances of the contrasts, and they can then be scaled so that all have the same variance and the same expectation (zero). We then have a series of contrasts that are independent and identically distributed, and to which standard statistical tests can be applied. The contrasts are not just between adjacent tip species; there are also contrasts that correspond to interior branches in the tree. In the duck-goose example we can compute a weighted average of the two duck phenotypes, minus a weighted average of the two goose phenotypes. If the weights are chosen correctly, that contrast is also independent of the among-ducks and among-geese contrasts. Without the branch lengths the weights cannot be computed correctly. I have explained elsewhere (25) the computation of the contrasts.

In looking at the regressions and correlations of the contrasts in different characters, we can examine the same questions originally addressed by regressions or correlations of the characters across the tip species. If we find that a positive contrast in brain weight is associated with a positive contrast in body weight, that is equivalent to showing that an increase in one is associated with an increase in another.

The method has scarcely ever been applied because it requires knowledge of the phylogeny. Sessions & Larson (50), in its first published application, used a phylogeny derived from a morphological and biochemical study to compute the contrasts and study the correlation between genome size and regenerative growth rate in plethodontid salamanders. The statistical validity of doing this is dependent on the precision of the phylogeny. Taking into account the possibility of error in the phylogeny is difficult. One way would be to use bootstrap sampling of the molecular data (26), to compute a series of bootstrap estimates of the tree. We could compute contrasts from each tree, bootstrapping the individual observations as well, so as to allow for their sampling variability. The resulting distribution of (say) correlation coefficients between brain weight and body weight could then be examined to see whether zero lay within its upper 95%.

Even with the uncertainty of the tree taken into account in this way, there must still remain uncertainty as to the validity of the Brownian motion model. If we were willing to assume that evolutionary changes in different branches of the tree are independent but not willing to make the rest of the assumptions of the Brownian motion model, we could still use a nonparametric test, although one that loses some power. This test, suggested to me by remarks of a graduate student (whose name unfortunately I do not know) at University College, London, is quite simple. We look at pairs of species on the tree, with the pairs chosen so that the paths from one partner to the other through the tree do not contain any branches in common. This can always be done so as to find $n/2$ such pairs if there are n species, with $n/2$ rounded down to the next integer. (Thus with 20 species there will be 10 pairs, with 19 species 9 pairs).

We can simply do a sign test as to whether the changes in our two characters match in sign. The sign test is robust against variations from the Brownian motion model, but it does require that change in different branches be independent, including successive branches in the same lineage.

THE LIMITS OF MORPHOLOGICAL SYSTEMATICS

The situation seems simple and depressing, but there is still an unresolved mystery. Given the tree, and accepting some model of change, we can estimate the covariances of changes in different characters and test hypotheses about correlations between characters. Given the characters and the covari-

ances of changes, we could transform them to equivalence and independence, and then use the new characters to estimate the tree. But could we estimate both covariances and trees? The answer seems to be that we cannot. There is a substantial confounding between our estimates of the tree and of the covariances. (I am indebted to Robin Thompson of the University of Edinburgh for demonstrating to me in the three-species case that the covariances and the tree are confounded).

This confounding is an uncertainty principle of sorts. What we would like to do is to take a series of characters, estimate a provisional tree, from that (using the contrasts dictated by that tree) revise our estimate of the covariances of the characters, from those get a new tree, and continue iteratively until the process converges. Thompson's result destroys any hope of this or any similar research program succeeding.

Molecular data can rescue us from the resulting dilemma. Molecular sequence data consists of information about evolutionary change at a number—usually a large number—of different sites. Molecular distance methods such as DNA hybridization will be successful if they can approximate this situation. The methods of phylogenetic inference, including both parsimony and likelihood methods, assume either implicitly or explicitly that change at the different sites is independent. So in effect we know the covariances (although, the characters being discrete, covariances are not directly involved) and are estimating the tree. If we accept the validity of this assumption for molecular data, we can then use the resulting trees, together with the morphology, to draw conclusions about the correlation of different morphological characters. This research program, one long favored by molecular evolutionists, is disliked by many morphologists on the grounds that it accepts the superiority of molecular data.

It also explicitly denies that morphological data can be used to infer phylogenies when the covariances of character change are unknown. Such a view seems mandated by the argument just given. Yet it is clearly not the whole story, for morphological data have been used with considerable success for the last century to infer phylogenies. It was morphology, not molecules, that convinced us long ago that we were closer kin to the gorilla than to the oak tree. How could this have been possible? What is wrong with the argument?

The inference about gorillas and oak trees was based on a reasonably large sample of characters. I think that it was not misled because covariances of character change were not tight between all pairs of characters. This is a property of inferences of this sort: if the average correlation of characters is not large, then the phylogeny will converge on the correct one if we accumulate enough characters. One can analogize the inference to observing a sequence of heads and tails in coin-tossing. Even if certain of the tosses have

correlated outcomes (as when the experimenter occasionally forgets to toss before writing down the next outcome), accumulation of enough tosses will still converge on the correct probability of heads.

The very fact that we have been able to use morphology to infer phylogenies thus tells us something about the pattern of character correlation. Although we have reason to believe the morphological estimate of the phylogeny, our ignorance of the details of the pattern of character correlation impedes our ability to use the data statistically. Any attempt to make a statistical statement about how much more we ought to believe one phylogeny than another would have to make use of the invalid assumption that all the characters have evolved independently, or else use the details of the (unknown) character correlations to compute the extent of uncertainty in our estimate.

For nucleic acid sequences the situation is different. We have prior reason to believe that different sites evolve nearly independently. Of course, nearby sites may evolve in a correlated way, especially if they are in the same codon. But even different sites in the same gene may well change independently, and if they are not in the same gene the changes that change is independent would seem to be quite high. We therefore have greater confidence in using the sites as if independent to make statistical statements. We can be more confident in the validity of our statistical statements about a tree built from molecular sequence data than from morphological data.

GEOGRAPHIC DIFFERENCES WITHIN SPECIES

I do not have space here to comment on the issues involved in interpreting patterns of geographic differentiation of characters. These issues are quite similar to those that arise in interpreting comparative data, except that the structure of relationships within species is radically different from that between species. Instead of a branching tree we may have a network of interconnections among populations, with neighbors in this network exchanging migrants. We may have to distinguish between similarity of phenotype due to the populations being in similar environments, and similarity due to gene flow. Sokal & Oden (56) have applied Cliff & Ord's (14) spatial autocorrelation methods to statistical testing of geographic variation, although without an explicit model of the biological processes generating the data. More conventional multivariate statistics are also frequently used, as reviewed by Thorpe (59). Sokal (57) provides some overview and further references to all these methods.

The difficulty with these methods is that, in the absence of a well-specified model of the population structure, the genetics of the characters, and the processes of evolutionary change, any method is inevitably arbitrary and exploratory. I have discussed (24) the difficulties of distinguishing between

historical branching events and later gene flow in the case of gene frequencies. No one has posed the corresponding problems for quantitative characters, and there have been hardly any attempts to solve the problem for either kind of data. Suffice it to say that many of the problems and models discussed in this paper have counterparts in geographic variation analysis.

THE CHARACTER CODING PROBLEM

Current systematic treatments of quantitative characters revolve around the problem of character coding. A brief review of the major approaches is in order here. I argue that the "problem" is a nonproblem; there is no compelling reason for coding quantitative characters into discrete states and hence that efforts to find the best way of doing so are misguided. For more detailed discussion of coding methods, the reader is referred to the paper of Archie (1).

Standardization

The first formal efforts at coding characters into discrete states were preceded by methods for scaling quantitative characters so as to make them equivalent. The intention was to make the characters equivalent in the sense that one unit of change in each was of equal evidential value in inferring phylogenies by parsimony. The earliest Wagner parsimony programs (notably WAGNER78) allowed the states of the characters to be on a continuous scale, but before they could be used one had to choose the equivalent scales for the characters.

Farris (18; see also 36) suggested standardizing characters so that their within-population variances are equal. This uses an estimate of the phenotypic variance, the within-population variance. Farris's suggestion is closely related to Mahalanobis's concept of generalized distance. However, it does not summarize the data in a single distance measure, but rather keeps the variables distinct. It is not equivalent to finding new scales that remove character correlations, because the standardization occurs character by character and does not attempt to correct for covariances. However, it is easily extended to do that. The standardization to equalize phenotypic variances is not guaranteed to equalize genetic variances. It is at best a rough approximation—there is no guarantee that genetic variance will be proportional to phenotypic variance, however much we might wish it to be. In the absence of genetic information it is perhaps the best we can do, but the roughness of the approximation must always be kept in mind.

A similar suggestion is Gower's (32) ranging, implicit in his generalized coefficient of similarity (see the discussion by Sneath & Sokal, 55), which can be used when there are no within-population samples. It consists simply of rescaling the characters so that their between-species ranges are equal and the two outermost species have the values 0.0 and 1.0. Once again, no

attempt is made to correct for character correlations (nor is it easy to see how this could be done). In this case the between-population variation is used for the standardization. This, as we have seen, is the result of the levels of genetic variation and (if selection is the force differentiating species) and the variation of the selection differential. Even if we ignore the need for correcting for correlations among characters, we have a possible problem with circularity. We saw above that we cannot infer both covariances of character change and a phylogeny from the same data without confounding the two inferences. The same confounding may exist when we use the same between-species data to infer phylogenies and infer variances of character change, for that is what we do when we use range coding to standardize characters. Unfortunately there have been no studies of the statistical properties of range coding nor of the extent to which it might avoid the problem of confounding.

Methods for Discrete Coding

Subsequent methods concentrate on a more restrictive task: coding the character into discrete states, usually as integers. The impetus for this comes from the current generation of parsimony computer programs, which require that the states be discrete. The assumption on which the programs operate is that changes of state in different characters are evidentially equivalent—in effect that probabilities of change in different characters are sufficiently equivalent. How much equivalence is required is a matter of debate—see my papers (20, 23) and my debate with Sober (28).

An assumption implicit in discrete coding is that there may be points on the scale of a quantitative character that natural selection would find it difficult to pull the population across. If so, change across these points would be less probable than change in between them, and a coding which emphasizes these improbable events would acquire some justification. The matter has not been discussed in any detail, but I believe that such a view underlies several of the character coding methods suggested in the literature.

The best known discrete character coding method is the "gap coding" of Mickevich & Johnson (45), according to which one examines the distribution of species means on the characters scale, looking for gaps. These are then interpreted as points of instability in the movement of the character, and they are the points that should be taken as the boundaries of the integer states. It is presumably harder, in some sense, for the population to cross those regions.

The difficulty that will be immediately apparent is that we may infer the existence of such regions when all that has happened is that species are absent from that portion of the scale as a result of random happenstance. The species at the tips of the tree are also not independently sampled—the very essence of a phylogeny is that they come in clusters. This leads to even greater clumping than their numbers would suggest. For instance, if we have two clusters of

rather closely related species, between which a great deal of evolution had taken place, under a Brownian motion model we would expect to see two clusters in the histogram of positions on the scale. Between them would be a gap. No matter how large this gap, a gap coding method would interpret it as a single evolutionary step. This would devalue the information if the gap were large and might inflate it if it were small. Archie (1) has proposed a generalization of gap coding that avoids the problem of collecting so many species that the gaps entirely disappear—another problem that can arise with gap coding.

Simon (52) has developed a coding method, generalized range coding, that has some similarity to scaling by within-population standard deviation. She uses within-species information to create groups of species that are homogeneous, that cannot be made larger without becoming heterogeneous to a statistically significant degree. She then proposes setting up a discrete character whose states reflect the membership in these (overlapping) homogeneous groups. This is an attempt to code in a way that loses as little as possible of the within-population information, without the interpretation that one is discovering truly discrete characters on a continuous scale. Like the within-population scaling method, it does not take into account covariances between characters. One difficulty with it is that how many states it creates in a character may be as much a function of the sample sizes (which may not be the same for all characters) as of the differentiation in that character.

It will be apparent from this brief treatment that I am skeptical of all schemes for discrete character coding. There is no requirement that phylogenies be based purely on data with discrete states. None of the authors on coding methods has yet faced the question of how we could test for the presence of underlying discrete states. Lacking such a test, there is no reason to discretize quantitative characters. The real "character coding problem" is that people insist on discretely coding quantitative characters that would better be left on continuous scales.

The Threshold Model—Continuous Coding of Discrete States

The possibility exists that even characters with discrete states may reflect underlying continuous characters! A standard model in quantitative genetics for discrete characters envisages discrete observable phenotypes but assumes that whether an individual is in one state or the other depends on many loci. This "threshold model" (see the treatment in 17) assumes that the underlying scale, called liability, shows polygenic variation. The individual has one state or the other, depending on whether its liability puts it beyond a threshold value. An pioneering application of this model was Sewall Wright's (62, 63) analysis of the genetics of the number of digits in the guinea pig hind foot.

When applied to a discrete phenotype observed in multiple species, the

threshold model leads to predictions different than those of the fully discrete Markovian model of transitions back and forth between states that has been used in phylogenetic inference. Under a threshold model with Brownian motion on the underlying liability scale, immediately after the population has changed from one state to another, it is more likely to change back again than it is later on, for the simple reason that its mean is still near the threshold. The frequencies of the two states in the population become important under the threshold model; they provide information on how much of the liability distribution overlaps the threshold, and thus how close the mean is to the threshold. Under the discrete Markovian model, the within-species variation is ignored.

The threshold model is not easily dealt with mathematically, but it seems likely to do an improved job of approximating the evolution of polygenic but discrete phenotypes. It is yet one more reason for dialog between evolutionary geneticists and systematists.

PALEONTOLOGICAL OBSERVATIONS

One method of inferring processes of evolutionary change of quantitative characters (and for that matter inferring phylogenies as well) is direct observation from the fossil record. Of course, observation is rarely as direct as we might wish, but an increasing number of paleontologists are seeking out data sets in which closely spaced observations on a single species can be made. It would be possible in principle to estimate covariances of character change by using measurements of successive populations. These could then be used to inform either neontological observations on different species (as described above), or to improve paleontological observations relevant to phylogeny, improving on the power of Gingerich's (31) "stratophenetic" approach to phylogeny.

When the paleontological observations are not infinitely dense, the gaps in the record demand inferences about descent that use logic similar to that needed for neontological data. It is customary for systematists of the "phylogenetic systematics" school to argue that fossil species ought not be given any special treatment, and that their date of occurrence should not inform inferences about the phylogeny. This is an overreaction to uncritical use of fossils. If we have a model of evolution, we should be able to use it to combine fossils and neontological information in an optimal way, without either losing the temporal ordering information or treating fossils invariably as ancestors.

For the moment statistical treatment of character change in paleontology is still at an early stage and is focused on single characters. The Brownian motion model has been used by Charlesworth (10) to test whether change in a

character shows evidence of directional selection. Bookstein (3, 4) has applied it to testing whether there is evidence of punctuational patterns of change in a character. There is undoubtedly much more to be done.

DOUBTS ABOUT THE MODEL

The Brownian motion model has been advocated here for two reasons: it corresponds well to what we expect if genetic drift is the mechanism of character change, and it is mathematically tractable. In the case of natural selection, the choice of Brownian motion is rather arbitrary, and other mathematically tractable models need to be investigated as well. Three candidates are the Ornstein-Uhlenbeck process, jump processes, and random walk of a population among adaptive peaks.

The Ornstein-Uhlenbeck Process

Uhlenbeck & Ornstein (60) defined a diffusion process that has a linear pressure returning it to a central point. At any instant the expected change is toward that point, at a rate proportional to the particle's distance from the point. The change varies around this in a way otherwise typical of Brownian motion. They intended it as a model for the velocity of motion of a particle that has momentum but is subjected to friction, which always tends to reduce the velocity toward zero. The process has become known as the Ornstein-Uhlenbeck (here called OU) process. A discrete-time analogue of the OU process is the random walk of an elastically bound particle, which is jumping back and forth at random but is tethered to a central point by an elastic band. It is possible to solve for transition probabilities of the OU process and to show that ultimately the particle is normally distributed around the central point, with variance determined by a balance between the strengths of the random noise and the returning force.

The OU process is a good model for the motion of a population which is wandering back and forth on a selective peak under the influence of genetic drift. Natural selection plays the role of the elastic band. Lande (37, 38) has used it to ask how far the population can wander from the selective peak given the strength of selection and the population size. He has discovered (38) that the reduction of fitness due to wandering part of the way down the peak is dependent on the population size but not on the curvature of the fitness curve. He points out the intriguing (and disquieting) implication that a finite population cannot tolerate optimum selection on a number of characters much larger than the population size. Species with small population size cannot be too finely adapted to their niches, and this puts them in additional jeopardy.

The OU process could also serve as the model for the wanderings of an adaptive peak in the phenotype space, where the optimum remains within a

relatively confined region. If the peak itself wanders according to an OU process, and the population mean is wandering by genetic drift while tethered to the peak, the resulting movement of the population will itself not be an OU process but can be well-approximated by one. A major feature of an OU model of character change is that it gradually "forgets" past history. A species' position on a character scale varies around the central point, and most of that variation comes from recent fluctuations. Older history of the species becomes less and less relevant, its influence erased by the steady pull toward the central point.

Comparative biologists tend to suspect comparisons of distantly related species; they hope to base their comparisons on recent evolutionary events that have not been overlaid by much subsequent change. No one has yet discovered how to carry out statistical comparative methods in the case of the OU process, but when this is done, it will be found that comparisons between distantly related species should be accorded much less weight than those between closely related species. The Brownian motion model instead argues that both should get equal weight.

However, in another respect the OU model is inadequate. Under it, there is no way to recover information about events of the distant past. All record of the ancient phenotypes is expected to have been erased by the pull toward the central point. Systematists believe, with apparent justification, that morphology does allow us access to information on ancient evolutionary events. So an OU model cannot be the whole story, any more than a Brownian motion model can be.

Other Models

The bestiary of tractable random processes that might serve as models of character change includes more than just the Brownian motion and OU processes. The "pure jump process" used by Gillespie (30) in his discussion of variable selection seems one candidate for a less smooth process that could approximate the discontinuity of much environmental change.

Another class of models that deserves more exploration is random wanderings from one adaptive peak to another, where the adaptive peaks are scattered about in the phenotype space. This process is reminiscent of Simpson's (53) image of evolution proceeding by changes of "adaptive zone," although he envisaged the phenotypes of competitors and changes in the environment causing continual movement, disappearance, and origination of the adaptive zones. Simpson's view was in turn heavily influenced by Sewall Wright's "shifting balance theory" of evolution (61, 65). A model of wandering among static peaks might, however, not be much different from a model of evolution by Brownian motion. If the peaks were spaced in a regular rectangular array, each peak with two neighbors in each dimension, then the movements from

peak to peak would closely resemble a random walk or Brownian motion if a sufficiently long time span were considered. Lande (37, 38) and I (21) have discussed the effect of genetic variation on the way a population moves near an adaptive peak. Kirkpatrick (35) discussed models for transition from one adaptive peak to another in the absence of genetic drift.

It might be thought that a model involving directional selection should be developed, but this becomes less obvious the more it is considered. If all populations are subjected to the same directional selection, then all are expected to change at the same rate in the same direction on the phenotype axis, and the selection will simply move them all along the axis together without affecting the distances between them. The only way to see the effect of the selection would be if it were different in some lineages than in others. One model that seems natural is to have each lineage subjected to a selection pressure that is constant throughout the life of the lineage and changes at the moment it gives rise to two daughter species. The net change in any branch of a phylogeny is then normally distributed, with a variance proportional not to the branch length but to the square of the branch length. An alternative model embodying punctuationist assumptions would have the position of the phenotypic optimum undergo a random walk, with a single step at the beginning of each branch of the phylogeny (or else at the beginning of only the daughter species, if the parent species is thought to continue in stasis).

The difficulty with these superficially attractive models is that they make an implicit assumption—that all the relevant speciation events in the phylogeny are visible to us. In reality, we may have examined only a fraction of the extant species and are probably ignorant of the very existence of most of the extinct species. If there have been many past extinctions within the group, the net change in any lineage will reflect changes in many smaller intervals between speciations. This makes its behavior much closer to a Brownian motion or OU process. There is a need for much more work on random models of character change or of movement of phenotypic optima.

Heterogeneity of Covariances

In the models I have mentioned above, the genetic covariances between characters are assumed to remain constant as the population mean changes. In reality there are many reasons they might change considerably (see, for example, the review by Mitchell-Olds and Rutledge, 46). Developmental constraints will be one cause of change in genetic variances and covariances. One easy way to visualize this is to think in terms of transformations of scale. If a character is being changed toward zero, at which point there is a developmental constraint (perhaps simply the fact that the character is the length of a bone, which can never be negative), then as the character mean approaches zero the genetic variation must necessarily finally be reduced to

zero. However, if we consider a transformation such as the logarithm of the character, as the character mean moves toward zero the mean of its logarithm moves toward negative infinity. The genetic variance on the logarithmic scale could remain constant. That corresponds to constancy on the original scale not of genetic variance but of the coefficient of variation.

Thus it is possible that some of the heterogeneity of genetic variances and covariances could be removed if we could transform the characters appropriately. The difficulty lies in knowing where are the barriers imposed by developmental constraints, and what are the transformations needed to correct for the compression of the effects of genetic variation as the barrier is approached. Frequently, we will not know enough to allow use of the kind of models envisioned here.

Another cause of heterogeneity of genetic variances and covariances is sampling variation. We may see different covariances within different lineages simply because we have small samples of each lineage. It will be important not to jump to the conclusion that there is heterogeneity when one has not checked whether it could be an artifact of small sample sizes.

Natural selection and genetic drift can both cause heterogeneity of genetic variances and covariances. The effect of the former is not easily distinguished from developmental constraints. As selection fixes one allele at a locus, the genetic variance is reduced. If there were a constraint (say a trivial one such as that the wing of a Drosophila could not be of negative size), then on selection for small wing size, once the alleles reducing the wing size had all fixed, if the wing had vanished there would then remain no further genetic variation that natural selection could use to reduce the size. The constraint shows up at the genetic level as an absence of existing genetic variation, as well as of mutational variation.

Genetic drift can also bring about heterogeneity of the variances and covariances in a way unconnected to the notion of developmental constraints. By moving the allele frequencies of the relevant genes closer to fixation, it reduces the genetic variation at these loci. If some populations have undergone more genetic drift than others, they are expected to have less genetic variation within them. However, this is only an expectation—in individual cases genetic variation may also increase in a lineage by the gene frequencies at some loci drifting closer to intermediate values. By a similar effect, genetic drift may make genetic covariances in different lineages heterogenous. These effects must be distinguished from mere sampling variation in the estimates of the genetic variances and covariances, as they cannot be made to disappear by taking a larger sample at the time the characters are measured. As a population evolves, with genetic variation disappearing at some loci and being introduced at others, the pattern of genetic covariation is expected to change through time and to be different in different lineages.

Realism and the Uses of Models

In the models used here, each character has been determined additively by a number of loci. Of course, real characters are not determined additively, but models such as this have long served as acceptable approximations for quantitative inheritance in animal and plant breeding. It is in the long run that quantitative genetic models are likely to do least well, primarily because the multivariate normal theory used implicitly assumes an infinite number of loci of infinitely small effect, and the exact number of loci becomes relevant to the long-term selection response. When the distribution of mutation effects is considered in the quantitative genetic theory, this too causes inaccuracy in long-run predictions.

We have seen that with developmental constraints there is some possibility that we can restore the validity of an additive model by an appropriate transformation. If we cannot, what point is there in using quantitative genetic theory? Questions of realism analogous to this are a common reaction to any theoretical work in biology. They have three possible answers.

First, by adding complications to the model we may hope to make it more fully realistic. There is some hope of this, but not much. Biological reality is so complex that we are very far from any reasonably mechanistic understanding of evolutionary processes. Nevertheless it is always necessary to keep trying to improve the realism of the models and not surrender to despair. Second, we may use the models to show how little we are able to do. If under the simplifying assumptions of quantitative genetics we are able to show that there is insufficient information to infer phylogenies from quantitative characters, then when those assumptions are removed the situation can only get worse. The models, even when they are not realistic, inform us of the limits of our abilities.

Models can also serve one other purpose. They give us tools to think more clearly about the interactions of the evolutionary forces and about the way these affect the quantities we observe. As heuristic devices they are always subject to skepticism based on their oversimplified nature, but they are nevertheless useful in honing our understanding and training our intuition. It is for this purpose that I am trying to call systematists' attention to quantitative genetic methods, and quantitative geneticists' attention to the data and problems waiting to be addressed in systematics. At the moment, most systematists discuss changes of character state in qualititative terms, as if these were instantaneous events happening in individuals rather than quantitative changes in gene frequencies. Quantitative geneticists simply do not discuss them at all.

I have no illusion that precise and powerful predictions will be possible within the limitations of our current knowledge. But there is something worth talking about, and members of each field must learn not to dismiss each others' work out of hand. It is time to start communicating.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant number BSR-8614807.

Literature Cited

1. Archie, J. W. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. *Syst. Zool.* 34:326–45
2. Atchley, W. R. 1983. Some genetic aspects of morphometric variation. In *Numerical Taxonomy*, ed. J. Felsenstein, pp. 346–63. NATO Advanced Sci. Inst. Ser. G, No. 1. New York: Springer-Verlag
3. Bookstein, F. L. 1987. Random walk and the existence of evolutionary rates. *Paleobiology* 13:446–64
4. Bookstein, F. L. 1988. Random walk and the biometrics of morphological characters. *Evol. Biol.* In press
5. Bulmer, M. G. 1980. *The Mathematical Theory of Quantitative Genetics*. Oxford: Clarendon
6. Cavalli-Sforza, L. L., Edwards, A. W. F. 1967. Phylogenetic analysis: models and estimation procedures. *Evolution* 22:550–70 (also published in *Am. J. Hum. Gen.* 19:233–57)
7. Chakraborty, R. 1978. Single-locus and multi-locus analysis of genetic differentiation of the races of man: a critique. *Am. Nat.* 112:1134–37
8. Chakraborty, R., Nei, M. 1982. Genetic differentiation of quantitative characters under optimizing selection, mutation, and drift. *Genet. Res.* 39:303–14
9. Chakraborty, R. 1982. Allocation versus variation: the issue of genetic differences between human racial groups. *Am. Nat.* 120:403–04
10. Charlesworth, B. 1984. Some quantitative methods for studying evolutionary patterns in single characters. *Paleobiology* 10:308–18
11. Cheverud, J. M. 1982. Phenotypic, genetic, and environmental morphological integration in the cranium. *Evolution* 36:499–516
12. Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* 110:155–71
13. Cheverud, J. M., Dow, M. M., Leutenegger, W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism of body weight among primates. *Evolution* 39:1335–51
14. Cliff, A. D., Ord, J. K. 1973. *Spatial Autocorrelation*. London: Pion
15. Edwards, A. W. F., Cavalli-Sforza, L. L. 1964. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification*, ed. V. H. Heywood, J. McNeill, pp. 67–76. London: Systemat. Assoc. Publ. No. 6.
16. Edwards, A. W. F. 1970. Estimation of the branch points of a branching-diffusion process. *J. Royal Statist. Soc. B* 32:155–74
17. Falconer, D. S. 1981. *Introduction to Quantitative Genetics*. London, New York: Longman. 2nd ed.
18. Farris, J. S. 1966. Estimation of conservatism of characters by constancy within biological populations. *Evolution* 20:587–91
19. Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Gen.* 25:471–92
20. Felsenstein, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27:401–10
21. Felsenstein, J. 1979. Excursions along the interface between disruptive and stabilizing selection. *Genetics* 93:773–95
22. Felsenstein, J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35:1229–42
23. Felsenstein, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linnean Soc.* 16:183–96
24. Felsenstein, J. 1982. How can we infer history and geography from gene frequencies? *J. Theoret. Biol.* 96:9–20
25. Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 126:1–25
26. Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–91
27. Felsenstein, J. 1986. Population dif-

ferences in quantitative characters and gene frequencies: a comment on papers by Lewontin and Rogers. *Am. Nat.* 127:731–32

28. Felsenstein, J., Sober, E. 1986. Parsimony and likelihood: an exchange. *Syst. Zool.* 35:617–26
29. Futuyma, D. J. 1988. *Sturm und drang* and the evolutionary synthesis. *Evolution* 42:217–26
30. Gillespie, J. H. 1978. A general model to account for enzyme variation in natural populations. V. The SAS-CFF model. *Theoret. Popul. Biol.* 14:1–45
31. Gingerich, P. D. 1979. The stratophenetic approach to phylogeny reconstruction in vertebrate populations. In *Phylogenetic Analysis and Paleontology,* ed. J. Cracraft, N. Eldredge, pp. 41–77. New York: Columbia Univ. Press.
32. Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–71
33. Harvey, P. H., Mace, G. M. 1982. Comparsions between taxa and adaptive trends: problems of methodology. In *Current Problems in Sociobiology,* ed. Kings College Sociobiol. Group, pp. 343–61. Cambridge: Cambridge Univ. Press
34. Huey, R. B., Bennett, A. F. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. *Evolution* 41:1098–1115
35. Kirkpatrick, M. 1982. Quantum evolution and punctuated equilibria in continuous genetic characters. *Am. Nat.* 119:833–48
36. Kluge, A. G., Farris, J. S. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1–32
37. Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–34
38. Lande, R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *Am. Nat.* 116:463–79
39. Lande, R. 1981. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402–16
40. Lande, R., Arnold, S. J. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–26
41. Lewontin, R. C. 1978. Single- and multiple-locus measures of genetic distance between groups. *Am. Nat.* 112:1138–39
42. Lewontin, R. C. 1984. Detecting population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* 123:115–24
43. Lewontin, R. C. 1986. A comment on the comments of Rogers and Felsenstein. *Am. Nat.* 127:733–34
44. Lynch, M., Hill, W. G. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40:915–35
45. Mickevich, M. F., Johnson, M. S. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Syst. Zool.* 25:260–70
46. Mitchell-Olds, T., Rutledge, J. J. 1986. Quantitative genetics in natural plant populations: a review of the theory. *Am. Nat.* 127:379–402
47. Ridley, M. 1983. *The Explanation of Organic Diversity. The Comparative Method and Adaptations for Mating.* Oxford: Clarendon
48. Rogers, A. R., Harpending, H. C. 1983. Population structure and quantitative characters. *Genetics* 105:985–1002
49. Rogers, A. R. 1986. Population differences in quantitative characters as opposed to gene frequencies. *Am. Nat.* 127:729–30

49a. Schluter, D. 1984. Morphological and phylogenetic relations among the Darwin's finches. *Evolution* 38:921–30

50. Sessions, S. K., Larson, A. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. *Evolution* 41:1239–51
51. Shaffer, H. B. 1986. Utility of quantitative genetic parameters in character weighting. *Syst. Zool.* 35:124–34
52. Simon, C. M. 1983. A new coding procedure for morphometric data with an example from periodical cicada wings. In *Numerical Taxonomy,* ed. J. Felsenstein, pp. 378–82. NATO Adv. Sci. Inst. Ser. G, No. 1. New York: Springer-Verlag
53. Simpson, G. G. 1953. *The Major Features of Evolution.* New York: Columbia Univ. Press
54. Smouse, P. E., Spielman, R. S., Park, M. H. 1982. Multiple-locus allocation of individuals to groups as a function of the genetic differentiation within and differences among human populations. *Am. Nat.* 119:445–63
55. Sneath, P. H. A., Sokal, R. R. 1973. *Numerical Taxonomy.* San Francisco: Freeman
56. Sokal, R. R., Oden, N. L., 1978. Spatial autocorrelation in biology 1. Methodology. *Biol. J. Linnean Soc.* 10:199–228
57. Sokal, R. R. 1983. Analyzing character variation in geographic space. In *Numerical Taxonomy,* ed. J. Felsenstein, pp.

383–403. NATO Adv. Sci. Inst. Ser. G, No. 1. New York: Springer-Verlag
57a. Stebbins, G. L. 1950. *Variation and Evolution in Plants*. New York: Columbia Univ. Press
58. Thompson, E. A. 1975. *Human Evolutionary Trees*. Cambridge: Cambridge Univ. Press
59. Thorpe, R. S. 1976. Biometric analysis of geographic variation and racial affinities. *Biol. Rev.* 51:407–52
60. Uhlenbeck, G. E., Ornstein, L. S. 1930. On the theory of Brownian motion. *Phys. Rev.* 36:823–41
61. Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. VI Int. Congr. Genet.* 1:356–66
62. Wright, S. 1934. An analysis of variability in number of digits in an inbed strain of guinea pigs. *Genetics* 19:506–36
63. Wright, S. 1934. The results of crosses between inbred strains of guinea pigs, differing in number of digits. *Genetics* 19:537–51
64. Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* 15:323–54
65. Wright, S. 1982. Character change, speciation, and the higher taxa. *Evolution* 36:427–43
66. Zeng, Z-B. 1988. Long-term correlated response, interpopulation covariation, and interspecific allometry. *Evolution* 42:363–74

Annual Review of Ecology and Systematics
Volume 19, 1988

CONTENTS

(*continued*)