# PHYLOGENIES AND THE COMPARATIVE METHOD

JOSEPH FELSENSTEIN

Department of Genetics SK-50, University of Washington, Seattle, Washington 98195

*Submitted November 30, 1983; Accepted May 23, 1984*

Recent years have seen a growth in numerical studies using the comparative method. The method usually involves a comparison of two phenotypes across a range of species or higher taxa, or a comparison of one phenotype with an environmental variable. Objectives of such studies vary, and include assessing whether one variable is correlated with another and assessing whether the regression of one variable on another differs significantly from some expected value. Notable recent studies using statistical methods of this type include Pilbeam and Gould's (1974) regressions of tooth area on several size measurements in mammals; Sherman's (1979) test of the relation between insect chromosome numbers and social behavior; Damuth's (1981) investigation of population density and body size in mammals; Martin's (1981) regression of brain weight in mammals on body weight; Givnish's (1982) examination of traits associated with dioecy across the families of angiosperms; and Armstrong's (1983) regressions of brain weight on body weight and basal metabolism rate in mammals.

My intention is to point out a serious statistical problem with this approach, a problem that affects all of these studies. It arises from the fact that species are part of a hierarchically structured phylogeny, and thus cannot be regarded for statistical purposes as if drawn independently from the same distribution. This problem has been noticed before, and previous suggestions of ways of coping with it are briefly discussed. The nonindependence can be circumvented in principle if adequate information on the phylogeny is available. The information needed to do so and the limitations on its use will be discussed. The problem will be discussed and illustrated with reference to continuous variables, but the same statistical issues arise when one or both of the variables are discrete, in which case the statistical methods involve contingency tables rather than regressions and correlations.

## THE PROBLEM

Suppose that we have examined eight species and wish to know whether their brain size ($Y$) is proportional to their body size ($X$). We may wish to test whether the slope of the regression of $Y$ on $X$ (or preferably of *log* $Y$ on *log* $X$) is unity. Figure 1 shows a scatter diagram of hypothetical data. It is tempting to simply do

Am. Nat. 1985. Vol. 125, pp. 1–15.

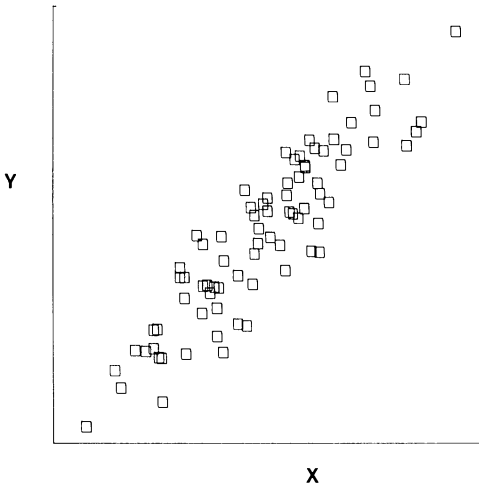


FIG. 1.—Scatter diagram of hypothetical data from 8 species, showing the relationship between $Y$ and $X$.

an ordinary regression and see whether the confidence limits on the slope include unity. (Since there is random error in both $X$ and $Y$ this is a questionable procedure, but we leave that issue aside here.)

If we were to do such a regression, what would be the implicit statistical model on which it was based? The simplest assumption would seem to be that the points in figure 1 were drawn independently from a bivariate normal (Gaussian) distribution. What evolutionary model could result in such a distribution? The simplest is shown in figure 2: the eight species resulted from a single explosive adaptive radiation. Along each lineage there were changes in both characters. If evolution in each lineage were independent, and the changes in the two characters were drawn from a bivariate normal distribution, then our distributional assumptions would be justified. The individual species could be regarded as independent samples from a single bivariate normal distribution.

Let us accept for the moment that the changes of a set of characters in different branches of a phylogeny can be reasonably well approximated as being drawn from a multivariate normal distribution, and that changes in distinct branches are independent. Even given those assumptions, a problem arises based on the unlikelihood that the phylogeny has the form shown in figure 2.

Consider instead the phylogeny shown in figure 3. In it, the eight species consist of four pairs of close relatives. Suppose that the changes in the two characters in each branch of the tree can be regarded as drawn from a bivariate normal distribution with some degree of correlation between the characters. We might not expect the same amount of change in short branches of the tree, such as the eight terminal branches, as in longer branches such as the four that arise from the original radiation. Let us assume that the variance of the distribution of change in a branch is proportional to the length in time of the branch, much as it would be if the characters were undergoing bivariate, and correlated, Brownian motion.

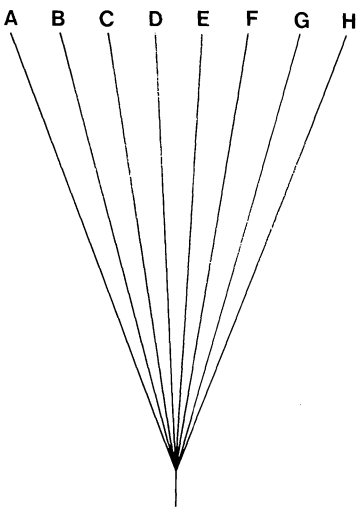


FIG. 2.—One phylogeny for the 8 species, showing a burst of adaptive radiation with each lineage evolving independently from a common starting point.

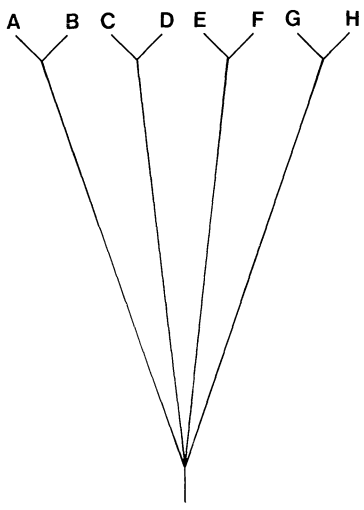


FIG. 3.—Another phylogeny for the 8 species, showing a radiation that gives rise to 4 pairs of closely related species.

This would produce a distribution like that shown in figure 4. It is apparent that instead of eight independent points we have four pairs of close relatives. If we were to carry out a statistical test based on the assumption of independence, say a test of the hypothesis that the slope of the regression of $Y$ on $X$ was zero, we would imagine ourselves to have 6 degrees of freedom (8 − 2). In fact, we very nearly have only four independent points, so that the effective number of degrees of freedom is closer to 2 (4 − 2). A test of the significance of the slope, or of the

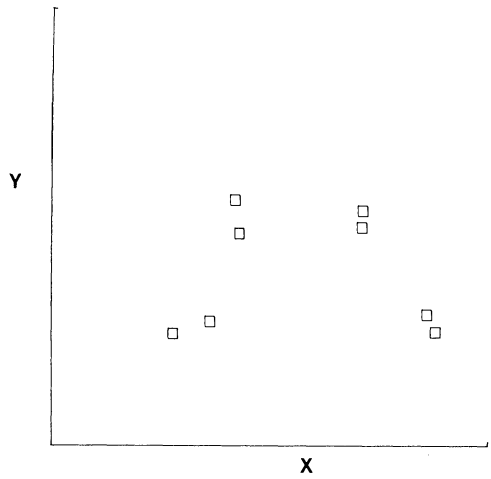


FIG. 4.—A data set simulated using the phylogeny of fig. 3, under a model of random, normally distributed, independent change in each character, where the change in each branch is drawn independently from a normal distribution with mean zero and variance proportional to the length of the branch.

extent of its difference from any preassigned value, will be excessively likely to show significance if the number of degrees of freedom is taken to be 6 rather than 2.

A worst case of sorts for the naive analysis is shown in figure 5, where the phylogeny shows that a large number of species actually consist of two groups of moderately close relatives. Suppose that the data turned out to look like that in figure 6. There appears to be a significant regression of $Y$ on $X$. If the points are distinguished according to which monophyletic group they came from (fig. 7), we can see that there are two clusters. Within each of these groups there is no significant regression of one character on the other. The means of the two groups differ, but since there are only two group means they must perforce lie on a straight line, so that the between-group regression has no degrees of freedom and cannot be significant. Yet a regression assuming independence of the species finds a significant slope ($P < .05$). It can be shown that there are more nearly 3 than 40 independent points in the diagram.

One might imagine that the problem could be avoided by use of robust nonparametric statistics. In fact, nonparametric methods, unless specifically designed to cope with the problem of nonindependence, are just as vulnerable to the problem as are parametric methods. For the data of figure 6, a Spearman rank correlation finds a nearly significant correlation ($P < .065$) between the two variables, showing that it has little better ability to cope with nonindependence than do parametric methods.

One might also imagine that we could escape from the problem simply by ensuring that we sample the species at random from the species that form the tips of the phylogeny, and thus somehow escape from the nonrandomness of the pool

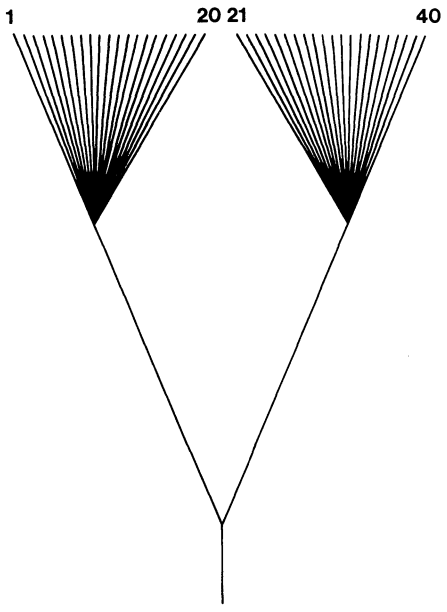


FIG. 5.—A "worst case" phylogeny for 40 species, in which there prove to be 2 groups each of 20 close relatives.

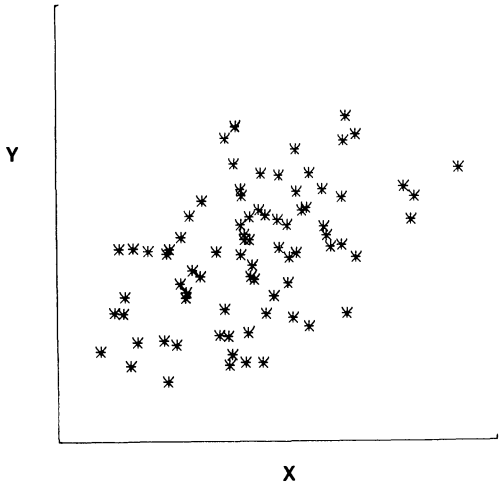


FIG. 6.—A typical data set that might be generated for the phylogeny in fig. 5 using the model of independent Brownian motion (normal increments) in each character.

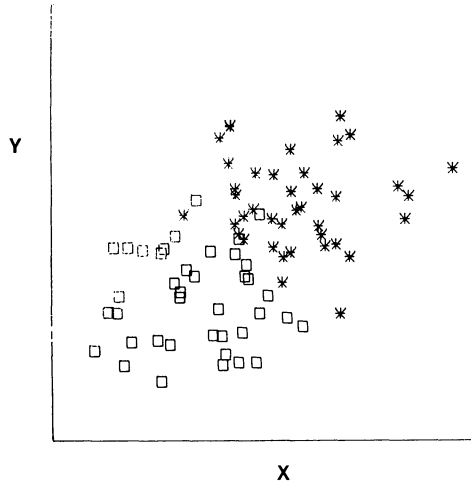


FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.

of species from which we are sampling. This does not work. Imagine two species that have diverged some time ago, and thus have diverged in both brain and body weight. Clearly the correlation between those characters cannot be significant, since there are only two points. Now if each species gives rise to a group of 100 daughter species, essentially identical to it, we now have two clusters of 100 species each. Sampling species from this pool of 200 species, we are actually sampling from two species, but do not know it. Correctly analyzed, no data from this group could possibly achieve significance, but if we draw (say) 50 species at random from the 200 and analyze that data as if the points were independent we will probably conclude that there is a significant correlation between brain and body weight.

There is one case in which the problem does not arise. That is when the characters respond essentially instantaneously to natural selection in the current environment, so that phylogenetic inertia is essentially absent. In that case we could correlate a phenotype with the environment. We could also correlate two characters with each other, provided that we realized that their correlation might simply reflect response to a common environmental factor. It may be doubted how often phylogenetic inertia is effectively absent. In any case the presumption of the absence of phylogenetic inertia should be acknowledged whenever it is proposed to do comparative studies without taking account of the phylogeny.

## PREVIOUS APPROACHES

The problem of correcting for nonindependent evolutionary origins has not gone unnoticed by previous workers in comparative biology. Clutton-Brock and Harvey (1977, pp. 6–8) pointed out that "if phylogenetic inertia is strong, the

potential adaptations that related species may evolve will be similarly constrained, with the effect that species cannot be regarded as independent of each other." They used a nested analysis of variance to find that taxonomic level (in their case it was genera) which accounted for as much of the variation as possible, and then tried to correct for nonindependence by using genera rather than species as the units of their statistical analysis. A more complete exposition of their approach is given by Harvey and Mace (1982). Baker and Parker (1979), who were analyzing bird coloration, also pointed out the problem, and tried to correct for it by seeing whether the same relationships held within different families. Sherman (1979) and Givnish (1982) discussed the problem, though without attempting to correct for it.

Gittleman (1981) used a parsimony method to reconstruct and count the number of times parental care had evolved independently on a phylogeny derived from the classification of bony fishes. Ridley (1983) has discussed the problem of nonindependence in considerable detail, also proposing that parsimony be used to reconstruct the placements of changes on the phylogeny, to enable tests of whether the occurrence of changes in two characters are independent.

There are two problems with using a parsimony approach as suggested by Gittleman and by Ridley. The most serious is that one is usually forced to rely on a presumed monophyly of taxa in the Linnean classification system, in the absence of external evidence as to the phylogenetic relationships between the groups. The traditional classification system is, of course, only partly a monophyletic one: only about half of the classes of the Chordata are thought to be monophyletic, for example, and even within the Mammalia orders such as Insectivora and Carnivora are believed not be monophyletic. The classification system is not sufficiently detailed to show all of the structure in the phylogeny, even if all its groups were monophyletic: the relationships between some orders of mammals are undoubtedly closer than between others. The phylogenetic meaning of a given category varies from group to group: the insect genus *Drosophila* is thought to be as old as the mammalian order Primates.

A second problem with parsimony assignments is that they have only partial statistical justification: I have argued elsewhere (Felsenstein 1978) that when parsimony is used to reconstruct the phylogeny, it can have undesirable statistical properties when evolutionary rates are not small and differ sufficiently in different lineages (for a review, see Felsenstein [1983]). Even when the phylogeny is known, and parsimony is used only to reconstruct the placement of changes of character state, it is well known that this can lead to biases; for example, if two changes occur in parallel in sister lineages, the reconstruction will instead show one change occurring in their immediate ancestor. If changes in one character occurred in parallel in the sister lineages, but the change in another character occurred in the immediate common ancestor, then, although the reconstructed changes appear coincident, the actual changes in those characters were not in fact coincident.

The seriousness of the additional statistical error and statistical biases that this may cause in comparative studies has never, to my knowledge, been investigated. Nevertheless, assigning changes of characters by parsimony on a known phy-

logeny would be immeasurably superior to simply treating the species as if independently evolved.

## A POSSIBLE SOLUTION

If we know the phylogeny and have a model of evolutionary change, it should be possible in principle to correct for the nonindependence of taxa. To see how, first let us consider the highly symmetrical phylogeny in figure 8, supposing that we know that it is the true phylogeny. Recalling that each character is being assumed to be evolving by a Brownian motion that is independent in each lineage, then taking $X_i$ to be the phenotype $X$ in species $i$ it is easy to see that differences between pairs of adjacent tips, such as $X_1 - X_2$ and $X_3 - X_4$, must be independent. This is so because the difference $X_1 - X_2$ depends only on events in branches 1 and 2, while $X_3 - X_4$ depends only on events in branches 3 and 4, and these two sets of events are independent.

Brownian motion is a random process modeling the wanderings of a molecule affected by thermal noise. If we measure the position of the molecule along one axis, its successive displacements are independent. This has the effect that the displacement after time $v$ has elapsed is the sum of a large number of small displacements, each of which is equally likely to be either positive or negative. The result is that the total displacement is drawn from a normal distribution with mean zero and a variance proportional to $v$. In the present model the different characters undergo Brownian motion at different rates, so that after one unit of time the change in $X$ has variance $s_X^2$ and the (possibly correlated) change in $Y$ has variance $s_Y^2$. After $v$ units of time their variances are, respectively, $s_X^2 v$ and $s_Y^2 v$.

Given this model, it is straightforward to show (Felsenstein 1973, 1981*b*) that the contrast $X_1 - X_2$ has expectation zero and variance $2s_X^2 v_1$. Since we assume that we know the $v_i$, we can scale the contrast by dividing by its standard deviation, obtaining a variate that should have expectation zero and unit variance. We can similarly scale the other three contrasts $X_3 - X_4$, $X_5 - X_6$, and $X_7 - X_8$ by dividing each by the square root of $2\,s_X^2\,v_1$. Even more contrasts are available. It will be less obvious, but nevertheless true, that $(X_1 + X_2)/2 - (X_3 + X_4)/2$ is a contrast independent of the others. It will have expectation zero and variance $s_X^2(v_1 + 2v_9)$. We can continue down the tree in similar fashion, obtaining two more contrasts, $(X_5 + X_6)/2 - (X_7 + X_8)/2$ and $(X_1 + X_2 + X_3 + X_4)/4 - (X_5 + X_6 + X_7 + X_8)/4$. Their expectations are also zero, and their variances are, respectively, $s_X^2(v_1 + 2v_9)$ and $s_X^2(2v_{13} + v_9 + v_1/2)$. They too can be scaled to have unit variance.

We have now extracted from this tree seven independent contrasts on the $X$ scale, each of which can be regarded as drawn from a normal distribution with mean zero and variance one. We can carry out the same process in the variable $Y$, and obtain seven independent contrasts in the same way. The $X$ contrasts will be independent of each other but not of the $Y$ contrasts. It can be shown that the coresponding contrasts $X_1 - X_2$ and $Y_1 - Y_2$ have covariance

$$\mathrm{Cov}[X_1 - X_2,\ Y_1 - Y_2] = 2v_1\, s_X\, s_Y\, r_{XY}, \qquad (1)$$

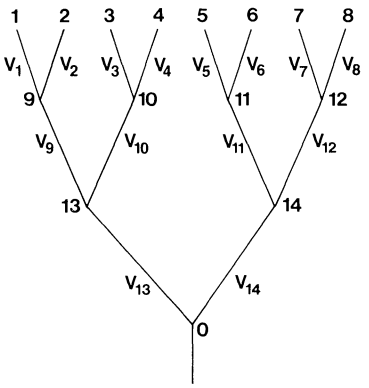


FIG. 8.—An example of a phylogeny, assumed known, from which we can define independent contrasts between taxa. This tree is highly symmetric, so that $v_1 = v_2 = v_3 = v_4 = v_5 = v_6 = v_7 = v_8$, $v_9 = v_{10} = v_{11} = v_{12}$, and $v_{13} = v_{14}$.

so that these two contrasts have the same correlation as the original variates. Since contrasts such as $X_1 - X_2$ and $X_3 - X_4$ are independent, a fortiori $X_1 - X_2$ will be independent of $Y_3 - Y_4$.

The quantities $s_X$ and $s_Y$ are important. It would be unreasonable to assume that characters $X$ and $Y$ had the same rates of evolution: $s_X$ and $s_Y$ are the scaling constants that convert from Brownian motion to the scales on which $X$ and $Y$ actually evolve. Thus $X$ is undergoing a Brownian motion, with variance $s_X^2$ accumulating per unit time, and $Y$ is undergoing a (possibly correlated) Brownian motion with variance $s_Y^2$ accumulating per unit time. We leave aside for the moment the problem of estimating $s_X$ and $s_Y$, and assume that they are known.

By dividing each contrast by its standard deviation, we have obtained from the original eight species seven pairs of contrasts that can be regarded as drawn independently from a bivariate normal distribution with means zero, variances unity, and an unknown correlation $r_{XY}$ between the members of a pair. Testing independence of the evolution of $X$ and $Y$ reduces to simply testing whether this correlation is zero. If instead we wanted to know the regression of changes in one variable on changes in another, we could use $s_X$ and $s_Y$ to compute

$$b_{Y.X} = s_Y r_{XY}/s_X \tag{2a}$$

and

$$b_{X.Y} = s_X r_{XY}/s_Y. \tag{2b}$$

These are not the usual equations for interconverting correlations and regressions, since $s_X$ and $s_Y$ are not observed standard deviations of $X$ and $Y$, but scaling constants that are merely proportional to the standard deviations of the variables $X$ and $Y$. Even though they are not standard deviations, they do allow us to correctly convert correlations into regressions. Other methods of analysis, such as principal components, can be carried out in similar fashion.

The preceding computation of contrasts depended on the phylogeny having a particular, and very unlikely, symmetric structure. Fortunately a more general procedure exists, of which the above was a special case. I have discussed its elements elsewhere (Felsenstein 1973) as part of a computational method for obtaining the likelihood of a given phylogeny. The general prescription for computing these contrasts is repeated applications of the following steps: (1) Find two tips on the phylogeny that are adjacent (say nodes $i$ and $j$) and have a common ancestor, say node $k$. (2) Compute the contrast $X_i - X_j$. This has expectation zero and variance proportional to $v_i + v_j$. (3) Remove the two tips from the tree, leaving behind only the ancestor $k$, which now becomes a tip. Assign it the character value

$$X_k = \frac{(1/v_i)\, X_i + (1/v_j)\, X_j}{1/v_i + 1/v_j}. \tag{3}$$

the weighted average of $X_i$ and $X_j$, the weights being proportional to the inverses of the variances $v_i$ and $v_j$. (4) Lengthen the branch below node $k$ by increasing its length from $v_k$ to $v_k + v_i v_j/(v_i + v_j)$. This lengthening occurs because the weighted average that computes $X_k$ in equation (3) does not compute the phenotype of the ancestor but only estimates it, and does so with an error that is statistically indistinguishable from an extra burst of evolution after node $k$.

After one pass through steps 1–4, we have found one contrast and reduced the number of tips on the tree by one. We continue to repeat steps 1–4 until there is only one tip left on the tree. This will extract $n - 1$ contrasts if there were originally $n$ species. Each contrast can be divided by the square root of its variance to bring them to a common variance. Since the $v_i$ are arbitrary, this procedure can be used on a phylogeny of any shape whatsoever, even on ones that contain multifurcations, since those can always be represented as a series of bifurcations having some branch lengths zero.

Figure 9 shows a nonsymmetric phylogeny, and table 1 the contrasts extracted from it by the above algorithm. The reader may want to try steps 1–4 on the symmetric phylogeny of figure 8, to verify the correctness of the contrasts and variances given above.

## DIFFICULTIES

One might imagine that, with the ability to compute independent contrasts from any phylogeny, we have an acceptable method of correcting for the presence of the phylogeny. Unfortunately, this is not the case. A number of difficulties intervene that leave us with much work remaining to be done.

### *How Do We Reconstruct the Phylogeny?*

In practice, we will hardly ever know the phylogeny in advance in sufficient detail to use it to obtain the contrasts. There are three sources of information likely to be used to reconstruct the phylogeny.

1. *Gene frequencies.*—A number of electrophoretic loci, blood group loci, or DNA restriction polymorphisms may be chosen and the frequencies of the alleles

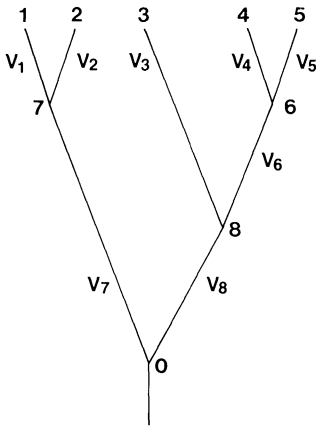


FIG. 9.—A less symmetrical phylogeny. The independent contrasts for this phylogeny are given in table 1.

TABLE 1

THE FOUR CONTRASTS EXTRACTED FROM THE PHYLOGENY SHOWN IN FIGURE 9, EACH WITH ITS VARIANCE, ALL COMPUTED USING STEPS 1–4 IN THE TEXT

| CONTRAST | VARIANCE |
|---|---|
| $X_1 - X_2$ | $v_1 + v_2$ |
| $X_4 - X_5$ | $v_4 + v_5$ |
| $X_3 - X_6$ | $v_3 + v'_6$ |
| $X_7 - X_8$ | $v'_7 + v'_8$ |

where

$$X_6 = \frac{v_4 X_5 + v_5 X_4}{v_4 + v_5}$$

$$v'_6 = v_6 + v_4 v_5/(v_4 + v_5)$$

$$X_7 = \frac{v_2 X_1 + v_1 X_2}{v_1 + v_2}$$

$$v'_7 = v_7 + v_1 v_2/(v_1 + v_2)$$

$$X_8 = \frac{v'_6 X_3 + v_3 X_6}{v_3 + v_6}$$

$$v'_8 = v'_7 + v_3 v'_6/(v_3 + v'_6)$$

at these loci determined in a sample from each species. This permits us to make an estimate of the phylogeny. The estimate has an error that must be taken into account when using it (see discussion below). It depends on the assumption that the evolutionary changes at the loci are predominantly due to genetic drift, and can therefore be modeled by Brownian motion of suitably transformed gene frequencies. This is more plausible the lower the taxonomic level at which we are working. A review of methods for inferring phylogenies from gene frequencies will be found in my recent paper on that subject (Felsenstein 1981*b*).

2. *Molecular sequences.*—As nucleic acid sequences become available for a wider range of organisms, it will become practical to infer the phylogeny from these, on the assumption that a "molecular clock" is valid for those sequences. I have described elsewhere (Felsenstein 1981*a*) a maximum likelihood method for inferring phylogenies from nucleic acid sequences.

3. *Quantitative characters.*—Most morphological taxonomic studies do not collect gene frequencies, but collect many morphological measurements besides those that we want to correlate. Could we use those quantitative characters to infer the phylogeny? We could do so in principle if we had a probability model for the evolution of the characters. The difficulty is that quantitative characters will evolve at different rates, and in a correlated fashion (as in the important case of size correlations). If we could find a transformation of the characters to a new set of coordinates that could be modeled as evolving independently by Brownian motion, with equal rates of accumulation of variance, we could apply the maximum likelihood methods developed for gene-frequency data. There would be no circularity involved: the new coordinates would be independent of each other and of the characters we were investigating. Unfortunately, there is no obvious way to get the information needed to untangle the skein of character correlations. Within-population samples, even when available, do not necessarily give us the required information. They allow us to estimate phenotypic covariances rather than additive genetic covariances. The latter control the covariances of evolutionary changes in characters if the characters change by natural selection or genetic drift. Even if the additive genetic covariances could be obtained by means of breeding experiments, they could not tell us whether the selection pressures for different characters were correlated. Cold weather, for example, might impose selection for large size and dark coloration while warm weather might select for small size and light coloration, leading to changes in these characters being correlated even if there were no genetic correlation between them. For the time being, transforming to remove within-population phenotypic covariance is the best that can be done, but this is necessarily an approximation, whose adequacy is unknown.

4. *The characters we are investigating.*—Sometimes the only characters available to us are the two whose relationship we are investigating. This is not only a severely limited amount of information with which to infer a phylogeny, but could result in some circularity since the phylogeny is inferred from some of the same information that is being used to reconstruct the changes of the characters along it. The matter needs a careful statistical investigation, but preliminary indications are that when we are trying to infer both the phylogeny and character correlations there is confounding between these, so that we can infer one or the other but not both.

### *How Do We Put Confidence Intervals on Our Inferences?*

If we were given a phylogeny known to be the true one, and were willing to trust the Brownian motion model of character change, we could obtain the appropriate contrasts from the phylogeny as outlined above, and use standard formulas to place confidence limits on the inferred correlations or slopes. But the phylogeny is never known without error. How are we to take the uncertainty of our knowledge of the phylogeny into account when constructing confidence intervals?

In principle this can be done by considering the phylogeny $T$ and the set of correlations (or slopes) $C$ to be a single multivariate quantity $(T, C)$ being estimated by maximum likelihood. The estimate is that pair $(T^*, C^*)$ resulting in the highest likelihood, and an approximate confidence interval is the set of all points $(T, C)$ whose likelihood is an acceptable fraction of the maximum likelihood, as judged by the likelihood ratio test. If a single correlation or slope is being estimated, the 95% confidence interval will be all values of $C$ for which there is a phylogeny $T$ such that their likelihood $L(T, C) \geq 0.1465\ L(T^*, C^*)$, since this is the ratio that would just reach significance in the likelihood ratio test with 1 degree of freedom.

It may be possible in some cases to discard the information about the phylogeny that we are least certain of, and use only those features in which we have reasonably high confidence. A student (whose name is unfortunately not known to me; see the ACKNOWLEDGMENTS section below) has pointed out to me that we could use contrasts between pairs of species that we were fairly sure had a common ancestor not shared with any member of another pair, and that these contrasts could then be safely assumed to be independent. For example, in a study of mammals we could use pairs consisting of two seals, two whales, two bats, two deer, etc. These contrasts would be independent, but would not necessarily have equal variances. We therefore could not simply compute correlations or regressions from the pairs. We could use certain nonparametric methods such as a sign test to test whether the changes in the two characters were correlated, but other nonparametric methods such as Spearman's rank correlation would not be valid because we could not assume that the pairs were drawn from a common distribution, even though they are independent. This sign test is essentially the same as that used by Baker and Parker (1979) to test whether regressions were similar in different bird families. It should be obvious that there is much statistical work remaining to be done on robust methods of using partial knowledge of phylogenies to make inferences about regressions and correlations of characters.

### *What If We Lack an Acceptable Statistical Model of Character Change?*

All of the above has been predicated on the acceptance of the Brownian motion model as a realistic statistical model of character change. There are certainly many reasons for being skeptical of its validity. Persistence of selection pressures over time may lead to correlation of changes in successive branches of the phylogeny, and common selective regimes experienced by different populations owing to common environmental factors such as weather or predators may lead to

changes in different lineages being correlated. Since the lengths of the branches of the phylogeny are given not in time units, but rather in units of expected variance of change (the $v_i$) the model already allows rates of change to differ in different lineages, and would allow, for example, change to be faster after speciation events than during later periods, as assumed to many punctuationists.

There is no reason to believe that the normal distribution is particularly plausible as the distribution from which changes in individual branches of the phylogeny are drawn, except insofar as the net change in a branch is the resultant of a series of bursts of change and thus might be approximately normal.

The matter of the model is an obvious point for future development and (to the extent that this is possible) empirical study. One rather serious problem that confronts comparative studies is that the relationship under study may change through time. Harvey and Mace (1982) have discussed the problem of change of the slope of the relationship between two variables with taxonomic level, which appears to be quite common. It should be possible to use the current model to study statistically whether there is any connection between the variance of a contrast and the slope of the regression of one variable on another.

### *What if We Do Not Take the Phylogeny into Consideration?*

Some reviewers of this paper felt that the message was "rather nihilistic," and suggested that it would be much improved if I could present a simple and robust method that obviated the need to have an accurate knowledge of the phylogeny. I entirely sympathize, but do not have a method that solves the problem. The best we can do is perhaps to use pairs of close relatives as suggested above, although this discards at least half of the data. Comparative biologists may understandably feel frustrated upon being told that they need to know the phylogenies of their groups in great detail, when this is not something they had much interest in knowing. Nevertheless, efforts to cope with the effects of the phylogeny will have to be made. Phylogenies are fundamental to comparative biology; there is no doing it without taking them into account.

## SUMMARY

Comparative studies of the relationship between two phenotypes, or between a phenotype and an environment, are frequently carried out by invalid statistical methods. Most regression, correlation, and contingency table methods, including nonparametric methods, assume that the points are drawn independently from a common distribution. When species are taken from a branching phylogeny, they are manifestly nonindependent. Use of a statistical method that assumes independence will cause overstatement of the significance in hypothesis tests. Some illustrative examples of these phenomena have been given, and limitations of previous proposals of ways to correct for the nonindependence have been discussed.

A method of correcting for the phylogeny has been proposed. It requires that we know both the tree topology and the branch lengths, and that we be willing to

allow the characters to be modeled by Brownian motion on a linear scale. Given these conditions, the phylogeny specifies a set of contrasts among species, contrasts that are statistically independent and can be used in regression or correlation studies. The considerable barriers to making practical use of this technique have been discussed.

ACKNOWLEDGMENTS

The suggestion that pairs of closely related organisms could be used in a way that avoided the need to know the full phylogeny was made by a student during discussion following my seminar at the Department of Genetics, University College, London. Unfortunately, I have been unable to discover her name. I wish to thank Ray Huey, John Gittleman, Robert Martin, and Mart Ridley for helpful discussions and/or access to their unpublished work. I am particularly grateful to Paul Harvey for bringing this particular piece of biological real estate to my attention and for many helpful conversations about it. I also wish to thank the reviewers of this paper for many helpful suggestions and for saving me from myself on at least one point. This work was supported by Task Agreement no. DE-AT06-76EV71005 of contract number DE-AM06-76RL02225 between the U.S. Department of Energy and the University of Washington.

LITERATURE CITED

Armstrong, E. 1983. Relative brain size and metabolism in mammals. Science 220:1302–1304.

Baker, R. R., and Parker, G. A. 1979. The evolution of bird colouration. Philos. Trans. R. Soc. Ser. *B* 287:63–130.

Clutton-Brock, T. H., and P. H. Harvey. 1977. Primate ecology and social organization. J. Zool. 183:1–39.

Damuth, J. 1981. Population density and body size in mammals. Nature 290:699–700.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492.

———. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

———. 1981*a*. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.

———. 1981*b*. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution 35:1229–1242.

———. 1983. Parsimony in systematics: biological and statistical issues. Annu. Rev. Ecol. Syst. 14:313–333.

Gittleman, J. 1981. The phylogeny of parental care in fishes. Anim. Behav. 29:936–941.

Givnish, T. J. 1982. Outcrossing versus ecological constraints in the evolution of dioecy. Am. Nat. 119:849–851.

Harvey, P. H., and G. Mace. 1982. Comparisons between taxa and adaptive trends: problems of methodology. Pages 343–361 *in* Current problems in sociobiology. King's College Sociobiology Group, ed. Cambridge University Press, Cambridge.

Martin, R. D. 1981. Relative brain size and basal metabolic rate in terrestrial vertebrates. Nature 293:57–60.

Pilbeam, D., and S. J. Gould. 1974. Size and scaling in human evolution. Science 186:892–901.

Ridley, M. 1983. The explanation of organic diversity. The comparative method and adaptations of mating. Clarendon, Oxford.

Sherman, P. W. 1979. Insect chromosome numbers and eusociality. Am. Nat. 113:925–935.