Chapter 3

Quantitative characters, phylogenies and morphometrics

Joseph Felsenstein

Department of Genetics
University of Washington
Box 357360, Seattle, WA 98195-7360

ABSTRACT

In spite of its title, the main subject of this paper will be to consider the use of quantitative characters in inference of phylogenies. Morphometrics can be viewed as a set of methods for extracting measurable traits from shapes. We come to morphometrics at the end, after first reviewing the way in which the resulting traits might be used. The great merit of morphometrics is that it automates the extraction of numerical measures from shapes, and thus presents evolutionary biologists with a torrent of quantitative characters, bringing the issues of how to treat them to the fore. In this article, I will use the term 'character' to refer to a feature of an organism, one that may assume a variety of 'states' or numerical values. In effect, a character is a column of the species x characters data matrix. Phylogenetic systematists often use the term 'character' differently: to refer only to the derived (apomorphic) states.

There have been represented at this symposium three main positions for how, and whether, quantitative characters may be used in inferring phylogenies:

Position 1 - That they cannot be used. This view was represented in this symposium, though not in this volume. It holds that if the states of the character are not inherently discrete, they are too problematic to use to infer phylogenies. References

to papers taking this position will be found in the paper in this volume by Humphries (2001).

Position 2 - That they can be used, but only after being coded into discrete states by an appropriate method. Swiderski et. al. (2001) exemplifies this view. Given this position, the solution to the 'character coding problem' becomes central to any use of quantitative characters.

Position 3 - That they can used, without necessarily being transformed into discrete characters first. Quantitative statistical methods should be employed. This review will take this view, with some important exceptions. As we will see, this view is not without its difficulties.

Before phylogenetic systematics became widespread, quantitative characters were often used by systematists. Frequently such characters were first reduced to discrete states such as 'long' and 'short'. Their use was not placed in any statistical context. It should be self-evident that valid information was extracted in this way, as the phylogenies of the last 100 years have held up remarkably well. What could not be done when quantitative characters were used in this way was to place any statistical interpretation on the results. One could infer that one phylogeny was better than another, but better by how much was not obvious. Twelve years ago I reviewed many of these same issues (Felsenstein 1988). My conclusions have not changed substantially since, but, as they have not been accepted by most morphological systematists, insistent and peevish repetition is in order. I ended that review doubting whether systematists will typically have the information necessary to use quantitative characters in a statistical treatment of phylogenetic inference. It thus seemed likely that molecular sequences would bear the brunt of such inference. But given an inferred phylogeny, we could then make statistical inferences about the evolution of quantitative characters.

In the years since that review, the use of quantitative comparative methods has become widespread. Statistical treatment of quantitative characters has made few inroads in the inference of phylogenies, but phylogenies have popularized statistical inferences about the evolution of the characters. Phylogenies and quantitative characters are getting together, though with the conversation going more one way than the other.

BROWNIAN MOTION AND CHARACTER CORRELATION

Attempts to model statistically the inference of phylogenies from quantitative characters have taken the Brownian motion model as their base. This was introduced as a model of gene frequency change by Edwards and Cavalli-Sforza (1964) in their pathbreaking paper on statistical inference of phylogenies. I applied it to quantitative characters (Felsenstein 1973). Lande (1976) also used a Brownian motion model for character change in his work on long-term evolution.

Brownian motion has an expected mean change of zero, and a variance of change that increases linearly with time. At the level of population genetics, the variability may arise from two sources: genetic drift or variable natural selection.

Brownian motion, drift and selection

A quantitative trait that has genetic variation controlled by a single locus will change as the gene frequencies at the locus undergo genetic drift. This process may be approximated by Brownian motion model. The approximation is imperfect, as the amount of change generated by Brownian motion is constant everywhere on the scale, while the amount generated by genetic drift becomes smaller as alleles near fixation. If the trait has additive genetic variance V_A, the variance of change due to genetic drift is V_A / N_e per generation. Interestingly, this relationship

for one locus can be extended to a trait that is the sum of effects from *n* loci with the same result. Thus Brownian motion is a reasonable approximation to change of a quantitative character by genetic drift, provided that V_A remains approximately constant. Quantitative genetic models of change in selectively neutral alleles by genetic drift have been introduced by Chakraborty and Nei (1982) and Lynch and Hill (1986). In these models the additive genetic variance is depleted by fixation, but continually replenished by new neutral mutations.

A second source of change of varying direction is natural selection. In a simple model of natural selection the change of gene frequency is

$$\Delta p \cong sp(1-p)\,, \qquad 3.1$$

If in different generations the selection coefficient *s* varies, including variation in its sign, the result can be a random walk that is difficult to distinguish from genetic drift. Cavalli-Sforza and Edwards (1967) suggested that varying selection at a single locus could be approximated by Brownian motion. I have (Felsenstein 1973, 1981) extended this to quantitative characters controlled by multiple loci, and argued that varying selection might be an important source of stochastic change in quantitative characters, particularly when neutrality is unlikely.

Response to selection

One of the central formulas of quantitative genetics gives the expected selection response as the product of the heritability (h^2) and the selection differential:

$$R = h^2 S\,, \qquad 3.2$$

The selection differential is the difference in mean phenotype between the selected parents and the population from which they were drawn. For natural selection, Lande (1981) has given a version of this formula in which the expected response is the product of the additive genetic variance and the slope of the gradient of log fitness:

$$R = V_A \frac{\partial \log \bar{w}}{\partial \bar{x}}, \qquad 3.3$$

The gradient term is simply the derivative of the logarithm of mean fitness, the derivative being taken with respect to the mean phenotype. The expressions above give the expected selection response. The actual selection response will also have a term from genetic drift added to this, a term whose expectation is zero.

Character correlation

These formulas are for the case of a single character. In morphological analysis we will be much concerned with character correlation, and want to know how to treat multiple characters. There are versions of these formulas for multiple characters, with matrices replacing these scalar quantities. For example, in the analogue to Lande's formulation, the vector of change in p characters $\Delta\mathbf{z}$ is the product of a p x p matrix of genetic covariances (**A**) and a p-dimensional vector **b** of the gradient of log fitness with respect to the means of all p characters (Lande 1981), plus a vector of terms for genetic drift (**e**):

$$\Delta\mathbf{z} = \mathbf{A}(\mathbf{b} + \mathbf{e}). \qquad 3.4$$

Taking expectations over generations in a lineage we can compute the covariance of changes in the different characters through time. We will assume for simplicity that the expectation of **b** is

zero, and we can make use of the fact that the genetic drift changes **e** have expectation zero and are uncorrelated with each other and with the changes in selection gradient. The expectation of the covariances of changes of characters over time is

$$\mathrm{E}\left[\Delta\mathbf{z}(\Delta\mathbf{z})^{\mathrm{T}}\right] = \mathbf{A}(\mathrm{E}[\mathbf{b}\mathbf{b}^{T}] + b\,\mathbf{I})\mathbf{A}^{T}. \qquad 3.5$$

The constant can easily be shown to be the inverse of the effective population size

$$b = \frac{1}{N_e}. \qquad 3.6$$

The term $\mathrm{E}[\mathbf{b}\mathbf{b}^{T}]$ is the covariance, across time, of the gradient of log fitness. We will call it **B**. Then

$$\mathrm{E}\left[\Delta\mathbf{z}(\Delta\mathbf{z})^{\mathrm{T}}\right] = \mathbf{A}(\mathbf{B} + b\,\mathbf{I})\mathbf{A}^{T}, \qquad 3.7$$

The covariances between characters thus come from three sources: genetic drift (b), additive genetic covariances (**A**), and the covariances of the selective pressures (**B**). This last source of covariation will be the least familiar. Nevertheless, it is not new. Stebbins (1950) discussed *selective correlation*, a term that came from Tedin (1925). Even if characters have no genetic covariance, their change along a phylogenetic lineage can covary owing to the covariance of the selection pressures on them. Imagine a set of species, some of which enter arctic habitats. Suppose that there is no genetic covariance among body size, relative limb length, and darkness of coloration in a mammal. In accordance with Bergmann's, Allen's, and Glogler's rules, natural selection may favor larger body size, smaller relative limb length, and darker coloration in arctic environments (as in Figure 3.1). Thus these characters will be expected to change in a

correlated manner: in the absence of genetic covariance, there would be a selective covariance in their changes. In the above equations, this is given by the covariance matrix **B**, which can create covariances even when the genetic covariance matrix **A** is a diagonal matrix.

The problem of estimation

Note that if we were to find a transformation that removed all additive genetic covariances, we would not remove all covariances between characters as long as there were also selective covariances. In order to make a statistical estimate of the phylogeny, we need to find a transformation that will remove the covariances of evolutionary change. We could then use the Brownian motion model to infer phylogenies. The difficulty lies in inferring the selective covariances. We can imagine doing, though perhaps with great effort, a quantitative genetic experiment to infer the additive genetic covariances in one or more species. We can hope that these additive genetic covariances stay roughly constant over a large enough span of time that we can use the results. But where are we to get an estimate of the selective covariances?

There are two possible sources:

- We may have paleontological data that follow a lineage through time, and enable us to infer the covariances of a set of characters through time. This does not give us a direct estimate of the selective covariances, but it does estimate the covariances of evolutionary change. If we also have an estimate of the additive genetic covariances, we can use Equation (3.7) to infer the selective covariances. Even if we do not have an estimate of the additive genetic covariances available, we at least then have an estimate of the covariances of evolutionary change, which is what we need to transform the characters to independence so that we can use the Brownian motion model.

- We can use molecular data to infer the phylogeny, and then observe the covariances of evolutionary change along that phylogeny. This is not done directly, as we cannot see the phenotypes of hypothetical ancestors. Instead we can use phylogenetic comparative

methods, which use the distribution of multiple characters on the tips of a known phylogeny to infer the covariances of evolutionary change (Felsenstein 1985; Harvey and Pagel 1991). Again, this does not give us the selective covariances directly.

DILEMMAS AND OPPORTUNITIES

Fossil and neontological data

The use of the comparative method (item 2 above) may seem beside the point: the objective is to infer the phylogeny, and we are assuming that we already have the phylogeny! But there are cases where we can make useful inferences. In particular, suppose that we have a group with both paleontological and neontological data. From the present-day species we infer a molecular phylogeny, and then use phylogenetic comparative methods to infer the covariances of evolutionary change of the quantitative characters. We then transform the characters to independence using those covariances. These new characters can be computed in both the living species and the fossils in the phylogeny. For each possible placement of the fossil species, the likelihood of the tree for the quantitative character data can be computed. The placement which maximizes this likelihood is to be preferred. This is in effect a Total Evidence approach (likelihood version), because the placement of the fossil species does not affect the likelihood of the tree on the molecular data. Taken together, the placement of all species by this method would maximize the overall likelihood, if we compute the overall likelihood as the product of the likelihoods of the molecular tree and the morphological tree.

This process is illustrated in Figure 3.2. In fact, only the part of the figure shown in bold lines is necessary, as hinted at by the double-headed arrow between the overall phylogeny and the covariances. The two adjust to each other in light of all data. The lighter lines in the diagram show steps that may be useful to make preliminary estimates.

This approach can also be useful when we have two groups of present-day organisms, and have a molecular data set for one of them. If we are willing to assume that the morphological characters had the same covariances of evolutionary change in both groups, we could infer the phylogeny from the molecular data in one group, infer the character covariances in that group, and then use those covariances to infer the phylogeny in the other group. This too can be seen as a Total Evidence approach (likelihood version). Sometimes we may want to apply this method when there are not two distinct groups, but instead where there is only a phylogeny for some of the species in the group. If we had a phylogeny from which some species were omitted, it could be used to infer character covariances. Then the missing species could be placed from their morphological characters.

Do we need molecules?

In the preceding argument, molecular inferences provided information about part or all of the phylogeny. That information was needed to obtain the covariances needed to make use of the quantitative characters. One can have serious doubts as to whether quantitative characters could be used in the absence of molecular data. This would at first sight seem to back Position I—that quantitative characters cannot be used in the inference of phylogenies. But it does differ from that position in one important respect. Adherents of Position I typically deny that statistical inference approaches using quantitative character data are possible. I am concerned about circularity in the inference—it may not be possible to infer both the phylogeny and the character covariances. But given that independent information is available about the phylogeny, one can use comparative methods to infer the covariances. If we have both we can use them, together with the morphological characters, to infer both the phylogenies and the covariances. The morphological characters together with their statistical model will have an effect, however small, on the phylogeny.

This is a statistical analysis. As always, it is subject to worries about the correctness of the model. But if our interest is in the evolution of these particular characters (rather than in the phylogeny itself), this position is closer to Position III than to Position I. In many cases the quantitative characters are collected because they are of intrinsic interest to the biologist, rather than simply as markers for inferring the phylogeny. As molecular data become easier to obtain, they tend to displace quantitative character data from the job of inferring phylogenies, so that more and more of the use of quantitative characters will be motivated by interest in the evolution of those characters. There will be less and less use of quantitative characters as arbitrary markers for inferring phylogeny.

Allowing for uncertainty

Of course, molecular data do not provide us with a precise picture of the phylogeny. The issue arises as to how to incorporate into the analysis the uncertainty about the phylogeny. There seem to be two ways of doing this. The harder (but slightly superior) way (Felsenstein 1985) would be to combine the probabilistic model of change of the molecules with the Brownian motion model of the quantitative characters, allowing for the covariances of the latter. One would could then compute a likelihood for all of the data, given both a tree and the covariances of evolutionary change of the quantitative characters. The collection of trees and covariances that were supported by the data would be those that had the highest likelihoods. If these did not have trees of different topologies, we could use asymptotic theory to choose the contour of the log-likelihood surface that defined the confidence interval—if there were n species and p characters it would be the 95% value of a χ^2 distribution with $2n-3+p(p+1)/2-1$ degrees of freedom. This is the number of quantities (branch lengths and covariances less one for a scaling between them that is confounded) being estimated. The combination of tree and covariances

that are acceptable can be based on the contours of the joint likelihood curve for the covariances and the tree. For an oversimplified picture see Figure 3.3. The actual tree is not a single variable, and the character covariances are also multidimensional. This approach would seem to resolve the question of whether there is some circularity involved in using the same characters to determine the tree as are used in inferring covariances in character evolution.

We could imagine using the method to infer just the character covariances. In that case the confidence interval on the covariances would be defined by the degrees of freedom restricted by defining the covariances (in this case, $p(p+1)/2$). The set of trees and covariances that lies within the likelihood contour for one-half the significant value of a χ^2 variate with that number of degrees of freedom would be found, and then the trees ignored, leaving the set of covariances. Similarly a confidence interval on the tree could be inferred by doing this and ignoring the resulting covariances, using $2n-3$ as the degrees of freedom. More specific hypotheses about the character covariances (such as that a covariance between two particular characters is zero) could be tested with even fewer degrees of freedom and consequently a tighter confidence interval. However at the moment none of this can be done, simply because present-day software is not designed for this task.

The other, and simpler, method is to estimate the tree solely from the molecular data. This gives us a slightly less precise estimate of the tree. However it is quite easy to allow for the uncertainty of the tree in inferring the covariances. I have pointed out (Felsenstein 1988) that for this one can use bootstrap sampling of the molecular sequences. For each bootstrap sample, one would infer the tree, and then use that tree to estimate the covariances of the quantitative characters. The resulting collection of estimates of the covariances would properly reflect the uncertainty about the tree. As the quantitative character data are derived from samples of

individuals in each species, one could add another level of bootstrapping, resampling individuals within species each time. This would be unnecessary if the within-species covariances were allowed for in inferring the phylogenetic covariances (Lynch 1991). Current versions of PHYLIP allow the bootstrapping of the molecular data to be carried out and the bootstrap sample estimates of the trees to be used to make multiple estimates of the covariances. Version 3.6 of PHYLIP will also allow for within-species components of variance (Lynch 1991) in inferring the covariances.

A way out?

One might wonder why we need to bother with the molecular data at all. Why not infer both the tree and the covariances from the same data set? One immediately wonders whether any such effort is totally circular. Interestingly, there is only a partial circularity, though it may be circular enough to make the whole effort mostly an academic exercise. We can get a good picture of this problem simply by counting degrees of freedom.

If there are n species and p characters, the data set has a total of np degrees of freedom. Of these, p are lost when we discard the means of the characters, leaving us with $p(n-1)$. There are $p(p+1)/2$ quantities to infer in the covariance matrix (the variances and the covariances). In the tree there are $2n-3$ branch lengths. However, we cannot use these quantities without taking into account that two of them are redundant. In particular, the total length of the tree is confounded with one of the parameters of the covariance matrix. If we double the length of the tree and halve all of the covariances, we leave the likelihood unchanged, since this leaves the covariances of the data unchanged. So we must remove one of the degrees of freedom.

This leaves us with a total of

$$p(n-1) - p(p+1)/2 - (2n-3) + 1 \quad = \quad np - 2n - \frac{1}{2}p^2 - \frac{3}{2}p + 4 \qquad 3.8$$

degrees of freedom. Simultaneous inference of the tree and the covariance matrix will be possible when the this number is positive. We can separate the terms in *n* and *p* to get a condition for simultaneous inference (assuming $p > 2$):

$$n \quad > \quad (\frac{1}{2}p^2 + \frac{3}{2}p - 4) \Big/ (p-2). \qquad 3.9$$

Table 3.1 shows the upper limit of the number of characters that satisfies this condition, for some values of *n*: Below 6 species, there is no whole number of characters that satisfies the conditions. As the number of species rises, the lower limit on characters is just above 2, and the upper limit can be shown to remain just below *2n-5*. One might wonder whether this is worth the effort. Given this upper limit on the number of characters, the inference of the tree cannot be made precise by increasing the number of characters without limit (I am indebted to Andrew Rambaut and Michael Charleston for pointing this out to me). On the other hand, one can make the inference of the covariances more and more precise by increasing the number of species sampled. This holds out some hope for the analysis of characters, but not much for the inference of the phylogeny. Even if we are willing to concentrate on the characters instead of the phylogeny, there is a limit to how many species we can find in the relevant group—it may be far easier to find new characters than new species.

With three species, there is no possibility of inferring both the phylogeny and the character covariances. It was this case that persuaded me (Felsenstein 1988) that the two were inextricably confounded and that any attempt to infer them separately was hopeless. As we can

see, this was not entirely true. They can be separated in principle, but the prospects for making practical use of this are not very encouraging.

Genomics to the rescue?

Ahead lies the terra incognita of genomics. Though difficult and expensive now, it is clear that in a decade it will be relatively easy to do genomics on characters of interest. We could find the loci that make the largest contribution to genetic variation of the characters within populations and, if we can cross individuals from different populations, also find the quantitative trait loci (QTLs) that make the largest contributions to differences between populations, and perhaps differences between species.

To the extent that we can do this, we transform the data into QTL gene frequencies in different populations. However, we can find only the loci of largest effect, leaving behind a residuum of polygenic variation at 'background' loci. Thus, until that residuum becomes small enough to be insignificant, quantitative genetic models will be useful. The transition from polygenic models to models that have known loci will be gradual. In general, to detect a locus with half the effect, we must quadruple the sample size.

In some cases the inability to detect loci of small effect may not be a serious problem. If the divergence of the loci were due primarily to natural selection, most of that divergence would be reflected in the gene frequencies of the loci of largest effect. In simple forms of selection (e.g., directional selection), changes in gene frequencies are proportional to the sizes of the genetic effects at the loci. A locus whose genetic variants have twice the effect of those at another locus will thereby accumulate genetic differences that are twice as large. That in turn means that the phenotypic differences caused by those loci will be four times as great, since both the genetic

effects and the gene frequency differences are twice as great. There is thus some prospect that the availability of genomics will rapidly illuminate cases where the differences are caused by natural selection, by detecting loci of large effect, which may be responsible for most of the differences.

No one has yet thought through how we can use QTL data, possibly in combination with a polygenic model for residual genetic variation, to infer phylogenies and to illuminate character covariation. The time for doing so is approaching. As I have suggested elsewhere (Felsenstein, 2000), genomic data do promise insights on whether natural selection has acted on the characters under study or on unobserved characters correlated with them. Given the possibility of escaping some of the constraints that have plagued analysis of morphology, genomic data seem worth investigating.

CHASING PEAKS

We have modelled natural selection as acting in randomly varying directions in different lineages. It is not self-evident that natural selection will vary randomly in direction from moment to moment. A more convincing model would be natural selection towards an optimum (cf. Lande 1976). Some of the possible variants on this model would be:

- A single optimum stays in one place with all species attracted to it. The species wander by random genetic drift (Lande 1976; Hansen and Martins 1996).

- Different species have different optima, the optima separating at the time of speciation. Each optimum wanders independently in the space, perhaps by Brownian motion. (Felsenstein 1988; Hansen and Martins 1996).

- Different species have different optima, the optima separating at speciation. The optima wander, but their positions are constrained so that they describe an Ornstein-Uhlenbeck process (random walk of an elastically bound particle) around a common point (Felsenstein 1988).

- Perhaps more realistically, each species has a different optimum, the optima separating at speciation, but optima of recently diverged species wander in a correlated fashion, the correlation declining the longer they are diverged.

A full treatment of the movement of a quantitative character under any of these models is difficult, but it is greatly simplified the longer a population remains under the influence of a peak. It is not hard to show that if a population is following a peak which is itself undergoing Brownian motion or the Ornstein-Uhlenbeck process, its distance from the peak settles down into a normal distribution with constant variance. In effect the population mean is towed along by the peak, but at the end of a somewhat flexible cable. The farther the peak wanders the more of the change of the character must be attributed to the movement of the peak and the less of it is accounted for by the cable.

If selection moves the population (say) 10% of the way toward the peak each generation, then the departure of the population from the peak will represent, events that have occurred in roughly the last 1/0.10=10 generations. If each lineage lasts much longer than that, and if genetic drift during the 10 generations is much smaller than the net movement of the peak over its existence, then the mean of the quantitative character is basically going where the peak goes.

Figure 3.4 shows a numerical example from a computer simulation of two characters (not all details of which are described here). The two characters are negatively genetically correlated, with a correlation coefficient of -0.9. They wander by genetic drift about a peak which is itself

moving. In the leftmost panel little time has elapsed; the peak has not moved much and the two characters show the negative correlation that is a consequence of their genetic covariation. The peak wanders with positive correlation between the two characters. Changes in the position of the optimum in one character have a correlation of 0.9 with the changes in the other character. Thus the genetic covariation 'wants' the characters to be negatively correlated, while selective covariation wants them to be positively correlated. In the center panel of the figure, 10 times as much time has elapsed and there has been some wandering of the peaks. This smears out the distribution of character values from lower-left to upper-right, resulting in a roughly circular distribution. In the rightmost panel we see the distribution over 100 times as much time as in the first panel (10 times as much as the center panel). Now the peak movement is the dominant influence, and the characters show a strong positive correlation.

This is reason to expect that selective covariances will be important—the covariation of character change will then mostly be a matter of the covariation of peak movements with respect to different characters. For example, provided selection favors large size and also tends to favor dark coloration in the same lineages, then there will be a correlated distribution of these characters that will override any genetic correlation.

PUNCTUATIONAL MODELS

In the models discussed here, it has been assumed that quantitative characters change continually along a branch of the tree. Under a punctuated equilibrium model, they would instead be expected to change mostly at the time a branch originates, and be static thereafter. If there were a burst of change (of roughly equal size) at the start of each branch, and no change thereafter, we might think that this could be accommodated by having the expected variance accumulated in each branch be equal. The tree would then consist of a series of branches, each

of unit length. Hansen and Martins (1996) have made calculations along these lines (see also Felsenstein 1988). If this were all that we needed to take into account, it would be straightforward to analyze data under the assumption of punctuation (though there would be the issue of which branch at each fork was the newly-originated one).

The difficulty with this tempting model is that we do not see all branches. Even if we can collect all extant species, there should be many forks at which the new species has persisted while the parent species has died out. That would show up in our tree as a burst of change in the middle of a branch. Branches that had undergone more of these bursts of change would be longer, so that not all branches would be of unit length. In addition to species that have become extinct, we may be omitting some extant species from our data set. If there are 200 beetles in our group, but we analyze only a capriciously-chosen sample of 40, there will be many places where a fork gave rise to one of our sampled species, with the parent species being the ancestor of ones we have omitted. This will create additional uncertainties about the branch lengths on the tree.

In short, a punctuational model may be harder to distinguish from a gradualist model than first appears. There is hope for doing so if many characters are analyzed, as under the assumption of puctuated equilibrium the parent species should not change while the daughter species changes in many characters. But the analysis is complex, and needs much further examination.

THE CHARACTER CODING PROBLEM

Many analyses of quantitative characters first reduce them to discrete characters. This is known as the ‘character coding problem’, and a variety of methods have been suggested for recoding the characters. Sometimes this is done under the assumption that parsimony methods require

discrete states. Most parsimony programs do have such a requirement, though in the early years of the parsimony literature methods were put forward that use the original quantitative scale (Farris 1970).

We might also want to recode the quantitative characters into discrete states if we believed that the continuous scale masked regions that had widely varying properties. For example, if a character can rather easily wander between values 4 and 10, and can also wander easily between 1 and 3, but has great difficulty changing from a value of 3 to a value of 4, we might want to approximate this by having two discrete states, one consisting of all values below 3.5, the other of all values above 3.5. If the change between these two ranges is sufficiently improbable, we want to weight it heavily. We would be losing some information by not distinguishing between values of (say) 6 and 10, but we would be gaining some information by taking into account the greater difficulty of change in certain regions of the scale.

I believe that many of the character coding methods, such as gap coding (Mickevich and Johnson 1976; see also Simon 1983 and Archie 1985) are implicitly trying to take account of situations like this, using the empirical distribution of character values among species as an indication of where the regions of difficult change are located. There are complications owing to the fact that species are not drawn independently from a distribution, but arise on a phylogeny in a highly clustered fashion. Thus, a gap in the distribution along the character scale may reflect, not a region which is rarely occupied, but the distinction between two clades. There is in addition the question of why coding is taking place one character at a time, when evolution at different characters may be correlated. These issues have never been given the serious statistical examination they deserve.

Given that there are ways to analyze quantitative characters on quantitative scales, there is no compelling reason to engage in character coding. Until we have a well-thought-out method for detecting regions of scales that ought to be treated differently, perhaps the best advice about character coding is to just say no.

THE CHARACTER UNCODING PROBLEM

In fact, one may want to do the opposite. It is possible for discrete characters to mask an underlying continuous scale. The threshold model of evolution has been around since the work of Wright (1934) on digit number in guinea pigs. It has been applied to human genetics by Falconer (1965). This model imagines an invisible underlying character (usually called 'liability') and a threshold value. The discrete trait results from a developmental system that monitors whether the liability exceeds the threshold value. The liability has the usual quantitative genetics. This class of models has some attractive features. We may compare it to a simple alternative, a simple Markov chain that alternates between two states, 0 and 1 (cf. Pagel 1994). In the threshold model, once a lineage changes from being largely of state 0 to being largely of state 1, its underlying liability is probably near the threshold. The longer the time that a lineage has remained in state 1, the farther the liability may have wandered beyond the threshold, and the less likely an immediate return to state 0. The simple Markov process model, by comparison, has the same probability of returning to state 0 however long it has been in state 1.

Figure 3.5 shows a depiction of the threshold model and a simulation of the change of a discrete character along a tree. Note that the threshold model has one other advantage over the simple Markov chain. It does not actually envisage a lineage changing instantaneously from one state to another. At any time, the lineage has both states present in it, their proportions depending on

where the threshold value lies in the distribution of the liability character. In many cases almost all of the phenotypes in the population will be the same, but as the mean of the liability crosses the threshold, there will be a period of polymorphism. This can be seen in the simulated tree.

The difficulty with the threshold model is its mathematical intractability. To compute the likelihood of a discrete character on a phylogeny, we would have to compute the probability that each individual lies above or below the threshold (depending on its observed phenotype). The probability density of the liabilities is a multivariate normal distribution, but the joint probability of the discrete phenotypes computes a corner of this distribution:

$$\begin{aligned} &\text{Prob}\,[1, 1, 0, 1, 1, 0, 0] \\ &\quad = \text{Prob}\,[x_1 > c,\ x_2 > c,\ x_3 < c,\ x_4 > c,\ x_5 > c,\ x_6 < c,\ x_7 < c] \qquad 3.10 \\ &\quad = \int_c^{\infty}\int_c^{\infty}\int_{-\infty}^{c}\int_c^{\infty}\int_c^{\infty}\int_{-\infty}^{c}\int_{-\infty}^{c} \text{Prob}\,[x_1, x_2, x_3, x_4, x_5, x_6]\, dx_1\, dx_2\, dx_3\, dx_4\, dx_5\, dx_6. \end{aligned}$$

Integrals of corners of normal distributions are hard to compute. It appears likely that they will yield only to Markov Chain Monte Carlo methods. These may make use of threshold models practical. This is in effect the ‘character uncoding problem’, and it seems more likely to be of interest than the character coding problem.

MORPHOMETRICS AT LAST

At the end, let us come full circle, back to morphometrics. Given all of this, where does it leave morphometrics? Morphometrics is a source of numeric characters. Morphometricians point out that it is much more than just another source of them, that it places individuals in a morphometric space that has particular desirable properties. Other numerical methods may choose coordinates that lead to absurd results when one extrapolates, or lead to misleading

covariation when there is measurement error. For the present discussion these distinctions are not important—we could as well be discussing any source of numeric characters.

Of the three positions on the use of quantitative characters in inferring phylogenies, Position I (that they cannot be used) would certainly lead to a lack of interest in using morphometrics. It might possibly be argued that this does not preclude the use of morphometrics retrospectively, using phylogenies to analyze the change of morphometric parameters. However that would require us to accept some model of change of these quantitative characters. If there were such a model possible, one could think of using it to infer phylogenies. Most practitioners of Position I do not believe that any such model is worth serious consideration.

Position II—that we can use quantitative characters only if discretely coded—leads to an interest in deriving discrete characters from morphometric parameters. Zelditch *et al.* (1995) have developed methods for doing so, and this has led to some controversy (for debate and earlier references see Rohlf 1998 and Zelditch *et al.* 1998). It becomes important to have the correct coding and the character coding problem becomes paramount.

Position III requires that we not only be able to derive numerical measurements from morphometric data, but that we ask about their genetic and selective correlations. Most of the morphometric literature has asked what parameterizations are best justified on geometric or mathematical grounds. Genetic correlations include developmental correlations. Asking about them should lead us toward a genetic and developmental morphometrics rather than a geometric morphometrics (Felsenstein 1992). As long as we do not have developmental models, we cannot construct developmental morphometrics from them. When they become available, they will lead to insights into the expected genetic correlations of morphometric parameters.

In inferring developmental models, we may be able to take the reverse route. Morphometric analyses along phylogenies may lead to insights into the genetic correlations, and thus may be a major source of insight into developmental models. The 'evo-devo' literature has yet to mine this lode. To do so will require the quantitative models of morphometrics, but also require us to relinquish a purely geometric approach.

The position taken in this essay has elements of similarity to Position III, but also to Position I. It argues that we typically do not have evidence as to the selective correlations, and often not for the genetic correlations either. Thus, most use of quantitative characters will be retrospective. However when this is possible, and when genetic correlations or developmental models are available, it should allow us to make interesting inferences about the selection pressures. We will then be making progress toward a functional morphometrics, even an ecological or behavioral morphometrics. If the genetical and/or developmental models are known, and the phylogenetic distribution also, we could make inferences about how selection is acting on the characters. Alternatively, if ecological information about selection is available, and also phylogenetic distribution, we might hope to infer genetic correlations and discriminate among developmental models. We can hope that the era of geometric morphometrics will be followed by an era in which developmental morphometrics exists in dynamic interaction with functional morphometrics, the interaction being mediated by modelling change of quantitative characters across phylogenies.

ACKNOWLEDGMENTS

This paper has been supported by NSF grants DEB-9815650 and BIR-9527687 with additional support from NIH grants R01 GM51929 and R01 HG01989. I am grateful to Michael

Charleston and Andrew Rambaut for pointing out the lack of consistency of tree estimates when done jointly with inference of character covariances, and I am indebted to W. Scott Armbruster for introducing me to G. L. Stebbins' use of the term selective correlation.

REFERENCES

Archie, J. W. (1985) Methods for coding variable morphological features for numerical taxonomic analysis. *Systematic Zoology,* **34**, 326–345.

Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967) Phylogenetic analysis: models and estimation procedures. *Evolution,* **32**, 550–570.

Chakraborty, R. and Nei, M. (1982) Genetic differentiation of quantitative characters under optimizing selection, mutation, and drift. *Genetical Research,* **39**, 303–314.

Edwards, A. W. F. and Cavalli-Sforza, L. L. (1964) Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification,* (eds V. H. Heywood and J. McNeill), London, Systematics Association Publication No. 6, pp. 67–76.

Edwards, A. W. F. (1970) Estimation of the branch points of a branching diffusion process. *Journal of the Royal Statistical Society, Series B,* **32**, 155–174.

Falconer, D. S. (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics,* **29**, 51–76.

Farris, J. S. (1970) Methods for computing Wagner trees. *Systematic Zoology,* **19**, 83–92.

Felsenstein, J. (1973) Maximum likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics*, **25**, 471–492.

Felsenstein, J. (1981) Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution,* **35**, 1229–1242.

Felsenstein, J. (1985) Phylogenies and the comparative method. *American Naturalist,* **125**, 1–15.

Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, **19**, 445–471.

Felsenstein, J. (1992) Review of Proceedings of the Michigan Morphometrics Workshop , *Quarterly Review of Biology*, **67**, 418–419.

Felsenstein, J. (2000) From population genetics to evolutionary genetics: a view through the trees. In *Evolutionary Genetics: From Molecules to Morphology* (eds R. S. Singh and C. B. Krimbas), Cambridge, Cambridge University Press, pp. 609–627.

Hansen, T. F. and Martins, E. P. (1996) Translating between microevolutionary process and macroevolutionary patterns: A general model of the correlation structure of interspecific data. *Evolution*, **50**, 1404–1417.

Harvey, P. H. and Pagel, M. D. (1991) *The comparative method in evolutionary biology*. Oxford: Oxford University Press.

Humphries, C. J. 2001. Homology, characters and continuous variables. In Morphometrics, Shape and Phylogeny (eds N. MacLeod and F. Forey), London, Taylor and Evans, pp. xxx-xxx

Lande, R. (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution*, **30**. 314–334.

Lande, R. (1981) Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution*, **33**, 402–416.

Lynch, M. and Hill, W. G. (1986) Phenotypic evolution by neutral mutation. *Evolution*, **40**, 915–935.

Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution*, **45**, 1065–1080.

Mickevich, M. F. and Johnson, M. S. (1976) Congruence between morphological and allozyme data in evolutionary inference. *Systematic Zoology*, **25**, 260–270.

Pagel, M. (1994) Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B Biological Sciences,* **255**, 37–45.

Rohlf, F. J. (1998) On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, **47**, 147–158.

Simon, C. M. (1983) A new coding procedure for morphometric data with an example from periodical cicada wings. In *Numerical Taxonomy* (ed. J. Felsenstein), New York, Springer-Verlag NATO Advanced Science Institutes Series G, No. 1, pp. 378–382

Stebbins, G. L. (1950) *Variation and Evolution in Plants.* New York, Columbia University Press.

Swiderski, D. L., Zelditch, M. L. and Fink, W. L. 2001. Comparability, morphometrics and phylogenetic systematics. In Morphometrics, Shape and Phylogeny (eds N. MacLeod and F. Forey), London, Taylor and Evans, pp. xxx-xxx

Tedin, O. (1925) Vererbung, Variation und Systematik in der Gattung Camelina. *Hereditas*, **6**, 275–386.

Wright, S. (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics*, **19**, 506–536.

Zelditch, M. L., Fink, W. L., and Swiderski, D. L. (1995) Morphometrics, homology, and phylogenetics - quantified characters as synapomorphies. *Systematic Biology,* **44**, 179–189.

Zelditch, M. L., Fink, W. L., Swiderski, D. L., and Lundrigan, B. L. (1998) On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology*, **47**, 159–167.

FIGURE CAPTIONS

Figure 3.1. An example of selective correlation. Mammalian lineages enter arctic environments, leading to correlated changes in body size, relative limb length, and coloration.

Figure 3.2. Flow chart showing the use of molecular phylogenies of present-day species to infer covariances of morphological characters, thereby allowing fossil data to be included.

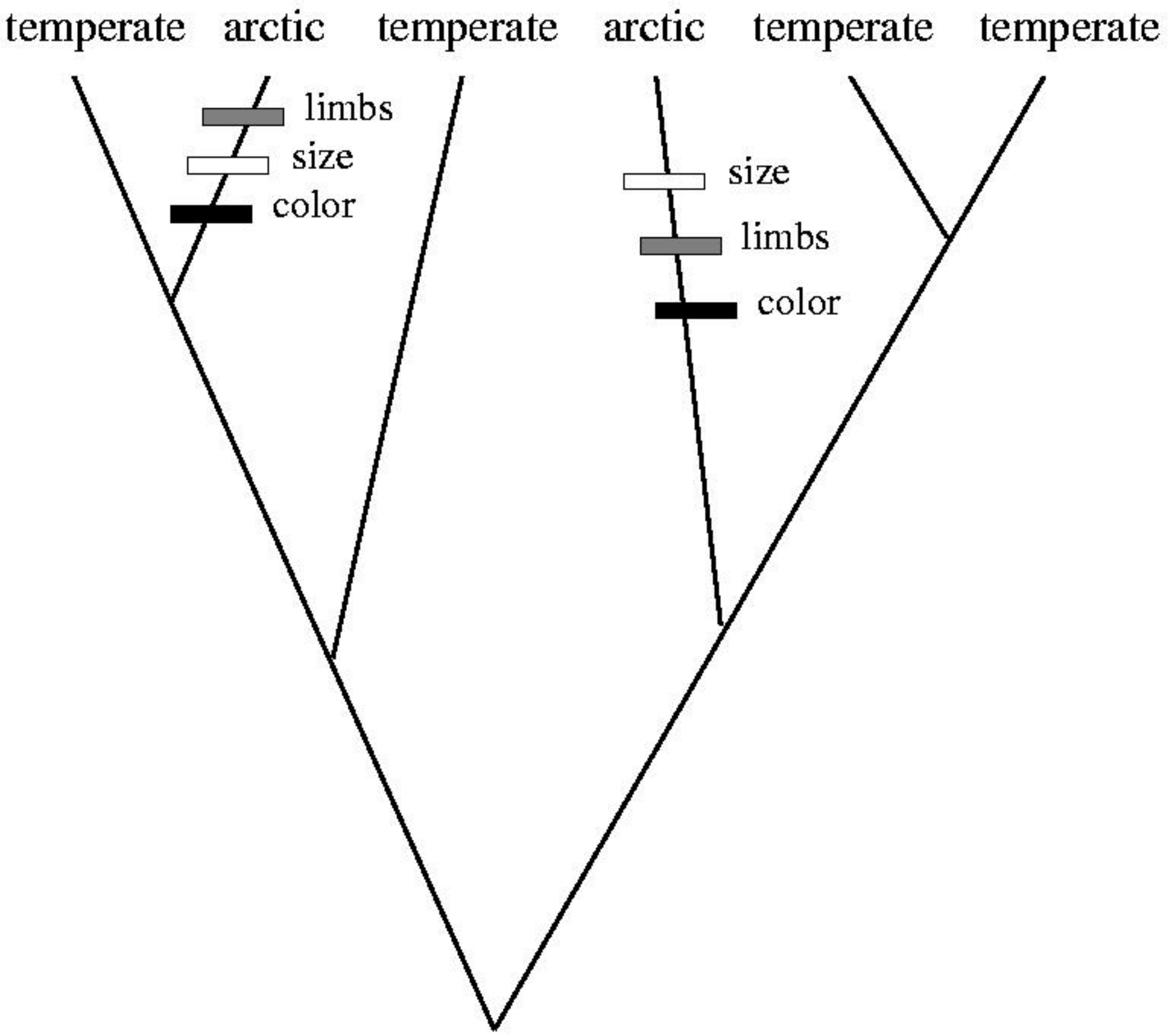
Figure 3.3. Simultaneous inference of the tree and the character correlations when probabilistic models for both molecular and morphological characters are available. Point estimates of the tree and a correlation are shown, and approximate likelihood-based confidence intervals for the individual parameters can be based on the profile likelihoods (contours with dark shading and two-headed arrows) and joint confidence intervals based on the contours of the full likelihood curve (lighter shading).

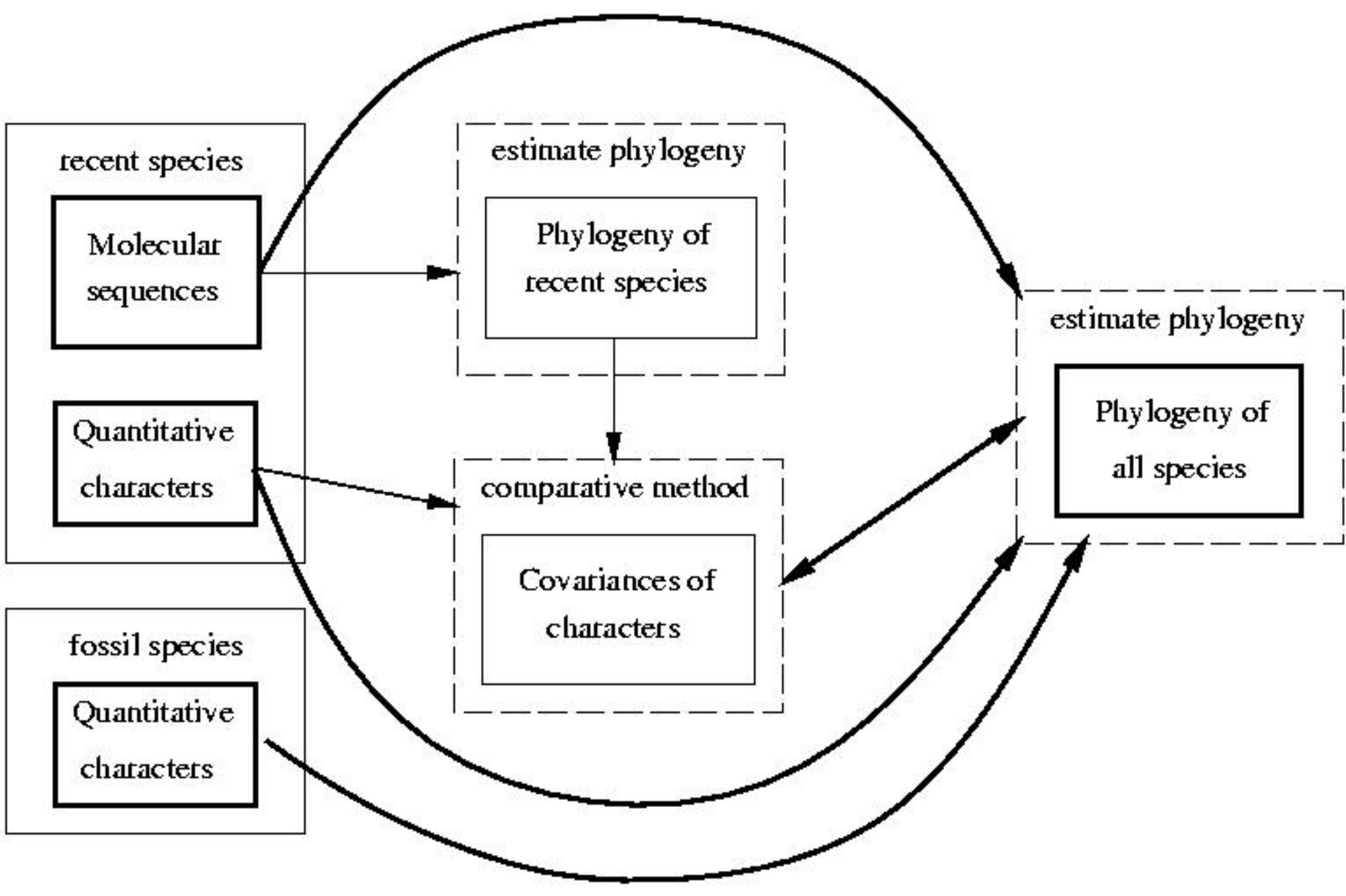
Figure 3.4. Covariation of two characters when genetic covariation between them is -0.9 but when they are attracted to optimum values that vary through time with a covariance of 0.9 in movements of the optima of the characters. For short periods of elapsed time the phenotypic covariation is negative, but as we wait 10 and 100 times longer, the movements of the optima tow the character values in positively correlated ways.

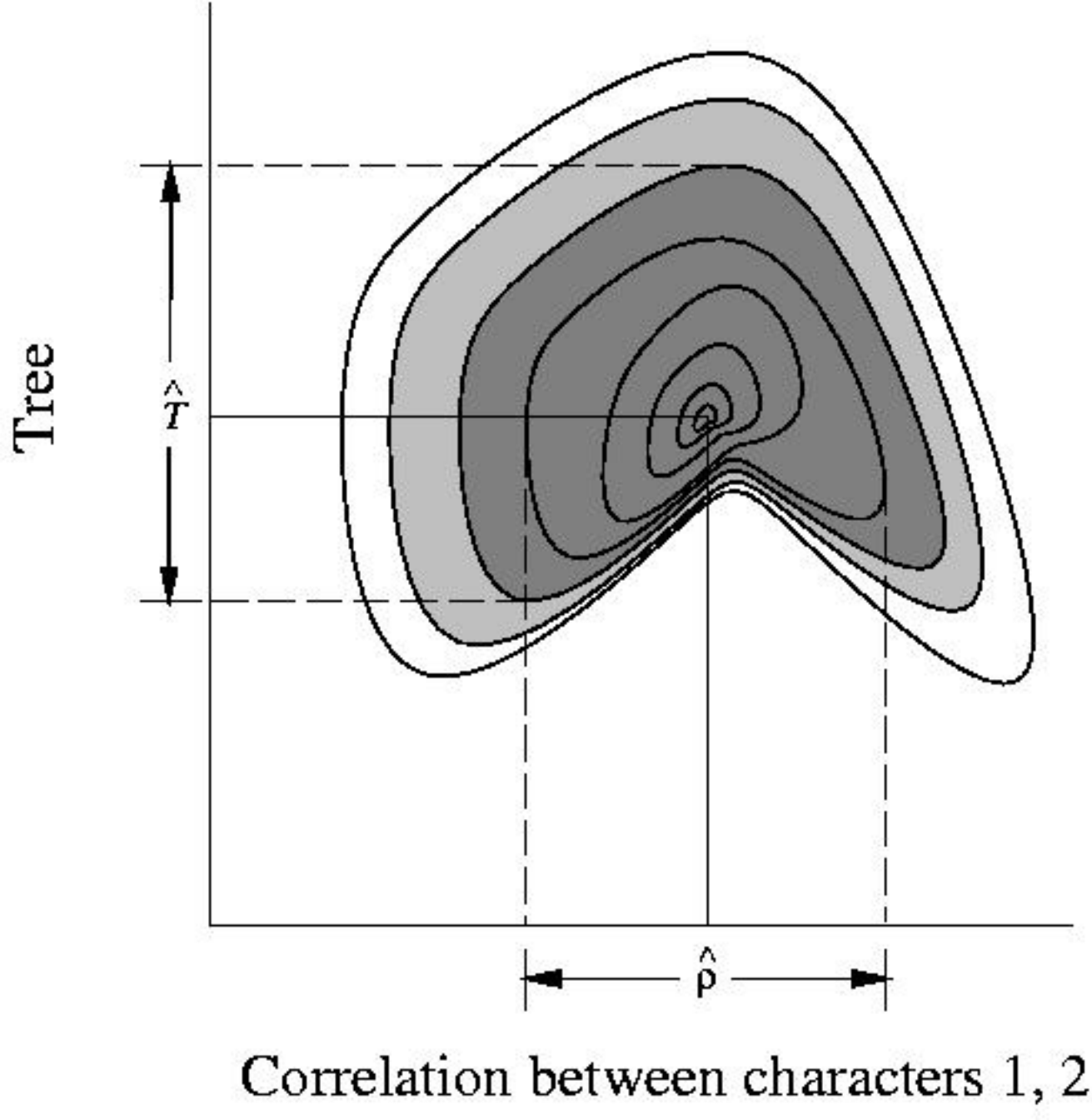
Figure 3.5. The threshold model, showing the role of the threshold and the underlying (unobserved) liability character, and the result of the simulation of the change of a threshold character along a simply phylogeny. The value of the underlying liability character is shown next to each node in the tree, and the shading in each branch shows the proportion of that population which has state 1.

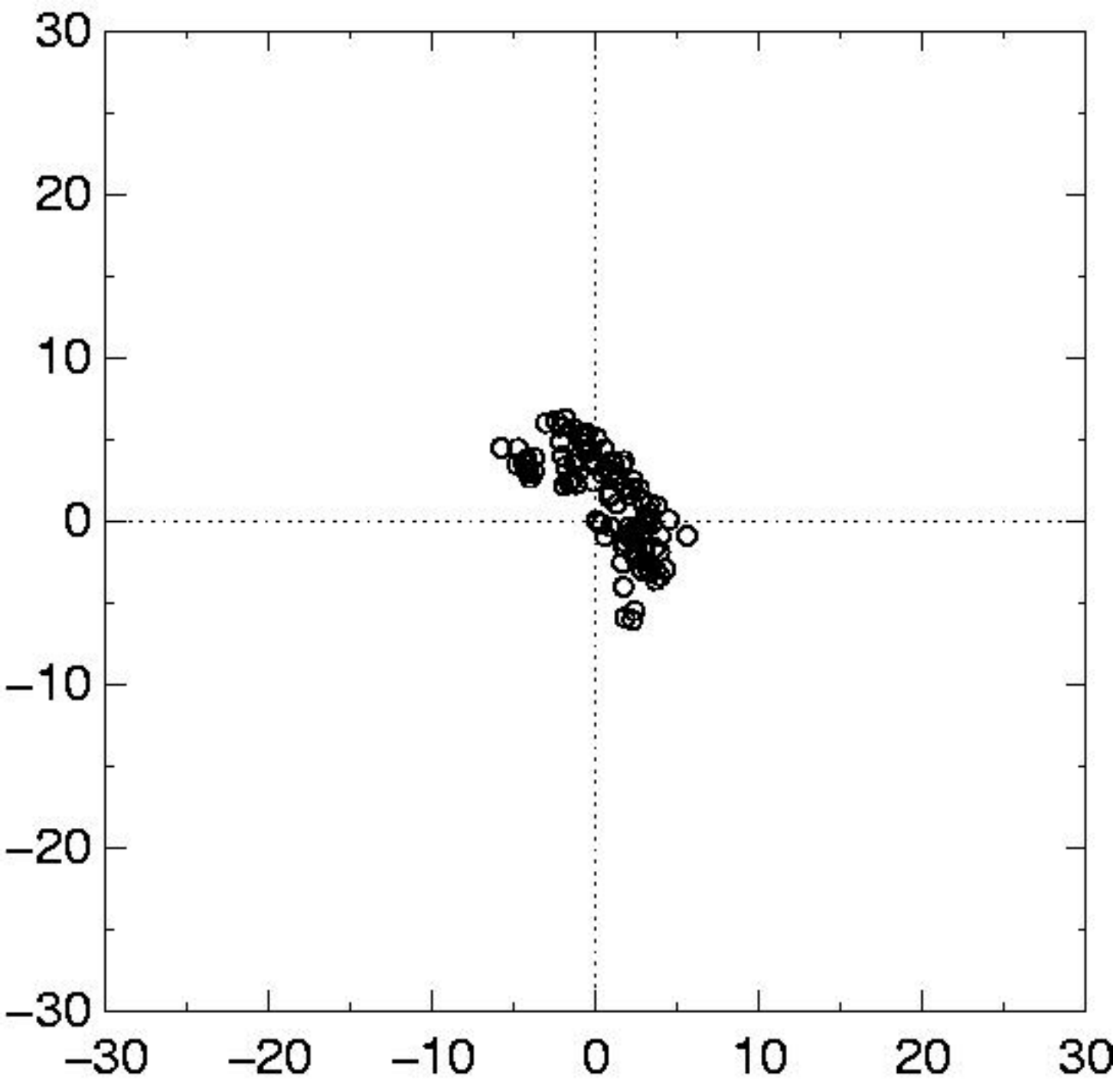
TABLE CAPTION

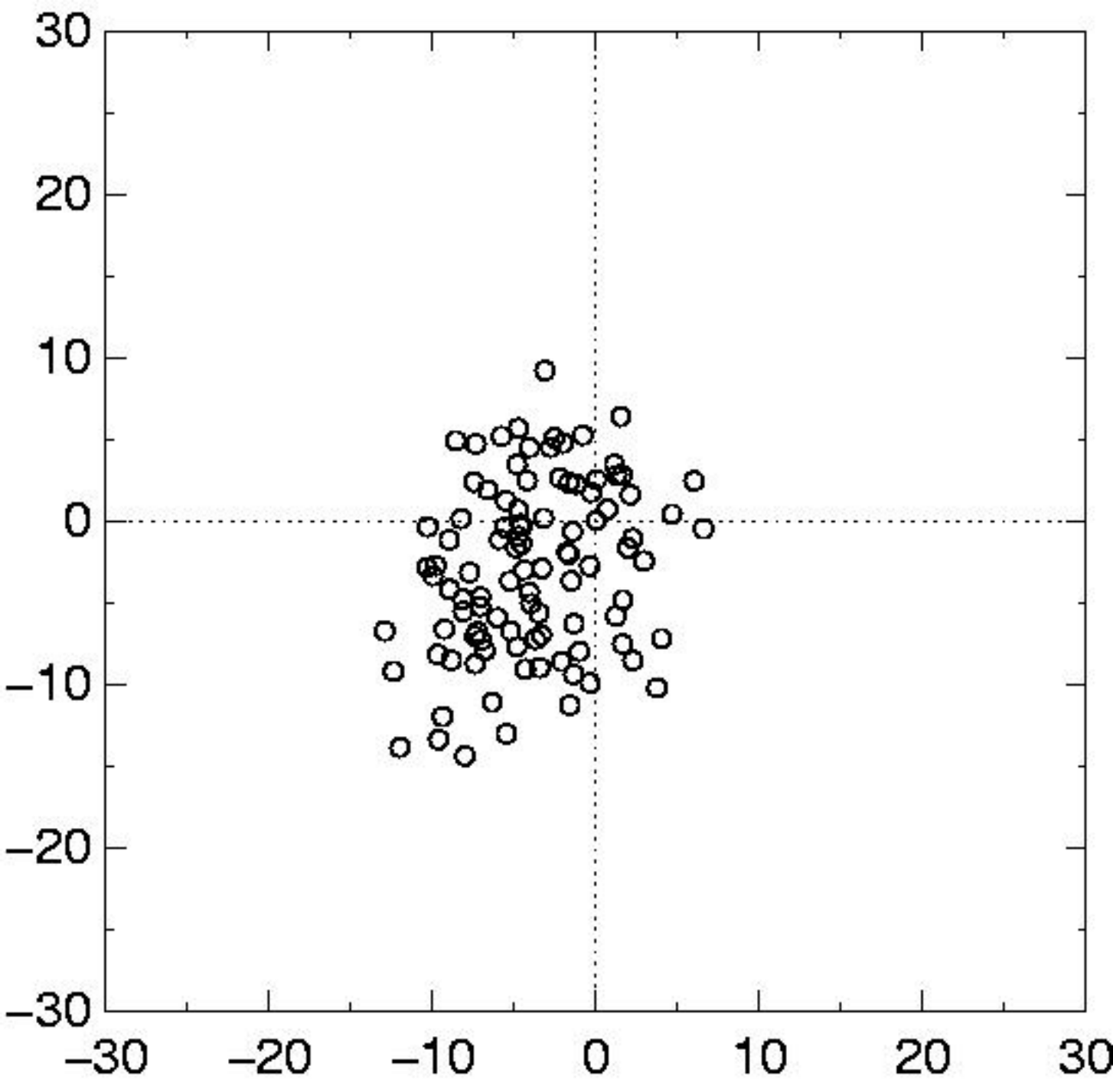
Table 3.1. For different numbers of species, the smallest and the largest number of characters for which there are enough degrees of freedom to simultaneously infer the tree and the covariances. Values obtained by solving the quadratic condition in p in Equation (3.9) are shown.

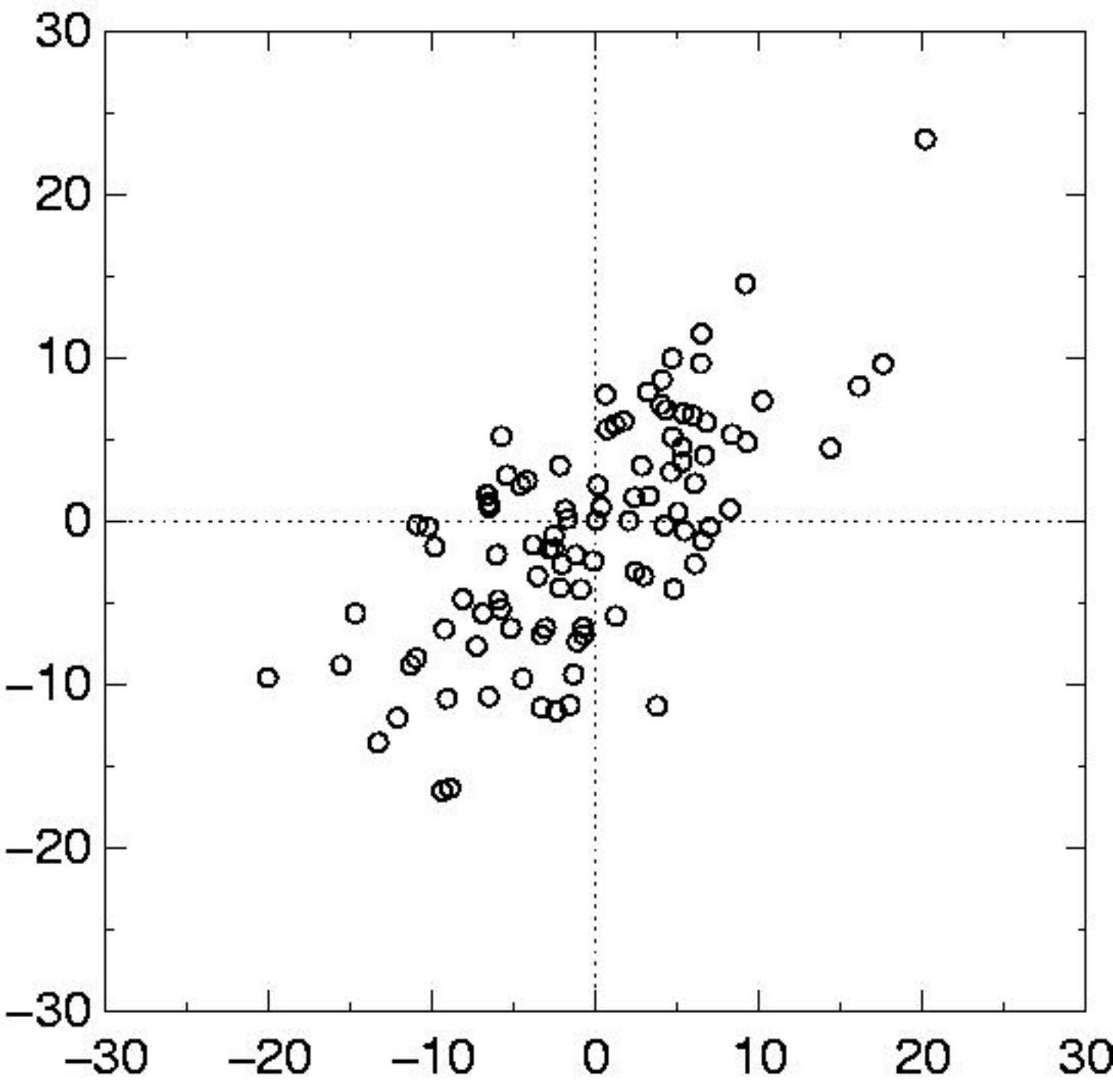

temperate
arctic
temperate
arctic
temperate
temperate
limbs
size
color
size
limbs
color

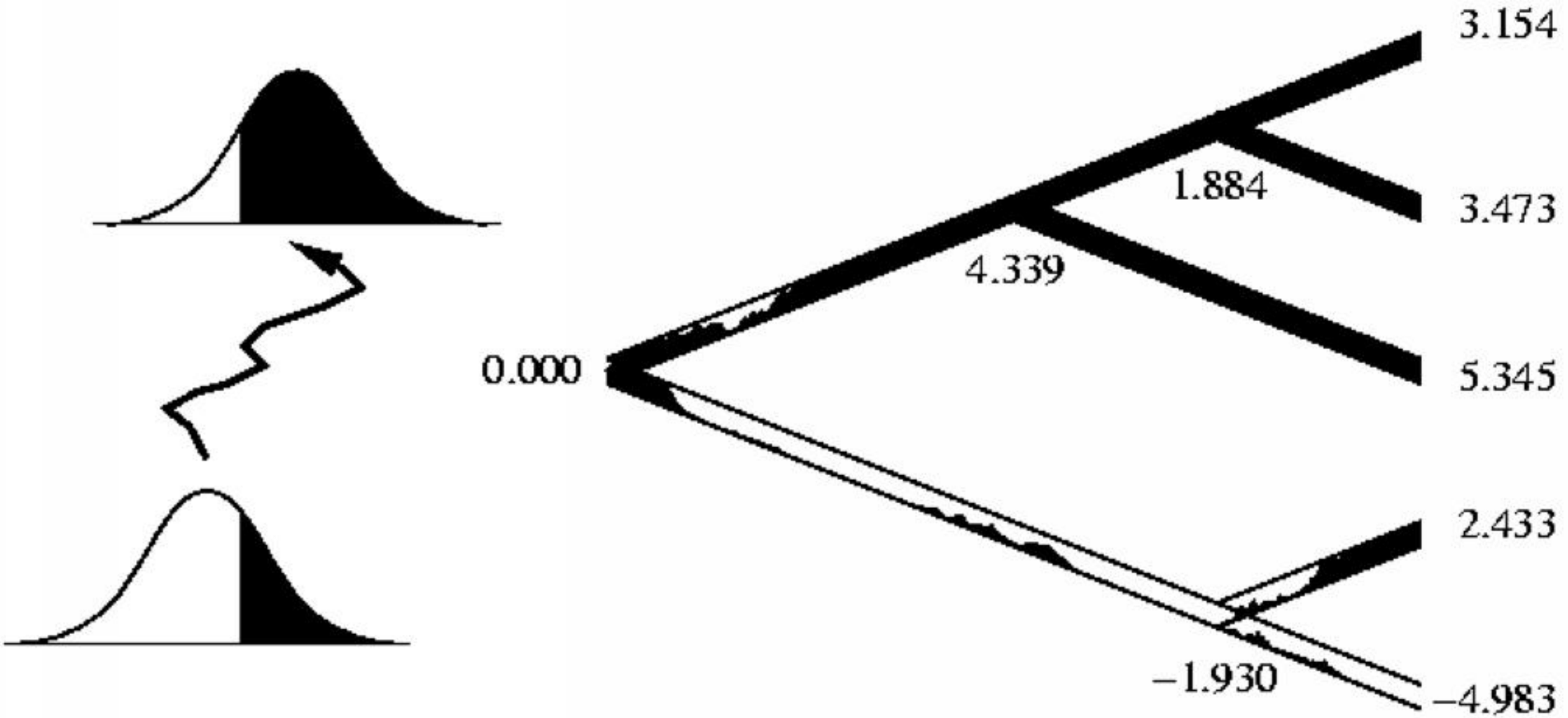

recent species
Molecular sequences
Quantitative characters
fossil species
Quantitative characters
estimate phylogeny
Phylogeny of recent species
comparative method
Covariances of characters
estimate phylogeny
Phylogeny of all species


Tree
$\hat{\tau}$
$\hat{\rho}$
Correlation between characters 1, 2


30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30


30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30


30
20
10
0
−10
−20
−30
−30
−20
−10
0
10
20
30


3.154
1.884
3.473
4.339
0.000
5.345
2.433
−1.930
−4.983

species	characters greater than	characters less than
6	2.43	6.56
7	2.30	8.70
8	2.23	10.77
9	2.18	12.82
10	2.16	14.84
11	2.13	16.87
12	2.12	18.88
13	2.11	20.89
14	2.10	22.90
15	2.09	24.91
16	2.08	26.92
17	2.07	28.93
18	2.07	30.93
19	2.06	32.94
20	2.06	34.94
25	2.05	44.95
30	2.04	54.96
35	2.03	64.97
40	2.03	74.97
50	2.02	84.98
100	2.01	194.99