A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model

Joseph Felsenstein*

Department of Genome Sciences and Department of Biology, University of Washington, Seattle, Washington 98195

Submitted October 14, 2010; Accepted October 4, 2011; Electronically published December 15, 2011

ABSTRACT: The threshold model developed by Sewall Wright in 1934 can be used to model the evolution of two-state discrete characters along a phylogeny. The model assumes that there is a quantitative character, called liability, that is unobserved and that determines the discrete character according to whether the liability exceeds a threshold value. A Markov chain Monte Carlo algorithm is used to infer the evolutionary covariances of the liabilities for discrete characters, sampling liability values consistent with the phylogeny and with the observed data. The same approach can also be used for continuous characters by assuming that the tip species have values that have been observed. In this way, one can make a comparative-methods analysis that combines both discrete and continuous characters. Simulations are presented showing that the covariances of the liabilities are successfully estimated, although precision can be achieved only by using a large number of species, and we must always worry whether the covariances and the model apply throughout the group. An advantage of the threshold model is that the model can be straightforwardly extended to accommodate within-species phenotypic variation and allows an interface with quantitative-genetics models.

Keywords: phylogeny, comparative method, threshold model, MCMC, discrete characters, continuous characters.

Introduction

Despite the widespread use of phylogenetic comparative methods for evolution of continuous characters and the development of some methods for evolution of discrete characters, there has been no fully developed method that could use both classes of characters. Harvey and Pagel (1991) suggested that discrete characters could be accommodated in analysis of phylogenetic contrasts by coding two states as 0 and 1 and then treating them as if continuous. They derived the means and variances of the contrasts from the means and variances of the two-state discrete stochastic model of change and proposed that these be used while treating the two-state character as one undergoing Brownian motion. This would allow analysis of both kinds of data, but the approximation involved is very rough.

For there to be a fully developed method that combines continuous and discrete characters, there must be a well-developed statistical phylogenetic method for discrete characters. Such a method has been proposed by Pagel (1994), with further development by Lewis (2001). It assumes that two discrete states exist, called 0 and 1, and that there is a continuous-time Markov process for changes between these two states. For two discrete characters, Pagel has shown how a likelihood ratio test can be done of the null hypothesis that the state of one character has no effect on the transition probabilities of the other character.

It would be possible to develop a mixed continuous/discrete model from Pagel's model, but there would be some difficulties. If we had (say) five discrete characters, we would need to specify what the continuous characters' processes were for all 2^5 possible combinations of the states of the discrete characters. This could involve effects on the means of changes of the continuous character or on the covariances of their changes. Some simplification could be achieved by making the assumption that the processes in the continuous characters depend on the states of the discrete characters through some function of their states, as is often done in statistics when log-linear models are used for contingency tables.

I have suggested an alternative (Felsenstein 1988, pp. 462–463; 2002, pp. 40–41; 2004, pp. 429–431; 2005), which is to use Sewall Wright's (1934) threshold model and to assume that the underlying characters change by covarying Brownian motion along a tree. In this article, I show that this also leads to a simple and natural treatment of data that has both discrete and continuous characters. The key is that the unobserved characters that underlie the discrete characters are assumed to have evolutionary covariation with the continuous characters, much as the latter do with each other. This has the advantage of simplicity and straightforwardness. It also naturally allows for within-species differences in the discrete character. However it is

* E-mail: joe@gs.washington.edu.

Am. Nat. 2012. Vol. 179, pp. 145–156.
0003-0147/2012/17902-52547$15.00.
DOI: 10.1086/663681

computationally difficult, requiring use of a Markov chain Monte Carlo integration to infer and test the covariation among the characters. How this can be done is explained after I briefly review the threshold model.

The Threshold Model

Sewall Wright (1934) developed the threshold model to fit data on the incidence of extra toes on the hind feet of guinea pigs in crosses among a series of inbred lines. He assumed that there was an underlying, unobservable continuous character, which has come to be called the "liability." On this scale there was a threshold value. Any individual whose liability was above that threshold developed state 1, and any individual below the threshold developed state 0 (it does not matter what happens when the liability is exactly at the threshold value, because such cases are infinitely improbable). The threshold model has been used in human genetics (most notably by Falconer 1965) to model discrete traits such as having type I diabetes, to fit incidence of the disease among relatives of affected individuals. Some further review of the model and these studies will be found in the text by Cavalli-Sforza and Bodmer (1971). The threshold model has come into regular use in pedigree analysis of discrete traits in quantitative genetics (Gianola 1982).

For phylogenies, I have suggested (Felsenstein 2002) that we treat the liability values of the population as undergoing a Brownian motion in their mean values. We assume that within each species the liability values follow a multivariate normal distribution, with common within-species covariances that do not change through time. The means of these normal distributions wander by Brownian motion along branches of the phylogeny, in the same way that means of continuous characters do. The covariances of changes in the species means along the phylogeny are the evolutionary covariances, which differ from the within-species covariances. At any moment, we could, in principle, calculate the number of within-species standard deviations between the population means and the threshold. For example, for one discrete character, if the mean were 1 SD above the threshold, we would expect that 0.8413 of the individuals in that population would show state 1 and that 0.1587 of them would show state 0. A similar but more complex calculation can be done if there are multiple discrete characters.

Although this model can be used to make integrated inferences for discrete characters within and between species, for the present let us ignore within-species variation and covariation and assume that in each population we observe the more frequent of the two states of the discrete character. If the population mean of the liability exceeds the threshold, the state we then observe is 1; otherwise, it is 0. The extension of this analysis to within-species variation in the discrete characters is straightforward and is discussed briefly below. For now, when we discuss the liability values in a lineage (or at a node on the tree), this is to be taken to be the population means of the liabilities, and the covariances of their changes are taken to be the evolutionary covariances.

Hadfield and Nakagawa (2010) have noted that all such models are equivalent to multivariate "mixed models" of quantitative genetics. For discrete traits, such as our 0/1 trait, they note the addition of "transfer functions" to accommodate them. No doubt this is true and worth exploring. For continuous characters, most of the examples they give involve only a single character with a scalar variance rather than multiple characters with a covariance matrix. For categorical characters such as 0/1 characters, they do discuss multivariate methods, although the transfer function that they use has the value of the categorical variable drawn from a Poisson distribution with the liability as its mean. In the present model, the transfer function should instead be a step function that rises from 0 to 1 at the threshold. Hadfield and Nakagawa argue that existing general-mixed-model software uses advanced algorithmics that would be vastly better than the specialized programs in use in comparative biology. This will have to be proven in particular cases. At a minimum, machinery would have to be added to present-day mixed-model programs to use a phylogeny to set up appropriate design matrices.

Ives and Garland (2010) also put forward a model in which there is logistic regression of discrete traits on continuous variables. In their model, the discrete traits change along the tree according to a two-state stochastic process similar to that used by Pagel (1994) and Lewis (2001). The continuous characters influence the discrete traits by logistic regression. However, the continuous characters are assumed to be known traits of the present-day species. They mention the possibility that these continuous characters could themselves evolve along the tree, but they do not develop methods for that case.

This article describes a method embodied in a Markov chain Monte Carlo (MCMC) program, Threshml, to infer the covariance matrix of changes in continuous characters and in the liabilities of the discrete characters along a phylogeny.

Sampling from the Unobserved Values

For continuous characters evolving by correlated Brownian motion, the joint distribution of the values at the interior nodes of the tree and at the tips is multivariate normal, with covariances that are easily calculated once the covariances of change in the means of continuous

characters are known. This applies equally to observed continuous characters and to the liabilities that underlie the discrete traits. A computation of the joint likelihood for all of the observed characters integrates over all possible character values at the interior nodes of the tree. For the discrete characters, it must also integrate over all of the liability values in the tips that fall on the correct sides of the thresholds. Thus, if $\mathbf{x}_c$ are the observed values of the continuous characters at the tips, if the elements of $\mathbf{y}$ are the phenotypes of the discrete characters, and if $\mathbf{x}_\ell$ are the (unknown) liability values at the tips, then the likelihood for tree $\mathbf{T}$ can be written in terms of the covariance matrix $\mathbf{A}$ of changes per unit branch length, and the expectation vectors $\boldsymbol{\mu}_c$ and $\boldsymbol{\mu}_\ell$, as the probability density

$$\begin{aligned} L(\mathbf{T}) &= f(\mathbf{x}_c, \mathbf{y}|\boldsymbol{\mu}_c, \boldsymbol{\mu}_\ell, \mathbf{A}, \mathbf{T}) \\ &= \int_{\mathbf{x}_\ell \in X(\mathbf{y})} \phi(\mathbf{x}_c, \mathbf{x}_\ell|\boldsymbol{\mu}_c, \boldsymbol{\mu}_\ell, \mathbf{A}, \mathbf{T}), \end{aligned} \tag{1}$$

where $X(\mathbf{y})$ is the region of liability values in which all of the liabilities are on the correct sides of their respective thresholds, so that they lead to the observed discrete characters.

The density ϕ is the joint multivariate normal density of the continuous characters and the liability values at the tips. Carrying out the integration in equation (1) involves finding the volume under this density in a multivariate corner of the density function (the corner in which all the liabilities are in region X). This is computationally very difficult. The objective of the method used here is not to compute the likelihood. It is assumed that the tree, including branch lengths, is supplied by the user (presumably having been inferred from molecular sequences). We want to make a maximum likelihood inference of the covariances $\mathbf{A}$ among characters of their changes along the branches of the tree. This will be done with an MCMC expectation-maximization (EM) algorithm (Guo and Thompson 1994). The algorithm used here for the discrete-character liabilities will differ from the more complicated one I previously proposed (Felsenstein 2005).

Stochastic EM Algorithm for the Covariances

If we somehow knew the values of the observed continuous characters and the unobserved liability characters at the tips of the tree and also their values at all interior nodes of the tree, we could make a maximum likelihood estimate of the evolutionary covariances $\mathbf{A}$. The covariance between characters i and j would be estimated by computing

$$\hat{a}_{ij} = \frac{1}{b}\sum_k \frac{(x_{ki} - x'_{ki})x_{kj} - x'_{kj}}{v_k}, \tag{2}$$

where x_{ki} is the value of character i at one end of branch k, and x'_{ki} is the value of character i at the other end (and similarly for character j), v_k is the length of branch k, and b is the number of branches in the tree. I show below that we can avoid the need to infer the values of the characters at the root of the tree.

The EM algorithm uses knowledge of the distributions to compute the expectation of this covariance formula over the distribution of the unobserved values of x, given the observed values and the current estimates of the parameters. These covariances become the new estimates, as they would if we were making maximum likelihood estimates of the covariances. This expectation and the maximization of the likelihood are repeated iteratively until the estimate converges (Dempster et al. 1977).

In the present case, we do not know the distribution of the values of x conditional on the observed data, but we can use MCMC sampling to draw a large sample of points from this distribution and average the estimates of the covariances over the points in that sample. This allows us to make an MCMC EM procedure. As such, it will not converge precisely to the maximum likelihood estimate but will come near it and then wander in that vicinity. At each stage, we run the Markov chain for as long as we can, to take a sufficiently large sample of points. How near the resulting estimate comes to the maximum likelihood estimate depends on how large a sample we are able to choose.

Under the Brownian motion model that we are using, the expectation of the change of each character in a branch is 0. Thus, no parameters are needed for the means of the changes.

Strategy of MCMC Sampling

A Gibbs sampler (Geman and Geman 1984; Gelfand and Smith 1990) will be used for the continuous characters and the liabilities at the interior nodes of the tree, and a Metropolis sampler (Metropolis et al. 1953) will be used for drawing the species mean liabilities at the tips. The continuous characters do not have to be sampled at the tips, because they are observed. A Gibbs sampler is preferable when it can be done, because it samples precisely from the desired distribution without any need to reject some of the samples and try again. For the liabilities at the tips, the distribution has a truncated normal density, and it is difficult to sample directly from that distribution. The Metropolis sampler is a good practical alternative for that case. I describe the details of the sampling below.

A provisional estimate of the covariances of the characters is maintained. At each stage, it is used to transform the characters so that, given that these were the true covariances, the changes of the new characters would be

independent. The MCMC sampler is then run with these new characters assumed to be independent, and an estimate of the covariances of these putatively independent characters is obtained. This is used to further update the estimates of the covariances of the original characters. For example, suppose that for the mth cycle of the MCMC EM algorithm our provisional estimate of the covariance matrix of changes is inferred to be **C**. If the vector of original continuous characters and liabilities were called **y** and had these covariances, then we could use a matrix square root **S** of **C** that satisfies $\mathbf{C} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$. We can then obtain a new vector of characters $\mathbf{z} = \mathbf{S}^{-1}\mathbf{y}$, which would have unit variance and would be uncorrelated.

If the MCMC sampling of unobserved values of **z** now infers that the vector of transformed characters **z** are not independent but actually have covariance matrix **B**, whose matrix square root is **R**, then it can be shown (as it is in app. A) that we can make a set of independent variables by computing $\mathbf{R}^{-1}\mathbf{z}$. The matrix square root that transforms the original characters is then modified from **S** to **SR**, so that its inverse is $\mathbf{R}^{-1}\mathbf{S}^{-1}$.

The MCMC run consists of a series of Markov chain runs (e.g., 30 chains). Each is run for a large number of steps (such as 100,000) After each chain, the covariance matrix **B** is inferred and the transform to independence is adjusted by premultiplying $\mathbf{S}^{-1}$ by $\mathbf{R}^{-1}$.

Gibbs Sampling at Interior Nodes

The sampling of character values at interior nodes in the tree is done by a Gibbs sampler. This is done for both the continuous characters and the liabilities of the discrete characters. At each stage of the EM iteration, the current estimate of the covariances is assumed to be known. If we consider characters **z**, transformed so that they have unit variances and zero covariances, we can update the value of each character at each interior node without considering any other character, and we can consider only the immediately neighboring nodes in the tree. Thus, if an interior node connects to three other nodes, numbered 1–3, we can draw a new value for a character based only on the values of that character in these three neighboring nodes in the tree.

Previously (Felsenstein 2005), I have given the algorithm for Gibbs sampling of interior nodes under a Brownian motion model. Appendix B derives these formulas. If a node has three neighbors, the ith one a branch length v_i away, then the Gibbs sampling draws a normally distributed value x that has expectation

$$\mathbb{E}[z] = \frac{(1/v_1)z_1 + (1/v_2)z_2 + (1/v_3)z_3}{(1/v_1) + (1/v_2) + (1/v_3)} \tag{3}$$

and variance

$$\sigma_z^2 = \frac{1}{(1/v_1) + (1/v_2) + (1/v_3)}. \tag{4}$$

This is done separately, and thus independently, in each of the transformed characters. For multifurcating nodes, the extension to more neighbors is obvious. The result is the same no matter where the tree is rooted.

Sampling the Liabilities at the Tips

While the values of the continuous characters at the tips of the tree are known, the values of the liabilities at the tips are not known, but they must be consistent with the observed discrete characters. In the previous article (Felsenstein 2005), a sampler was proposed, together with a rather elaborate reweighting method. This has been reconsidered and replaced by a simpler Metropolis sampler that makes small changes in the liabilities, accepting or rejecting them according to whether they cause the liability to conflict with the discrete character. A Metropolis sampler is like a Gibbs sampler, except that it does not draw directly from the conditional distribution of the quantity but adds an acceptance-rejection step formulated to produce the desired conditional distribution.

The sampling of the independent (transformed) characters is very simple: each is changed by an amount drawn independently from a normal distribution whose mean is 0 and whose variance is a parameter set by the user. This is made more complicated by the covariances among the characters. The characters **z** that are sampled on the transformed scales, where they are expected to be independent, must be examined to see whether they result in a conflict between any of the liabilities and the observed discrete characters. If the current square root of the covariance matrix of the original continuous characters and the liabilities is called **S**, then $\mathbf{y} = \mathbf{S}\mathbf{z}$ undoes the transformation and returns us to the original character scale. We can then check the variable **y**, or at least the coordinates in it that are liabilities, to see whether they are on the wrong side of their thresholds.

The computation is made much easier if, in the vector **y**, the continuous characters are placed first and then followed by the discrete-character liabilities. The matrix square root **S** that we use is lower-triangular, so that in returning to the original scale, the jth character is a linear transformation of the independent characters $z_1, z_2, \ldots, z_j$. We do not allow any change in the continuous characters at the tips of the tree. These are affected only by the first p_c of the independent characters, and so those are not

changed. The next $p_d = p - p_c$ affect the liabilities of the discrete characters, and these must be changed. As each of these p_d independent characters is sampled, we can compute another one of the liabilities and immediately check it with its threshold value. The entire set of changes is rejected if any one of the liabilities is on the wrong side of its threshold value. Thus, when there are seven continuous characters and z_{10} is being sampled, after new values z'_8 and z'_9 have been drawn successfully, we draw a new value of z_{10}, called z'_{10}, and immediately compute

$$x'_{10} = s_{10,1}z_1 + s_{10,2}z_2 + s_{10,3}z_3 + s_{10,4}z_4 + s_{10,5}z_5 + s_{10,6}z_6 + s_{10,7}z_7 + s_{10,8}z'_8 + s_{10,9}z'_9 + s_{10,10}z'_{10}. \quad (5)$$

We then check x'_{10} to see on which side of its threshold it is. If it is on the wrong side, then the whole set of tip liabilities is rejected and the process starts over with the observed tip values of the continuous characters and the choice of a new value of z_8, and it again proceeds to z_9, z_{10}, and so on, again rejecting the whole set when any of the x'_i is found to be on the wrong side of the threshold. Rejection rates can be monitored, and the parameter that is the variance of the proposed normally distributed changes of the z_i can be adjusted to be smaller if there is too much rejection and larger if there is too little.

If the new values of the independent characters pass this test, so that the resulting liabilities imply the correct discrete character values, they still must be checked as to whether they have too low a density of the normal distribution of the independent characters. Appendix C shows this calculation, which is straightforward. If any of the values of the independent characters is rejected (say $p_c + j$), then we start over at independent character $p_c + 1$ and draw new values of the independent characters $p_c + 1$, $p_c + 2$, and so forth, until we succeed in drawing all p_d of them. The user-defined parameter for the size of the changes in the independent characters allows us to keep the acceptance rate of the proposals from being either too high or too low.

Testing Hypotheses about the Covariances

Hypotheses of interest about the covariation of the characters include whether characters are independent of one another in their evolution. There is some question about how to test this and what questions are meaningful. Testing whether one particular covariance, say the one between characters 6 and 8, is nonzero seems of little biological relevance, as the two could still be connected by patterns of covariation with other characters. A more reasonable hypothesis to test would be whether a set of characters evolves independently of all of the other characters.

A likelihood ratio test can be constructed from the Markov chain Monte Carlo sampling. The probability (or probability density) of the data under a hypothesis whose parameter values are Θ can be written as

$$\Pr(D|\Theta) = \int_{X(\mathbf{y})} f(\mathbf{x}; \mathbf{C}), \quad (6)$$

where $\mathbf{x}$ is the vector of values of the continuous characters and liabilities at all nodes of the tree, including interior nodes, $X(\mathbf{y})$ is the set of points for which $\mathbf{x}$ agrees with the observed discrete phenotypes and $\mathbf{C}$ is the matrix of covariances of the characters. The matrix $\mathbf{C}$ is affected by the parameters Θ. The quantity $\Pr(D|\Theta)$ should be understood as a probability if all characters are discrete and as a probability density otherwise.

The MCMC sampler draws from an importance-sampling density $\Pr(D|\Theta_0)$. For the likelihood ratio test of whether one set of variables does not covary with the other variables, the covariance matrix under the null hypothesis is $\mathbf{C}_0$, in which the covariances between the two sets of variables are constrained to remain 0. The density function of the $\mathbf{x}$ values is given by equation (6), with $\mathbf{C}$ equal to $\mathbf{C}_0$. The basic importance-sampling equation in this case becomes

$$\Pr(D|\Theta) = \mathbb{E}\left[\frac{f(\mathbf{x}|\mathbf{C})}{f(\mathbf{x}|\mathbf{C}_0)/\Pr(D|\Theta_0)}\right], \quad (7)$$

which is easily rearranged into

$$\frac{\Pr(D|\Theta)}{\Pr(D|\Theta_0)} = \mathbb{E}\left[\frac{f(\mathbf{x}|\mathbf{C})}{f(\mathbf{x}|\mathbf{C}_0)}\right]. \quad (8)$$

When likelihood ratio testing of covariances is carried out, Threshml does an extra set of sampling chains, sampling with the covariance matrix $\mathbf{C}$ constrained to force the covariances between sets of variables to 0. For each point at which samples are taken for the test, the density function of the continuous characters and liabilities at all nodes on the tree are computed under the null hypothesis and under its alternative. The ratio of these densities is averaged over all of the samples. The result is a likelihood ratio that can be used in a likelihood ratio test. If there are p characters, divided into two sets with p_1 and p_2 characters, the number of degrees of freedom for the test is p_1p_2.

Restricted Maximum Likelihood

A subtlety is what we have done regarding the mean vector $\boldsymbol{\mu}$ in the above expressions. You may have noticed that we did not infer it. We sampled from the distribution of char-

acter values at the interior nodes of the unrooted tree, which did not include a root node below the rootmost fork. The expectations and variances of the continuous-character value at a fork are, as we have seen, influenced by the values at the neighboring nodes, weighted inversely by the branch lengths to those nodes. By allowing that root node no influence, we assumed, in effect, that it was infinitely far removed in the past. That, in turn, has the interesting effect that the location of the root on the tree does not matter. No matter where in the tree we connect the root, the character values at the other interior nodes will be unaffected by where that root is.

The result is something like REML (restricted maximum likelihood) estimation. The joint distribution of the character values at the tips of the tree will depend only on the unrooted form of the phylogeny. In the case where there are no discrete characters, the inferences converge on the results of an ordinary analysis using contrasts—and those are REML estimates. However, when there are discrete characters, the inferences of root liability are not dependent only on the differences in character values between tips of the tree, as they would be in REML estimation. The matter deserves more careful attention than I can give it here.

Issues of Power

One serious limitation of the analyses proposed here is that there is very limited power for inference of covariation of liabilities with each other or with continuous characters. If we have a phylogeny with 100 species at the tips, then the usual contrasts method for inferring the evolutionary covariances makes that inference from only 99 independent quantities. That would give the correlation between two characters a standard deviation (if the true correlation is small) of 0.101535, which is fairly large for a quantity that is constrained to be between -1 and 1.

However, the situation is even worse when we have thresholded continuous characters and observe only on which side of the threshold they are. Then, two sister species will often be on the same side of the threshold, and thus comparison of their phenotypes provides us with little information. In the continuous-characters case, if there is no within-species sampling variation, we can hope to use the small difference between the sibling species and scale that by dividing by the intervening branch length—but for discrete characters that will not work. So the effective amount of information is considerably less than 99 independent data points.

A similar problem affects the evolutionary covariation of continuous characters when there is also within-species covariation. The signal of change between closely related pairs of species tends in that case to be swamped by the noise of within-species variation. If there were no within-species covariation, we could take the mean phenotypes of the two species and make contrasts between them in each character. Their covariances would then be proportional to the branch length on the path connecting these two species. The branch length would be small, which would make even small differences between sibling species potentially informative. But in the presence of within-species covariance, the covariances of the contrasts are mostly affected by that sampling error, and they convey very little information about the covariances of evolutionary changes.

One might hope to "make it up in volume" by using less closely related species to get a much larger tree. The difficulty with this is that we are relying on a very crude evolutionary model, and as we deal with a broader range of species, we are correspondingly less confident that the model holds throughout the tree and that the covariation can be considered to be the same.

The lesson from all this is that we have a limited amount that we can discover, and we may have to learn how to be satisfied with that. In particular, inferences about the phenotypes and genotypes of ancestral species, inferences that are beloved of popular science media, have large and irreducible errors. Finding ways of propagating that uncertainty through the further analyses that we do will be a major challenge.

The Program

A computer program, Threshml, has been written to infer covariances of threshold characters as well as covariances of both continuous characters and threshold characters. It uses the MCMC algorithm outlined here. The program, which will also be included in version 3.7a of the PHYLIP package, can, until that release, be downloaded at http://evolution.gs.washington.edu/phylip/download/threshml/. It is available as C source code and as Windows, Linux, and Mac OS X executables, with HTML documentation. After the release of version 3.7a, Threshml will be available with the PHYLIP package at its usual Web site, http://evolution.gs.washington.edu/phylip.html.

Simulations

Some computer simulations of the behavior of the method have been done for a single tree with 100 species (shown in fig. 1). Changes of three characters were simulated under a Brownian motion model, where the true covariance matrix (which remained unknown to Threshml) was taken to be

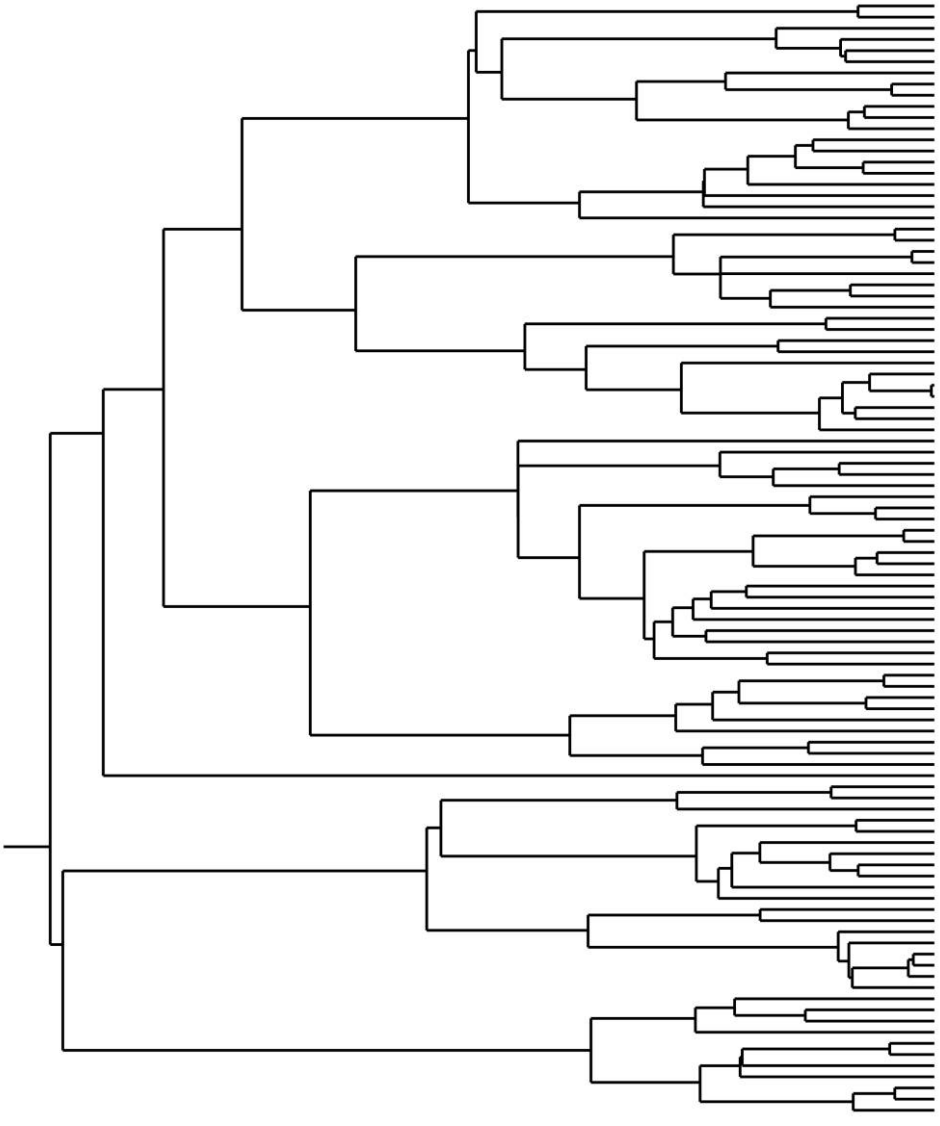


Figure 1: Phylogeny of 100 species used for the simulations. The phylogeny was generated by a pure birth process, with a birth rate of 1 per unit time.

$$\begin{bmatrix} 1.64 & 0.8 & 0 \\ 0.8 & 1.36 & -0.6 \\ 0 & -0.6 & 1.0 \end{bmatrix}.$$

The Brownian motion started at phenotypes of (0, 0, 0) and was simulated in 100 replicates, each replicate generating one data set.

Discrete-Character Simulation

In this case, the data set was taken and all three characters thresholded, so that each became 0 or 1. As noted above, the scale of the threshold characters is arbitrary (as long as there is no within-species variation). As the variance of those liability characters was constrained to remain 1, we were, in effect, inferring only the correlation coefficients between the liabilities. There were three such coefficients in the covariance matrix, which was then expected to be

$$\begin{bmatrix} 1 & 0.535672 & 0 \\ 0.535672 & 1 & -0.514496 \\ 0 & -0.514496 & 1 \end{bmatrix}.$$

Figure 2 shows the histogram of the 100 values of the

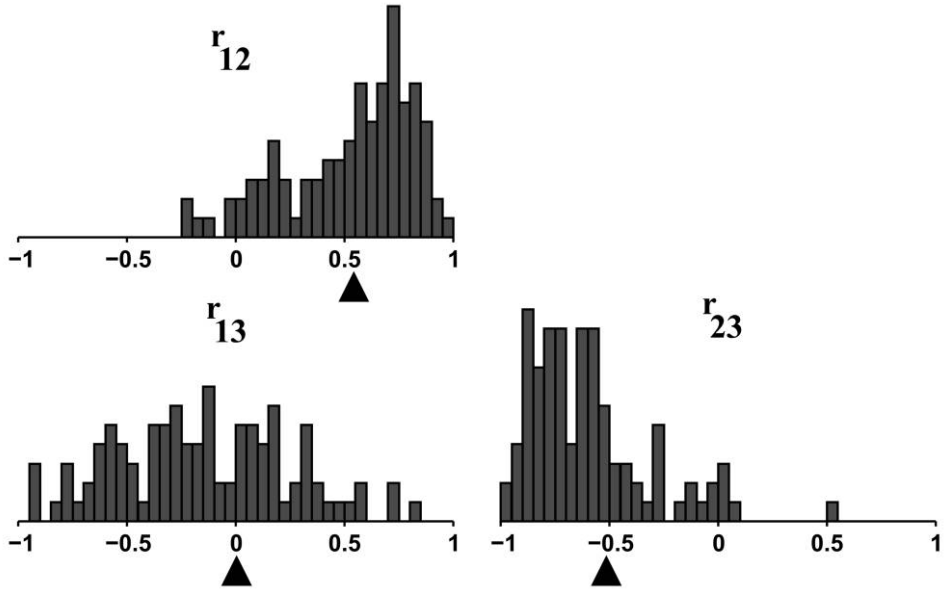


Figure 2: Histograms of correlation coefficients r_{12}, r_{13}, and r_{23} for the characters that covaried as described in the text. One hundred data sets were simulated. Their true correlation coefficients were 0.535672, 0, and −0.514496, respectively. These values are shown by the triangles.

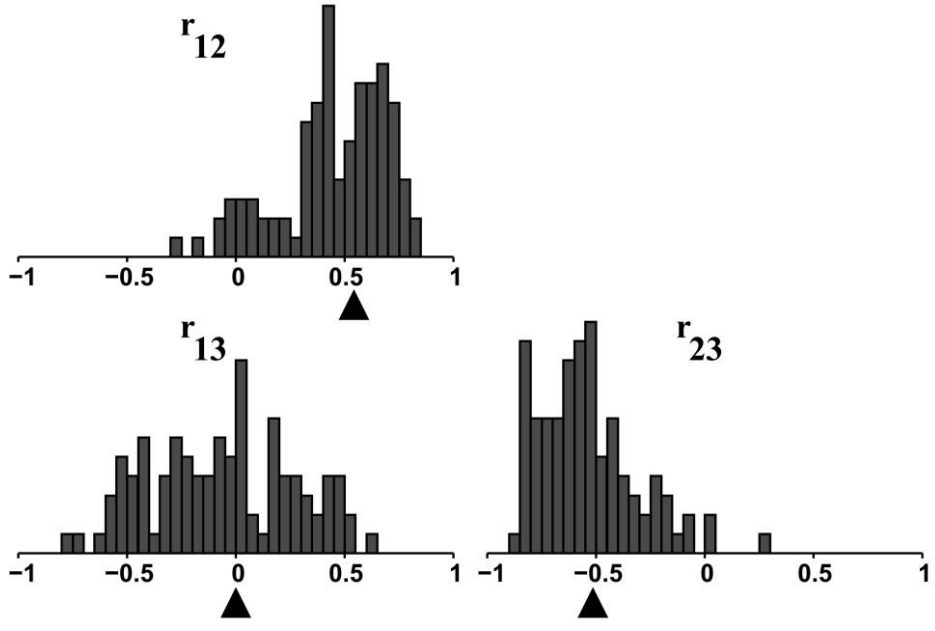


Figure 3: Histograms of correlation coefficients r_{12}, r_{13}, and r_{23} with the same simulation of 100 data sets as in figure 2, except that only character 2 was thresholded and the exact numerical values of characters 1 and 3 were used as quantitative characters. The true values are shown by the triangles.

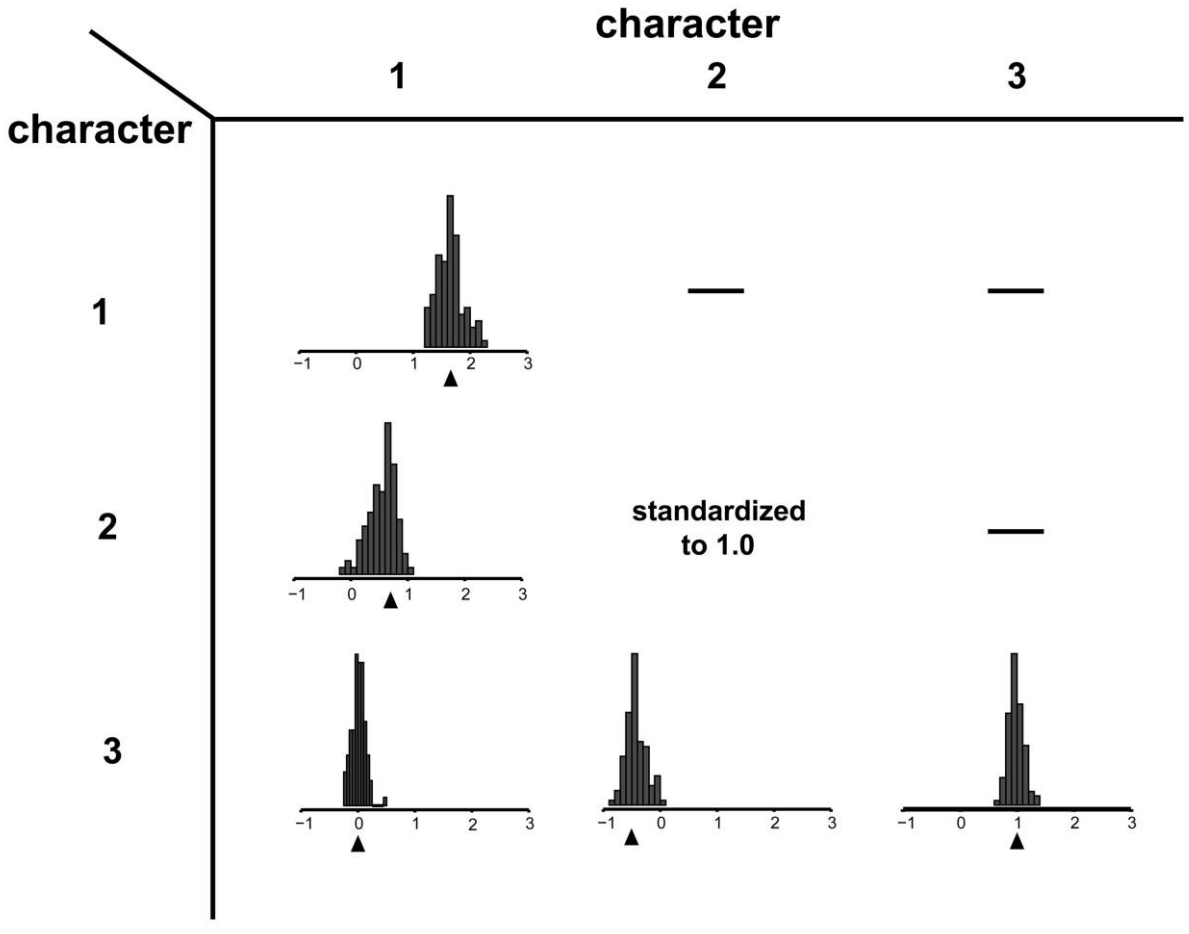


Figure 4: Histograms of covariance estimates for the 100 replicates of the mixed continuous/discrete simulation (the one also described by fig. 3). The histograms for the elements of the lower triangle of the covariance matrix are shown. Character 2, the discrete character, is always standardized so that its variance is 1. The true values are shown by the triangles.

three correlation coefficients. The triangles show the true values. There is little sign of bias, although the values are not narrowly clustered around the true values. When the truth is a positive or a negative correlation, the inference is able to infer only a little more than the sign of the correlation correctly.

Discrete and Continuous Characters

In this case, the same Brownian motion simulation was used, but only character 2 was thresholded to become a discrete character. Figure 3 shows the histogram of the 100 correlation coefficients. Again, there is no noticeable bias, and again the inference of the correlations is very rough and can tell us little more than the sign of the correlation.

The covariances for these three characters are shown in figure 4 for that simulation. Character 2 (the discrete-character liability) had its inferred variance standardized to 1, so that its covariance with the other two characters was correspondingly affected. The triangles that show the true values are, in the case of covariances with character 2, adjusted for this standardization.

Within-Species Covariances

If we consider the variation and covariation of the liabilities within species, we can make an analysis of both the within-species and the between-species character covariation. We would then have two covariance matrices to infer, one the evolutionary covariances and the other the within-species phenotypic covariances. The data would no longer correspond to the species means (or the discrete

characters implied by the species mean liability) but would consist of discrete and continuous phenotypes recorded from a sample of individuals from each species. Sample sizes could vary from species to species, being as small as 1 for some species.

I have not yet implemented or tested such a model, but inferences could be made for it with an MCMC strategy quite similar to that outlined here. The liabilities (and continuous-character values) for the species means would be sampled, as well as the liabilities and continuous-character values for the individuals and for the hypothetical ancestral nodes. The individuals in each within-species sample would lie at the tips of branches radiating from the population mean, with branch length 1. Thus, if the sample size for a species was 4, there would be a node on the tree that represented the species mean, and a quadrifurcation from this would lead to the nodes for the individuals, with branch lengths 1. The within-species covariances would be estimated, in the MCMC EM iteration, from the changes of the liabilities and continuous-character values along those branches. The evolutionary covariances would, as before, be estimated from the changes along the other branches, the ones connecting the different species and their ancestors.

The analyses described in this article do not attempt to take within-species variation into account but instead represent each discrete character of a species by the most frequent state and each continuous character by the species mean of that character. When there are no discrete characters, the estimates of the within-species covariances and the evolutionary covariances from the MCMC EM procedure should be close to those obtained for the corresponding statistical model in a comparative-method analysis with sampling error (Felsenstein 2008). It has not escaped my attention that a quite similar strategy could be used when some species are represented by samples from multiple populations, especially when we also have estimates of the rates and pattern of migration between those populations and of their population sizes.

Connection to Quantitative-Genetics Experiments

Another fruitful area for development of these models is to connect them as well to quantitative-genetics experiments that estimate the additive genetic covariances between characters. I have already discussed this with respect to within- and between-species analyses of continuous characters (Felsenstein 2008). The same applies to data sets that contain discrete characters, when the threshold model can be used for them. A combined quantitative-genetics and comparative analysis seems the only way to infer how much of the evolutionary covariation reflects additive genetic covariances and how much reflects selective covariances, which describe the covariances of selection pressures (Felsenstein 1988). Covariation of changes in phenotypes along a phylogeny may reflect either or both, and a pure comparative-methods analysis cannot tease them apart.

Acknowledgments

I thank C. Geyer for suggesting the use of the Metropolis sampler for updating the tip liabilities and B. Giansiracusa, L. Harmon, and J. McGill for helpful comments on the manuscript. I also thank S. Blomberg and the other participants in the review of this article for very useful suggestions. The initial work on this project (2003–2008) was funded by National Science Foundation grant DEB 0316632. I am grateful to the Department of Genome Sciences of the University of Washington for "life-support" funding to cover part of my salary after that period until November 2010. Work after that date was supported by National Science Foundation grant DEB 1019583, principal investigators J. Felsenstein and F. L. Bookstein.

APPENDIX A

Updating Transformations of Variables

If a vector $\mathbf{y}$ of variables whose means are 0 is inferred to have a covariance matrix $\mathbf{C}$, where $\mathbf{C} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$, then if we can invert matrix $\mathbf{S}$ we can obtain new variables $\mathbf{z} = \mathbf{S}^{-1}\mathbf{y}$. The covariances of $\mathbf{z}$ would then be $\mathbb{E}[\mathbf{z}\mathbf{z}^{\mathrm{T}}]$. This is then $\mathbb{E}\left[\mathbf{S}^{-1}\mathbf{y}\mathbf{y}^{\mathrm{T}}(\mathbf{S}^{-1})^{\mathrm{T}}\right]$, which is $\mathbf{S}^{-1}\mathbb{E}\left[\mathbf{y}\mathbf{y}^{\mathrm{T}}\right](\mathbf{S}^{-1})^{\mathrm{T}}$. Since the expectation of $\mathbf{y}\mathbf{y}^{\mathrm{T}}$ is $\mathbf{S}\mathbf{S}^{\mathrm{T}}$, we can easily see (substituting this) that the covariance matrix of $\mathbf{z}$ is supposed to be the identity matrix $\mathbf{I}$.

Suppose that we obtain $\mathbf{z}$ using this transformation but then find, on further sampling, that its covariance matrix is actually $\mathbf{B}$. If we obtain the matrix square root $\mathbf{R}$ of $\mathbf{B}$, such that $\mathbf{B} = \mathbf{R}\mathbf{R}^{\mathrm{T}}$, then we can make a further transform, $\mathbf{u} = \mathbf{R}^{-1}\mathbf{z}$. The covariances of $\mathbf{u}$ would then be

$$\mathbb{E}\left[\mathbf{u}\mathbf{u}^{\mathrm{T}}\right] = \mathbf{R}^{-1}\mathbb{E}\left[\mathbf{z}\mathbf{z}^{\mathrm{T}}\right](\mathbf{R}^{-1})^{\mathrm{T}} = \mathbf{R}^{-1}\mathbf{B}(\mathbf{R}^{-1})^{\mathrm{T}}$$
$$= \mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{\mathrm{T}}(\mathbf{R}^{-1})^{\mathrm{T}} = \mathbf{I}.$$

In the Metropolis algorithm for sampling liabilities, we make use of the matrix square root $\mathbf{S}$ of the covariance matrix $\mathbf{C}$. Once we have computed $\mathbf{R}$ and find that it is not the identity matrix, the transform is $\mathbf{u} = \mathbf{R}^{-1}\mathbf{S}^{-1}\mathbf{x} = \mathbf{S}\mathbf{R}^{-1}\mathbf{x}$, so that $\mathbf{S}$ must now be replaced by $\mathbf{S}\mathbf{R}$.

APPENDIX B

Gibbs Sampler for Interior-Node Character Values under Brownian Motion

Suppose that we have a character x that evolves on a tree by Brownian motion with variance 1 per unit time. Consider using a Gibbs sampler to choose a new value for the character at one node, where this node has three neighboring nodes whose values are x_1, x_2, and x_3. Suppose that the variances of change on branches 1, 2, and 3 are v_1, v_2, and v_3, respectively. The value chosen in the Gibbs sampler will have a normal distribution, the same as the conditional distribution of x, given x_1, x_2, and x_3. If we take node 1 as the immediate ancestor of our node and nodes 2 and 3 as its descendants, then the joint distribution of x_2, x_3, and x is normal, with means all x_1 and covariance matrix

$$\begin{bmatrix} v_1 + v_2 & v_1 & v_1 \\ v_1 & v_1 + v_3 & v_1 \\ v_1 & v_1 & v_1 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{B1}$$

where the blocks are

$$\begin{aligned} \Sigma_{11} &= \begin{bmatrix} v_1 + v_2 & v_1 \\ v_1 & v_1 + v_3 \end{bmatrix}, \\ \Sigma_{12} &= \begin{bmatrix} v_1 \\ v_1 \end{bmatrix} = \Sigma_{21}^{\mathrm{T}}, \\ \Sigma_{22} &= [v_1] \end{aligned} \tag{B2}$$

For the multivariate normal distribution, if we compute the expectation of x conditional on the values of x_2 and x_3, this is

$$\mathbb{E}[x] = x_1 + \Sigma_{21}\Sigma_{11}^{-1}\begin{bmatrix} x_2 - x_1 \\ x_3 - x_1 \end{bmatrix} \tag{B3}$$

(e.g., Rao 1973, pp. 522–523), and the variance of x is

$$\mathrm{Var}\,[x] = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \tag{B4}$$

These expressions can easily be worked out in straightforward fashion, and they lead to the results in equations (3) and (4).

APPENDIX C

Rejection Rule for Independent Characters at Tips

When new values are sampled for the independent characters at the tips, these are for independent characters $m + 1$, $m + 2$, ..., n. As each is drawn, it is checked for whether the corresponding discrete character's liability is on the wrong side of its threshold. After all of them pass this test, we must also check whether the density of the distribution of this set of independent characters is too low. If the new values of the independent characters are x'_{m+1}, x'_{m+2}, ..., x'_n, then the density of this set of independent characters, conditional on the nearest (interior) node, node j, which is a branch length v from this tip, is

$$\prod_{i=m+1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{v}} \exp\left[-\frac{1}{2}\frac{(x'_i - x_i^{(j)})^2}{v}\right]. \tag{C1}$$

When this is compared this to the density at the old values x_{m+1}, x_{m+2}, ..., x_n, their ratio simplifies to

$$\exp\left[-\frac{\sum_{i=m+1}^{n}(x'_i - x_i)(x_i + x'_i - 2x_i^{(j)})}{2v}\right]. \tag{C2}$$

The acceptance rule is, as usual, that a uniform random number be less than this ratio.

Literature Cited

Cavalli-Sforza, L. L., and W. F. Bodmer. 1971. The genetics of human populations. W. H. Freeman, San Francisco.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39:1–38.

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

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

———. 2002. Quantitative characters, phylogenies, and morphometrics. Pages 27–44 in Morphology, shape, and phylogenetics. N. MacLeod, ed. Systematics Association Special Volume 64. Taylor & Francis, London.

———. 2004. Inferring phylogenies. Sinauer, Sunderland, MA.

———. 2005. Using the quantitative genetic threshold model for inferences between and within species. Philosophical Transactions of the Royal Society B: Biological Sciences 360:1427–1434.

———. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. American Naturalist 171:713–725.

Gelfand, A. E., and A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721–741.

Gianola, D. 1982. Theory and analysis of threshold characters. Journal of Animal Science 54:1079–1096.

Guo, S. W., and E. A. Thompson. 1994. Monte Carlo estimation of mixed models for large complex pedigrees. Biometrics 50:417–432.

Hadfield, J. D., and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and

multi-trait models for continuous and categorical characters. Journal of Evolutionary Biology 23:494–508.

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

Ives, A. R., and T. Garland Jr. 2010. Phylogenetic logistic regression for binary dependent variables. Systematic Biology 59:9–26.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50: 913–925.

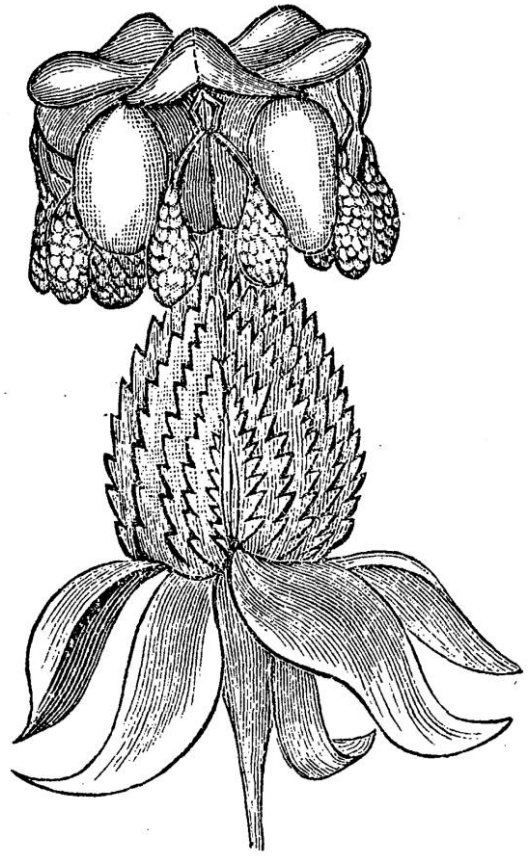
Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21:1087–1092.

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

Rao, C. R. 1973. Linear statistical inference and its applications. 2nd ed. Wiley, New York.

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

Associate Editor: David D. Ackerly
Editor: Mark A. McPeek



Milkweed *Asclepias cornuti* flower with the hoods cut away. "The visitor, attracted by the odor, alights to suck the nectar secreted in the hoods. In its progress over the blossom some one of the hairs of its legs is sure to slide into the slit between the hoods. Pursuing his way by drawing up his leg, the hair will be guided by two flanges at the sides into the upper and narrower part of the slit, and there become fast. Feeling a detention, the captive will pull to release himself, and, if possessed of sufficient force, will bring out of the sacs at the sides two pear-shaped pollinia, each fastened to the lamina, or gland, by a short appendage." From "The Milkweeds" by Joseph F. James (*American Naturalist*, 1887, 21:605–615).