The Phylogenetic Mixed Model

Elizabeth A. Housworth,[1,2,*] Emília P. Martins,[2,†] and Michael Lynch[2,‡]

1. Department of Mathematics, Indiana University, Bloomington, Indiana 47405;
2. Department of Biology, Indiana University, Bloomington, Indiana 47405

Submitted November 11, 2002; Accepted June 27, 2003; Electronically published January 28, 2004

Online enhancements: figure, appendix, relationship matrix.

ABSTRACT: The phylogenetic mixed model is an application of the quantitative-genetic mixed model to interspecific data. Although this statistical framework provides a potentially unifying approach to quantitative-genetic and phylogenetic analysis, the model has been applied infrequently because of technical difficulties with parameter estimation. We recommend a reparameterization of the model that eliminates some of these difficulties, and we develop a new estimation algorithm for both the original maximum likelihood and new restricted maximum likelihood estimators. The phylogenetic mixed model is particularly rich in terms of the evolutionary insight that might be drawn from model parameters, so we also illustrate and discuss the interpretation of the model parameters in a specific comparative analysis.

Keywords: comparative method, mixed model, phenotypic evolution, phylogenetic analysis, phylogenetic heritability, quantitative genetics.

It is now well appreciated that interspecific analyses can be compromised if they fail to account for the statistical dependence resulting from shared evolution along a phylogeny. A number of researchers have provided solutions to this problem by developing statistical approaches for incorporating phylogenetic information (see Martins and Hansen 1996*b* for review). More recently, a few researchers (e.g, Charnov 1993; Westoby et al. 1995; Price 1997) have pointed out that the more popular of these phylogenetic comparative methods (Felsenstein's [1985] independent contrasts) may sometimes overcorrect for phylogenetic effects when, in truth, selection has led to comparative data that are not strongly influenced by phylogenetic history. One solution in these cases is to ignore phylogeny completely, but doing so can also lead to poor results when selection has not been as strong as assumed. Alternatively, a handful of researchers have developed methods that explicitly include the effects of selection or other evolutionary assumptions (see Martins 2000 for review).

Recall that Felsenstein's (1985) independent contrasts method is based on an assumption that the traits of interest have evolved via Brownian motion along the phylogeny. Under this model, the phenotype randomly increases or decreases each generation, as is usually expected for phenotypes undergoing random genetic drift or fluctuating directional selection. The resulting correlation in the trait values between two species is the proportion of time the two species shared a common ancestor. Thus the correlation structure of the trait data is given pictorially by the phylogeny.

Many other forms of selection result in sets of comparative data that are not well described by a Brownian motion model (Hansen and Martins 1996). The phylogenetic generalized least squares approach described in Martins and Hansen (1997) provides an extension of Felsenstein's (1985) independent contrasts method that allows the researcher to choose among a variety of explicit evolutionary alternatives to the Brownian motion model (e.g., those summarized in Hansen and Martins 1996). In particular, Martins and Hansen (1997) recommended use of an Ornstein-Uhlenbeck model, applied by population geneticists to describe phenotypes undergoing stabilizing selection, selective response to fluctuating environments, and other types of long-term constraints (for further discussion and examples, see Felsenstein 1988; Hansen and Martins 1996; Hansen 1997). The resulting correlation in the trait value between two species is a function of an evolutionary model–specific transformation of the branch lengths of the original phylogeny.

Other authors have encouraged researchers to apply statistical approaches to this problem. Some propose modifying Felsenstein's basic Brownian motion approach (e.g., Diaz-Uriarte and Garland 1996) by transforming the

* Corresponding author; e-mail: ehouswor@indiana.edu.

† E-mail: emartins@indiana.edu.

‡ E-mail: mlynch@bio.indiana.edu.

Am. Nat. 2004. Vol. 163, pp. 84–96.

branch lengths in a statistical rather than an evolutionary model–specific way, such as by taking the logarithm of all the branches. The resulting correlation in the trait value between two species is then a function of this transformation of the branch lengths of the original time-based phylogeny. Others invent entirely new statistical methods (e.g., Cheverud et al. 1985; Diniz-Filho et al. 1998).

In introducing the phylogenetic mixed model, Lynch (1991) recognized that in addition to the gradual accumulation of evolutionary changes envisioned by Felsenstein's independent contrasts and phylogenetic generalized least squares approaches, traits sometimes evolve so readily that aspects of their current states are essentially unconstrained by their phylogenetic past. For example, many *Daphnia* species are able to produce protective head extensions (helmets and neck teeth) when growing in the presence of invertebrate predators, whereas other species do not possess this ability. The ability to produce protective head extensions has some genetic component (Spitze and Sadler 1996) and appears to evolve quickly and independently as species shift from ephemeral ponds to lakes or vice versa (Colbourne et al. 1997). Thus, interspecific variation may be best explained by rapid, reversible evolutionary changes (perhaps in response to environmental shifts) than by long-lasting, gradual change. The phylogenetic mixed model (PMM) estimates the relative contribution of these two types of evolutionary change. The resulting correlation in the trait value between two species comes partially (with proportion h^2, which is mathematically defined below) from the phylogenetic relationship between species and partially (with proportion $1 - h^2$) from an independent, species-specific contribution. The transformation to the original phylogeny that pictorially describes the correlation structure in the trait values between species is given in figure 1.

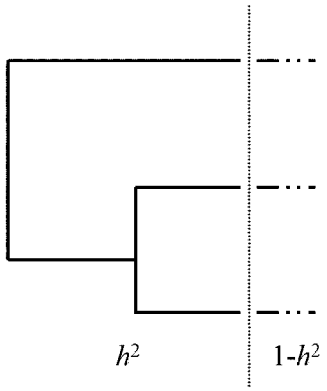

Figure 1: Illustration of the phylogenetic mixed model (PMM) drawn as a phylogeny in units of expected variance of character change. The PMM envisions extant taxon phenotypes to be the result of a linear combination of gradually accumulated evolutionary changes occurring along a true species phylogeny and short-lived evolutionary changes (possibly selective responses to rapid environmental shifts) occurring in each taxon independently and not passed on between ancestor and descendant taxa.

The PMM is an analog of the mixed model from quantitative genetics, which partitions phenotypes of individuals related by a pedigree into additive genetic (heritable) and residual (nonheritable) components (Henderson 1984; Lynch and Walsh 1998). The mixed model describes the trait phenotype (y_i for the ith individual or ith taxon mean) as the sum of a grand mean (μ), a heritable factor (a_i), and a residual deviation (e_i); that is, $y_i = \mu + a_i + e_i$ (Lynch 1991). The grand mean is a scaling term that can be interpreted in the phylogenetic context as the genotypic state of the ancestor at the root of a phylogeny.

For a pedigree from quantitative genetics, the correlation structure of the heritable components is given by a relationship matrix describing genetic similarities. For instance, an individual has 100% genetic similarity with himself and 50% similarity with a parent or sibling (assuming no inbreeding). For species, the correlation structure of the phylogenetically heritable components is given by the proportion of time that taxa share a common ancestor in the phylogeny. In the phylogenetic context, the heritable component contains not only genetic changes but also nongenetic contributions to the phenotype, such as environmental or cultural contributions, that are described by the phylogenetic relationship among the taxa.

Similarly, for a pedigree, the residual, nonheritable component to an individual's phenotype is often considered to be the environmental component and is often modeled as being independent for each individual in the pedigree. This component is the part of the phenotype not explained by the relationship between individuals in the pedigree. In the phylogenetic context, the nonheritable component to a taxon's phenotype is modeled as being independent for each taxon and is the part of the taxon phenotype not explained by the phylogenetic relationship between taxa. The phylogenetically nonheritable component includes phenotypic plasticity, rapid genetic response to the environment or to fluctuating selection, and measurement error. Because the phylogenetically heritable component is modeled via Brownian motion (following Felsenstein's [1985] example), the PMM might also be described as an extension of the independent contrasts method that incorporates the possibility of species-specific evolutionary change.

In quantitative genetics, the heritability of a trait is the proportion of the variance in the trait explained by the relationship between individuals as given in the pedigree. Similarly, we define the phylogenetic heritability of a character as the proportion of variance in the character ex-

plained by the relationship among taxa as given by the phylogeny. In both cases, the mathematical formula is $h^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2)$. Although there are clear similarities between the quantitative genetic and the phylogenetic definitions of heritability, there are important differences in the interpretation due to the fact that, in the phylogenetic context, genetic, environmental, and other factors contribute to both the heritable and nonheritable components and, as explained further in the "Discussion," the heritability of a trait will likely depend on the depth of the phylogeny.

We note that the univariate phylogenetic heritability estimator is mathematically equivalent to the phylogenetic correlation estimator λ examined recently in Freckleton et al. (2002) despite those authors' claim to the contrary. The Freckleton et al. article makes a very nice companion to our current work because those authors determined the value of $\lambda = h^2$ for numerous real phylogenies and data sets from the biology literature. In this article, we discuss these ideas in the context of bivariate analyses and discuss the interpretation of the parameters for one worked example (fig. 7 in the online edition of the *American Naturalist*).

In principle, the PMM has the potential to be more informative than many other phylogenetic comparative methods that for the most part have been used only for estimating phylogenetically corrected correlations between traits. The PMM performs roughly as well as other statistically flexible phylogenetic comparative methods, such as phylogenetic generalized least squares (PGLS; Grafen 1989, 1992; Martins and Hansen 1997) or Cheverud et al.'s (1985) autoregressive model, in terms of estimating a correlation between two characters under evolutionary scenarios that differ from the PMM assumptions (Martins et al. 2002). The PMM can also be used to estimate ancestral states on a phylogeny (breeding values of all members of the phylogeny) and the fraction of the total interspecific phenotypic variance that is associated with phylogenetically heritable effects (the phylogenetic heritability, h^2). For bivariate or multivariate analyses, the PMM also allows the separation of the phenotypic correlation between two traits (ρ) into components associated with phylogenetically heritable (ρ_a) and nonheritable (ρ_e) contributions to the phenotype. The PMM makes efficient use of the data, is unbiased by phylogenetically uninformative contributions to the mean phenotypes, and corrects for phylogenetic dependence of the data only to the extent that the observed variation has a phylogenetically heritable basis (Lynch 1991).

However, practical applications of the method have been hampered by technical difficulties with estimating model parameters. Although the mixed model is commonly used in the analysis of large complex pedigrees (Meyer 1989, 1991; Thompson and Shaw 1990, 1992), it was not clear whether the model would work well even with large phylogenies, which are considerably smaller than most pedigrees. As illustrated by the worked example in the original description of the method (Lynch 1991), small sample sizes often lead to bivariate correlation estimates (ρ_a and ρ_e) of -1.0 or 1.0, regardless of the data used in the study.

We begin by providing an overview of the PMM and by discussing the problems that have been encountered in estimating the parameters of the model. We then propose a shift in focus to parameters that are better estimated with the small sample sizes typical of phylogenetic comparative analyses. We also present a new algorithm for parameter estimation that provides a simple and efficient alternative to the expectation-maximization (EM) algorithm proposed by Lynch (1991; following Thompson and Shaw 1990, 1992). We use computer simulation to explore the statistical properties of PMM and present a worked example using our algorithm and interpreting the PMM parameters in a phylogenetic context. Although we largely focus on phylogenetic analyses, our results are also relevant to quantitative genetic analysis of pedigrees.

Small Sample Sizes and the PMM

The mixed model was designed for pedigree analyses, which often consist of hundreds or even thousands of independent families. In quantitative genetics (and in Lynch 1991), emphasis is usually placed on partitioning out the effects of additive genetic (heritable) versus environmental (nonheritable) components. For bivariate analyses, for example, we might be particularly interested in the correlation between heritable components of the variation (ρ_a) and how that might differ from a correlation between nonheritable components (ρ_e). Unfortunately, these two parameters can be very poorly estimated with the sample sizes commonly found in comparative analyses. The problem is that small sample sizes often lead to negative variance estimates (for σ_a^2 or σ_e^2 for one or both traits) because the mixed model does not mathematically constrain both of these variance components to be positive. Negative variance components in turn lead to heritabilities outside their natural range (0, 1) and to correlations outside their natural range (-1, 1). Negative variance estimates are a well-known difficulty in quantitative genetics and can occur even with very large pedigrees when considering multiple traits and the correspondingly greater numbers of parameters (Hill and Thompson 1978).

Computer simulation can be used to illustrate this small sample size problem. To begin, we developed four symmetric phylogenies of 32 (fig. 2*A*), 64, 128, and 256 taxa. For each phylogeny, we used SAS (SAS Institute 1990) to generate 1,000 sets of two traits for each taxon

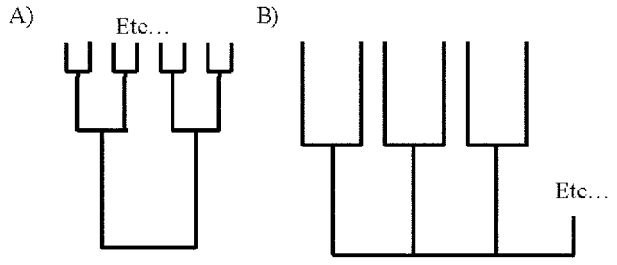

Figure 2: Two hypothetical relationship structures used in theoretical and computer simulation consideration of the phylogenetic mixed model. Each phylogeny was generated using a repeated branching process to add as many taxa as were needed (usually 32, but sometimes more). In all cases, lengths were set so that the total distance from root to tips equals 1. *A*, Symmetrical phylogeny created using an idealized branching process model, such that times between speciation events are inversely proportional to the number of extant taxa. *B*, Bifurcating star phylogeny equivalent to a parent-offspring pedigree.

that satisfy $h_1^2 = h_2^2 = 0.5$, $\rho_a = \rho_e = 0.0$. That $\rho_a = \rho_e = 0.0$ means that both components of the two traits were evolving independently of each other. The expression $h_1^2 = h_2^2 = 0.5$ means that, for both traits, 50% of the variation among extant taxa was explained by the phylogenetic relationship (heritable effects) and the remaining 50% is explained by abrupt evolutionary changes occurring at or since the last speciation event (nonheritable effects). We then used the PMM to estimate the heritable (ρ_a) correlations between the two traits, bounding the correlation by -1 and 1. Figure 3 shows the results of these simulations. For 32 taxa, estimates of the heritable correlations (ρ_a) were usually pegged at -1 or 1, only rarely falling anywhere near the true value of 0. The situation improved with larger numbers of taxa. With 256 taxa, for example, the heritable correlation estimates (ρ_a) at least followed a bell-shaped distribution. Even so, the distribution was very broad. Further, we determined analytically that for data generated along a 32-taxon bifurcating star phylogeny (fig. 2*B*; leaving all other assumptions the same as above), at least one of the parameter estimates was out-of-bounds 72% of the time. Far more than 32 taxa will be required to obtain reasonable estimates of all the PMM parameters.

Note that although the problem of out-of-bounds parameters occurs with both phylogenies and pedigrees, the likelihood of obtaining out-of-bounds parameters depends on tree or pedigree shape as well as the number of measured taxa or individuals. Given the same number of taxa or individuals, data arising from a phylogeny are usually slightly less likely to yield out-of-bounds heritabilities than are data from a pedigree. As explained in the appendix (available in the online edition of the *American Naturalist*), the possible range for the heritability, h^2, is determined by the eigenvalues of the genetic or phylogenetic relationship matrix. Taxa related by a phylogeny (e.g., fig. 2) are generally less independent than individuals related by a pedigree, simply because phylogenies usually bifurcate from the root, whereas pedigrees may originate with several unrelated families. Thus most phylogenetic matrices have a wide range of eigenvalues with a few very large values and many smaller values, some of which are very close to 0. This leads to a range of possible heritabilities close to the proper range 0 to 1. When fewer historical effects are shared (e.g., in starlike phylogenies), the phylogenetic relationship matrix approaches the identity matrix, and all of its eigenvalues approach 1. The result is a wider range of possible heritability estimates and hence a greater chance of obtaining a value outside the natural range (0, 1).

Another technical difficulty with applying the PMM to sample sizes common in phylogenetic analyses is that the standard error and likelihood ratio tests proposed in Lynch (1991) rely on large sample estimator properties that are not likely to be true. Thus we also do not recommend the use of the variance estimators recommended in Lynch (1991). Instead, researchers should apply simulation techniques to obtain confidence intervals and hypothesis tests (e.g., as suggested in Martins and Garland 1991). To do so, researchers can generate large numbers of data sets under the PMM for a particular phylogeny, estimate the desired parameters using each of these data sets, and use the resulting sampling distribution to conduct hypothesis tests or generate confidence intervals. This will result in Type I error rates that are necessarily correct given the assumptions of the mixed model.

Reparameterization and a New Algorithm

Fortunately for comparative analysis, small sample size affects some PMM parameter estimates more than others, and the parameters of most interest may differ between phylogenetic and genetic studies. For example, instead of focusing on the heritable and nonheritable correlations separately, in a phylogenetic context we are often interested in calculating total phenotypic correlation properly (but not overly) corrected for the phylogeny, $\rho = h_1 h_2 \rho_a + [(1 - h_1^2)(1 - h_2^2)]^{1/2} \rho_e$, where the letter subscripts refer to heritable (a) and nonheritable (e) effects and the number subscripts refer to traits 1 and 2. Although this parameter was not introduced in the original description of the PMM (Lynch 1991), it is more directly comparable to the correlations estimated by other phylogenetic methods. Estimators of this parameter are also more robust to small sample sizes than are the separate estimators for the heritable (ρ_a) and nonheritable (ρ_e) correlations. When ap-

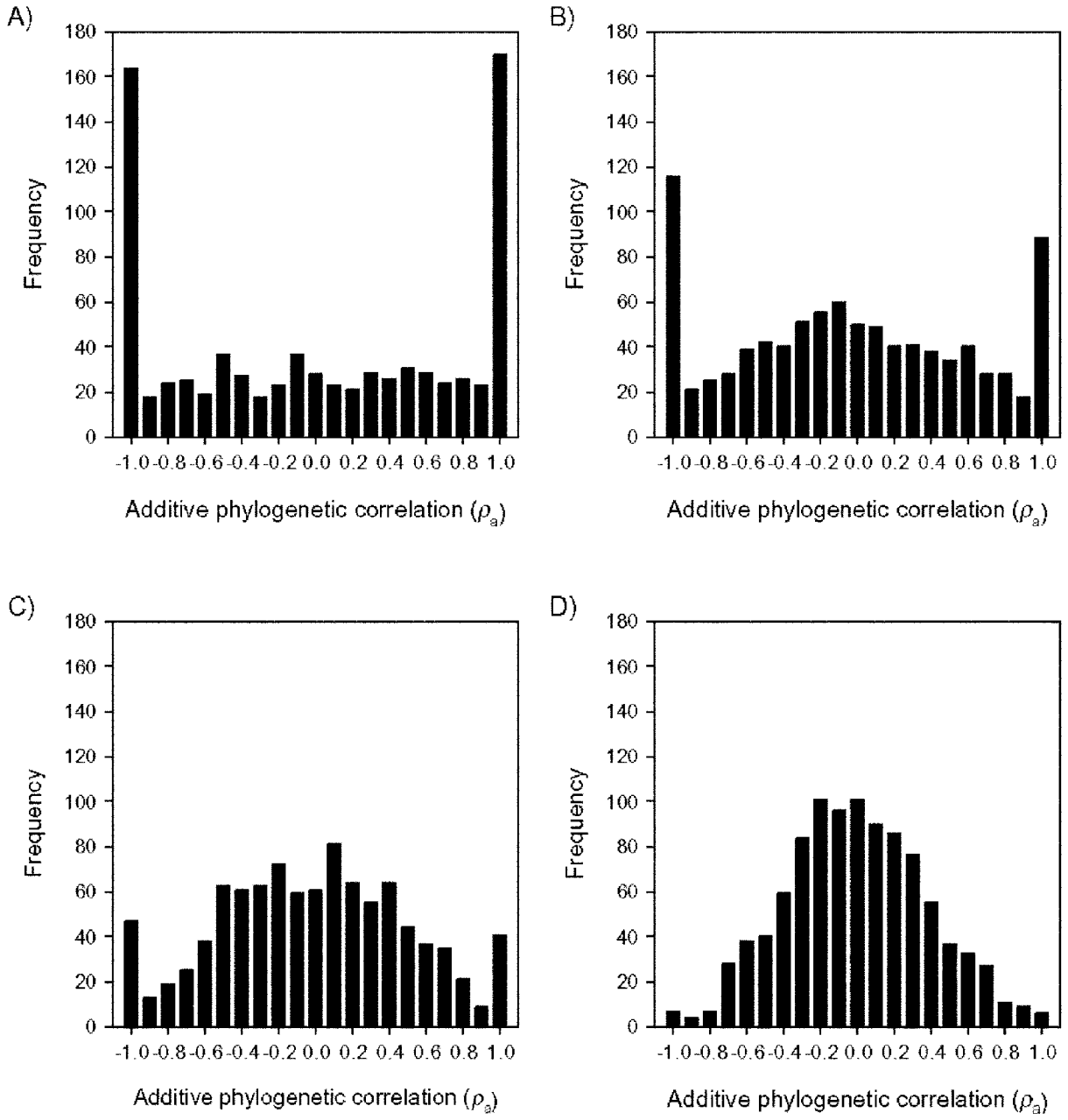

Figure 3: Example of frequency distributions of heritable correlation estimates (ρ_a) obtained through analyses of bivariate simulations on symmetric phylogenies (fig. 2*A*) of different size: *A*, 32 taxa; *B*, 64 taxa; *C*, 128 taxa; and *D*, 256 taxa. Results are for 1,000 sets of bivariate data generated via computer simulation with both means set to 0, both variances set to 1, the true values of correlations set equal to 0 ($\rho_a = \rho_e = \rho = 0$), and phylogenetic heritabilities set to 1/2 ($h_1^2 = h_2^2 = 0.5$). Additive correlations were not calculated whenever h_1^2 or h_2^2 was estimated to be 0, so results are presented for 822 data sets in *A*, 932 data sets in *B*, 972 data sets in *C*, and 995 data sets in *D*.

plied to the above simulated data, even with only 32 taxa, the sampling distribution of the estimator for the total phenotypic correlation (ρ) was bell shaped around the value it was intended to estimate (0).

We also recommend a shift in focus from the heritable (σ_a^2) and nonheritable (σ_e^2) variance components used in the original description (Lynch 1991) to the total variance ($\sigma^2 = \sigma_a^2 + \sigma_e^2$) and the phylogenetic heritability ($h^2 = \sigma_a^2/\sigma^2$). For univariate analyses, for a given heritability, the most likely mean and total variance can be found explicitly.

We use the golden section algorithm (Cheney and Kincaid 1980) to obtain the maximum likelihood heritability estimator. A major advantage of this parameterization is that one can also obtain the likelihood profile as a function of h^2 over the full (0, 1) interval, verifying that the true maximum likelihood estimator, and not a secondary peak in the likelihood surface, has been located. This maximum in the likelihood profile identifies the joint estimates for the phylogenetic mean (μ), total variance (σ^2), and phylogenetic heritability (h^2) that best explain the observed data. If desired, estimates of σ_a^2 and σ_e^2 can be obtained directly from the latter two quantities. For two traits, we apply the golden section algorithm iteratively in random directions, calculating the maximum likelihood estimators for the means and variances given each possible value of the heritabilities and correlations.

Finally, we propose the use of restricted maximum likelihood (REML; as in Meyer 1989, 1991) rather than maximum likelihood estimators for model parameters. The most burdensome part of finding the maximum likelihood estimators in bivariate analyses is solving for the means (μ_1 and μ_2) numerically. REML estimators do not require estimation of the mean and also take into consideration the loss of degrees of freedom associated with estimating the means, often yielding less-biased estimators for variances than those obtained by straight maximum likelihood methods (e.g., Lynch 1991). The popular independent contrasts method (Felsenstein 1985; Grafen 1989, 1992) is an REML procedure that transforms n phenotypes with mean μ into $n - 1$ contrasts with mean 0 (Rohlf 2001). To develop REML estimators for the PMM parameters, we begin with descriptions of independent contrast estimators in matrix form and extend these to include a nonheritable component. All mathematical details are provided in the online appendix.

Relative Statistical Performance

Because the phylogenetic heritability is obtained numerically, the distribution of its estimator is neither analytically known nor easily approximated analytically. The estimators for the grand mean, variances, and covariances depend on the heritability, so their distributions are also not known. Thus, we used limited computer simulations to explore the statistical performance of the PMM and the model's dependence on the details of the phylogenetic structure. We did so (as above) by using SAS (SAS Institute 1990) to generate comparative data that corresponded directly to the assumptions of the PMM, given a particular relationship structure and set of parameters. We then applied our algorithms to these data to estimate both the restricted maximum likelihood and straight maximum likelihood versions of the PMM parameters. While we report results only for the symmetric and bifurcating star phylogenies in figure 2, we also examined a comb phylogeny and a pedigree. The general conclusions reported below are the same for all these structures.

We estimated the bias and root mean square error (RMSE) for each estimated parameter as has been done in earlier simulation comparisons of phylogenetic methods (e.g., Martins et al. 2002). We focus on RMSE rather than on Type I error rates simply because for sample sizes typical of interspecific analyses, we recommend that hypothesis tests for the PMM parameters be conducted against a null distribution obtained through simulations, and such a distribution will necessarily yield the correct Type I error rates.

In Martins et al. (2002), data for two traits were generated under various Ornstein-Uhlenbeck models of constraining selection for which the PGLS method was adapted. Many phylogenetic comparative methods, including the PMM, were used to estimate the evolutionary (phenotypic) correlation between the two traits. The PMM performed roughly as well as other statistically flexible methods (e.g., Cheverud et al.'s [1985] autoregressive model, Diniz-Filho et al.'s (1998) phylogenetic eigenvector regression), regularly outperforming independent contrasts and the nonphylogenetic method. The simulations conducted for this article contribute to these previous results by comparing the relative performance of three of these methods (PMM, independent contrasts, and the nonphylogenetic method) for estimating correlations when confronted with data generated under the mixed model assumptions. The results below are based on simulated data for 32 taxa.

With only 32 taxa, the bounded estimates of the phylogenetic heritability were highly variable and tended to underestimate the true heritability (fig. 4). Unbounded forms of both maximum likelihood (ML) and restricted maximum likelihood (REML) heritability estimates (not constrained to fall between 0 and 1) were generally less biased than their bounded counterparts (results not shown). REML estimates were more reliable than ML estimates (lower RMSE) when the true heritability was large (figure 5). Our results also confirm that the efficient asymptotic properties of ML estimators have not uniformly taken effect with only 32 taxa. The bifurcating star phylogeny (fig. 2*B*) is equivalent to a parent-offspring pedigree, but neither the ML nor the REML estimator was significantly better than the regression-based heritability estimator (twice the slope of the parent-offspring regression line) when the true heritability was large (fig. 5*B*).

Bias for estimates of the grand means (the ancestral states at the root of the tree, μ_1 and μ_2) was negligible (bias ≤ 0.01 in all cases, results not shown). Estimates of the grand mean were also pretty good in an absolute sense,

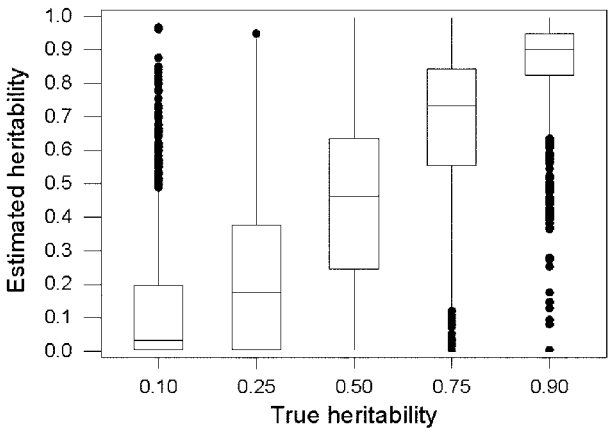

Figure 4: Box plots showing the distribution of phylogenetic heritability (h^2) estimates obtained using the bounded restricted maximum likelihood estimator at true heritabilities of 0.10, 0.25, 0.50, 0.75, and 0.90. Results are for 1,000 sets of data generated via computer simulation on the symmetric phylogeny (fig. 2*A*) with $\mu = 0$ and $\sigma = 1$. Lines in the boxes show the median, edges show twenty-fifth and seventy-fifth percentiles, whiskers extend a distance of 1.5 times the interquartile range, and dots depict outlying points.

with RMSE ranging between 0.2 and 0.6 (increasing with the phylogenetic heritability and the dependency of the tree structure) for a true ancestral state of 0.0. Estimates of the standard errors for the ancestral mean based on the estimate for the variance in the trait were also quite respectable, never differing by more than 0.02 from those calculated using the known heritability and variance (results not shown).

Although ML yielded underestimates of the total phenotypic variance (σ^2) consistently, this bias was relatively small and was generally corrected by the REML estimator. However, REML estimates had larger sampling variances, making ML and REML estimators of this parameter roughly comparable in terms of RMSE (results not shown).

Total phenotypic correlation (ρ) estimates were within the correct bounds and had negligible bias in all the cases we examined. In general, the PMM yielded estimates that were roughly as good as the other methods in many cases and substantially better than Felsenstein's independent contrasts (equivalent to $h_1^2 = h_2^2 = 1$) or the nonphylogenetic method (equivalent to $h_1^2 = h_2^2 = 0$) when the assumptions of those two methods were far from being met (fig. 6). As expected, independent contrasts gave the best performance (lowest RMSE) when heritabilities were large and $\rho_a = \rho_e$ (fig. 6*A*, 6*C*). Independent contrasts also did remarkably well when heritabilities were small and $\rho_a = \rho_e$ (fig. 6*A*, 6*C*), despite the violation of its assumptions. When $\rho_a \neq \rho_e$, however, the independent contrasts method overcounts the nonheritable contribution, ρ_e, leading to very biased estimates (fig. 6*B*, 6*D*) for larger values of the heritabilities. The nonphylogenetic method yielded the lowest RMSE when heritabilities were small, but the improvement is not substantial. The nonphylogenetic approach did not suffer the same bias as independent contrasts when $\rho_a \neq \rho_e$. When the heritabilities were large, however, it sometimes resulted in very poor performance (high RMSE values; fig. 6). Overall, independent contrasts noticeably outperformed the PMM only in really extreme cases with high heritabilities and similar heritable and nonheritable correlation values. The nonphylogenetic approach only barely outperformed the PMM when the true heritabilities were small. The better performance comes from the assumptions of the other models being approximately met and those models having fewer parameters to estimate from the data.

An Example

To illustrate our new approach to the phylogenetic mixed model, we offer an analysis of body length and geographic range size in a group of 50 platyrhine primate taxa using data and phylogeny as compiled by Diniz-Filho et al. (2000; phylogeny, relationship matrix, and data are in the online appendix). We use these data purely as an illustration and offer no opinions on the validity of either data or phylogenetic hypotheses as real descriptions of primate biology. We apply the mixed model to these data to estimate the correlation between body length and geographic range size for these primates. For comparison, we report results also for the nonphylogenetic correlation, Felsenstein's (1985) independent contrasts method, Cheverud et al.'s (1985) spatial autoregressive model, and the PGLS method described in Martins and Hansen (1997; with an exponential model involving estimation of a single α parameter for the two traits, as in Martins et al. 2002). All calculations were conducted in COMPARE 4.4 (Martins 2001). COMPARE finds the REML estimates of bivariate mixed model parameters.

First, there appears to be little if any evidence of a relationship between body length and geographic range size for these primates (table 1). Even the strongest correlation estimate, the nonphylogenetic Pearson correlation of the raw data, is not significantly different from 0. The PMM estimates of the total phenotypic correlation (ρ) was -0.025, falling at the lower end of the range of those obtained for other phylogenetic methods (-0.04, 0.17; tables 1, 2).

With the PMM, however, we can take the results a few steps further to gain insight into the evolution of these two traits. The estimate of phylogenetic heritability for body length in these primates was essentially 1.0, indicating that heritable change accumulating along the entire

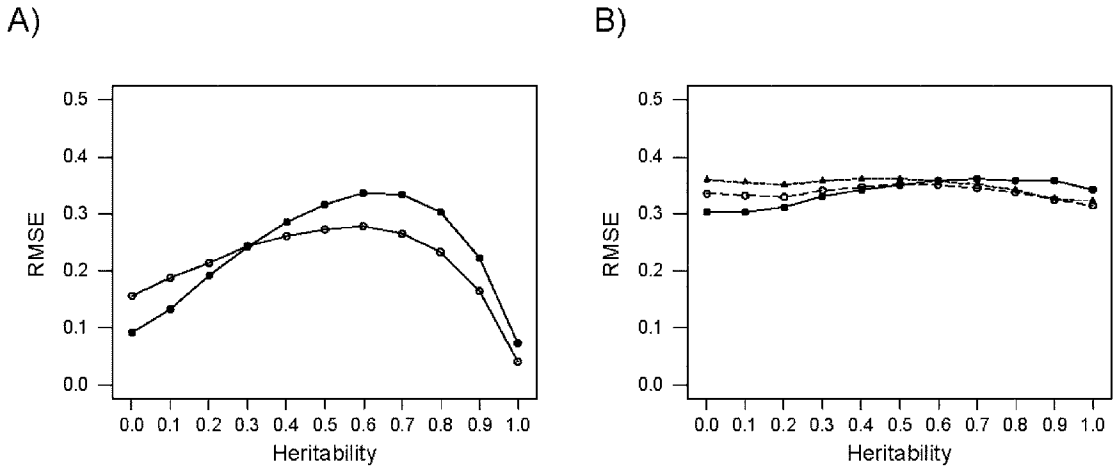

Figure 5: Root mean square error (*RMSE*) for estimates of the phylogenetic heritability (h^2). One graph is shown for each of the tree structures (letter labels correspond to those in fig. 2). Solid circles correspond to results for the maximum likelihood estimator, and open circles depict results for the restricted maximum likelihood estimator. Open triangles indicate results for the heritability estimate obtained from a parent-offspring regression. Results are for 10,000 sets of data generated via computer simulation with $\mu = 0$, $\sigma = 1$, and $h^2 = 0, 0.1, 0.2, \ldots, 0.9, 1$.

length of the phylogeny explains 100% of the interspecific phenotypic variation in body length and that recent, short-lived changes do not contribute much, if anything, to explaining interspecific variation in this trait.

The second trait, geographic range size, also exhibited a relatively high estimate of phylogenetic heritability ($h^2 = 0.6$). Randomization tests showed that this value is significantly <1 ($P = .027$ using 1,000 computer-generated data sets) and significantly >0 ($P = .04$, again using 1,000 computer-generated data sets). Although some might argue that geographic range size is a property of a species rather than of individual organisms and that it is not, therefore, expected to evolve along a species phylogeny, the high estimate of the heritability indicates that about 60% of the phenotypic variation in geographic range is explained by changes accumulating along the phylogeny. Since the heritability is <100%, it would probably be a mistake to analyze geographic range size using independent contrasts, which assumes that $h^2 = 1.0$ and is thus likely to overestimate the importance of phylogeny for this trait. On the other hand, there seems to be phylogenetic dependence in the data, and it is probably also a mistake to ignore phylogeny completely.

Note that these heritability estimates are quite different from measures of phylogenetic inertia estimated by Cheverud et al.'s (1985; Gittleman and Kot 1990) spatial autoregressive method (table 2). Although both the autoregressive parameters ρ (autocorrelation) and R^2 (phylogenetic inertia) are larger for body length than for geographic range size, actual values for the phylogenetic inertia for both traits are quite small ($R^2 \leq 10.1\%$). Following the advice of this method's proponents (Cheverud et al. 1985; Gittleman and Kot 1990), we would probably decide that phylogenetic transformation of either variable is not necessary. Thus the autoregressive method and the mixed model approach come to nearly opposite conclusions about the importance of the phylogeny in explaining these data.

Although both the autoregressive model and the mixed model partition phenotypic variation into phylogenetic and specific components, the mathematical differences between them are profound. With a little algebra (similar to that in Martins and Hansen 1996*a* and Rohlf 2001), we can rewrite the measure of phylogenetic inertia proposed by Cheverud et al. (1985; R^2) in mixed model terms. Doing so, we obtain

$$R^2 = 1 - \frac{\mathbf{y}^{\mathrm{T}}\mathbf{K}[h^2\mathbf{G}_{\mathbf{K}} + (1 - h^2)\mathbf{I}]^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{K}\mathbf{K}^{\mathrm{T}}\mathbf{y}},$$

where $\mathbf{K}^{\mathrm{T}}\mathbf{y}$ are the REML standardized trait values and $\mathbf{G}_{\mathbf{K}}$ is the corresponding transformed relationship matrix. See the online appendix for the details of the choice for $\mathbf{K}$. Clearly, the autoregressive phylogenetic inertia parameter, R^2, is really quite different from the mixed model measure of phylogenetic heritability, h^2. For the mixed model, R^2 measures the percent reduction in the variance estimate for the trait assuming the trait evolves according to the mixed model with heritability h^2 versus assuming that the trait evolves independently of the phylogeny. Even if the heritability of a trait is 1 and there is no nonphylogenetic component to the trait, R^2 may well not be large unless the mixed model variance estimate is much, much

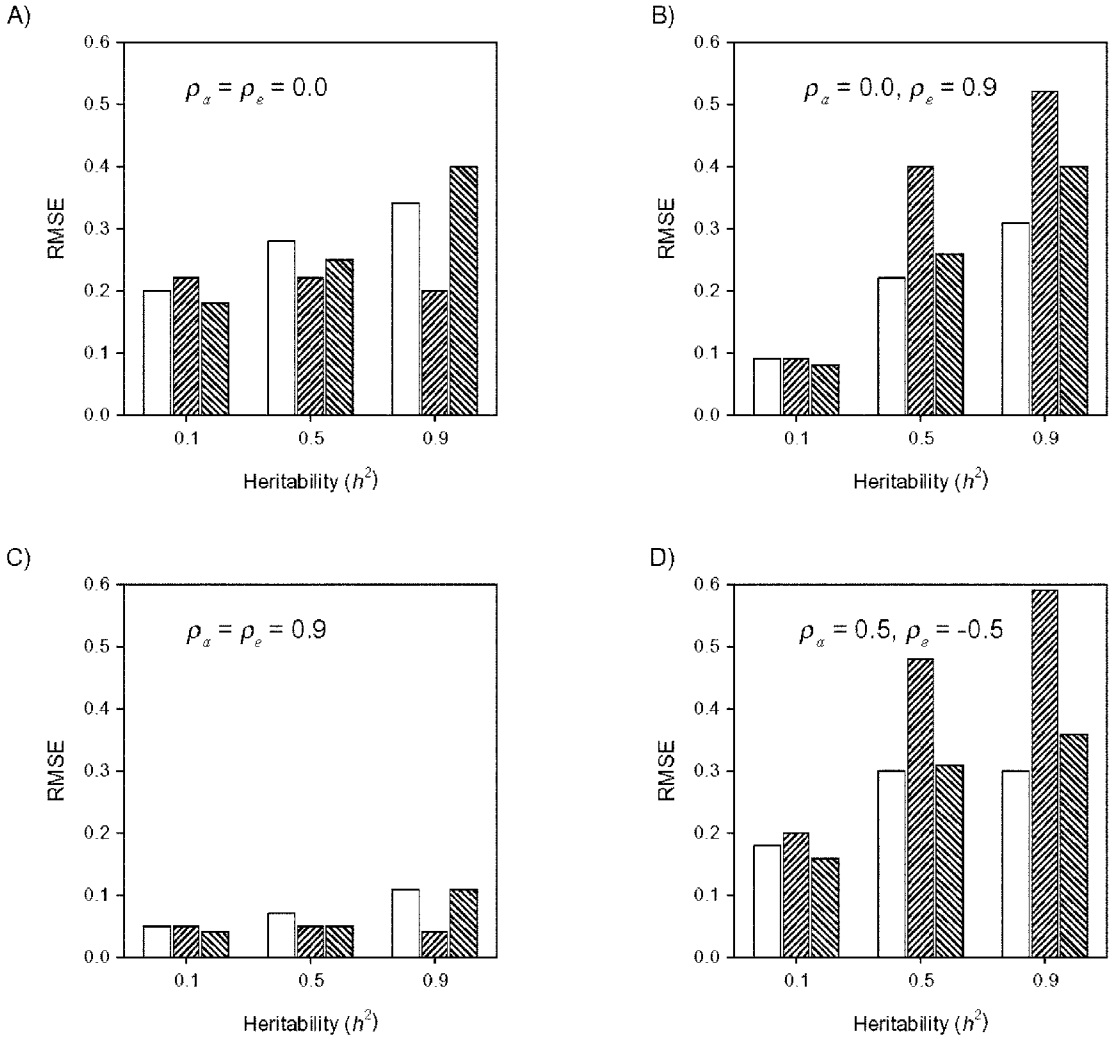

Figure 6: Root mean square error (*RMSE*) for the total phenotypic correlation from bivariate analyses using the phylogenetic mixed model (*open bars*), Felsenstein's (1985) contrasts (assumes $h_1^2 = h_2^2 = 1$; *bars with left-to-right upward hatching*), and a nonphylogenetic approach (assumes $h_1^2 = h_2^2 = 0$; *bars with left-to-right downward hatching*) to analyze data generated via computer simulation along the symmetric phylogeny (fig. 2*A*, 32 taxa). The results are for 1,000 sets of bivariate data generated via computer simulation with both means set to 0, both total phenotypic variances set to 1, both heritabilities set to the value listed on the horizontal axes, and heritable and nonheritable correlations set to the values given in the figures.

smaller than the nonphylogenetic variance estimate for the trait.

Although sample sizes of 50 taxa are probably not large enough to get very good estimates of the heritable and nonheritable correlation components, the partitioning of the bivariate correlation between body length and geographic range size may also be of interest. In this case, however, an estimated phylogenetic heritability of 1.0 for body length means that there is no relevant nonheritable correlation estimate involving body length. Thus, estimates of the nonheritable correlation component between body length and geographic range size (e.g., ρ_e) are meaningless. The total phenotypic correlation (ρ) is determined entirely by the correlation between the phylogenetic components of the two traits (ρ_a) and the heritability of the geographic range size (table 1).

Discussion

The phylogenetic mixed model offers greater evolutionary insight and more realistic evolutionary assumptions than many of the other existing methods for conducting com-

Table 1: Results from a worked example using the data and phylogeny in figure 7

	PMM	FIC	TIPS
Trait 1: geographic range:			
Phylogenetic heritability h_1^2	.63	1	0
Total phenotypic variance σ_1^2	4.56×10^5	8.65×10^5	4.11×10^5
Trait 2: body length:			
Phylogenetic heritability h_1^2	1	1	0
Total phenotypic variance σ_1^2	3.66×10^6	3.67×10^6	7.77×10^6
Bivariate results:			
Correlation between heritable effects	−.02		
Correlation between nonheritable effects	NA		
Total phenotypic correlation	−.025	−.06	.17

Note: All calculations were done in COMPARE (Martins 2001). Figure 7 is available in the online edition of the *American Naturalist*. PMM = phylogenetic mixed model; FIC = Felsenstein's (1985) independent contrasts method; TIPS = nonphylogenetic correlation.

parative studies. It envisions phenotypic evolution as being the result of a complex of forces including some that are retained over long periods of time, forming patterns in trait variation that reflect the underlying phylogenetic structure, and others that act more quickly, in bursts of change that are lost easily at new speciation events. Our model reparameterization and new algorithms provide good estimates of several evolutionarily interesting parameters, including a phylogenetic heritability measuring the relative importance of phylogenetic versus nonphylogenetic evolutionary effects.

Comparison to the Phylogenetic Generalized Least Squares Method

The choice between the PMM and the PGLS method is one of underlying evolutionary assumptions. Both the PGLS method and the PMM are extensions of the independent contrasts method. As mentioned above, independent contrasts is a special case of the mixed model with phylogenetic heritability (h^2) assumed to equal 1. Independent contrasts can also be viewed as a special case of the generalized least squares method in which the constraining force (α) tends to 0. The additional parameters are what give PGLS and the PMM the flexibility to avoid the poor statistical performance of independent contrasts and the nonphylogenetic method when the underlying evolutionary model is not known (Martins et al. 2002). In terms of evolutionary interpretation, however, the two parameters (α and h^2) are quite different and extend independent contrasts in different directions. The mixed model phylogenetic heritability parameter measures the relative importance of long-lasting versus short-lived change in explaining interspecific variation, whereas the PGLS parameter α estimates the strength of evolutionary constraints acting throughout the phylogeny. Although either model could be expanded to include the parameters of the other, this does not seem practical given the small sample sizes typical of phylogenetic analyses and the already observed challenges for the mixed model parameter estimation.

Instead, the choice between these two models should depend on the types of traits and historical processes thought to be important for a particular set of data. For example, the mixed model may be particularly effective with large clades evolving over very long periods of time, which might exhibit considerable phenotypic plasticity at the tips of the phylogeny (e.g., the *Daphnia* head extension example above). PGLS may be more effective with smaller clades when a single constraining force or optimum is thought to have acted throughout the history of the clade (e.g., forcing overall body length not to become too large or too small). Cheverud et al.'s (1985; Gittleman and Kot 1990) autoregressive method and Diniz-Filho et al.'s (1998) phylogenetic eigenvector regression are more similar to the mixed model than to independent contrasts and PGLS, at least in spirit, because they also partition phenotypic variation into phylogenetic and specific components. But as shown above, they are mathematically very different models than the mixed model, and direct comparisons may not be possible. In the end, a multipronged approach comparing the results of all or several of these methods on the same data set may be the most informative.

Interpretations and Applications

Although the PMM decomposition of the total phenotypic variance is analogous to the traditional quantitative genetic approach, the interpretation of its parameters is quite different. In quantitative genetics, variance within a species is partitioned into components associated with heritable and nonheritable effects, and the heritability explicitly estimates the proportion of observed variation that is due

Table 2: Results from analysis of data and phylogeny in figure 7

	Body length	Geographic range
ARM:		
Phylogenetic autocorrelation (ρ)	.04	.64
Phylogenetic inertia (R^2)	10.10%	.02%
Correlation of phylogenetic effects	.76	
Correlation of specific effects (comparable to other methods)	.15	
PGLS:		
Total phenotypic correlation	−.04	

Note: All calculations were done in COMPARE (Martins 2001). Figure 7 is available in the online edition of the *American Naturalist*. ARM = autoregressive model (Cheverud et al. 1985); PGLS = phylogenetic generalized least squares (Grafen 1989, 1992; Martins 1999).

to additive genetic causes. In contrast, at the interspecific level, the heritable component includes not only gradual genetic changes accumulated over the phylogeny but also any nongenetic response to an environment that is shared by an entire clade of organisms. The phylogenetic nonheritable component includes any short-lived change in phenotype, whether they are the result of genetic change, phenotypic plasticity, or response to changes in the environment. In the phylogenetic context, the distinction between heritable and nonheritable effects is one of time rather than mechanism, and the phylogenetic heritability is a function of the depth of the phylogeny. Since the phylogenetic component to the phenotype is modeled as evolving via Brownian motion, the heritable variation in the trait satisfies $\sigma_a^2 = mt$, where m is the rate of origin of heritable variation and t is the length of time spanned by the phylogeny. This leads to the heritability parameter satisfying $h^2 = mt/(mt + \sigma_e^2)$. All other things being equal, estimated phylogenetic heritabilities will be larger when estimated using phylogenies that span greater lengths of evolutionary time. Thus comparisons between heritabilities will be most relevant when they involve traits evolving along the same tree (with the same total time, t).

In addition to providing reasonable parameter estimates of the phylogenetic heritability and the usual bivariate correlations, the mixed model can be used to estimate several other parameters (e.g., σ^2, ρ_a, ρ_e) that would be of considerable evolutionary interest. For example, it would be very interesting to compare two component correlations (ρ_a and ρ_e), which respectively provide insight into the constraints on long- versus short-term phenotypic evolution. Also, the grand mean of the mixed model is an estimate of the ancestral state at the root of the phylogeny, with similarities to the maximum likelihood estimator of Schulter et al. (1997) and to the generalized least squares estimator of Martins and Hansen (1997) but with a different underlying model (see the online appendix for explanation of how to estimate the phenotypes of internal nodes).

Practical Considerations

Unfortunately, it is already well known that the accurate estimation of genetic and environmental correlations generally requires phenotypic data on several hundreds of independent families (Van Vleck and Henderson 1961; Brown 1969; Lynch and Walsh 1998). Thus the poor behavior of the mixed model with small sample sizes has little to do with phylogenies per se. As mentioned above, the required numbers of taxa for phylogenetic heritability analyses between species are actually smaller than the required numbers of individuals for quantitative genetic heritability analyses, due to the architectural differences between typical phylogenies and pedigrees. Although good estimates of the total phenotypic correlation (ρ) can be obtained with the small sample sizes typical of most phylogenetic studies, good estimates of the component correlations in bivariate phylogenetic analyses may require information on the mean phenotypes of hundreds of taxa. We see this as a strong reason to conduct large-scale comparative studies rather than as a weakness of the method because of the unprecedented opportunity to gain new insight into the phenotypic architecture of interspecific data.

While a simple test of whether the phylogenetic heritability is significantly different from either 0 or 1 may seem a useful way to justify formally the use of a nonphylogenetic approach or Felsenstein's (1985) independent contrasts method, we must be cautious because with the sample sizes typical of most recent phylogenetic analyses, the statistical power for rejecting these sorts of null hypotheses is generally quite limited. If h^2 is substantially lower than 1.0, independent contrasts can perform poorly and if h^2 is only a little larger than 0.0, the nonphylogenetic approach will also perform poorly. Furthermore, it is not even sufficient to consider values of h^2 close to 1.0 or 0.0. For example, when ρ_a is very different from ρ_e, independent contrasts can yield very poor estimates of the evolutionary relationship between traits (ρ), even when h^2 is relatively

large but unequal to 1.0. Thus neither independent contrasts nor the nonphylogenetic method should be applied unless the researcher is very certain that the PMM view of evolution does not apply to the traits of interest.

Within-Taxon Variation

In outlining the PMM, we treated the mean phenotype of a taxon as being a linear sum of phylogenetically heritable and nonheritable components, both of which are assumed to apply to all members of the associated taxon. If taxon-specific means are estimated with error (due, for example, to the measurement of a finite number of individuals and populations within species), there will also be a third contribution to the taxon-specific mean—a residual deviation resulting from sampling error, yielding the mathematical model $y_{ij} = \mu + a_i + e_i + \epsilon_{ij}$. As outlined in the online appendix, this third term is readily added to the mixed model. The primary modification to the analysis is the use of measures of individuals rather than of population mean phenotypes, and application of the modified mixed model equations then yields three variance components, the third of which is a measure of evolutionarily irrelevant sampling variance. In the absence of this modification, measurement-error variance will be confounded with the estimate of the phylogenetically nonheritable (but biologically relevant) component of variance, yielding slightly downward-biased estimates of the phylogenetic heritability as well as likely small biases in other parameter estimates.

A related approach was taken in PHYLIP version 3.6 (Felsenstein 2000), in which Felsenstein adapted the ideas in Lynch (1991) to develop a test for the phylogenetic effect present in a set of comparative data by incorporating within-species variation into his independent contrasts method. Specifically, the program requires data from individually measured organisms (rather than taxon mean phenotypes) and then applies the model $y_{ij} = \mu + a_i + \epsilon_{ij}$, where μ is the grand mean, a_i is the heritable effect for taxon i, and ϵ_{ij} is the independent residual for individual j in taxon i. Christman et al. (1997) used this model in their analysis of morphological data on four populations of amphipods. Felsenstein's extended model and the Christman et al. (1997) analyses still differ from the PMM in that they do not include a term for nonheritable effects (e.g., fast genetic changes) that apply to all individuals within a taxon.

Acknowledgments

We thank J. Felsenstein for helpful discussions, J. Rohlf for comments on an earlier version of this work and for sharing early copies of his manuscript, and two anonymous reviewers for their comments on the manuscript. E.A.H. also thanks J. Colbourne for extremely helpful discussions in interpreting the model in the context of *Daphnia* evolution. The work was funded by National Science Foundation grants DMS-0075143 to E.A.H., DEB-9720641 to E.P.M., and DEB-9903920 to M.L. and by National Institutes of Health grant GM36827 to M.L.

Literature Cited

Brown, G. H. 1969. An empirical study of the distribution of the sample genetic correlation coefficient. Biometrics 22:63–72.

Charnov, E. L. 1993. Life history invariants. Oxford University Press, Oxford.

Cheney, W., and D. Kincaid. 1980. Numerical mathematics and computing. Brooks/Cole, Monterey, Calif.

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39:1335–1351.

Christman, M. C., R. W. Jernigan, and D. Culver. 1997. A comparison of two models for estimating phylogenetic effect on trait variation. Evolution 51:262–266.

Colbourne, J. K., P. D. N. Hebert, and D. J. Taylor. 1997. Evolutionary origins of phenotypic diversity in *Daphnia*. Pages 163–188 *in* T. J. Givnish and K. J. Sytsma, eds. Molecular evolution and adaptive radiation. Cambridge University Press, New York.

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: 138.

Diaz-Uriarte, R., and T. Garland, Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Systematic Biology 45:27–47.

Diniz-Filho, J. A. F., C. E. Ramos De SantAna, and L. M. Bini. 1998. An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262.

Diniz-Filho, J. A. F., A. S. G. Coelho, and C. E. R. Sant'Ana. 2000. Statistical inference of correlated evolution between macroecological variables using phylogenetic eigenvector regression. Ecologia Austral 10:27–36.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:115.

———. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 192:445–471.

———. 2000. PHYLIP (Phylogeny Inference Package), version 3.6. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data. American Naturalist 160:712–726.

Gittleman, J. L., and M. Kot. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. Systematic Zoology 39:227–241.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society of London B 326: 119–157.

———. 1992. The uniqueness of the phylogenetic regression. Journal of Theoretical Biology 156:405–424.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

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

Henderson, C. R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario.

Hill, W. G., and R. Thompson. 1978. Probabilities of nonpositive definite between group or genetic covariance matrices. Biometrics 34:429–439.

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

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, Mass.

Martins, E. P. 1999. Estimation of ancestral states of continuous characters: a computer simulation study. Systematic Biology 48:642–650.

———. 2000. Adaptation and the comparative method. Trends in Ecology & Evolution 15:296–299.

———. 2001. COMPARE 4.4: for the statistical analysis of comparative data. http://compare.bio.indiana.edu. Department of Biology, Indiana University, Bloomington.

Martins, E. P., and T. Garland. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

Martins, E. P., and T. F. Hansen. 1996*a*. A microevolutionary link between phylogenies and comparative data. Pages 273–288 *in* P. Harvey, J. Maynard-Smith, and A. Leigh-Brown, eds. New uses for new phylogenies. Oxford University Press, Oxford.

———. 1996*b*. The statistical analysis of interspecific data: a review and evaluation of phylogenetic comparative methods. Pages 22–75 *in* E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford University Press, Oxford.

———. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149:646–667; erratum, American Naturalist 153:448.

Martins, E. P., J. A. Diniz-Filho, and E. A. Housworth. 2002. Adaptation and the phylogenetic comparative method. Evolution 56:1–13.

Meyer, K. 1989. Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. Genetics Selection Evolution 21:317–340.

———. 1991. Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. Genetics Selection Evolution 23:67–83.

Price, T. 1997. Correlated evolution and independent contrasts. Philosophical Transactions of the Royal Society of London B 352:519–529.

Rohlf, F. J. 2001. Geometric interpretations of comparative methods for the analysis of continuous variables. Evolution 55:2143–2160.

SAS Institute. 1990. SAS, version 6.1. SAS Institute, Cary, N.C.

Schulter, D., T. Price, A. O. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Spitze, K., and T. D. Sadler. 1996. Evolution of a generalist genotype: multivariate analysis of the adaptiveness of phenotypic plasticity. American Naturalist 148(suppl.): S108–S123.

Thompson, E. A., and R. G. Shaw. 1990. Pedigree analysis for quantitative traits: variance components without matrix inversion. Biometrics 46:399–413.

———. 1992. Estimating polygenic models for multivariate data on large pedigrees. Genetics 131:971–978.

Van Vleck, L. D., and C. R. Henderson. 1961. Empirical sampling estimates of genetic correlations. Biometrics 17:359–371.

Westoby, M., M. R. Leishman, and J. M. Lord. 1995. On misinterpreting the phylogenetic correction. Journal of Ecology 83:531–534.

Associate Editor: James M. Cheverud