General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters

J. D. HADFIELD* & S. NAKAGAWA†

*School of Biological Sciences, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK
†Department of Zoology, University of Otago, Dunedin, New Zealand

Abstract

Although many of the statistical techniques used in comparative biology were originally developed in quantitative genetics, subsequent development of comparative techniques has progressed in relative isolation. Consequently, many of the new and planned developments in comparative analysis already have well-tested solutions in quantitative genetics. In this paper, we take three recent publications that develop phylogenetic meta-analysis, either implicitly or explicitly, and show how they can be considered as quantitative genetic models. We highlight some of the difficulties with the proposed solutions, and demonstrate that standard quantitative genetic theory and software offer solutions. We also show how results from Bayesian quantitative genetics can be used to create efficient Markov chain Monte Carlo algorithms for phylogenetic mixed models, thereby extending their generality to non-Gaussian data. Of particular utility is the development of multinomial models for analysing the evolution of discrete traits, and the development of multi-trait models in which traits can follow different distributions. Meta-analyses often include a nonrandom collection of species for which the full phylogenetic tree has only been partly resolved. Using missing data theory, we show how the presented models can be used to correct for nonrandom sampling and show how taxonomies and phylogenies can be combined to give a flexible framework with which to model dependence.

Introduction

The work of Fisher, Haldane and Wright not only established the field of quantitative genetics but made substantial contributions to the field of statistics (Falconer, 1983). These statistical tools are still routinely used in comparative biology, although with a few notable exceptions (Lynch, 1991; Felsenstein, 2005; Naya *et al.*, 2006) the connection with quantitative genetics seems to have been largely lost. In this paper, we aim to reconnect quantitative genetics with comparative biology via the mixed model, highlighting solutions developed in quan-

Correspondence: Jarrod D. Hadfield, School of Biological Sciences, Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. Tel.: +44 131 650 7782; fax: +44 131 650 6564;
e-mail: j.hadfield@ed.ac.uk

titative genetics for problems that appear not to have been addressed or resolved in comparative biology.

Although used across the sciences, mixed models have their origin in quantitative genetics where a large and sophisticated, but perhaps inaccessible literature exists (Lynch & Walsh, 1998; Sorensen & Gianola, 2002; Thompson, 2008). Given their origin, it is perhaps not surprising that an early application of mixed models was to the analysis of data collected on individuals linked through a pedigree – an analysis now known as the 'animal model' (Henderson, 1976). In an important paper, Lynch (1991) showed that this same model can be applied to problems in phylogenetic comparative biology despite the difference in timescales over which shared ancestry is measured. Although Lynch's (1991) paper had received little attention until relatively

recently (Housworth et al., 2004; Felsenstein, 2008), an equivalent model (Pagel, 1999) was developed independently in the intervening period (Housworth et al., 2004).

A perceived difficulty of Lynch's (1991) original phylogenetic mixed model was that finding the maximum likelihood (ML) estimate was too computer intensive to make it a practical tool (e.g Martins, 1996; Diniz-Filho et al., 1998). However, a great deal of quantitative genetic literature had accumulated for efficiently fitting a range of large complex models (for a review, see Thompson et al., 2005) and by at least 1996 this theory had a general implementation in the program ASReml (Gilmour et al., 2002). For many data sets, Lynch's (1991) model could have been fitted in a matter of seconds using restricted maximum likelihood (REML), which became the method of choice in quantitative genetics relatively early (Patterson & Thompson, 1971). By contrast, the ML and generalized least squares (GLS) procedures advocated by Lynch (1991) and Pagel (1999) have largely been superseded in quantitative genetics due to their inherent bias and inflexibility. This bias arises because the methods fail to take into account the uncertainty in the fixed effects, resulting in downwardly biased variance components. The bias is likely to be severe in the context of phylogenetic comparative analyses because the fixed effects are associated with the ancestral state, and the ancestral state usually has high sampling error.

In this paper we start by showing that the relationship between the animal model and the phylogenetic mixed model is deeper than had been noted. The original phylogenetic mixed model was derived by making the analogy between the matrix of phylogenetic distances and the relatedness matrix defined by a pedigree. However, by expanding the phylogenetic covariance matrix to include ancestral nodes we show that these matrices also share several structural properties. More specifically, we show that a phylogeny is mathematically equivalent to an inbred pedigree, where the inbreeding coefficients are equal to the branch lengths. This relationship can be exploited in order to develop algorithms that are more accurate and orders of magnitude faster for large problems.

We go on to emphasize that general solutions and software are already available for dealing with many aspects of comparative analysis for which comparative biologists often flag as future avenues of research. We illustrate this by taking three recently published comparative papers (Ives et al., 2007; Adams, 2008; Felsenstein, 2008) and show that they can all be considered phylogenetic meta-analyses in a mixed model framework. By doing this we highlight that the original phylogenetic meta-analysis (Adams, 2008) is implemented incorrectly, and that REML estimates could have been obtained for all three models over a decade ago without the need to develop new statistical tools or software. As a worked example, we re-analyse data collected by Adams (2008) in order to test Bergmann's (1847) rule – an ecological rule predicting a positive intraspecific correlation between body size and latitude.

We go on to discuss mixed model procedures for dealing with imperfect data in the context of comparative biology. In particular, the problem of missing data has received a great deal of attention in quantitative genetics and general methods that correct for nonrandom sampling are available and well understood (e.g Im et al., 1989; Hadfield, 2008). These results are particularly important in the context of meta-analysis and comparative analysis because they may be able to correct for the publication bias that arises through nonrandom sampling of taxa, for example when common or 'fluffy' species are over-represented (Fisher et al., 2003; Nakagawa & Freckleton, 2008). In a similar vein, the availability of a complete phylogeny may not be available for all taxa, and we show how taxonomic models (Clutton-Brock & Harvey, 1977) and phylogenetic models can be combined relatively simply using standard methodology. Although not an ideal solution, the method does provide a flexible work-around for analysing data where phylogenetic information is currently incomplete.

We end by discussing phylogenetic generalized linear mixed models for non-Gaussian traits, as standard REML methods are known to be unreliable due to the intractability of the likelihood. Markov chain Monte Carlo (MCMC) methods have proved to be useful tools for solving this problem both in quantitative genetics (Sorensen & Gianola, 2002) and phylogenetics (Pagel et al., 2004; Felsenstein, 2005) and we show how efficient Gibbs samplers from quantitative genetics can be directly used for a wide range of phylogenetic methods. In particular, we discuss in detail a model where the trait can be one of $J > 2$ nominal states, as this type of model does not appear to have been used in quantitative genetics or comparative biology. The model allows the analysis of continuous and discrete characters to be brought under the same framework by shifting emphasis from evolutionary jumps between states to continuous evolution of the probability for expressing a state. In the context of phenotypic evolution, the proposed model seems to have an easier biological interpretation than currently available alternatives derived from substitution models of DNA (e.g. Pagel, 1994) because it allows for the fact that a whole host of developmental pathways are often required for the expression of complex categorical phenotypes. For example, a flightless stick insect is inherently more likely to produce a flying descendant than a flightless rodent.

Phylogenetic covariance matrix vs. the additive relationship matrix

We give a brief description of the phylogenetic covariance matrix and the additive genetic relationship matrix

because it was the link between these two concepts that allowed Lynch (1991) to develop the phylogenetic mixed model. We then show that if the phylogenetic covariance matrix is expanded to include ancestral species then these matrices are also similar in form. This may seem like a technical aside (the technical details are left to the Appendix), but we include it for two reasons. First, it makes the link between the phylogenetic mixed model and the animal model more explicit. Second, the structural properties of the additive genetic relationship matrix have played a key role in the development of robust and efficient algorithms in quantitative genetics. As phylogenetic comparative analyses become larger in scale, it will be useful and perhaps even necessary to exploit these properties.

The additive genetic relatedness matrix (**A**) is a square matrix equal in dimension to the number of individuals in the pedigree. Element A_{ij} is twice the probability that two alleles drawn at random, one from individual i and the other from individual j, are identical by descent. In the absence of inbreeding this is equivalent to the expected proportion of genes shared by two individuals (i.e. 1 if $i = j$, 0.5 if i and j are parent and offspring, 0.5 if full-sibs, 0.25 if half-sibs and so on).

In a phylogenetic context the equivalent matrix is equal in dimension to the number of species at the tips of the phylogeny. In this case the elements A_{ij} are equal to the length of the path from the most recent common ancestor of species i and j to the root of the phylogeny. Generally, the length of the path from the tips to the root of the phylogeny is scaled to unity so that the matrix is the correlation matrix with all the diagonal elements being 1.

However, in most statistical applications it is not **A** that is required, but its inverse $\mathbf{A}^{-1}$. For pedigrees this matrix can be very large and efficient ways of obtaining the inverse made the fitting of these models practical (Henderson, 1976; Quaas, 1976; Meuwissen & Luo, 1992). These algorithms usually start by inserting 'phantom parents' into the pedigree so that all individuals can be traced back to a set of unrelated parents. By analogy, we can extend the concept of the phylogenetic covariance matrix to include all ancestral nodes:

$$\mathbf{S} = \begin{bmatrix} \mathbf{F} & \mathbf{B} \\ \mathbf{B}' & \mathbf{A} \end{bmatrix} \qquad (1)$$

where **F** is a square matrix of dimension $n - 2$ (the number of internal nodes, excluding the root, where n is the number of tips). In the Appendix, we show why **S** has the same form as the complete pedigree matrix **A**, which allows us to apply Henderson's (1976) results directly to the problem of inverting **S** (i.e. $\mathbf{S}^{-1}$). This equivalence of form is due to the fact that a phylogeny has the same graph structure as a pedigree without fathers, and the branch lengths between parent and child nodes are equivalent to inbreeding coefficients.

In later sections we will often use models parametrized in terms of $\mathbf{A}^{-1}$ rather than $\mathbf{S}^{-1}$ so that the connection with earlier work on comparative analysis is clearer. However, we emphasize that it is usually better to work with the $\mathbf{S}^{-1}$ parametrization, even though this involves including $n - 2$ missing records. There are three reasons for this. First, $\mathbf{S}^{-1}$ can be formed without the need to use direct inversion techniques such as Gauss–Jordan elimination. **A** has to be inverted this way which will be slow for large phylogenies, and may suffer from numerical problems as the matrix becomes ill conditioned, which is more likely with phylogenies than pedigrees as the variation in eigenvalues is generally higher (Housworth *et al.*, 2004). Second, $\mathbf{S}^{-1}$ has reduced storage requirements because the number of nonzero elements is linear in n (i.e. $6(n - 1)$) and the matrix is said to be 'sparse'. By contrast, $\mathbf{A}^{-1}$ is dense with the number of nonzero elements nonlinear in n (i.e. n^2). Last and most importantly, the pattern of zeros in $\mathbf{S}^{-1}$ allows GLS/mixed model equations to be re-ordered in such a way that the number of arithmetic operations needed to solve them is drastically reduced (for an introduction to sparse matrix methods, see Davis, 2006).

In Appendix S1, we simulate phenotypic data for the recently published mammal super-tree of 4510 species (Bininda-Emonds *et al.*, 2007) and show that depending on the method used, fitting a model with the $\mathbf{A}^{-1}$ parametrization either fails completely or takes up to a month of computing time. By contrast, the $\mathbf{S}^{-1}$ method takes between 0.2 s and 8 min depending on the method used.

Phylogenetic meta-analysis

Using **A** to represent the phylogenetic relatedness matrix, we show that several recent comparative papers are variations of a common theme – phylogenetic meta-analysis. However, our main point is that this model is also a relatively minor variation of the basic mixed model, for which software has been available for some time.

We consider the simplest, but sufficiently general phylogenetic meta-analysis where the study test statistic (y) for study i has the form:

$$y_i = \mu + a_i + e_i + m_i \qquad (2)$$

where μ is the intercept (the ancestral state of the root node under a Brownian model), a are phylogenetic effects, e are residuals and m are study-specific measurement errors. The random effects are assumed to follow normal distributions:

$$\begin{aligned} \mathbf{a} &\sim N(\mathbf{0}, \sigma_a^2 \mathbf{A}) \\ \mathbf{e} &\sim N(\mathbf{0}, \sigma_e^2 \mathbf{I}) \end{aligned} \qquad (3)$$

where σ_e^2 is the residual variance and σ_a^2 is the phylogenetic variance. The identity matrix **I** represents the

assumption that the residuals are independent and identically distributed. σ_e^2 and σ_a^2 are generally unknown and must be estimated; however, the distribution of m are often assumed to be known, and can usually be represented by a diagonal matrix $\mathbf{M}$, where each diagonal element is the sampling variance of the published test statistic (often assumed to be the square of the standard error):

$$\mathbf{m} \sim N(\mathbf{0}, \sigma_m^2 \mathbf{M}) \qquad (4)$$

where for generality we include the measurement error variance σ_m^2, although this is set to one when the sampling variances are known. In meta-analysis, it is common to talk about weighting, and the weight matrix ($\mathbf{W}$) is equal to the inverse of $\mathbf{M}$.

The (co)variance structure of the data $\mathbf{V}$ is therefore of the form:

$$\mathbf{V} = \sigma_a^2 \mathbf{A} + \sigma_e^2 \mathbf{I} + \sigma_m^2 \mathbf{M} \qquad (5)$$

In GLS, an estimate of the fixed effects ($\hat{\mu}$) is obtained using:

$$\hat{\mu} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \qquad (6)$$

where $\mathbf{X}$ is the fixed effects design matrix (in this case an $n \times 1$ vector of ones). $\mathbf{V}^{-1}$ is assumed to be known, up to proportionality, and when this is satisfied $\hat{\mu}$ is equivalent to the best linear unbiased estimate (BLUE) of μ in a REML analysis.

The method of Adams (2008)

Garland & Ives (2000) in their non-meta-analytic approach assume that $\mathbf{M} = \mathbf{0}$, naturally, but also assume no residual error (i.e. $\sigma_e^2 = 0$). Under these assumptions, $\mathbf{A}^{-1}$ and $\mathbf{V}^{-1}$ are proportional because $\mathbf{V}^{-1} = (1/\sigma_a^2)\mathbf{A}^{-1}$. Because of these assumptions Garland & Ives (2000) are able to project $\mathbf{X}$ and $\mathbf{y}$ using the matrix $\mathbf{\Psi}$ which allows an ordinary least squares parametrization of the model. This is possible because $\mathbf{\Psi}$ is defined as the non-normalized eigenvectors of $\mathbf{A}^{-1}$, and has the property $\mathbf{\Psi}'\mathbf{A}\mathbf{\Psi} = \mathbf{I}$ and so $\mathbf{A} = (\mathbf{\Psi}')^{-1}\mathbf{\Psi}^{-1} = (\mathbf{\Psi}\mathbf{\Psi}')^{-1}$, giving:

$$
\begin{aligned}
\hat{\mu} &= ((\mathbf{\Psi}'\mathbf{X})'\mathbf{\Psi}'\mathbf{X})^{-1}(\mathbf{\Psi}'\mathbf{X})'\mathbf{\Psi}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{\Psi}\mathbf{\Psi}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Psi}\mathbf{\Psi}'\mathbf{y} \qquad (7) \\
&= (\mathbf{X}'\mathbf{A}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{-1}\mathbf{y}
\end{aligned}
$$

which is equivalent to eqn 6 only because $\mathbf{A}^{-1} \propto \mathbf{V}^{-1}$. This relationship can only be satisfied when there are no additional sources of random variation. This may be unrealistic, and particularly so in the context of meta-analysis where the aim is to take into account the variation in the precision of study-specific estimates, using weighted statistical models.

Even if $\sigma_e^2 = 0$ then the expected (co)variances under a phylogenetic meta-analytic model would have the form $\mathbf{V} = \mathbf{A}\sigma_a^2 + \sigma_m^2\mathbf{W}^{-1}$ and $\mathbf{V}$ cannot be decomposed unless σ_a^2 and σ_m^2 are known. Adams (2008), equation 6) proposes the GLS estimator:

$$
\begin{aligned}
\hat{\mu} &= ((\mathbf{\Psi}'\mathbf{X})'\mathbf{W}\mathbf{\Psi}'\mathbf{X})^{-1}(\mathbf{\Psi}'\mathbf{X})'\mathbf{W}\mathbf{\Psi}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{\Psi}\mathbf{W}\mathbf{\Psi}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Psi}\mathbf{W}\mathbf{\Psi}'\mathbf{y}
\end{aligned} \qquad (8)
$$

which implies that $(\mathbf{A}\sigma_a^2 + \sigma_m^2\mathbf{W}^{-1})^{-1} \propto \mathbf{\Psi}\mathbf{W}\mathbf{\Psi}'$. However, if this is the case then:

$$
\begin{aligned}
(\mathbf{A}\sigma_a^2 + \sigma_m^2\mathbf{W}^{-1})^{-1} &= c\mathbf{\Psi}\mathbf{W}\mathbf{\Psi}' \\
\mathbf{W}^{-1} &= c(\mathbf{I}\sigma_a^2 + \sigma_m^2\mathbf{\Psi}'\mathbf{W}^{-1}\mathbf{\Psi})
\end{aligned} \qquad (9)
$$

where c is some constant (for the full derivation, see Appendix). As $\mathbf{W}^{-1}$ is a diagonal matrix, eqn 9 can only hold if $\mathbf{W}^{-1}$ can be diagonalized by $\mathbf{\Psi}$, implying (amongst other things) that the taxa are unrelated. Clearly, such a technique is not useful in a phylogenetic context.

The method of Ives et al. (2007)

Ives et al. (2007) derive an analogous model which is correct when $\sigma_e^2 = 0$ and so the data have the expected variance structure:

$$\mathbf{V} = \sigma_a^2\mathbf{A} + \mathbf{M} \qquad (10)$$

where $\mathbf{M}$ is assumed known. Ives et al. (2007) did not explicitly mention the connection with meta-analysis, as $\mathbf{M}$ for them represents the precision of species mean estimates rather than the precision of some arbitrary effect size statistic, but eqn 10 shows the models to be analogous to meta-analysis. Ives et al. (2007) give ML, REML and estimated GLS (EGLS) procedures that estimate σ_a^2 to give estimates of μ. Ives et al. (2007) also consider the problem of the measurement error variance σ_m^2 for those cases where it is unknown but did not provide a strategy for estimating σ_m^2.

The method of Felsenstein (2008)

Felsenstein (2008), rather than working with species means, considers data points collected on individuals where multiple individuals may have been measured per species. If $\mathbf{Z}_s$ is the random effect design matrix relating records (rows) to species (columns), then the expected (co)variance between data points due to measurement error (species) is equal to $\sigma_s^2\mathbf{Z}_s\mathbf{Z}_s'$ where $\mathbf{Z}_s\mathbf{Z}_s'$ is a square block diagonal matrix which is equivalent to $\mathbf{M}$, and σ_s^2 is the within-species variance which can be interpreted as σ_m^2. The main differences between this model and Ives et al. (2007) is that the expected within-species variance is assumed to be constant across species, and that the variance (σ_s^2) is unknown and must be estimated. In Ives et al. (2007) the within-species variances are estimated on a species-by-species level, and are incorporated into the model as known quantities.

Again, it is implicitly assumed that $\sigma_e^2 = 0$ and we disagree with Felsenstein (2008) that this model is equivalent to the original phylogenetic mixed model of Lynch (1991) because the two models only coincide when the phylogenetic heritability $[H^2 = \sigma_a^2/(\sigma_a^2 + \sigma_e^2),$

which is also known as lambda (Pagel, 1999) is equal to 1]. Lynch (1991) did not assume that species means could be completely explained by Brownian motion down a phylogeny, and this is why a residual term was included to model deviations of species means from those expected. Although measurement error may be an important source of these deviations, there are a range of other processes that could cause them, and the inclusion of a residual term could be seen as a robust alternative (Lynch, 1991; Freckleton *et al.*, 2002; Housworth *et al.*, 2004; Revell *et al.*, 2008).

The method of Gilmour (*c.* 1996)

All of the above methods are standard in statistical quantitative genetics and a great deal of effort has gone into developing efficient computational strategies and understanding the properties of the REML estimators. Since 1996 at least, all of the above models could have been fitted using ASReml (Gilmour *et al.*, 2002). In the Supporting Information, we give the ASReml-R syntax for fitting the models presented above, and in each case we also include a model with a residual term. We do not present the theory, nor the algorithms involved, as this information is widely available (Gilmour *et al.*, 2002, and references therein), but note that all analyses can be fitted using a single line of code.

Worked example: a re-analysis of Bergmann's rule

Adams (2008) analysis of Bergmann's rule found a mean effect size of 0.2883 (SE 0.0301) using conventional fixed effects meta-analysis, and a mean effect size of 0.672 (SE 0.4745) using phylogenetic meta-analysis.

Using ASReml we fitted the standard meta-analysis and obtained exactly the same result as Adams (2008). However, we prefer to use random effect meta-analysis where the assumption is relaxed that the correlation between latitude and body size would tend to the same value if replication within each species was very large. This type of assumption has been criticized even in the context of controlled clinical trials (Higgins *et al.*, 2009), and would seem untenable when the data are associated with different species (West & Sheldon, 2002). Using this technique the log-likelihood increased by more than 100 indicating the model was much better. Although the mean effect size was broadly similar (0.2271), the standard error increased substantially (SE 0.1156) because the original analysis underestimated the variability considerably. The correlation was not significant at the nominal 0.05 threshold ($P = 0.057$).

Using a phylogenetic meta-analysis under the assumption that $\sigma_e^2 = 0$ we obtained a different answer from Adams (2008) due to problems with the calculations (mean effect size 0.1729 (SE 0.0995)). Relaxing the assumption that $\sigma_e^2 = 0$ we obtained a mean effect size of 0.2271 (SE 0.1156) which coincides with the nonphylogenetic random meta-analysis because the REML estimate of the phylogenetic variance was zero. It is worth noting that the standard non-meta-analytic phylogenetic model gives a different mean effect size (0.4454, SE 0.2364) and indicates a reasonable phylogenetic signal ($H^2 = 0.444$).

In the context of this study we suggest a more direct and powerful approach would be to fit a model of body size with latitude as either an additional response variable or as a fixed effect. Each study would be a separate data point with species fitted as an additional random effect. This analysis also allows species with only a single data point to be included, which are excluded from the former analysis because the correlation coefficient cannot be estimated. If these species are localized, and Bergmann's rule is the result of local adaptation, then these species may well be a nonrandom sample due to reduced gene flow, which in widespread species may weaken the relationship between body size and latitude. In later sections, we discuss multi-response models and biases resulting from missing data, both of which are relevant to meta-analysis.

Taxonomic mixed model

For many comparative analyses, the taxonomic scope is often focused enough that a complete phylogeny is available. However, meta-analyses often include a heterogeneous collection of species for which the full phylogenetic tree has only been partly resolved (for example, Kingsolver *et al.*, 2001). To accommodate this, we show how the classic nested taxonomic model (Clutton-Brock & Harvey, 1977) can be combined with the phylogenetic mixed model (Lynch, 1991; Housworth *et al.*, 2004) to give a flexible framework with which to model phylogenetic dependence. Combining these models is a direct extension of Felsenstein (2008).

In the nested taxonomic model (Clutton-Brock & Harvey, 1977) we can model the data vector (**y**) as:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}_c\mathbf{c} + \mathbf{Z}_o\mathbf{o} + \mathbf{Z}_f\mathbf{f} + \mathbf{Z}_g\mathbf{g} + \mathbf{Z}_s\mathbf{s} + \mathbf{e} \qquad (11)$$

where **c**, **o**, **f**, **g**, **s** are vectors of random effects for class, order, family, genus and species respectively. These can be thought of as the expected mean statistics for the different taxa, and these are related to the data by the random effects design matrices **Z**. These design matrices have dimensions equal to the number of data points (rows) and the number of taxa at the subscripted taxonomic level (columns). The design matrix for the vector of residuals (**e**) is assumed to be an identity matrix (**I**) and is therefore omitted. In addition, we assume a simple fixed effects structure where only an intercept is estimated. The taxonomic random effects are assumed to be identically and independently distributed (i.i.d.) and also normally distributed with a mean of zero and a

taxonomic level-specific variance. For example, the distribution of genus effects follows:

$$\mathbf{g} \sim N(\mathbf{0}, \sigma_g^2 \mathbf{I}) \qquad (12)$$

Taxonomic and phylogenetic mixed model

The taxonomic mixed model can be represented as a phylogenetic mixed model allowing taxonomy and phylogeny to be incorporated into a single analysis. Although the two types of model have different biological interpretations, they can be combined to give a description of the data that are consistent with an evolutionary process under certain, perhaps restrictive assumptions.

To illustrate the interpretational differences between the two types of model, we will start with a hypothetical taxonomy with two special properties. First, the taxonomy is an accurate description of the true phylogenetic topology which is polytomous when more than two representatives of a taxon exist. Second, the different taxonomic levels are assumed to be equally spaced in evolutionary time, and thus the taxonomy also captures phylogenetic branch lengths accurately. This second assumption will be relaxed later.

An example of such a taxonomy/phylogeny is depicted in Fig. 1 and can also be represented by the correlation (because the tree is ultrametric) matrix $\mathbf{A}$:
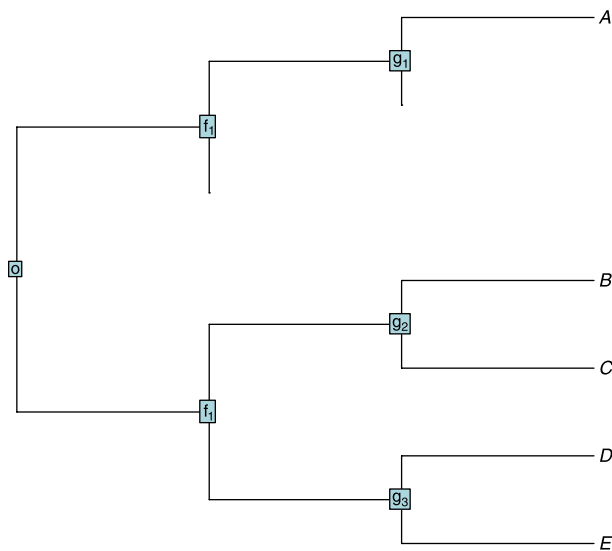
Fig. 1 A pictorial representation of a phylogeny where branching events occur at regular intervals and each set of branching events represent a different taxonomic grouping (order, family or genus). Under these conditions the phylogeny and taxonomy can be represented by scaled versions of the same picture/correlation matrix (see eqn 13).

$$\mathbf{A} = \frac{1}{3}\mathbf{Z}_f\mathbf{Z}_f' + \frac{1}{3}\mathbf{Z}_g\mathbf{Z}_g' + \frac{1}{3}\mathbf{Z}_s\mathbf{Z}_s'$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$+ \frac{1}{3}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad (13)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 1 & 2/3 \\ 0 & 1/3 & 1/3 & 2/3 & 1 \end{bmatrix}$$

With the inclusion of a residual term the expected (co)variance between data points is

$$\mathbf{V} = \sigma_f^2 \mathbf{Z}_f\mathbf{Z}_f' + \sigma_g^2 \mathbf{Z}_g\mathbf{Z}_g' + \sigma_s^2 \mathbf{Z}_s\mathbf{Z}_s' + \sigma_e^2 \mathbf{I} \qquad (14)$$

As $\mathbf{Z}_s\mathbf{Z}_s'$ is an identity matrix, species effects are confounded with the residual component (each species is measured only once) and so an identifiable taxonomic model would be:

$$\mathbf{V} = \sigma_f^2 \mathbf{Z}_f\mathbf{Z}_f' + \sigma_g^2 \mathbf{Z}_g\mathbf{Z}_g' + (\sigma_s^2 + \sigma_e^2)\mathbf{I}$$
$$\sigma_f^2 \mathbf{Z}_f\mathbf{Z}_f' + \sigma_g^2 \mathbf{Z}_g\mathbf{Z}_g' + \sigma_{e^s}^2 \mathbf{I} \qquad (15)$$

where we use the superscript s in $\sigma_{e^s}^2$ to indicate that the term includes variation at the species level (s) through to the residual level (e).

Alternatively, the model can be recast as the phylogenetic model, under the assumption that $\sigma_f^2 = \sigma_g^2 = \sigma_s^2$:

$$\mathbf{V} = \sigma_f^2 \mathbf{Z}_f\mathbf{Z}_f' + \sigma_g^2 \mathbf{Z}_g\mathbf{Z}_g' + (\sigma_s^2 + \sigma_e^2)\mathbf{I}$$
$$= \sigma_a^2 (\tfrac{1}{3}\mathbf{Z}_f\mathbf{Z}_f' + \tfrac{1}{3}\mathbf{Z}_g\mathbf{Z}_g' + \tfrac{1}{3}\mathbf{I}) + \sigma_e^2 \mathbf{I} \qquad (16)$$
$$= \sigma_a^2 \mathbf{A} + \sigma_e^2 \mathbf{I}$$

In reality, a taxonomy is unlikely to coincide with a phylogeny exactly, except perhaps in topology, and so it may make more sense to estimate the taxon-specific variances. If it is found that the taxon-specific variances are not equal, then there are two equally valid interpretations. First, it could be that the taxonomy and the phylogeny do coincide and that the different variances represent temporal variation in phylogenetic inertia (Fig. 2). Or alternatively, the phylogenetic signal may be constant over time and the different variances indicate that the taxonomic branch lengths must be rescaled in order to coincide with the real phylogeny (Fig. 3).

Either interpretation is equally valid without prior information (Paradis, 2006). However, there may be parts of a taxonomy for which the phylogeny is available,
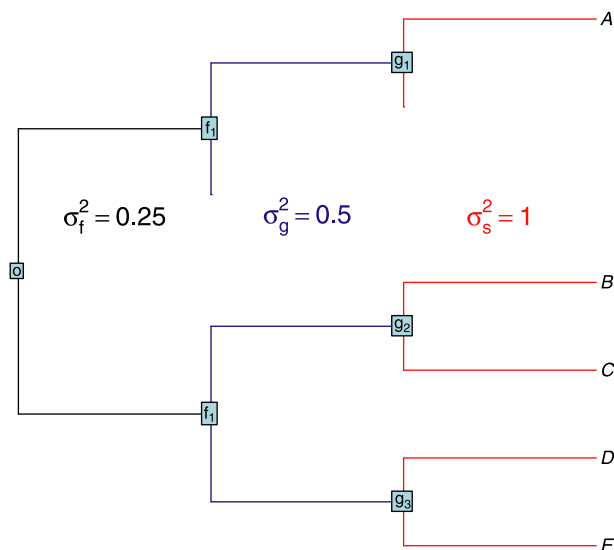
Fig. 2 A representation of a taxonomic model fitted to a phylogeny (Fig. 1) where the variance explained by each taxonomic level is found to vary.
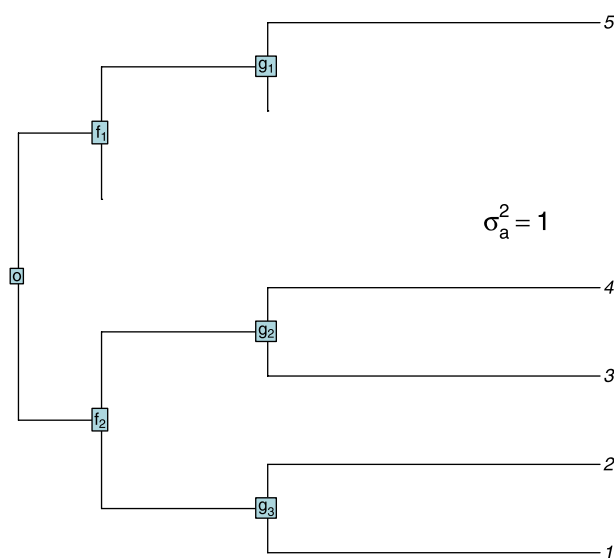
Fig. 3 A reparametrization of the taxonomic model in Fig. 2 so that the variance explained by each taxonomic level is constant, but the branch lengths connecting taxonomic levels are rescaled to give a different phylogeny.

and the assumption of a common variance across those taxonomic levels may be valid. For example, it may be that a phylogeny exists at the family level, but the classification of species into genera is by taxonomy. As in the standard taxonomic model, we can superscript phylogenetic effects with the region that they span, from the lowest taxonomic level to the highest, where $\mathbf{a}^{o:f}$

indicates that the phylogenetic effects are associated with the complete phylogeny up to the family level (assuming all taxa belong to the same order).

We could then fit the model:

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{Z}_a \mathbf{a}^{o:f} + \mathbf{Z}_g \mathbf{g} + \mathbf{Z}_s \mathbf{s} + \mathbf{e} \qquad (17)$$

assuming that there are multiple species within genera, and multiple individuals within species, so that the generic and specific variances are estimable. $\mathbf{g}$ and $\mathbf{s}$ both follow the assumption of identical and independent distribution as in the taxonomic model. However, the phylogenetic effects ($\mathbf{A}$) have expected (co)variances proportional to the phylogenetic covariance matrix ($\mathbf{A}$) from the standard phylogenetic mixed model. However, rather than the tips of the phylogeny being species, the tips represent families (for the ASReml and MCMC syntax for fitting this model, see Supporting Information).

Publication bias and missing data

An important part of meta-analysis is to assess the sensitivity of the parameter estimates to possible biases in the way the data were collected. This bias is often referred to as publication bias (Rothstein *et al.*, 2005) and it can occur at multiple stages of publication (e.g. submission, review or editorial decision). The main cause of publication bias is that statistically significant results are more likely to be published than nonsignificant results which are often relegated to the 'file drawer' (Rosenthal, 1979). This essentially results in missing data, where the probability of missingness depends on both effect size and sample size. In a comparative context, however, it is also possible that nonrandom sampling of species may happen due to species' biology and status (e.g. accessibility, abundance or conservation status), which can be referred to as taxonomic bias (Nakagawa & Freckleton, 2008). In quantitative genetics this type of bias is known as selection bias (Lush & Shrode, 1950), and can occur when individuals in the pedigree have missing phenotypes, usually because they died before they could be measured or before they expressed the trait (Im *et al.*, 1989; Hadfield, 2008). If the missing phenotypes are a nonrandom sample, then biased estimates are possible. However, the problem can be alleviated if other data have been collected that determine the relationship between phenotype and the probability of missingness (Rubin, 1976). If this is the case the data are said to be missing at random (MAR) rather than missing completely at random (MCAR), which covers the intuitive concept of randomness (Little & Rubin, 2002; Nakagawa & Freckleton, 2008). In the context of comparative analysis the condition of MAR would be satisfied if a phylogeny is available that covers those species for which trait data are unavailable, and complete measurements are available for those traits that determine the probability of missingness. For example,

information on life-history traits may be more likely to be unavailable for rare species than common species. Then, any association between abundance and life history can cause biases if the missing data are not taken into account. However, if abundance is available for all species then by including abundance in the analysis, either as a covariate or an additional response variable, unbiased estimates are possible. This is achieved by updating the missing life-history data conditional on the information provided by abundance using data augmentation, imputation or EM techniques (Fisher *et al.*, 2003; Nakagawa & Freckleton, 2008).

It is important to note that in phylogenetic meta-analysis, effect size statistics are often considered to be the species' trait, for example when the relationship between two variables or difference between two groups are to be summarized (e.g. correlation coefficient or Cohen's *d*). Therefore, phylogenetic meta-analysis of summary statistics will often suffer from both taxonomic bias and publication bias, and this can be much harder to correct for. The difficulty arises because of uncertainty in the number of missing studies, and the complicated decisions that govern the process of publication. Numerous methods have been developed to detect and to correct for publication bias, but there appears to be no consensus on a general method (Smith *et al.*, 2000; Congdon, 2003; Rothstein *et al.*, 2005). A full review is outside the scope of this paper, but we briefly discuss some simple heuristic techniques for dealing with publication bias.

Although various correlation- or regression-based methods for detection of publication bias have been suggested, they all suffer from the problem of statistical power (Macaskill *et al.*, 2001; Sterne *et al.*, 2005). This is due to the fact that publication bias is more likely to occur and to cause incorrect estimates as the number of studies used in a meta-analysis becomes smaller (Moller & Jennions, 2001). Therefore, visual inspection of publication bias such as funnel plots (Sterne & Egger, 2005) is generally more preferable (but see Kulinskaya *et al.*, 2008). Once publication bias is (visually) detected, the correction of such bias may be necessary, and there are several easy-to-use sensitivity tests available. One of these is the 'trim and fill' method (Duval & Tweedie, 2000a,b; Duval, 2005) which relies on visual assessment of funnel plot asymmetry and then, adding data points that make the plot more symmetrical by utilizing existing data points. The trim and fill method has been successfully applied to meta-analysis in ecology and evolution (Jennions & Moller, 2002). In addition, the fail-safe *N* (file-drawer number; Rosenthal, 1979) and related statistics (e.g. Orwin, 1983; Rosenberg, 2005) have often been reported as a means of assessing the validity of mean effect size estimates in meta-analysis in evolution and ecology (Moller & Jennions, 2001), although more recently these statistics have come under heavy criticism (Becker, 2005).

Markov Chain Monte Carlo techniques for non-Gaussian traits

Phylogenetic mixed models have mainly been applied to traits which are assumed to be normally distributed (for exceptions, see Felsenstein, 2005; Naya *et al.*, 2006). Generalized linear mixed models extend the linear mixed model to non-Gaussian responses, although model fitting has proved more difficult because the likelihood cannot be obtained in closed form. MCMC techniques solve this problem by breaking the high-dimensional joint distribution into a series of lower dimensional conditional distributions which are easier to sample from. By repeatedly sampling from these conditional distributions it is possible to very accurately approximate the complete joint distribution, and thereby extract things of interest (often marginal distributions).

For a thorough description of MCMC methods in the context of quantitative genetics we refer the interested reader to Sorensen & Gianola (2002). Here, we describe an MCMC algorithm for fitting a basic phylogenetic model and highlight those aspects which differ from already published results. Due to lack of space we are not able to cover the complete range of models that can be fitted using this technique, and so we restrict ourselves by highlighting a model with a nominal multinomial response, as this model does not seem to have been applied in quantitative genetics.

The general MCMC algorithm is described in detail elsewhere (Hadfield, 2010), and we also provide an R library (MCMCglmm) which is accompanied by a user manual in which the full range of supported models are discussed in greater detail. In brief, Gaussian, Poisson, Zero-inflated Poisson, Binomial, Multinomial, Ordinal and Exponential distributions are supported. More than one response variable is allowed, and the multiple responses can follow different distributions. Missing and censored data for the responses are tolerated under the assumption of a MAR process (Rubin, 1976). Any number of fixed and/or random effects can be fitted, and the random effects can be i.i.d. (as in species effects) or correlated (as in phylogenetic effects). The routines are fast, as all posterior simulations are done in compiled C++ using direct methods for sparse linear systems (Davis, 2006).

The basic model

With a Gaussian response the linear model is applied to *y*:

$$\mathbf{y} = \mathbf{W}\theta + \mathbf{e} \qquad (18)$$

where **W** is a design matrix which relates the predictors to the data, θ is a vector of location effects (fixed and random effects), and **e** is vector of residuals.

With non-Gaussian data a latent variable (*l*) is introduced which is the canonical parameter of some distribution on the link scale. For example, if datum y_i is

Poisson distributed and a log link specified then the assumption is:

$$y_i \sim \text{Pois}(\lambda = \exp(l_i)) \quad (19)$$

where λ is the canonical parameter of the Poisson distribution (often called the rate parameter or mean parameter) and exp is the inverse link function.

In this case the linear model is applied to l:

$$\mathbf{l} = \mathbf{W}\theta + \mathbf{e} \quad (20)$$

There is little distinction between fixed and random effects in a Bayesian analysis (hence we represent both with θ). The 'fixed' effects are usually assumed *a priori* to be independently distributed about zero with *specified* variance (σ_B^2):

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \sigma_B^2\mathbf{I}) \quad (21)$$

Usually σ_B^2 is set to something large in order to represent diffuse prior knowledge. The 'random' effects are also assumed to come from a normal prior distribution with zero mean, but the variance (σ_a^2) is usually inferred *a posteriori*:

$$\boldsymbol{a} \sim N(\mathbf{0}, \sigma_a^2\mathbf{A}) \quad (22)$$

and in the case of phylogenetic effects the identity matrix ($\mathbf{I}$) is replaced by a relationship matrix ($\mathbf{A}$). In the simplest case, the residuals are assumed to be i.i.d.:

$$\boldsymbol{e} \sim N(0, \sigma_e^2\mathbf{I}) \quad (23)$$

where the residual variance σ_e^2 is also estimated.

Parameter estimation

There are three types of parameter to estimate: the latent variables (l), location effects (θ) and the variance components σ_a^2 and σ_e^2. We describe the basic sampling schemes using the example of a single trait following a Poisson distribution.

The latent variables (l) do not have a recognizable conditional distribution and so we sample them one at a time using Metropolis–Hastings updates (Metropolis *et al.*, 1953; Hastings, 1970). The conditional probability of the latent variable is proportional to the product of two terms: the Poisson likelihood of the data given l and the normal likelihood of e given a mean of zero and variance σ_e^2.

The location effects are multivariate normal, conditional on the latent variables and variance components, and can be Gibbs sampled (Geman & Geman, 1984). We use the Gibbs sampling method of Garcia-Cortes & Sorensen (2001) which updates all the location effects in a single pass and avoids inverting the mixed model coefficient matrix.

The variance components follow scaled inverted chi-squared distributions with scale equal to the cross-product of the 'random' location effects, or the cross-product of the residuals in the case of σ_e^2. As the

variance components come from a known distribution they can be Gibbs sampled also. If the random effects are correlated, as they will be for the phylogenetic effects, the scale matrix is obtained using $\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}$ rather than the direct cross-product $\mathbf{a}'\mathbf{I}^{-1}\mathbf{a} = \mathbf{a}'\mathbf{a}$.

Extensions to multiple responses

Multiple response models are not widely used in comparative biology but can be useful in many situations. Extending the model to the multivariate case is straightforward by concatenating the data vectors and latent variables for each trait and structuring the mixed model equations accordingly. For simplicity, we consider a bivariate Poisson analysis with response vector

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^{(1)} \\ \mathbf{Y}^{(2)} \end{bmatrix}$$

and latent variable vector

$$\mathbf{l} = \begin{bmatrix} \mathbf{l}^{(1)} \\ \mathbf{l}^{(2)} \end{bmatrix}.$$

In multi-response models it is usual to replace the variance components with their multivariate analogues, (co)variance matrices, which denote the variance within each trait and between each trait for the designated source of variance. For example, if $\mathbf{V}_a$ is a 2×2 matrix with the phylogenetic variance for the two traits along the diagonal, and the covariance between the phylogenetic effects for the two traits in the off-diagonals, then the complete set of phylogenetic effects

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}^{(1)} \\ \mathbf{a}^{(2)} \end{bmatrix}$$

have the expected distribution $N(\mathbf{0}, \mathbf{V}_a \otimes \mathbf{A})$ where $\otimes$ is the Kronecker product.

The latent variables in the multi-response model are sampled using Metropolis–Hastings updates as before, although they are updated in blocks corresponding to observational units (i). In the bivariate case, this involves updating the latent variables $l_i^{(1)}$ and $l_i^{(2)}$ together, where the likelihood is proportional to the product of three densities: the Poisson density of $y_i^{(1)}$ given $l_i^{(1)}$, the Poisson density of $y_i^{(2)}$ given $l_i^{(2)}$, and the density of the residuals

$$\begin{bmatrix} e_i^{(1)} \\ e_i^{(2)} \end{bmatrix}$$

from a bivariate normal distribution with null mean vector and (co)variance matrix $\mathbf{V}_e$ (the multivariate analogue of σ_e^2).

The location effects remain multivariate normal as before and so can be Gibbs sampled in a single pass following the multivariate extension to the Gibbs sampling method of Garcia-Cortes & Sorensen (2001) (Korsgaard *et al.*, 2003). The covariance matrices are Gibbs sampled using the multivariate analogue of the

scaled inverse chi-squared distribution: the inverse Wishart distribution. The scale matrix in the multivariate case has a similar form: $[\mathbf{a}^{(1)}\mathbf{a}^{(2)}]'\mathbf{A}^{-1}[\mathbf{a}^{(1)}\mathbf{a}^{(2)}]$.

Although similar in many respects to Felsenstein's (2005) application of MCMC to the comparative analysis of binary traits, the method differs in three respects. First, the model is constructed explicitly as a generalized linear model so that the concept of the latent variable can be extended beyond binary traits to other types of distribution. Second, the model can be more easily generalized to multi-trait models where the different traits can follow different distributions, and lastly all fixed and random effects are sampled simultaneously from their multivariate conditional distribution rather than one at a time from univariate conditional distributions. This is expected to improve the efficiency of the algorithm substantially (Roberts & Sahu, 1997).

Multinomial phylogenetic mixed model

We briefly describe some results additional to those described above that are required to fit a multinomial logit model for more than two nominal categories. The model does not appear to have been used in quantitative genetics but is quite widely used in econometrics and political science (Congdon, 2005). However, fitting such a model is a direct extension of a multivariate binary mixed model, with additional constraints on parameter space and a slight modification to the latent variable likelihood. If the categories are ordered, then it is possible to work with a different parametrization of the model presented below (Hedeker, 2003), or alternatively the ordered multinomial probit model can be used, which has been used in quantitative genetics (Gianola & Foulley, 1983).

In a binary model, a single data point can be one of two categories ($J = 2$), and this can be expressed as the univariate binary variable. Likewise, if $J > 2$ then it is usual to use a transformation that reduces the problem to $J - 1$ dimensions (Daganzo, 1979). In the binary model, the motivation for the dimension reduction is obvious; if a variable increased the probability of expressing the first category by 10%, it must by necessity reduce the probability of expressing the second category by 10% because an individual cannot express both categories simultaneously. The dimension reduction essentially constrains the probability of expressing the first or the second trait to unity [$\Pr(Y_i = j_1) + \Pr(Y_i = j_2) = 1$]. For the three-trait case, we will think of the three colours: red, black and white. Denoting α_{ij} as the probability that species i is colour j, the unit sum constraint has the form $\sum_{j=1}^{J} \alpha_{ij} = 1$. To reduce the problem into $J - 1$ dimensions, it is usual to work with the parametrization in terms of a logs odd ratio with respect to an arbitrary baseline category (we will use the first category – red) so that

$$l_{ij} = \log\left(\frac{\alpha_{ij}}{\alpha_{i1}}\right) = \log(\alpha_{ij}) - \log(\alpha_{i1}).$$

l_{ij} is the latent variable for individual i and colour j. The problem can be represented using the contrast matrix Δ (Bunch, 1991):

$$\mathbf{\Delta} = \begin{bmatrix} -1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{24}$$

For a simple fixed effects model:

$$\exp\left((\mathbf{\Delta\Delta'})^{-1}\mathbf{\Delta X}_i\boldsymbol{\beta}\right) \propto E \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \alpha_{i3} \end{bmatrix} \tag{25}$$

where $\mathbf{X}_i$ is the design matrix for species i and has $J - 1$ rows. Likewise, the residual ($\mathbf{V}_e$) and any random effect (e.g. $\mathbf{V}_a$) covariance matrices are for estimability purposes estimated on the $J - 1$ space: $\mathbf{V}_a = \mathbf{\Delta'V}_{\tilde{a}}\mathbf{\Delta}$ and $\mathbf{V}_e = \mathbf{\Delta'V}_{\tilde{e}}\mathbf{\Delta}$, where the tilde indicates a covariance matrix on the scale of the three $\log(\alpha)$'s. Moreover, as there is only a single realization from the multinomial, then no element of $\mathbf{V}_e$ is estimable and must be fixed. A choice for $\mathbf{V}_e$ is essentially arbitrary, and we choose to fix $\mathbf{V}_e = \frac{1}{J}(\mathbf{I} + \mathbf{J})$ where $\mathbf{J}$ is the unit matrix. We can visualize the unit sum constraint for three categories as a model parametrized on the simplex (Fig. 4).

Discussion

In this paper, we develop both classic and new results from quantitative genetics in the context of phylogenetic comparative analysis. We show how the relationship between pedigree structure and phylogeny structure can be used to exploit efficient algorithms already developed for 'animal model' analyses. We also suggest that many of the techniques currently being developed in comparative biology have already existed as standard REML tools in quantitative genetics for at least a decade. More recently, Bayesian models have gained popularity in quantitative genetics for certain types of problem for which standard techniques were known to behave poorly (Sorensen & Gianola, 2002; O'Hara et al., 2008). Many of these problems also exist for the comparative method, and we show how quantitative genetic MCMC algorithms can be implemented for phylogenies. Taken together, these results not only demonstrate that the theory and software for fitting the original phylogenetic mixed model is well developed, but also that extensions dealing with meta-analysis, measurement error, missing data, multi-trait models, non-Gaussian data, small-sample inference and uncertainty in derived quantities such as phylogenetic heritability are readily available. In addition, we extend the quantitative genetic models to give a multinomial phylogenetic mixed model for analysing categorical traits: a type of model that does not appear to have
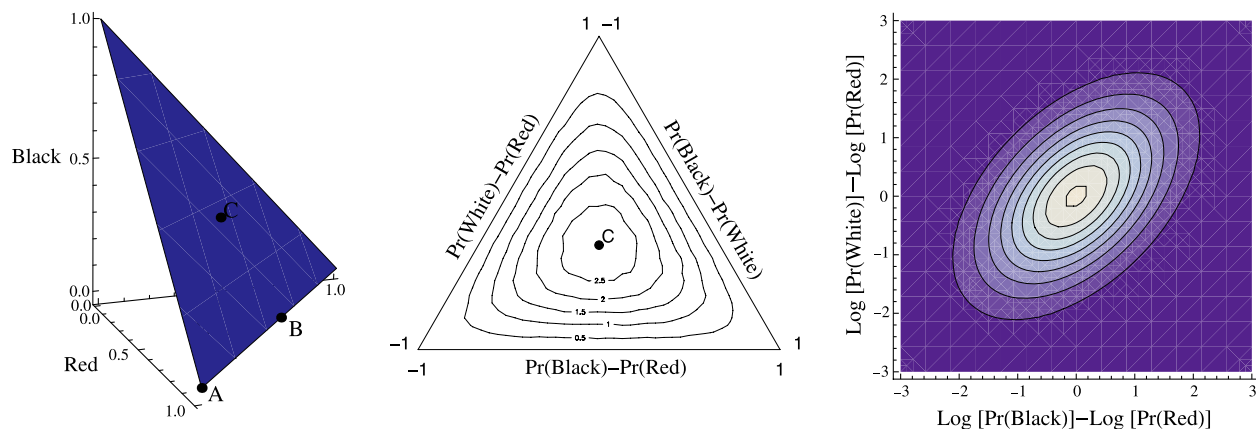
Fig. 4 The three axes in the left panel represent the probability of belonging to one of the three colours: red, back and white. In this example, we assume that there are only three possible colours and so the two-simplex (triangle) represents the parameter space where the probability of being one of the three colours is equal to one. With more than three categories, the model is difficult to represent, but the parameter space of a four-category model would be a three-simplex (tetrahedron) in a four-dimensional subspace. Three points have been plotted in parameter space (A, B and C) and for ease of interpretation, let us assume that these are the probabilities associated with an ancestral state. If the phylogenetic heritability is high (or the phylogenetic distance between ancestor and descendant small) then these probabilities should predict well the probabilities associated with their descendants. For example, descendants of species A should be red with high probability, and descendants of species B should be red or white with equal probability but are unlikely to be black. Finally, descendants of species C are equally likely to express any one of the three colours. It should be understood that the uncertainty regarding the expressed colours of species B and particularly species C, does not reflect uncertainty in the underlying probability; species C really does have an equal propensity to express any of the colours and the uncertainty is associated with exactly which colour is manifest. Although predictions of individual random effects (BLUPS) could be plotted as above, it is more usual to interpret the distribution of the random effects. When these distributions are multivariate normal they can be represented by ellipses circumscribing regions of trait space that have equal density. The right panel shows the hypothetical distribution $\frac{1}{3}(\mathbf{I} + \mathbf{J})$ on the scale of the log contrasts, the scale on which the assumption of normality was made (which includes the scale of the log probabilities log(α) also). The distribution is nonstandard on the inverse link (probability) scale, and so cannot be represented by ellipses on the simplex. However, the probability distribution can be approximated using posterior simulation as shown in the central panel. The distribution is seen to have triangular axes of symmetry and this is the motivation behind the choice of $\frac{1}{3}(\mathbf{I} + \mathbf{J})$ for $\mathbf{V}_e$.

been used in either field, quantitative genetics or comparative biology.

The multinomial model developed in this paper follows a logic similar to the threshold model of Wright (1934a,b) for binary characters in that a continuous polygenic probability is postulated that underlies which state is manifest. The model received some of the earliest Bayesian treatments in quantitative genetics (Foulley *et al.*, 1983; Gianola & Foulley, 1983), and was followed by the development of MCMC procedures (Sorensen *et al.*, 1995) which are discussed at length in Sorensen & Gianola (2002). Felsenstein (2005) discusses this threshold model in the context of phylogenies and provides alternative computational strategies for estimating the relevant parameters also, using MCMC. Alternative methods exist, mainly based around models of substitutions in DNA sequences (Pagel, 1994; Huelsenbeck *et al.*, 2003; Pagel & Meade, 2006), but Felsenstein (2005) gives a persuasive argument as to why the quantitative genetic approach is to be preferred for the comparative analysis of categorical phenotypes. From a statistical perspective, a multi-trait threshold model can often be parametrized with far fewer parameters, which given the small amount of information in phylogenies reduces the chances of trying to fit over identified models (Felsenstein, 2005). From a biological perspective, we believe the substitution model is harder to interpret in terms of phenotypic evolution because it assumes that the probability from jumping from a zero to one state is constant over the phylogeny. When focus shifts to the underlying probability, species that have zero states, but that are found in clades with many species in the one state, have an increased chance of flipping to the one state relative to a species found in clades dominated by zero states. This reasoning seems natural because the expression of phenotypes, even categorical ones, are often dependent on a whole range of developmental and biochemical processes being in place. For example, the re-appearance of wings in a wingless lineage of stick insects should be much less surprising than the appearance *de novo* of wings on guinea pigs. This being said, both the threshold model and the substitution model may be equivalent in the univariate case, and the differences may be merely interpretational; the substitution model would, no doubt, infer a wingless ancestral state to the rodents with very high confidence.

As Housworth *et al.* (2004) noted, although phylogenies tend to be more informative than equally sized pedigrees, typically phylogenies are orders of magnitude smaller than pedigrees in evolutionary biology. As measures of uncertainty and hypothesis testing are often conducted using Fisher information or likelihood ratio tests, both of which rely on large-sample asymptotic behaviour, statistical inference from small sample sizes should be treated with some caution. Although resampling techniques are often used to overcome this problem, there seems to be little formal justification for the resampling strategies employed (but see Lapointe & Garland, 2001), which is surprising given that resampling methods for dependent data are usually nontrivial to develop (Shao & Tu, 1996). Furthermore, naive application of basic resampling methods often give incorrect results (Rao & Wu, 1988). For example, the validity of permuting species data over the phylogeny in a phylogenetic meta-analysis is unclear given that permutation tests require the data to be exchangeable under the null hypothesis. In Adams (2008) model, the stated null hypothesis was a zero effect size, not a lack of phylogenetic inertia; so, it seems unlikely that the data are expected to be independent under the null hypothesis. It would seem that the permutation test will give the distribution of the test statistic in the absence of phylogenetic inertia, irrespective of whether Bergmann's rule exists or not. Bayesian inference makes no large-sample approximations and the resulting posterior distributions are an accurate description of uncertainty given the probability model (Gelman *et al.*, 2004). However, with small phylogenies the relative importance of prior information may increase, and effort should be made in obtaining accurate prior information and checking the sensitivity of the results to alternative prior specifications.

Although the generalized phylogenetic mixed model is a flexible tool for comparative analysis, having a range of other models as special cases (e.g. independent contrasts, Felsenstein, 1985; nested taxonomic model, Clutton-Brock & Harvey, 1977), there are alternative comparative methods that are based on different variance structures (Hansen & Martins, 1996; Martins & Hansen, 1997). These alternative models are often identical to well-known models in time-series analysis and geostatistics (Ives & Zhu, 2006); for example, the simple Ornstein–Uhlenbeck model (Hansen & Martins, 1996) is equivalent to the isotropic exponential; the continuous time analogue of the first-order autoregressive model. We are not familiar with these fields but given they are both large and have a long history we suspect that many of the problems involved with model fitting, over-identification, missing data and prediction already have good and well-tested solutions for these types of model. Following Ives & Zhu (2006), we stress that although phylogenies are unique and inherently interesting to biologists, statistically they are little different from space, time or pedigrees, all of which have been the focus of much statistical research.

References

Adams, D.C. 2008. Phylogenetic meta-analysis. *Evolution* **62**: 567–572.

Becker, B.J. 2005. Fail safe n or file-drawer number. In: *Publication Bias in Meta-analysis : Prevention, Assessment and Adjustments* (H. Rothstein, A.J. Sutton & M. Borenstein, eds.), pp. 111–125. John Wiley, Chichester.

Bergmann, C. 1847. Über die verhältnisse der wärmeökonomie der thiere zu ihrer grosse. *Göttinger Studien* **1**: 595–708.

Bininda-Emonds, O.R.P., Cardillo, M., Jones, K.E., MacPhee, R.D.E., Beck, R.M.D., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. & Purvis, A. 2007. The delayed rise of present-day mammals. *Nature* **446**: 507–512.

Bunch, D.S. 1991. Estimability in the multinomial probit model. *Transport. Res. B-Meth.* **25**: 1–12.

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

Congdon, P. 2003. *Applied Bayesian Modelling. Wiley Series in Probability and Statistics*. Wiley, Chichester.

Congdon, P. 2005. *Bayesian Models for Categorical Data. Wiley Series in Probability and Statistics*. Wiley-Blackwell, Chichester.

Daganzo, C. 1979. *Multinomial Probit: The Theory and Its Application to Demand Forecasting*. Academic Press, New York.

Davis, T.A. 2006. *Direct Methods for Sparse Linear Systems*. Fundamentals of Algorithms. SIAM, Philadelphia, PA.

Dempster, A.P., Laird, N.M. & Rubin, D.B. 1977. Maximum likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **39**: 1–38.

Diniz-Filho, J.A.F., De Sant'ana, C.E.R. & Bini, L.M. 1998. An eigenvector method for estimating phylogenetic inertia. *Evolution* **52**: 1247–1262.

Duval, S. 2005. The trim and fill method. In: *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments* (H. Rothstein, A.J. Sutton & M. Borenstein, eds.), pp. 127–144. John Wiley, Chichester.

Duval, S. & Tweedie, R. 2000a. A nonparametric ''trim and fill'' method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.* **95**: 89–98.

Duval, S. & Tweedie, R. 2000b. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**: 455–463.

Falconer, D. 1983. *Introduction to Quantitative Genetics*. Longman Group, Harlow, UK.

Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* **125**: 1–15.

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**: 1427–1434.

Felsenstein, J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* **171**: 713–725.

Fisher, D.O., Blomberg, S.P. & Owens, I.P.F. 2003. Extrinsic versus intrinsic factors in the decline and extinction of Australian marsupials. *Proc. R. Soc. Lond. B Biol. Sci.* **270**: 1801–1808.

Foulley, J.L., Gianola, D. & Thompson, R. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth-weight and pelvic opening. *Genet. Sel. Evol.* **15**: 401–423.

Freckleton, R.P., Harvey, P.H. & Pagel, M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**: 712–726.

Garcia-Cortes, L.A. & Sorensen, D. 2001. Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genet. Sel. Evol.* **33**: 443–452.

Garland, T. & Ives, A.R. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* **155**: 346–364.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004. *Bayesian Data Analysis. Texts in Statistical Science.* Chapman & Hall, Boca Raton, 2nd edn.

Geman, S. & Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEET. Pattern Anal.* **6**: 721–741.

Gianola, D. & Foulley, J.L. 1983. Sire evaluation for ordered categorical-data with a threshold-model. *Genet. Sel. Evol.* **15**: 201–223.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J. & Thompson, R. 2002. ASReml User Guide Release 1.0 [www document]. URL http://www.VSN-Intl.com.

Hadfield, J.D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. Biol. Sci. Ser. B* **275**: 723–734.

Hadfield, J. 2010. MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* (in press).

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

Hastings, W.K. 1970. Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.

Hedeker, D. 2003. A mixed-effects multinomial logistic regression model. *Stat. Med.* **22**: 1433–1446.

Henderson, C.R. 1976. Simple method for computing inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**: 69–83.

Higgins, J.P.T., Thompson, S.G. & Spiegelhalter, D.J. 2009. A re-evaluation of random-effects meta-analysis. *J. R. Stat. Soc. Ser. A-Stat. Soc.* **172**: 137–159.

Housworth, E.A., Martins, E.P. & Lynch, M. 2004. The phylogenetic mixed model. *Am. Nat.* **163**: 84–96.

Huelsenbeck, J.P., Nielsen, R. & Bollback, J.P. 2003. Stochastic mapping of morphological characters. *Syst. Biol.* **52**: 131–158.

Im, S., Fernando, R.L. & Gianola, D. 1989. Likelihood inferences in animal breeding under selection – a missing-data theory view point. *Genet. Sel. Evol.* **21**: 399–414.

Ives, A.R. & Zhu, J. 2006. Statistics for correlated data: phylogenies, space, and time. *Ecol. Appl.* **16**: 20–32.

Ives, A.R., Midford, P.E. & Garland, T. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* **56**: 252–270.

Jennions, M.D. & Moller, A.P. 2002. Relationships fade with time: a meta-analysis of temporal trends in publication in ecology and evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **269**: 43–48.

Kingsolver, J.G., Gomulkiewicz, R. & Carter, P.A. 2001. Variation, selection and evolution of function-valued traits. *Genetica* **112**: 87–104.

Korsgaard, I.R., Lund, M.S., Sorensen, D., Gianola, D., Madsen, P. & Jensen, J. 2003. Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Sel. Evol.* **35**: 159–183.

Kulinskaya, E., Morgenthaler, S. & Staudte, R.B. 2008. *Meta Analysis: A Guide to Calibrating and Combining Statistical Evidence.* John Wiley, Chichester.

Lapointe, F.J. & Garland, T. 2001. A generalized permutation model for the analysis of cross-species data. *J. Classif.* **18**: 109–127.

Little, R.J.A. & Rubin, D.B. 2002. *Statistical Analysis with Missing Data.* Wiley-Interscience, Hoboken, NJ, 2nd edn.

Lush, J.L. & Shrode, R.R. 1950. Changes in milk production with age and milking frequency. *J. Dairy Sci.* **33**: 338–357.

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

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits.* Sinauer, Sunderland, MA.

Macaskill, P., Walter, S.D. & Irwig, L. 2001. A comparison of methods to detect publication bias in meta-analysis. *Stat. Med.* **20**: 641–654.

Martins, E.P. 1996. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. *Evolution* **50**: 1750–1765.

Martins, E.P. & Hansen, T.F. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**: 646–667.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**: 1087–1092.

Meuwissen, T.H.E. & Luo, Z. 1992. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* **24**: 305–313.

Moller, A.P. & Jennions, M.D. 2001. Testing and adjusting for publication bias. *Trends Ecol. Evol.* **16**: 580–586.

Nakagawa, S. & Freckleton, R.P. 2008. Missing inaction: the dangers of ignoring missing data. *Trends Ecol. Evol.* **23**: 592–596.

Naya, H., Gianola, D., Romero, H., Urioste, J.I. & Musto, H. 2006. Inferring parameters shaping amino acid usage in prokaryotic genomes via Bayesian MCMC methods. *Molecular Biology and Evolution* **23**: 203–211.

O'Hara, R.B., Cano, J.M., Ovaskainen, O., Teplitsky, C. & Alho, J.S. 2008. Bayesian approaches in evolutionary quantitative genetics. *J. Evol. Biol.* **21**: 949–957.

Orwin, R.G. 1983. A fail-safe *n* for effect size n meta-analysis. *J. Educ. Stat.* **8**: 157–159.

Pagel, M. 1994. Detecting correlated evolution on phylogenies – a general-method for the comparative-analysis of discrete characters. *Proc. R. Soc. Lond. B Biol. Sci.* **255**: 37–45.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.

Pagel, M. & Meade, A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *Am. Nat.* **167**: 808–825.

Pagel, M., Meade, A. & Barker, D. 2004. Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.* **53**: 673–684.

Paradis, E. 2006. *Analysis of Phylogenetics and Evolution with R.* Springer-Verlag, New York.

Patterson, H.D. & Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545–554.

Quaas, R.L. 1976. Computing diagonal elements and inverse of a large numerator relationship matrix. *Biometrics* **32**: 949–953.

Rao, J.N.K. & Wu, C.F.J. 1988. Resampling inference with complex survey data. *J. Am. Stat. Assoc.* **83**: 231–241.

Revell, L.J., Harmon, L.J. & Collar, D.C. 2008. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **57**: 591–601.

Roberts, G.O. & Sahu, S.K. 1997. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **59**: 291–317.

Rosenberg, M.S. 2005. The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution* **59**: 464–468.

Rosenthal, R. 1979. The ''file drawer problem'' and tolerance for null results. *Psychol. Bull.* **86**: 638–641.

Rothstein, H., Sutton, A.J. & Borenstein, M. (eds.), 2005. *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments*. John Wiley, Chichester.

Rubin, D.B. 1976. Inference and missing data. *Biometrika* **63**: 581–590.

Shao, J. & Tu, S. 1996. *The Jackknife and Bootstrap. Springer Series in Statistics*. Springer, New York.

Smith, D., Givens, G.H. & Tweedie, R. 2000. Adjusting for publication and quality bias in Bayesian meta-analysis. In: *Meta-analysis in Medicine and Health Policy* (D. Stangl & D. Berry, eds.), pp. 277–304. Marcel Dekker, New York.

Sorensen, D. & Gianola, D. 2002. *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics. Statistics for Biology and Health*. Springer-Verlag, New York.

Sorensen, D.A., Andersen, S., Gianola, D. & Korsgaard, I. 1995. Bayesian-inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* **27**: 229–249.

Sterne, J.A.C. & Egger, M. 2005. Regression method to detect publication and other bias in meta-analysis. In: *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments* (H. Rothstein, A.J. Sutton & M. Borenstein, eds), pp. 99–125. John Wiley, Chichester.

Sterne, J.A.C., Becker, B.J. & Egger, M. 2005. The funnel plot. In: *Publication Bias in Meta-analysis : Prevention, Assessment and Adjustments* (H. Rothstein, A.J. Sutton & M. Borenstein, eds), pp. 75–98. John Wiley, Chichester.

Tanner, M.A. & Wong, W.H. 1987. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* **82**: 528–540.

Thompson, R. 2008. Estimation of quantitative genetic parameters. *Proc. R. Soc. Biol. Sci. Ser. B* **275**: 679–686.

Thompson, R., Brotherstone, S. & White, I.M.S. 2005. Estimation of quantitative genetic parameters. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**: 1469–1477.

West, S.A. & Sheldon, B.C. 2002. Constraints in the evolution of sex ratio adjustment. *Science* **295**: 1685–1688.

Wright, S. 1934a. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* **19**: 0506–0536.

Wright, S. 1934b. The results of crosses between inbred strains of guinea pigs, differing in number of digits. *Genetics* **19**: 0537–0551.

Appendix

To obtain (E)GLS, REML or Bayesian estimates, it is necessary to form $\mathbf{A}^{-1}$ or $\mathbf{S}^{-1}$ where

$$\mathbf{S} = \begin{bmatrix} \mathbf{F} & \mathbf{B} \\ \mathbf{B}' & \mathbf{A} \end{bmatrix} \qquad (26)$$

$\mathbf{F}$ is a square matrix of dimension $n - 2$ (the number of internal nodes, excluding the root, where n is the number of tips). Following Henderson (1976), the lower triangle matrix ($\mathbf{L}$) of the Cholesky decomposition $\mathbf{S} = \mathbf{L}\mathbf{L}'$ can be formed recursively:

$$L_{ti} = \begin{cases} L_{p^t i}, & \text{for } i = 1, \ldots, p^t, \\ 0, & \text{for } i = p^t + 1, \ldots, t - 1, \\ \sqrt{f_t}, & \text{if } i = t, \end{cases} \qquad (27)$$

where p^t indexes the parental node of t and f_t is the length of the branch connecting node t to p^t. It is interesting to note that f_t is equivalent to the inbreeding coefficient in the context of a pedigree.

Note that $\mathbf{L} = \mathbf{TD}$ where $\mathbf{D}$ is a diagonal matrix with diagonal elements equal to those of $\mathbf{L}$ ($\sqrt{f_t}$), and $\mathbf{T}$ has the same nonzero pattern as $\mathbf{L}$, but with all nonzero elements equal to one. Henderson (1976) shows that $\mathbf{T}$ is easy to invert, with $\mathbf{T}^{-1}$ being a lower triangle matrix with all diagonal elements equal to 1, and all nonzero elements left to the diagonal of the tth row being -1 for columns corresponding to the node t^p. This is useful because:

$$\begin{aligned} \mathbf{S}^{-1} &= (\mathbf{L}\mathbf{L}')^{-1} \\ &= (\mathbf{TDDT}')^{-1} \qquad (28) \\ &= (\mathbf{T}')^{-1}\mathbf{D}^{-2}\mathbf{T}^{-1} \end{aligned}$$

and because $\mathbf{D}$ is diagonal, $\mathbf{D}^{-2}$ is also diagonal, each diagonal element of $\mathbf{D}$ taken to the power -2. This leads to a simple recursive algorithm for the inverse:

$$S_{ti}^{-1} = \begin{cases} -\frac{1}{f_t}, & \text{if } i = p^t \text{ or } i \in \mathcal{O}_t, \\ \frac{1}{f_t} + \sum_o^{\mathcal{O}_t} \frac{1}{f_o}, & \text{if } t = i, \\ 0, & \text{otherwise,} \end{cases} \qquad (29)$$

where $\mathcal{O}_t$ are the set of t's child nodes.

Using expectation–maximization (EM) (Dempster *et al.*, 1977) techniques for REML or data augmentation (Tanner & Wong, 1987) techniques for a Bayesian analysis, we can treat the ancestral states as missing data and work with the $\mathbf{S}^{-1}$ parametrization rather than the usual $\mathbf{A}^{-1}$ parametrization.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Additional Supporting Information may be found in the online version of this article:
General quantitative genetic methods for comparative biology.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.