Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2015, **6,** 706–714 doi: 10.1111/2041-210X.12354

Increasing the efficiency of MCMC for hierarchical phylogenetic models of categorical traits using reduced mixed models

Jarrod D. Hadfield*

Institute of Evolutionary Biology, The University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK

Summary

1. Integrating out the random effects in generalised linear mixed models (GLMM) cannot be done analytically unless the response is Gaussian. Many stochastic, deterministic or hybrid algorithms have been developed to perform the integration. With categorical data and probit link (aka the threshold model), the random effect structure can be partitioned into a part that can be easily integrated deterministically (the R-structure) and a part that cannot (the G-structure).

2. We show that in the context of phylogenetic mixed models, part of the G-structure (the phylogenetic effects at the tips) can be moved into the R-structure and integrated out deterministically. This result follows directly from the concept of the reduced animal model from quantitative genetics (*Journal of Animal Science*, 51, 1980, 1277) and its implications for discrete data (*Genetics Selection Evolution*, 42, 2010, 1).

3. Although the conditional distribution of the phylogenetic variance is no longer in standard from, it does provide a stable and efficient 2-block MCMC algorithm for situations when the phylogenetic heritability is assumed to be one. We show that a GLMM with such an assumption is equivalent to the model proposed by Felsenstein (*American Naturalist*, 179, 2005, 145). Extensions to multivariate models are straightforward and a 3-block algorithm can be constructed when there is only a single categorical trait but multiple Gaussian traits. With ≥2 categorical traits, an additional non-Gibbs update is required for the correlation (sub)matrix.

4. An implementation of these algorithms is distributed in the R package MCMCglmm and is up to several orders of magnitude faster than published alternatives.

Key-words: quantitative genetics, population genetics, software, bioinformatics, comparative analysis, evolutionary biology

Two major modelling frameworks are currently used in phylogenetic comparative biology to analyse the evolution of categorical characters. The first, and currently the most popular, was introduced by Pagel (1994) and was inspired by the Jukes & Cantor (1969) model of base substitutions in DNA. Under this model, the state of daughter lineages is conditionally independent given the parental *state*. The second framework was developed at the turn of last century by biometricians (Pearson 1900) and first applied in a phylogenetic context by Felsenstein (2005). Under this model, the state of daughter lineages is conditionally independent given the parental *probability* of being in a particular state. In this manuscript, we concentrate on this second class of models that focus on the evolution of the underlying probability.

Pearson (1900) developed the foundations of what became to be known as Wright's (1934a,b) threshold model whereby the probability of being in a particular state is assumed to be normally distributed on the probit scale (Bliss 1935). In a quantitative genetic framework, these normally distributed latent variables (also called liabilities) are further assumed to be the sum of normally distributed breeding values and environmental deviations (Falconer 1960). Later, the model (with fixed effects only) was subsumed within the class of generalised linear models (Nelder & Wedderburn 1972), and with the addition of random effects, Thompson (1979) anticipated the generalised linear mixed model (Gianola & Foulley 1983; Harville & Mee 1984). Unlike the linear mixed model, however, the likelihood cannot be calculated analytically because the random effects cannot be integrated out. Consequently, various methods have been used for approximating the necessary integrals including algebraic approximations such as Laplace approximations (Harville & Mee 1984; Gilmour, Anderson & Rae 1985; Schall 1991; Breslow & Clayton 1993; Breslow & Lin 1995), numerical approximations such as deterministic methods like (adaptive) Gaussian quadrature (Hinde 1982; Anderson & Aitkin 1985; Liu & Pierce 1994) and stochastic Monte Carlo methods such as Markov chain Monte Carlo (MCMC) (Albert & Chib 1993) (see Breslow 2004, for a review). These approximations have also been applied to situations where the random effects are correlated through a

*Correspondence author. E-mail: j.hadfield@ed.ac.uk

pedigree: Laplace approximations (Gilmour, Anderson & Rae 1985), Gaussian quadrature (Im & Gianola 1988) and MCMC (Sorensen *et al*. 1995).

In a phylogenetic context, several types of (multivariate) threshold model have been used. Felsenstein (2005) developed a model of multiple binary traits where the underlying latent variables evolved according to correlated Brownian motion, and later extended this model to scenarios where the response variables could be both binary and Gaussian (Felsenstein 2012). Ives & Garland (2010) developed a model for a single binary response where the underlying latent variables evolved according to an Ornstein–Uhlenbeck process (Hansen & Martins 1996). Hadfield (2010) & Hadfield & Nakagawa (2010) developed a similar model to Felsenstein (2012) where the response variables could come from various different distributions (including binary, Gaussian and (ordered) polychotomous) but assumed that the underlying latent variables were partly structured by correlated Brownian motion as in Felsenstein (2012), but could also be unstructured. The magnitude of these structured phylogenetic effects and non-structured non-phylogenetic effects was estimated, as in similar models developed for Gaussian traits (Lynch 1991; Pagel 1999). In these models, the relative strength of these two processes is often expressed as a phylogenetic heritability [Lynch (1991), also known as Pagel's λ (Pagel 1999)], the proportion of the total variance explained by the phylogenetic effects. Here, we show that the models suggested by Felsenstein (2005) and Felsenstein (2012) are special cases of GLMM with the phylogenetic heritability set to one.

In order to fit these models, various algorithms have been used to integrate over the random effects in both Frequentist [PQL: Ives & Garland (2010); Ho & Ane (2014), MCMC: Felsenstein (2005, 2012)] and Bayesian [MCMC: Hadfield (2010); Hadfield & Nakagawa (2010); Revell (2014)] frameworks, but here, we focus on MCMC algorithms. Felsenstein (2005) used a single-site Gibbs sampler to update the phylogenetic effects at internal nodes and importance sampling was used to update the phylogenetic effects at the tips. In Felsenstein (2012), importance sampling was replaced with Metropolis–Hastings updates. Hadfield (2010) used a redundant parameterisation in order to employ the blocked Gibbs sampler of Garcia-Cortes & Sorensen (2001) for all phylogenetic effects at the tips, with the option of Metropolis–Hasting updates or slice sampling for the tip non-phylogenetic effects associated with non-Gaussian traits. In this algorithm, the phylogenetic effects at internal nodes were integrated out analytically which is computationally very expensive when the problem is presented in this form (Freckleton 2012). An alternative strategy was proposed where the problem is expanded to include augmented random effects associated with internal nodes (called the $\mathbf{S}^{-1}$, rather than the $\mathbf{A}^{-1}$ parameterisation; Hadfield & Nakagawa 2010). Although augmented and tip phylogenetic effects can be Gibbs-sampled in a block, the $\mathbf{S}^{-1}$ parameterisation still has poorer mixing properties compared to the $\mathbf{A}^{-1}$ parameterisation. However, the CPU per iteration is dramatically decreased and in general the $\mathbf{S}^{-1}$ parameterisation outperforms the $\mathbf{A}^{-1}$ parameterisation, except for the smallest phylogenies. For ordered polychotomous traits, where thresholds have to be inferred, Cowle's (1996) Hastings-with-Gibbs joint update scheme was employed. Revell (2014) used a sequential Metropolis–Hastings update for all parameters, including the thresholds for models of ordered polychotomous traits.

Hadfield (2010) misunderstood the earlier work of Albert & Chib (1993) who presented results implying that the linear mixed model can be applied to the *truncated* latent variables directly. Consequently, the blocked Gibbs sampler can be applied without the need for a redundant parameterisation, and the non-phylogenetic effects at the tips are in standard form (truncated normal) and can be sampled efficiently using the rejection algorithm of Robert (1995). Here, we show that this result can be combined with work on the reduced animal model from quantitative genetics (Quaas & Pollak 1980) in order to develop an efficient Gibbs sampler for phylogenetic models in which the phylogenetic heritability is set to one. Under these assumptions, most GLMM software would run into numerical problems as the distribution of the data tends to degeneracy, and the algorithm proposed by Felsenstein (2005, 2012) neatly sidesteps this issue. We compare our related algorithm with those proposed by Felsenstein (2012) and Revell (2014) for a range of models and find substantial improvements in speed and efficiency.

The threshold model as a GLMM

The linear mixed model is formulated as:

$$\mathbf{l} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \qquad \text{eqn 1}$$

where $\mathbf{l}$ are the data, $\boldsymbol{\beta}$ and $\mathbf{X}$ are the fixed effects and their design matrix, $\mathbf{u}$ and $\mathbf{Z}$ are the random effects and their design matrix, and $\mathbf{e}$ are the residuals. $\mathbf{u}$ and $\mathbf{e}$ are assigned prior distributions which are usually multivariate normal with zero mean and covariance matrix $\mathbf{G}$ and $\mathbf{R}$, respectively. In a Bayesian analysis, the fixed effects are also assigned a prior. $\mathbf{G}$ and $\mathbf{R}$ are usually parameterised by a (small) set of hyper-parameters, which in a Bayesian setting would be assigned hyper-priors. Here, we will consider a simple phylogenetic model where $\mathbf{G} = \sigma_a^2\mathbf{A}$ and $\mathbf{R} = \sigma_e^2\mathbf{I}$ where $\mathbf{A}$ is the phylogenetic relatedness matrix, and σ_a^2 and σ_e^2 are the hyper-parameters (the phylogenetic and residual variances, respectively).

After marginalising the random effects and residuals, the marginal distribution of the data is multivariate normal:

$$\mathbf{l} \sim N(\mathbf{X\beta}, \sigma_a^2\mathbf{ZAZ}' + \sigma_e^2\mathbf{I}) \qquad \text{eqn 2}$$

The threshold model assumes that some latent variable has these properties and that values of l less than the threshold γ are associated with outcomes (y) being zero and values of l greater than γ are associated with outcomes being one.

Evaluating the n-dimensional integral to find the probability that $\mathbf{l}$ falls in some subspace that is compatible with the observations would be difficult when the number of observations (n) is large. However, conditioning on the random

effects, and because $\mathbf{R}$ is diagonal, the integration is easier because only univariate cumulative distribution functions have to be evaluated:

$$l_i \sim N(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}, r_{ii}) \quad \text{eqn 3}$$

and so

$$\begin{aligned}
&\Pr(Y_i = 1|\mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{u}, \mathbf{G}, \mathbf{R}, \gamma) \\
&\quad = \int_{e_i=-\infty}^{\infty} \mathbf{I}(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u} + e_i > \gamma) \\
&\qquad f_N(e_i, 0, r_{ii})de_i \\
&\quad = \int_{e_i=\gamma-(\mathbf{x}_i\boldsymbol{\beta}+\mathbf{z}_i\mathbf{u})}^{\infty} f_N(e_i, 0, r_{ii})de_i \\
&\quad = 1 - F_N(\gamma - (\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}), 0, r_{ii}) \\
&\quad = F_N(-\gamma + (\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}), 0, r_{ii}) \\
&\quad = F_N(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{z}_i\mathbf{u}, \gamma, r_{ii})
\end{aligned} \quad \text{eqn 4}$$

where f_N is the normal density function and F_N the normal cumulative distribution function. The first line on the RHS is the representation in terms of thresholds, and the final line is the representation in terms of a glm with probit link (Nelder & Wedderburn 1972). This equivalence is widely known and in the Appendix S1 we give a graphical explanation. The derivation relies on the facts that

$$1 - F_N(\eta, \gamma, \sigma^2) = F_N(-\eta, -\gamma, \sigma^2) \quad \text{eqn 5}$$

and

$$F_N(\eta, \gamma, \sigma^2) = F_N(\eta + c, \gamma + c, \sigma^2) \quad \text{eqn 6}$$

for any constant c . From the latter, it is clear that the distinction between threshold and latent variable is arbitrary, since they are just reflections of each other:

$$\begin{aligned}
F_N(\eta, \gamma, \sigma^2) &= F_N(\eta + (-\eta - \gamma), \gamma + (-\eta - \gamma), \sigma^2) \\
&= F_N(-\gamma, -\eta, \sigma^2)
\end{aligned} \quad \text{eqn 7}$$

Also, because

$$F_N(\eta, \gamma, \sigma^2) = F_N(\eta\sqrt{c}, \gamma\sqrt{c}, \sigma^2 c) \quad \text{eqn 8}$$

the scale of the latent variable is also arbitrary. By convention, the scale is set in a GLMM framework with 'standard' probit link by having $r_{ii} = 1\forall i$ and the 'threshold' is set to zero ($\gamma = 0$) and essentially absorbed into the intercept of the latent variable, β_1. Alternative parameterisations exist and have often been published as new models. For example, Curnow (1972) developed a model in which the threshold was not abrupt, but Bulmer (1985) showed its equivalence. Hazel, Smock & Johnson (1990) developed the environmental threshold model, where the linear predictor is split across the latent variable and threshold, but Roff (1996) pointed out its equivalence. More recently, Buoro, Gimenez & Prevost (2012) presented the 'latent' environmental threshold model which we show in the Appendix S2 is a special case of the standard GLMM. Some alternative parameterisations have been used in full knowledge of their equivalence, but have been used because of their algorithmic properties (e.g. Sorensen *et al*. 1995).

The extension to ordered categorical data falling into k ($k > 2$) categories is straightforward. If we denote $\boldsymbol{\gamma}$ as a vector of thresholds where $\boldsymbol{\gamma} = [-\infty, 0, \gamma_3, \gamma_4, \ldots \gamma_k, \infty]$, then:

$$\Pr(Y_i = y_i|\boldsymbol{\eta}, \mathbf{R}, \boldsymbol{\gamma}) = F_N(\boldsymbol{\gamma}_{y_i+1}, \eta_i, r_{ii}) - F_N(\gamma_{y_i}, \eta_i, r_{ii}) \quad \text{eqn 9}$$

where y takes on values from 1 to k. In the binary case (where $Y_i = 2$ if outcome i is a success), we can see that it is equivalent:

$$\begin{aligned}
\Pr(Y_i = 2|\boldsymbol{\eta}, \mathbf{R}, \gamma) &= F_N(\gamma_3, \eta_i, r_{ii}) - F_N(\gamma_2, \eta_i, r_{ii}) \\
&= F_N(\infty, \eta_i, r_{ii}) - F_N(0, \eta_i, r_{ii}) \\
&= 1 - F_N(0, \eta_i, r_{ii}) \\
&= F_N(0, -\eta_i, r_{ii}) \\
&= F_N(\eta_i, 0, r_{ii})
\end{aligned} \quad \text{eqn 10}$$

The reduced GLMM and phylogenies

The marginal distribution of the random effect predictions is:

$$\mathbf{Zu} \sim N(0, \sigma_a^2\mathbf{ZAZ}') \quad \text{eqn 11}$$

In some instances, we can find some $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{A}}$ such that:

$$\sigma_a^2\mathbf{ZAZ}' = \sigma_a^2\widetilde{\mathbf{Z}}\widetilde{\mathbf{A}}\widetilde{\mathbf{Z}}' + \sigma_a^2\mathbf{D} \quad \text{eqn 12}$$

where $\mathbf{D}$ is diagonal and (hopefully) $\widetilde{\mathbf{Z}}$ and/or $\widetilde{\mathbf{A}}^{-1}$ are smaller, sparser and less complex than $\mathbf{Z}$ and $\mathbf{A}^{-1}$ (Quaas & Pollak 1980; Lynch & Walsh 1998). We can therefore calculate the probability of $\mathbf{l}$, conditional on $\widetilde{\mathbf{u}}$ by evaluating the univariate cdf's:

$$l_i \sim N(\mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}}, r_{ii} + \sigma_a^2 d_{ii}) \quad \text{eqn 13}$$

and

$$\Pr(Y_i = 1|\mathbf{X}, \boldsymbol{\beta}, \widetilde{\mathbf{Z}}, \widetilde{\mathbf{u}}, \mathbf{G}, \mathbf{R}, \mathbf{D}) = F_N(\mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}}, \gamma, r_{ii} + \sigma_a^2 d_{ii}) \quad \text{eqn 14}$$

We could evaluate eqn 14 by evaluating the linear predictor $\widetilde{\eta}_i = \mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}}$ in a datum-specific cumulative distribution function with $\sigma_{\tilde{e}_i}^2 = r_{ii} + \sigma_a^2 d_{ii}$ or we could rescale each $\widetilde{\eta}_i$ by $\sqrt{r_{ii} + \sigma_a^2 d_{ii}}$ and evaluate using the standard inverse probit function. The method generalises to multiple categories, and under these situations, the thresholds must also be scaled if the second option is taken.

For a REML animal model with Gaussian data, Quaas & Pollak (1980) derived a useful set of reduced mixed model equations where columns of $\widetilde{\mathbf{Z}}$ index the parent(s) of individuals that have no offspring, or the individuals themselves if they do have offspring. For individuals with offspring $d_{ii} = 0$, but for individuals without offspring d_{ii} is proportional to the Mendelian sampling variation, the variation due to segregation and recombination that is independent of parental breeding values. d_{ii} take on values $0{\cdot}5(\delta_{d_i} + \delta_{s_i} - f_{d_i} - f_{s_i})$ where δ_{d_i} is one if the dam of individual i is not known and zero otherwise, and f_{d_i} is the inbreeding coefficient of individual i's dam. Terms subscripted with s denote sires. $\widetilde{\mathbf{A}}$ is then the additive genetic relationship matrix of individuals that have at least one

offspring and can be inverted using standard techniques (Henderson 1976; Quaas 1976; Meuwissen & Luo 1992).

Here, we show that a similar decomposition can be obtained for phylogenies, where columns of $\widetilde{\mathbf{Z}}$ index the interior nodes immediately ancestral to the tips, and $\widetilde{\mathbf{A}}$ is the phylogenetic relatedness matrix between the immediate ancestral species. d_{ii} are the branch lengths from the tips to the immediate ancestral species. Generally, $\widetilde{\mathbf{A}}$ has a dense inverse that must be solved directly. However, as in pedigree analysis, phantom parents (all remaining internal nodes) can be included in the analysis so that the inverse is sparse with simple structure. Following Hadfield & Nakagawa (2010), we will denote this augmented phylogenetic relatedness matrix as $\widetilde{\mathbf{S}}$. It is worth noting that $\widetilde{\mathbf{S}}$ is of dimension $n-2$, whereas $\mathbf{S}$ is of dimension $2n-2$. In addition, whereas $\mathbf{S}$ always requires augmenting with $n-2$ phantom parents, the degree of augmentation under the $\widetilde{\mathbf{S}}$ parameterisation varies with tree topology: with perfectly balanced trees $(n-4)/2$ phantom parents are required, but with perfectly pectinate trees augmentation can be dispensed with. The reduced parameterisation is therefore likely to be particularly efficient for unbalanced trees.

In order to illustrate these properties of the reduced parameterisation, eqn 15 describes the model for a perfectly balanced tree and eqn 16 describes the model for a perfectly pectinate tree. The trees are represented in Fig. 1.

$$\begin{bmatrix} l_{t_1} \\ l_{t_2} \\ l_{t_3} \\ l_{t_4} \\ l_{t_5} \\ l_{t_6} \\ l_{t_7} \\ l_{t_8} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} l_{n_1} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} l_{n_3} - l_{n_1} \\ l_{n_4} - l_{n_1} \\ l_{n_6} - l_{n_1} \\ l_{n_7} - l_{n_1} \\ l_{n_2} - l_{n_1} \\ l_{n_5} - l_{n_1} \end{bmatrix} + \begin{bmatrix} \tilde{e}_{t_1} \\ \tilde{e}_{t_2} \\ \tilde{e}_{t_3} \\ \tilde{e}_{t_4} \\ \tilde{e}_{t_5} \\ \tilde{e}_{t_6} \\ \tilde{e}_{t_7} \\ \tilde{e}_{t_8} \end{bmatrix} \quad \text{eqn 15}$$

$$\begin{bmatrix} l_{t_1} \\ l_{t_2} \\ l_{t_3} \\ l_{t_4} \\ l_{t_5} \\ l_{t_6} \\ l_{t_7} \\ l_{t_8} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} l_{n_1} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} l_{n_2} - l_{n_1} \\ l_{n_3} - l_{n_1} \\ l_{n_4} - l_{n_1} \\ l_{n_5} - l_{n_1} \\ l_{n_6} - l_{n_1} \\ l_{n_7} - l_{n_1} \end{bmatrix} + \begin{bmatrix} \tilde{e}_{t_1} \\ \tilde{e}_{t_2} \\ \tilde{e}_{t_3} \\ \tilde{e}_{t_4} \\ \tilde{e}_{t_5} \\ \tilde{e}_{t_6} \\ \tilde{e}_{t_7} \\ \tilde{e}_{t_8} \end{bmatrix} \quad \text{eqn 16}$$

The first incidence matrix (the column of ones) is the fixed effect design matrix $\mathbf{X}$, and the second incidence matrix is the random effect design matrix $\tilde{\mathbf{Z}}$. In both cases, all latent variables ($l_{t_1} \ldots l_{t_8}$ and $l_{n_1} \ldots l_{n_8}$) can be formed from combinations of the fixed effect (l_{n_1}), the random effects ($\tilde{\mathbf{u}}$: the vector after $\tilde{\mathbf{Z}}$) and the residuals ($\tilde{\mathbf{e}}$). However, with the balanced tree, $l_{n_2} - l_{n_1}$ and $l_{n_5} - l_{n_1}$ are included as augmented random effects and the final two columns of $\tilde{\mathbf{Z}}$ are null. The augmented effects are separated from the non-augmented effects by a solid vertical line, and it should be noted that in previous versions of MCMCglmm (≤2·21), the response vector also had to be augmented because null-column sparse matrices were not handled.

In quantitative genetics, σ_a^2 is estimated and σ_e^2 set to one. In a phylogenetic context, Felsenstein (2005, 2012) assumed $\sigma_a^2 = 1$ and $\sigma_e^2 = 0$ which would give:

$$\Pr(Y_i = 1) = F_N(\mathbf{x}_i\boldsymbol{\beta} + \tilde{\mathbf{z}}_i\tilde{\mathbf{u}}, \gamma, d_{ii}) \quad \text{eqn 17}$$

Because in a phylogenetic context $d_{ii} > 0 \forall i$, we can formulate the model as a reduced phylogenetic model and allow σ_e^2 to be set to zero. In most other applications, $d_{ii} = 0$ for some i, which would result in eqn 14 being degenerate if σ_e^2 was set to zero. Felsenstein (2012) expressed concern that the GLMM analogue of eqn 17 (i.e. eqn 14) did not have a transfer function that was a step. However, with $\sigma_e^2 = 1$ and σ_a^2 becoming large, we can see that

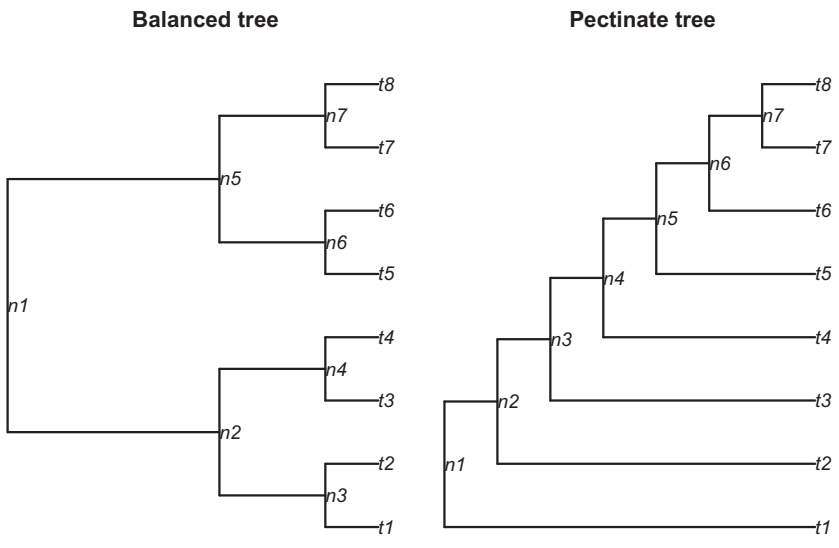

Fig. 1. Perfectly balanced (left) and perfectly pectinate (right) trees with interior nodes and tips labelled. The model equations for these trees are given in eqns 15 and 16, respectively.

$$\begin{aligned}\Pr(Y_i = 1) &= F_N(\mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}}, 0, 1 + \sigma_a^2 d_{ii}) \\ &= F_N((\mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}})/\sigma_a, 0, (1 + \sigma_a^2 d_{ii})/\sigma_a^2) \\ &= \lim_{\sigma_a^2 \to \infty} F_N((\mathbf{x}_i\boldsymbol{\beta} + \widetilde{\mathbf{z}}_i\widetilde{\mathbf{u}})/\sigma_a, 0, d_{ii})\end{aligned} \quad \text{eqn 18}$$

Dividing fixed and random effects by σ_a is used to obtain transformed location effects on some alternative (and non-identifiable) scale (e.g. Hadfield 2009): in this case a scale in which $\sigma_a = \sigma_a + \sigma_e = 1$. The model of Felsenstein (2012) is therefore a special case of a phylogenetic GLMM with $h^2 = 1$. However, it should be emphasised that most applications of GLMM would break as σ_a becomes large and the probit function under/overflows. Felsenstein's (2012) method neatly sidesteps this issue when it can be assumed that $h^2 = 1$, as does the closely related reduced mixed model formulation presented here.

MCMC algorithm for the (multivariate) reduced phylogenetic GLMM

In the univariate categorical model with $h^2 = 1$, the only parameters are the fixed intercept β and the random phylogenetic effects (the ancestral states $\tilde{\mathbf{u}}$ and the tip + phantom parent states $\mathbf{l}$). Conditional on $\mathbf{l}$, β and $\tilde{\mathbf{u}}$ can be Gibbs-sampled in a block using the algorithm of Garcia-Cortes & Sorensen (2001). Conditional on β and $\tilde{\mathbf{u}}$, $\mathbf{l}$ can be sampled from independent truncated normals with variances proportional to the diagonal elements of $\mathbf{D}$. Robert (1995) describes an efficient rejection algorithm for sampling from a truncated normal.

In the univariate Gaussian model with $h^2 = 1$, β and $\tilde{\mathbf{u}}$ can be Gibbs-sampled in a block as in the binary case, but the tip states are observed and do not need to be updated. However, unlike the binary case, the scale is identifiable and does need to be estimated. Noting that

$$\mathbf{u} = \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{l} - \mathbf{x}\boldsymbol{\beta} \end{bmatrix} \sim N(\mathbf{0}, \sigma_a^2\mathbf{S}) \quad \text{eqn 19}$$

the posterior distribution of σ_a^2 is inverse-Wishart (with a conjugate inverse-Wishart prior) and can be Gibbs-sampled (Sorensen & Gianola 2002). With 0 prior degrees of freedom, the posterior degrees of freedom is $\nu = 2(n-1)$ and the scale matrix $\boldsymbol{\Psi}$ is:

$$\begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{l} - \mathbf{X}\boldsymbol{\beta} \end{bmatrix} \mathbf{S}^{-1} \begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{l} - \mathbf{X}\boldsymbol{\beta} \end{bmatrix}^{\top} \quad \text{eqn 20}$$

An alternative but equivalent parameterisation which has an easier implementation is:

$$\begin{bmatrix} \tilde{\mathbf{u}} \\ \mathbf{l} - \tilde{\mathbf{Z}}\tilde{\mathbf{u}} - \mathbf{X}\boldsymbol{\beta} \end{bmatrix} \sim N\big(\mathbf{0}, \sigma_a^2(\tilde{\mathbf{S}} \oplus \mathbf{D})\big) \quad \text{eqn 21}$$

where $\oplus$ is the direct sum. This is preferable because $(\tilde{\mathbf{S}} \oplus \mathbf{D})^{-1} = \tilde{\mathbf{S}}^{-1} \oplus \mathbf{D}^{-1}$ such that the inverse is easier to obtain than $\mathbf{S}^{-1}$ and is sparser with simpler structure.

Extending the reduced mixed model to models where $h^2 < 1$ is less straightforward. In standard GLMMs, the (co)variance components can be Gibbs-sampled if their only constraint is non-negative definiteness. However, the conditional distribution of the variance components in the reduced mixed model is in non-standard form because the reduced 'residuals' ($\tilde{e}$) are a mixture distribution of residual (e) and phylogenetic effects. In this case, the variance component(s) would have to be updated by some alternative method (Bink *et al.* 1998).

For multivariate models, the tip states for categorical traits can be Gibbs-sampled conditional on the remaining tip states for that species. For example, if we subscript terms by 1 and 2 to differentiate between the two traits of a bivariate analysis, then the latent variable for trait 2 in species i can be updated conditional on the current value of trait 1 for that species:

$$\begin{aligned}&\Pr(l_{2i}|\mathbf{l}_{/2i}, \tilde{\mathbf{Z}}, \tilde{\mathbf{u}}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{d}, r) \\ &\quad = f_{TN}(l_{2i}|\mathbf{X}_2\boldsymbol{\beta}_2 + \tilde{\mathbf{Z}}_2\tilde{\mathbf{u}}_2 \\ &\quad + r(l_{1i} - \mathbf{X}_1\boldsymbol{\beta}_1 - \tilde{\mathbf{Z}}_1\tilde{\mathbf{u}}_1), \\ &\quad d_{ii} - r^2 d_{ii}^2, \gamma_{y_{2i}}, \gamma_{y_{2i}+1})\end{aligned} \quad \text{eqn 22}$$

where the subscript $/i$ denotes a vector with element i omitted.

In the multivariate model (following the second parameterisation in the univariate section),

$$\begin{bmatrix} \tilde{\mathbf{u}}_1 \\ \mathbf{l}_1 - \tilde{\mathbf{Z}}_1\tilde{\mathbf{u}}_1 - \mathbf{X}_1\boldsymbol{\beta}_1 \\ \tilde{\mathbf{u}}_2 \\ \mathbf{l}_2 - \tilde{\mathbf{Z}}_2\tilde{\mathbf{u}}_2 - \mathbf{X}_2\boldsymbol{\beta}_2 \end{bmatrix} \sim N\big(\mathbf{0}, \mathbf{V} \otimes (\tilde{\mathbf{S}} \oplus \mathbf{D})\big) \quad \text{eqn 23}$$

where $\mathbf{V}$ is the phylogenetic effect covariance matrix. When both traits are Gaussian, $\mathbf{V}$ is unconstrained and we can sample directly from the inverse-Wishart as in the univariate case. When both traits are categorical, the scale for each trait (the diagonal elements of $\mathbf{V}$) is not identifiable and can be constrained to one. The resulting correlation matrix can be updated using methods outlined in Liu & Daniels (2006). If there is a single categorical trait and multiple non-categorical traits, then a single diagonal element of $\mathbf{V}$ should be constrained at one, which can be achieved using the method of Korsgaard, Andersen & Sorensen (1999). These updating schemes were available in MCMCglmm v1.00 (2009) and v2.02 (2010), respectively. Additional categorical traits could be included in the analysis although the correlation between them would have to be constrained to zero under versions <2.20. Here, we present a hybrid algorithm to overcome this restriction, by updating the correlation submatrix using methods outlined in Liu & Daniels 2006 and then conditioning on it while updating the remaining components using a modified version of the method presented in Korsgaard, Andersen & Sorensen (1999). For example, if we divide the traits into two categories where category 1 contains ≥1 non-categorical trait and category 2 contains ≥2 categorical traits, then:

$$\begin{aligned}&\Pr\left(\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{11} \\ \mathbf{V}_{12} & \mathbf{V}_{22} \end{bmatrix} \Big| \begin{bmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{11} \\ \boldsymbol{\Psi}_{12} & \boldsymbol{\Psi}_{22} \end{bmatrix}, \nu\right) \\ &\quad = \Pr(\mathbf{V}_{22}|\boldsymbol{\Psi}, \nu)\Pr(\mathbf{V}_{11}, \mathbf{V}_{12}|\mathbf{V}_{22}, \boldsymbol{\Psi}, \nu)\end{aligned} \quad \text{eqn 24}$$

From Korsgaard, Andersen & Sorensen (1999) and references therein, $\mathbf{V}_{22}$ can be drawn directly from its marginal distribution $\mathbf{V}_{22} \sim IW(\mathbf{\Psi}_{22}, \nu - d_1)$ where d_1 is the dimension of $\mathbf{V}_{11}$ and the inverse-Wishart is parameterised in terms of the scale matrix, rather than its inverse $\mathbf{\Sigma} = \mathbf{\Psi}^{-1}$ as in Korsgaard, Andersen & Sorensen (1999). However, in this example, the diagonal elements of $\mathbf{V}_{22}$ are not identifiable in the likelihood and so $\mathbf{V}_{22}$ is constrained to be a correlation matrix which does not have a standard distribution. However, assuming the identity prior specification in Barnard, McCulloch & Meng (2000), we use the method of Liu & Daniels (2006) with $\nu - d_1$ degrees of freedom. $\nu = \nu_p + 2n - 2$ under the parameterisation given in eqn 19, but may be less under the parameterisation given in eqn 21 depending on tree topology. ν_p is the prior degrees of freedom and under the identity prior (Barnard, McCulloch & Meng 2000) controls the degree to which $\mathbf{V}_{22}$ is shrunk to an identity matrix. Korsgaard, Andersen & Sorensen (1999) provide a strategy to sample from $\Pr(\mathbf{V}_{11}, \mathbf{V}_{12}|\mathbf{V}_{22}, \mathbf{\Psi}, \nu)$ when $\mathbf{V}_{22} = \mathbf{I}$ although their result generalises directly to cases where $\mathbf{V}_{22} \neq \mathbf{I}$ since 10.4.3 from Hoffman-Jorgensen (1994) holds generally. Consequently, $\mathbf{T}_1$ and $\mathbf{T}_2$ [defined in Korsgaard, Andersen & Sorensen (1999)] can be sampled and $\mathbf{T}_3$ set to $(\mathbf{V}_{22})^{-1}$ (Korsgaard, Andersen & Sorensen 1999). Substituting into the first equation on p180 of Korsgaard, Andersen & Sorensen (1999) draws $\mathbf{V}$ from the required distribution. Note that there is a small typological error in Korsgaard, Andersen & Sorensen (1999): the $\mathbf{T}_3^{\top}$ in the upper right submatrix of the first equation on p180 should be $\mathbf{T}_3^{-\top}$. We use the notation $\mathbf{A}_{11}^{-1}$ and $(\mathbf{A}_{11})^{-1}$ to differentiate between a submatrix of the inverse of $\mathbf{A}$ and the inverse of a submatrix of $\mathbf{A}$.

Comparison with other MCMC algorithms

We tested our algorithm against those proposed in Revell (2014) and implemented in the R package phytools (Revell 2012) using three models: a) the ancestral states in a univariate binary model, b) the parameters of a bivariate binary Gaussian model (the two intercepts, the Gaussian variance and the correlation) and c) the threshold in a 3-category ordinal model. In each case, we sampled 500 random tips from the recently published complete avian phylogeny (Jetz *et al.* 2012) and then simulated trait values with $\beta = 0$ and $\sigma_a^2 = 1$. In the bivariate analysis (b), we used two schemes where the phylogenetic correlation was either set to zero or 0·5. In the ordinal model, the non-constrained threshold (γ_3) was set to one. In order to gauge the efficiency of the algorithms, we calculated the MCMC effective sample size per iteration, either averaged over ancestral states (a) or for the phylogenetic correlation (b) or the threshold (c). In addition, we calculated the amount of time taken to complete one iteration (defined as a complete update of all parameters). All models were fitted for 13 000 iterations with 3000 iterations as burn-in using the algorithm presented here (referred to as MCMCglmm) and that proposed in Revell (2014) (referred to as phytools). The code to perform the analyses can be found in Appendix S3. The full results are presented in Table 1 and show that MCMCglmm is between ten thousand and 1 million times more efficient than phytools, depending on the model and parameters assessed. In order to visualise the better mixing properties of the MCMCglmm algorithm, we also plot traces of the MCMC chain for model b) with a true phylogenetic correlation of 0·5 (Fig. 2).

Table 1 . Profiling statistics for (a) a univariate binary model, (b) a bivariate binary Gaussian model and (c) a three-category ordinal model fitted with different MCMC software. The statistics are the effective sample size (ESS), the time (in minutes) per iteration (Time) and the number of effective samples per minute. The ESS statistics are a) the average for ancestral states, (b) for the phylogenetic correlation and (c) for the threshold

	Model a	Model b ($r = 0$)	Model b ($r = 0{\cdot}5$)	Model c
Phytools				
ESS	282·25	281·78	192·7	17·32
Time (min)	9970	3665	3684	9982
ESS/Time	0·02831	0·07689	0·05231	0·00174
MCMCglmm				
ESS	2360·96	669·53	418·46	334·13
Time (min)	0·1551	1·0593	1·0476	0·168
ESS/Time	15217·3	632·1	399·4	1988·7

The algorithm developed by Felsenstein (2012) and implemented in phylip (Felsenstein 1989) uses MCMC to obtain the likelihood, rather than as a technique to sample from the complete posterior distribution of the model. However, under model b) (model a) is not implemented in phylip) the MCMC algorithm is sampling from the posterior distribution of latent variables conditional on the covariance matrix of the current iteration. Consequently, we fixed the covariance matrix at its true value (with phylogenetic correlation equal to 0·5) in both MCMCglmm and phylip and compared the latent variables for the categorical trait in the same manner as was done for the phytools/MCMCglmm comparison in model a) above. Modified phylip code for writing the latent variables to file is available from the author. MCMCglmm collects 5·2 more effective samples per iteration than phylip but iterates 2·38 times slower resulting in comparable speeds: MCMCglmm collects 2·18 more effective samples per minute.

In order to gauge how much tree shape influences the speed and efficiency of our algorithm, we also simulated balanced and pectinate trees (with 512 tips) using the stree function from the R package ape (Paradis, Claude & Strimmer 2004) and ran model a). The efficiency of our algorithm depends on tree shape with the algorithm collecting 23 more effective samples per iteration for pectinate trees than equivalent sized balanced trees but iterating at an equivalent speed (1·05 times faster).

Discussion

In this paper, we show that the original phylogenetic threshold model (Felsenstein 2005) is equivalent to a GLMM with the phylogenetic 'heritability' set to one. We go on to develop an MCMC algorithm based on the reduced animal model from quantitative genetics that is robust under this restriction. The new algorithm can collect adequate effective samples ($\sim$1000)

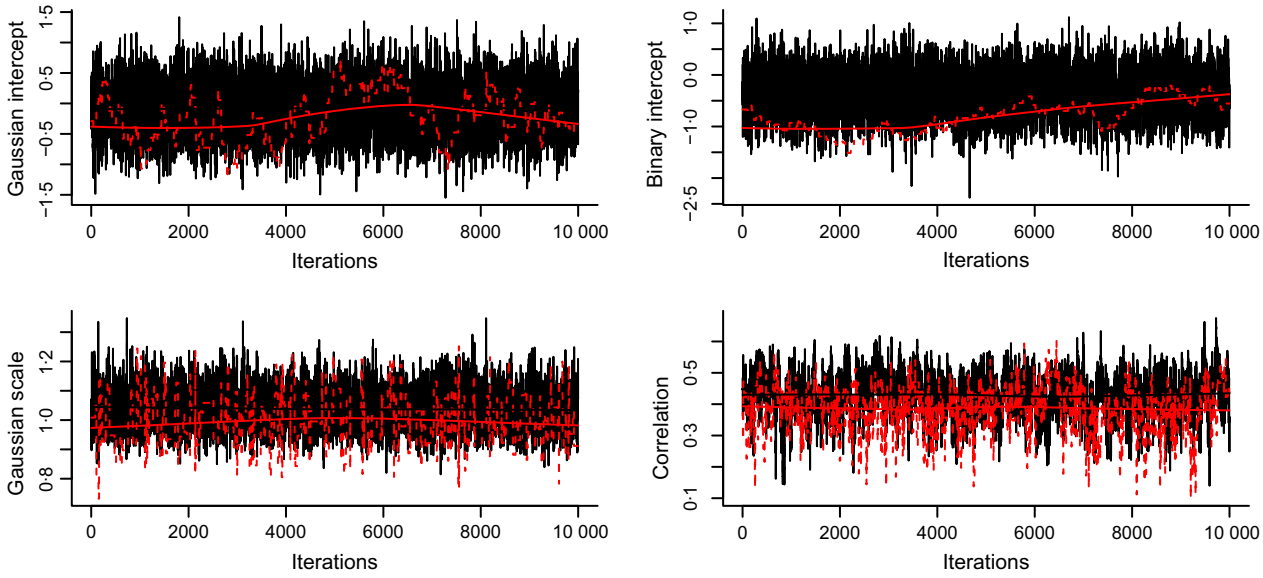

Fig. 2. Traces of Markov chain output for the binary Gaussian model from phytools (red) and MCMCglmm (black).

from the posterior distribution in approximately a minute for a 500 tip phylogeny for a variety of threshold models. In contrast, the only published alternative, the naive sampler developed by Revell (2014) may require weeks or even years depending on the model. In part, this could have been improved upon by single-site Gibbs sampling of the latent variables at interior nodes, as suggested by Felsenstein (2005). For sampling latent variables, the algorithms in Felsenstein (2005, 2012) are comparable in speed to those developed here where the reduced efficiency per iteration is balanced by the increase in speed per iteration.

The algorithm presented by Revell (2014) shows particularly slow mixing for the three-category ordinal model. The Metropolis–Hastings updates for the threshold could be replaced by a Gibbs sampling step, since their conditional distribution is simply uniform between the maximum and minimum latent variables associated with data in adjacent categories (Albert & Chib 1993). However, Cowles (1996) showed that conditional updates of the thresholds followed by the latent variables result in very poor mixing because the interval between the minimum and maximum latent variables from the previous iteration is likely to be very narrow, especially when there are many observations. To alleviate this problem, Cowles (1996) proposed a clever Hastings-with-Gibbs joint update that was implemented in Hadfield (2010) and here. Nandram & Chen (1996) extended this idea and demonstrated that the Hastings step could be omitted for three-category problems, and for >3-category problems, Dirichlet proposal densities outperform the truncated normal proposal densities of Cowles (1996). Similarly, Sorensen et al. (1995) showed that Gibbs sampling for a three-category problem is possible by fixing all thresholds and estimating σ^2_e. We did not pursue these extensions since the current implementation has adequate mixing properties.

Odegard *et al.* (2010) presented a Gibbs sampling scheme for binary data collected on a pedigree. The central idea is that because the Mendelian sampling effect is confounded with the residual for individuals without offspring, then it is also not identifiable. Consequently, the conditional distribution of the additive genetic variance on parental breeding values is equivalent to the distribution after marginalising offspring breeding values. They therefore use the standard (non-reduced) mixed model equations but only use the parental random effects to update variance components. This result was a motivating factor behind the algorithms developed in this paper, despite the result not being applicable to models where the heritability is set to one, as here. Nevertheless, the result is likely to prove useful for the more general phylogenetic model where the heritability is estimated.

Most of the ideas presented in this paper are straightforward extensions to a large body of work that has accumulated in statistics and quantitative genetics. We would like to reiterate that many phylogenetic comparative models are special cases of this broader work and that understanding and utilising known results from these fields may prove as fruitful as developing algorithms *de novo* (Ives & Zhu 2006; Hadfield & Nakagawa 2010). An exception to this is the class of Markov models initiated by Pagel (1994) which are fundamentally different from the generalised linear mixed models employed by quantitative geneticists. For some traits, the Markov model is probably a more realistic model of how discrete traits evolve (e.g. states that conform to Dollo's law (Goldberg & Igić 2008), but for others a threshold model is probably more realistic (See the arguments put forward in Felsenstein 2005; Hadfield & Nakagawa 2010; Revell 2012). The common practice of discretising a continuous trait in order to model the evolution of the resulting states (e.g. Fitzpatrick *et al.* 2009) is an extreme and clear example of character evolution that would not conform to a Markov model; Pearson (1900) actually used the discretisation of human heights to explain the concept of the threshold model. However, given the Markov model's ability to say something about the temporal order of evolutionary events, and the direction of causality, such

practices are likely to persist. Consequently, a better understanding of the relationship between Markov models and threshold models, and the development of models that contain aspects of both would be a very welcome addition to the literature on the evolution of discrete characters.

Acknowledgements

Thanks to Ian White, Bill Hill, Joe Felsenstein, Asko Mäki-Tanila, Liam Revell, Albert Phillimore, Darren Obbard, Tony Ives, Simon Blomberg, Laura Ross & two anonymous reviewers for their input through programming advice, useful discussions and comments on previous versions of this manuscript. This work was funded by a Royal Society fellowship.

References

Albert, J.H. & Chib, S. (1993) Bayesian-analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, **88**, 669–679.

Anderson, D.A. & Aitkin, M. (1985) Variance component models with binary response – interviewer variability. *Journal of the Royal Statistical Society Series B-Methodological*, **47**, 203–210.

Barnard, J., McCulloch, R. & Meng, X.L. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, **10**, 1281–1311.

Bink, M., Quaas, R.L. & Van Arendonk, J.A.M. (1998) Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects. *Genetics Selection Evolution*, **30**, 103–125.

Bliss, C.I. (1935) The calculation of the dosage-mortality curve. *Annals of Applied Biology*, **22**, 134–167.

Breslow, N. (2004) Whither PQL? *Proceedings of the Second Seattle Symposium in Biostatistics*, pp. 1–22. Springer.

Breslow, N.E. & Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.

Breslow, N.E. & Lin, X.H. (1995) Bias correction in generalized linear mixed models with a single-component of dispersion. *Biometrika*, **82**, 81–91.

Bulmer, M.G. (1985) *The Mathematical Theory of Quantitative Genetics*. Oxford University Press, New York, NY.

Buoro, M., Gimenez, O. & Prevost, E. (2012) Assessing adaptive phenotypic plasticity by means of conditional strategies from empirical data: the latent environmental threshold model. *Evolution*, **66**, 996–1009.

Cowles, M.K. (1996) Accelerating Monte Carlo Markov chain convergence for cumulative – link generalized linear models. *Statistics and Computing*, **6**, 101–111.

Curnow, R.N. (1972) Multifactorial model for inheritance of liability to disease and its implications for relatives at risk. *Biometrics*, **28**, 931–946.

Falconer, D.S. (1960) *Introduction to Quantitative genetics*. Oliver & Boyd, Edinburgh/London.

Felsenstein, J. (1989) PHYLIP-phylogeny inference package (version 3.2). *Cladistics*, **5**, 164–166.

Felsenstein, J. (2005) Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1427–1434.

Felsenstein, J. (2012) A comparative method for both discrete and continuous characters using the threshold model. *American Naturalist*, **179**, 145–156.

Fitzpatrick, J.L., Montgomerie, R., Desjardins, J.K., Stiver, K.A., Kolm, N. & Balshine, S. (2009) Female promiscuity promotes the evolution of faster sperm in cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1128–1132.

Freckleton, R.P. (2012) Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*, **3**, 940–947.

Garcia-Cortes, L.A. & Sorensen, D. (2001) Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genetics Selection Evolution*, **33**, 443–452.

Gianola, D. & Foulley, J.L. (1983) Sire evaluation for ordered categorical-data with a threshold-model. *Genetics Selection Evolution*, **15**, 201–223.

Gilmour, A.R., Anderson, R.D. & Rae, A.L. (1985) The analysis of binomial data by a generalized linear mixed model. *Biometrika*, **72**, 593–599.

Goldberg, E.E. & Igić, B. (2008) On phylogenetic tests of irreversible evolution. *Evolution*, **62**, 2727–2741.

Hadfield, J.D. (2009) *MCMCglmm Course Notes*.

Hadfield, J.D. (2010) MCMC methods for multi–response generalised linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, **33**, 1–22.

Hadfield, J.D. & Nakagawa, S. (2010) General quantitative genetic methods for comparative biology: Phylogenies, taxonomies, meta-analysis and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, **23**, 494–508.

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

Harville, D.A. & Mee, R.W. (1984) A mixed-model procedure for analyzing ordered categorical-data. *Biometrics*, **40**, 393–408.

Hazel, W.N., Smock, R. & Johnson, M.D. (1990) A polygenic model for the evolution and maintenance of conditional strategies. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **242**, 181–187.

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

Hinde, J. (1982) Compound poisson regression models. *Lecture Notes in Statistics*, **14**, 109–121.

Ho, L.S.T. & Ane, C. (2014) A linear-time algorithm for gaussian and non-gaussian trait evolution models. *Systematic Biology*, **63**, 397–408.

Hoffman-Jorgensen, J. (1994) *Probability with a View Towards Statistics*, volume II. CRC Press, London.

Im, S. & Gianola, D. (1988) *Mixed models for binomial data with an application to lamb mortality*. Applied Statistics, , pp. 196–204.

Ives, A.R. & Garland, T. (2010) Phylogenetic logistic regression for binary dependent variables. *Systematic Biology*, **59**, 9–26.

Ives, A.R. & Zhu, J. (2006) Statistics for correlated data: Phylogenies, space, and time. *Ecological Applications*, **16**, 20–32.

Jetz, W. Thomas, G.H., Joy, J.B., Hartmann, K. & Mooers, A.O. (2012) The global diversity of birds in space and time. *Nature*, **491**, 444–448.

Jukes T. H. & Cantor, C.R. (1969) Evolution of protein molecules. *Mammalian Protein Metabolism III*, (ed. H.N. Munro), pp. 21–132. Academic Press, New York.

Korsgaard, I.R., Andersen, A.H. & Sorensen, D. (1999) A useful reparameterisation to obtain samples from conditional inverse Wishart distributions. *Genetics Selection Evolution*, **31**, 177–181.

Liu, X.F. & Daniels, M.J. (2006) A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, **15**, 897–914.

Liu, Q. & Pierce, D.A. (1994) A note on Gauss-Hermite quadrature. *Biometrika*, **81**, 624–629.

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

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

Meuwissen, T.H.E. & Luo, Z. (1992) Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, **24**, 305–313.

Nandram, B. & Chen, M.H. (1996) Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence. *Journal of Statistical Computation and Simulation*, **54**, 129–144.

Nelder, J.A. & Wedderburn, R.W. (1972) Generalized linear models. *Journal of the Royal Statistical Society Series A*, **135**, 70–384.

Odegard, J., Meuwissen, T.H.E., Heringstad, B. & Madsen, P. (2010) A simple algorithm to estimate genetic variance in an animal threshold model using Bayesian inference. *Genetics Selection Evolution*, **42**, 1–7.

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

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

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pearson, K. (1900) Mathematical contributions to the theory of evolution. VII. on the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society A*, **195**, 1–47.

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

Quaas, R.L. & Pollak, E.J. (1980) Mixed model methodology for farm and ranch beef-cattle testing programs. *Journal of Animal Science*, **51**, 1277–1287.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Revell, L.J. (2014) Ancestral character estimation under the threshold model from quantitative genetics. *Evolution*, **68**, 743–759.
Robert, C.P. (1995) Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121–125.
Roff, D.A. (1996) The evolution of threshold traits in animals. *Quarterly Review of Biology*, **71**, 3–35.
Schall, R. (1991) Estimation in generalized linear-models with random effects. *Biometrika*, **78**, 719–727.
Sorensen, D. & Gianola, D. (2002) *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. *Statistics for Biology and Health*. Springer-Verlag, New York, NY.
Sorensen, D.A., Andersen, S., Gianola, D. & Korsgaard, I. (1995) Bayesian-inference in threshold models using gibbs sampling. *Genetics Selection Evolution*, **27**, 229–249.
Thompson, R. (1979) Sire evaluation. *Biometrics*, **35**, 339–353.
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.

Received 20 October 2014; accepted 27 January 2015
Handling Editor: Robert B. O'Hara

Supporting Information

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

Appendix S1. The relationship between GLMM and the threshold model.

Appendix S2. The relationship between GLMM and the latent environmental threshold model.

Appendix S3. Code for fitting example models.