General Quantitative Genetic Methods for Comparative Biology: supplementary material

November 10, 2009

Section: Phylogenetic Meta-analysis

The method of Adams (2008) implies that $(\mathbf{G}\sigma_a^2 + \sigma_m^2\mathbf{W}^{-1})^{-1} \propto \mathbf{\Psi W \Psi}^{'}$, which means:

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

where c is some constant.

Worked example

To illustrate the range of phylogenetic meta-analyses that can be performed using mixed models we give `asreml`, `rbugs` and `MCMCglmm` code.

```
> library(MCMCglmm)
> library(asreml)
```

To demonstrate the possibilities we re-analyse Adams (2008) analysis of Bergmann's rule.

```
> Adams.phylo <- read.tree("~/Work/Shinichi/Data/Raw/AdamsP.tre")
> Adams.data <- read.csv("~/Work/Shinichi/Data/Raw/AdamsD.csv")
```

where `Adams.phylo` is a `phylo` object from `ape` (Paradis 2006) for 40 mammal species included in the analysis (See Figure 1):

```
> plot(Adams.phylo)
```

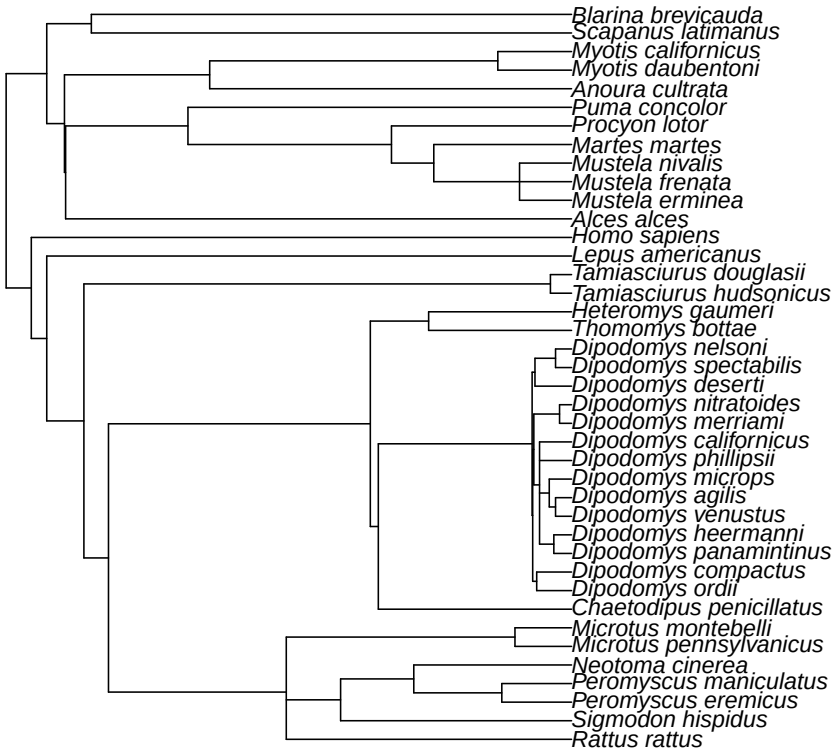

Figure 1: Phylogeny used by Adams (2008) to test Bergmann's rule

The data consist of a species identifiers (`animal`) the Z-transformed correlation (`corr`) between latitude and body size (`Z`) and the replication (number of populations) for each species (`N`). The sampling variance of the Z-scores (`mev`) were approximated using 1/(N-3), with the weights (`weights`) as 1/`mev` and the standard errors (`mesd`) as $\sqrt{\texttt{mev}}$.

The most general meta-analytic model that we entertained was:

$$y_i = \mu + a_i + e_i + m_i \tag{2}$$

where y are the data, a phylogenetic effects, e residuals and m measurement errors. μ is the intercept and is interpreted as the mean effect size.

The three random effects are assumed to be normally distributed:

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

and by setting some variance components to zero or one, many published models can be derived as special cases (Table 1)

Model	Phylogenetic	Meta-analysis	σ_a^2	σ_e^2	σ_m^2	Reference
m1	✘	✘	0	E	0	
m2	✘	fixed	0	0	1	
m3	✘	fixed	0	0	E	
m4	✘	random	0	E	1	
m5	✘	random	0	E	E	
m6	✔	✘	E	0	0	(Felsenstein 1985)
m7	✔	✘	E	E	0	(Lynch 1991; Pagel 1999)
m8	✔	fixed	E	0	1	(Ives *et al.* 2007; Adams 2008)
m9	✔	fixed	E	0	E	(Felsenstein 2008)
m10	✔	random	E	E	1	
m11	✔	random	E	E	E	

Table 1: Table of phylogenetic and meta-analysis models, where variance components are either estimated (E), set to zero (0) or set to one (1). References are given for those models developed in the comparative literature.

All of these non-phylogenetic models can be fitted in ASReml directly, but for phylogenetic models $\mathbf{A}^{-1}$ needs to be formed and formatted correctly. This can be done using functions in the `MCMCglmm` package:

```
> IA <- inverseA(Adams.phylo, nodes = "TIPS")
> IAasreml <- sm2asreml(IA$Ainv, IA$node.names)
```

where the argument `nodes="TIPS"` specifies that $\mathbf{A}^{-1}$ is to be formed, rather than `nodes="ALL"` which would form $\mathbf{S}^{-1}$.

The 11 models are fitted below, and the results presented in Table 2.

```
> m1 <- asreml(Z ~ 1, data = Adams.data)
> m2 <- asreml(Z ~ 1, weights = weights, family = asreml.gaussian(dispersion = 1),
+     data = Adams.data)
> m3 <- asreml(Z ~ 1, weights = weights, data = Adams.data)
> m4 <- asreml(Z ~ 1, random = ~units, weights = weights, family = asreml.gaussian(dispersion = 1),
+     data = Adams.data)
> m5 <- asreml(Z ~ 1, random = ~units, weights = weights, data = Adams.data)
> m6 <- asreml(Z ~ 1, random = ~giv(animal, var = T), family = asreml.gaussian(dispersion = 1e-06),
+     data = Adams.data, ginverse = list(animal = IAasreml))
> m7 <- asreml(Z ~ 1, random = ~giv(animal, var = T), data = Adams.data,
+     ginverse = list(animal = IAasreml))
> m8 <- asreml(Z ~ 1, random = ~giv(animal, var = T), weights = weights,
+     family = asreml.gaussian(dispersion = 1), data = Adams.data,
+     ginverse = list(animal = IAasreml))
> m9 <- asreml(Z ~ 1, random = ~giv(animal, var = T), weights = weights,
+     data = Adams.data, ginverse = list(animal = IAasreml))
> m10 <- asreml(Z ~ 1, random = ~giv(animal, var = T) + units,
```

```
+     weights = weights, family = asreml.gaussian(dispersion = 1),
+     data = Adams.data, ginverse = list(animal = IAasreml))
> m11 <- asreml(Z ~ 1, random = ~giv(animal, var = T) + units,
+     weights = weights, data = Adams.data, ginverse = list(animal = IAasreml))
```

Model	Phylogenetic	Meta-analysis	σ^2_a	σ^2_e	σ^2_m	μ	P-value	loglik
m1	✘	✘	0	0.615	0	0.293±0.124	0.0231	-11.853
m2	✘	fixed	0	0	1	0.288±0.030	< 0.0001	-132.926
m3	✘	fixed	0	0	9.184	0.288±0.091	0.0031	-16.578
m4	✘	random	0	0.414	1	0.227±0.116	0.0565	-10.352
m5	✘	random	0	1.546	0.362	0.215±0.115	0.0678	-10.254
m6	✔	✘	4.037	0	0	0.471±0.689	0.4986	-21.386
m7	✔	✘	0.320	0.402	0	0.445±0.236	0.0670	-10.371
m8	✔	fixed	6.280	0	1	0.173±0.100	0.0902	-20.090
m9	✔	fixed	0.572	0	5.432	0.179±0.100	0.0797	-11.988
m10	✔	random	0.000	0.414	1	0.227±0.115	0.0564	-10.352
m11	✔	random	0.067	0.344	1.616	0.211±0.114	0.0731	-10.245

Table 2: Parameters estimates for the 11 models defined in Table 1 fitted to data from Adams (2008). Model `m4` has the lowest AIC.

In general it is not possible to set the residual variance to zero in `MCMCglmm` because the Markov chain is irreducible. In practice it can be set to some small value, but mixing may be a problem and the chain may have to be run for many iterations. Below we give code for models where a residual variance is estimated.

```
> m1.mcmc <- MCMCglmm(Z ~ 1, data = Adams.data)
> m4.mcmc <- MCMCglmm(Z ~ 1, mev = Adams.data$mev, data = Adams.data)
> m5.mcmc <- MCMCglmm(Z ~ 1, random = ~us(mesd):units, data = Adams.data)
> m7.mcmc <- MCMCglmm(Z ~ 1, random = ~animal, data = Adams.data,
+     pedigree = Adams.phylo)
> m10.mcmc <- MCMCglmm(Z ~ 1, random = ~animal, mev = Adams.data$mev,
+     data = Adams.data, pedigree = Adams.phylo)
> m11.mcmc <- MCMCglmm(Z ~ 1, random = ~animal + us(mesd):units,
+     data = Adams.data, pedigree = Adams.phylo)
```

Section: Taxonomic & Phylogenetic Mixed Model

In model `m9` we reference Felsenstein (2008) under the assumption that the repeated measurements have been averaged over and a weighted analysis fitted. However Felsenstein (2008) actually presents a model, which although similar in purpose, uses more information by not averaging repeated measurements on the same species, but by fitting species as a random effect. If `Adams.data` had repeated measurements per species, the model of Felsenstein (2008) could also have been fitted in `asreml`:

```
> m9b <- asreml(Z ~ 1, random = ~giv(animal, var = T) + species,
+     data = Adams.data, ginverse = list(animal = IAasreml), family = asreml.gaussian(dispersion = 1e-08))
```

or `MCMCglmm` with a residual term:

```
> m9b.mcmc <- MCMCglmm(Z ~ 1, random = ~animal + species, data = Adams.data,
+     pedigree = Adams.phylo)
```

In a similar vein, the taxonomic model defined by equation 15 could be fitted as :

```
> asreml(Z ~ 1, random = ~giv(animal, var = T) + genus + species,
+     data = Adams.data, ginv = list(animal = IAasreml))
> MCMCglmm(Z ~ 1, random = ~animal + genus + species, data = Adams.data,
+     pedigree = Adams.phylo)
```

Section: Multivariate Models

For multivariate models the random terms can be interacted with the reserved variable `trait` which indexes each response variable:

```
> asreml(cbind(y1, y2) ~ trait - 1, random = ~us(trait):giv(animal,
+     var = F), rcov = ~us(trait):units, ginverse = list(animal = IAasreml),
+     data = Adams.data)
> MCMCglmm(cbind(y1, y2) ~ trait - 1, random = ~us(trait):animal,
+     rcov = ~us(trait):units, pedigree = Adams.phylo, data = Adams.data)
```

where the term `us(trait)` specifies a covariance matrix for which all elements are to be estimated, and the dimension of the matrix is equal to the number of traits, in this case 2. The interaction with animal specifies a matrix that describes the expected (co)variances within and between different traits due to phylogeny, which we designate as $\mathbf{V}_a$ in the manuscript. Equivalent matrices exist for other sources of random variation, an in particular `rcov=` $\sim$ us(trait):units specifies the 2×2 matrix $\mathbf{V}_e$ which is the expected (co)variances of the residuals in different traits.

Section: R-BUGS code

We also provide R-BUGS code for fitting the meta-analytic model using MCMC:

```
> model {
+
+   for (j in 1:NSPE){                        # NSPE = no. of species
+     Zr[j] ~ dnorm (Zr.hat[j], tau.Zr[j])  # Zr = Fisher's Z
+     tau.Zr[j]<-1/VarZr[j]                 # VarZr = Variance for Zr
+     Zr.hat[j]<-mu+SPE.hat[Species[j]]     # mu = grand mean
+   }
+                                           # Species = species list
+   mu ~ dnorm(0, 0.0001)
+
+   for (i in 1:NSPE){
+     SPE.hat[i]<-PHY[i]+ERR[i]             # PHY = phylogenetic effect
+     ERR[i] ~ dnorm(0, tau.ERR)            # ERR = residual
+   }
+
+   tau.ERR<-pow(sigma.ERR, -2)
+   sigma.ERR ~ dunif(0, 1000)
+   PHY[1:NSPE] ~ dmnorm(zeros[ ], Tau.PHY[ , ]) # zeros = a vector of 0s
+   Tau.PHY[1:NSPE,1:NSPE]<-inverse(Sigma.PHY[ , ])
+
+   for(i in 1:NSPE){
+     for(j in 1:NSPE){
+       Sigma.PHY[i, j]<-sigma.PHY*sigma.PHY*Tree[i, j]
+     }                          # Tree = correlation matrix for phylogeny
+   }
+
```

```
+   sigma.PHY ~ dunif(0, 1000)
+
+ }
```

and R-BUGS code for the non-meta-analytic model:

```
> model {
+
+   for (j in 1:NSPE){
+     Zr[j] ~ dnorm (Zr.hat[j], tau.ERR)
+     Zr.hat[j]<-mu+PHY[Species[j]]
+   }
+
+   tau.ERR<-pow(sigma.ERR, -2)
+   sigma.ERR ~ dunif(0, 1000)
+   mu~dnorm(0, 0.0001)
+   PHY[1:NSPE]~dmnorm(zeros[],Tau.PHY[,])
+   Tau.PHY[1:NSPE,1:NSPE]<-inverse(Sigma.PHY[ , ])
+
+   for(i in 1:NSPE){
+     for(j in 1:NSPE){
+       Sigma.PHY[i, j]<-sigma.PHY*sigma.PHY*Tree[i, j]
+     }
+   }
+
+   sigma.PHY ~ dunif(0, 1000)
+
+ }
```

Two additional sections removed because of lack of space

Prediction of Ancestral States

As Lynch (1991) pointed out, prediction of ancestral state is analogous to the prediction of breeding value in animal breeding. In animal breeding best linear unbiased prediction (BLUP) is the most widely used method for prediction of breeding value, and is used as a method for selecting on a range of commercially important traits. However, it is well known that the desirable properties of BLUP depend on the variance components being known without error (Lynch & Walsh 1998). There may be some justification for this in animal breeding where variance components may be estimated from hundreds of thousands of records, but in the context of phylogenies precise variance estimates are unlikely. Bayesian estimation of ancestral states avoids these problems by integrating over the uncertainty in variance components to give the marginal distributions of ancestral state (see Blasco (2001) for an in depth discussion regarding BLUP and Bayesian estimators, and also Pagel *et al.* (2004) in the context of phylogenies). Similar arguments hold for the prediction of ancestral states in models of discrete traits that work with transition probabilities rather than variance components (Ronquist 2004). In addition, although BLUP offers unbiased measures of ancestral state when the variance components are known, inferences about the distribution of ancestral states based on the distribution of BLUPs can be severely biased. These biases can be severe (Postma 2006), particularly when the distribution of BLUP's is used to test for deviations from model assumptions (Hadfield 2008). For example, using BLUP ancestral states to measure directional evolutionary change will be biased towards zero when the model assumes Brownian motion (Webster & Purvis 2002). Moreover, using BLUP ancestral states to test the significance of such deviations is prone to difficulty because of the uncertainty within, and dependence between predictions is often not taken into account.

Uncertain Phylogenies

Although, phylogenies in comparative analysis are often treated as known fixed quantities, there is often considerable error in their estimation. Conclusions drawn from the analysis of single phylogenies may therefore lack robustness because they fail to incorporate the uncertainty in phylogeny estimation into uncertainty in the parameter of primary interest (e.g. heritability). With frequentist techniques it can be difficult to incorporate these two sources of error: one arising because the phylogeny is not known with certainty, and one arising because the phylogeny is finite in size (Martins 1996). On the other hand, MCMC techniques are ideally suited for integrating these two sources of information by evaluating the distribution of the parameters conditional on a new phylogeny drawn from the posterior distribution of phylogenies at each iteration (Ronquist 2004; Pagel *et al.* 2004). Given that reconstruction of phylogenies is becoming increasingly Bayesian (Huelsenbeck & Ronquist 2001; Drummond & Rambaut 2007), this method should become a common practical tool for making comparative analysis robust to uncertainty in the phylogeny (Ronquist 2004). However, this method does assume that the trait data do not provide additional information to the genotype data regarding the phylogeny. It can be shown, again using the analogy with pedigree reconstruction that this can only hold if the genetic data are so informative that the phylogeny is resolved without error and/or the phylogenetic heritability of the trait is zero (Hadfield *et al.* 2006). Failing this, estimates of phylogenetic heritability will be biased towards zero unless a model is constructed that simultaneously estimates the phylogeny and the parameters of the comparative analysis.

References

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

Blasco, A., 2001. The Bayesian controversy in animal breeding. *J. Anim. Sci.* **79**: 2023–2046.

Drummond, A. J. & Rambaut, A., 2007. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**.

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

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

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

Hadfield, J. D., Richardson, D. S. & Burke, T., 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* **15**: 3715–31.

Huelsenbeck, J. P. & Ronquist, F., 2001. Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

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

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.

Martins, E. P., 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. *Evolution* **50**: 12–22.

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

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.

Postma, E., 2006. Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. *J. Evol. Biol.* **19**: 309–320.

Ronquist, F., 2004. Bayesian inference of character evolution. *Trends Ecol. Evol.* **19**: 475–481.

Webster, A. J. & Purvis, A., 2002. Testing the accuracy of methods for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B Biol. Sci.* **269**: 143–149.