

Chapter 11– *Electronic Supplement* General Quantitative Genetic Methods for Comparative Biology

Pierre de Villemereuil and Shinichi Nakagawa

This Online Practical Material (OPM) is intended to provide working examples of the different models described in the corresponding chapter (hereby referred to as the main text). As stated in the main text, we will focus on the R package MCMCglmm (Hadfield, 2010a) to fit phylogenetic mixed models.

11.1 A simple model in MCMCglmm

Sources

`ape`{R} (Paradis et al, 2004)
`MCMCglmm`{R} (Hadfield, 2010a)

Data

`data_simple.txt` Data frame containing the phenotypic data
`phylo.nex` Phylogeny file (NEXUS file)

We will begin by fitting the simple comparative model described in main text section 2.1, using simulated data. Assume we have measurements of a phenotype `phen` (say the body size) and a `cofactor` variable (say the temperature of the environment), for several species:

```
> library(ape)
> library(MCMCglmm)
> phylo<-read.nexus("phylo.nex")
> data<-read.table("data_simple.txt",header=TRUE)
> head(data)
```

	phen	cofactor	phylo
1	107.06595	10.309588	sp_1
2	79.61086	9.690507	sp_2
3	116.38186	15.007825	sp_3
4	143.28705	19.087673	sp_4
5	139.60993	15.658404	sp_5
6	68.50657	6.005236	sp_6

We want to investigate a possible relationship between the phenotypes and the cofactor, while controlling for any phylogenetic dependency in the dataset.

de Villemereuil P, Nakagawa S (2014) General Quantitative Genetic Methods for Comparative Biology. In: Garamszegi LZ (ed.), Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Springer, Heidelberg, p 000-000

To achieve our goal, we want to use the phylogenetic mixed model of the OPM section 11.1, implemented in the `MCMCglmm` package. Note the `phylo` column containing the name of the species in our dataset: it corresponds to the phylogenetic effect we are going to include in our model. In order to do so, we need to calculate the inverse of the Σ matrix of phylogenetic correlation:

```
> inv.phylo<-inverseA(phylo,nodes="TIPS",scale=TRUE)
```

The `inverseA()` function accepts R phylogenetic objects. Here, we state that we want to calculate Σ^{-1} using the argument `nodes='TIPS'`. If, as explained in the Introduction of the main text, we want to calculate the inverse of the bigger matrix Ω (for a much larger phylogeny, it would be much more efficient), we would have used `nodes='ALL'` to include ancestral nodes into the calculation. The `scale` argument yields a correlation matrix (scaling total branch length, from root to tips, to one).

Now, we have the inverse of our matrix, but because it is using a Bayesian framework, this package needs prior distributions for the fixed and random effects. The default prior for the fixed effects is suitable for our needs. However, regarding the random effects, we need to define a set of priors for the variance components of each random effect:

```
> prior<-list(G=list(G1=list(V=1,nu=0.02)),R=list(V=1,nu=0.02))
```

These priors (`G` for the random effect(s), and `R` for the residual variance) correspond to an inverse-Gamma distribution with shape and scale parameters equal to 0.01^1 , which is relatively canonical, though not without drawbacks (see Gelman, 2006, for more information). The model is then defined as follows:

```
> model_simple<-MCMCglmm(phen~cofactor,random=~phylo,
  family="gaussian",ginverse(phylo=phylo.inv$Ainv),prior=prior,
  data=data,nitt=5000000,burnin=1000,thin=500)
```

Here, we assume a linear relationship between `phen` and `cofactor`, with a random effect `phylo` corresponding to the phylogenetic effect. The argument `ginverse` allows us to include a custom matrix for our random effect `phylo`, using the results of the `inverseA` function (above). We used the `prior` variable defined above. The variables `nitt` and `burnin` are used to calibrate the MCMCM algorithm: it will iterate for `burnin` iterations before recording samples (to ensure convergence), and then iterate `nitt` times. The parameter `thin` helps us to save memory by saving only every '`thin`' value and thus,

¹ `MCMCglmm` univariate prior formulation is such that it corresponds to an inverse-Gamma with shape parameter $\alpha = \frac{\text{nu}}{2}$ and scale parameter $\beta = \frac{\text{nu} \times \mathbf{V}}{2}$. It is important to note that this inverse-Gamma could become unwantedly 'informative' when variance components are close to 0 so that it is always recommended running models with different prior specifications. For which prior should be used, see discussion in Hadfield (2010b) and also one can find more recent discussion on this topic online within the correspondences in the `r-sig-mixed-modes` mailing list.

dropping highly auto-correlated values². Note that the use of `nodes='TIPS'` or `nodes='ALL'` in the `inverseA` function can have a noticeable impact on auto-correlation: whereas the latter would speed up computation, it can result in higher auto-correlation. Whether to use one or the other would thus depend mainly on the size of the phylogeny (very large phylogenies would probably need `nodes='ALL'` to allow `MCMCglmm` to run at all). Finally, note that this example can take up to a few hours to run.

After checking for convergence (for example using `heidel.diag()` function), we can look at the output summary:

```
> summary(model_simple)

Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998

DIC: 1515.822

G-structure: ~phylo

      post.mean 1-95% CI u-95% CI eff.samp
phylo      210    92.27   333.8    9998

R-structure: ~units

      post.mean 1-95% CI u-95% CI eff.samp
units      85.83    60.78   112.8   10986

Location effects: phen ~ cofactor

      post.mean 1-95% CI u-95% CI eff.samp pMCMC
(Intercept)  39.714  26.086  53.602   9998 <1e-04 ***
cofactor      5.179   4.902   5.444   9998 <1e-04 ***
```

The first part shows a summary of MCMC sampling parameters, and gives the Deviance Information Criterion (DIC) of the model. The DIC is a model selection criterion comparable to Akaike's Information Criterion (AIC)³. Following are the results for the random effect variances (`G-structure`, contain-

² Because the MCMC sampling is Markovian, it is a time-series, which often appears to be auto-correlated: closely following iterations tend to resemble each other. The "thinning" help to save memory when running the MCMC for longer. For phylogenetic mixed model, this auto-correlation can be large and problematic: always make sure your effective sample size is large enough and that auto-correlation is low.

³ Here we do not provide explanations on how information criteria can be used for model selection. For a detailed discussion on this topic, the reader is referred to Chapter 12.

ing the variance of the `phylo` effect) and the residual variance (`R-structure`, the residual variance is called `units` in `MCMCglmm`). We have information about the posterior mean of the estimate, its 95% credible interval⁴ and its effective sample size. The latter is a measure of the auto-correlation within the parameter sample: it should be close to the MCMC sample size above, or failing that, it should be at least large enough (say more than 1,000). The summary of the fixed effects (`intercept` and `cofactor`) are similar, except we also have a “pMCMC” value for significance testing if the parameter is different from zero⁵. By using `plot(model_simple)`, we can obtain the “trace” of the sampling (to check for convergence and auto-correlation) and posterior density of each parameter (Fig. 11.1).

Finally, we can easily calculate the posterior probability of the phylogenetic signal λ (see section 2 in the main text) using:

```
> lambda <- model_simple$VCV['phylo']/
(model_simple$VCV['phylo']+model_simple$VCV['units'])
```

We can calculate the posterior mean (mean of the posterior distribution), posterior mode (most likely value regarding the posterior distribution) and the 95% credible interval of λ :

```
> mean(lambda)
[1] 0.6961347
> posterior.mode(lambda)
      var1
0.7442347
> HPDinterval(lambda)
      lower      upper
var1 0.5267074 0.8522012
```

11.2 Multiple measurements model in `MCMCglmm`

Sources

`ape`{R} (Paradis et al, 2004)
`MCMCglmm`{R} (Hadfield, 2010a)

Data

`data_repeat.txt` Data frame containing the phenotypic data (with multiple measurement)
`phylo.nex` Phylogeny file (NEXUS file)

⁴ Credible interval can be considered as the Bayesian version of confidence intervals, and also it is known as the highest posterior density (see Hadfield, 2010b).

⁵ If we are strictly Bayesian, we should not do significance testing because such a concept belongs to the frequentists’ paradigm. However, we use “pMCMC” as if frequentists’ *p-values* for convenience.

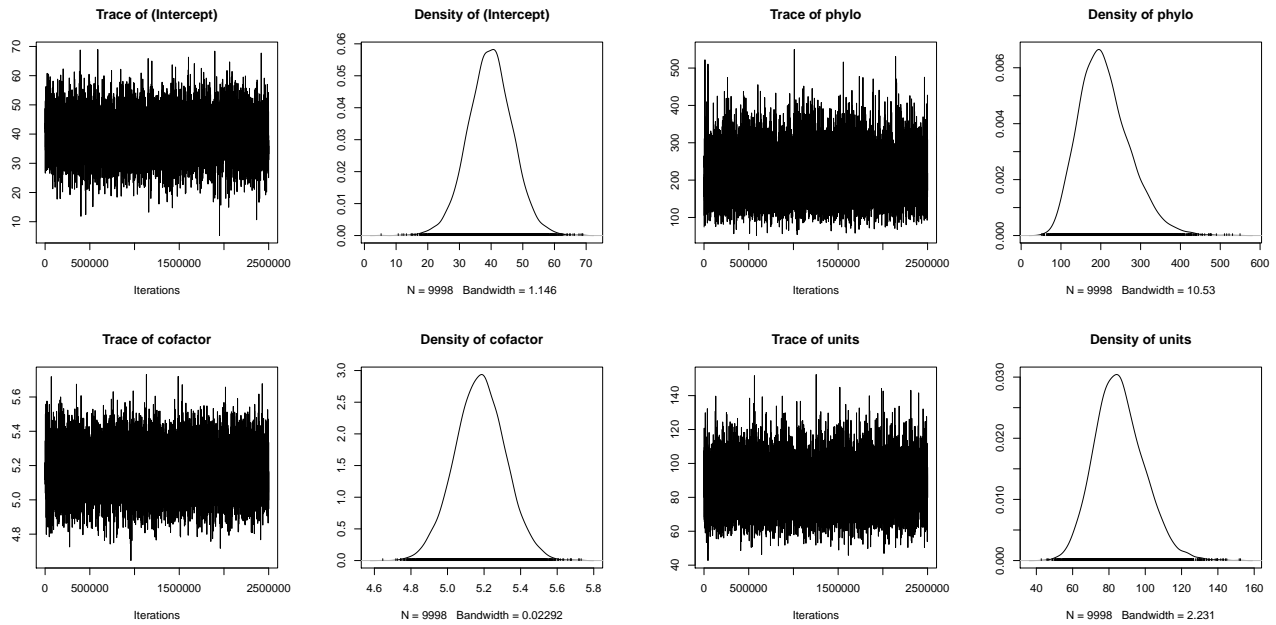


Fig. 11.1 Plot of trace and posterior density for fixed effects (left) and variance parameters (right).

Let's build up on the example from the last section by adding multiple measurements per species:

```
> data<-read.table("data_repeat.txt",header=TRUE)
> head(data)
      phen  cofactor species phylo
1 107.41919 11.223724  sp_1  sp_1
2 109.16403  9.805934  sp_1  sp_1
3  91.88672 10.308423  sp_1  sp_1
4 121.54341  8.355349  sp_1  sp_1
5 105.31638 11.854510  sp_1  sp_1
6  64.99859  4.314015  sp_2  sp_2
```

How can we analyse such a dataset using a phylogenetic mixed model? We only have to add a new random effect taking into account the fact that each species has a “multiple measurement effect”. Note the new column `species`, which is identical to the `phylo` one and will be used for that purpose. First, we will try to fit the model described in Eqn. 6 (main text), using the specific mean of the cofactor:

```
> data$spec_mean_cf<-sapply(split(data$cofactor,
data$phylo),mean)[data$phylo]
```

The code is a bit complex because we need to overcome the fact that R is trying to sort alphabetically the species name (which we don't want!), but we are simply calculating the mean of the cofactor for each species and pulling it in a new entry of the data frame. Now, because we have a new random effect, we also need to change our priors. The model would thus be:

```
> prior2<-list(G=list(G1=list(V=1,nu=0.02),G2=list(V=1,nu=0.02)),
  R=list(V=1,nu=0.02))
> model_repeat1<-MCMCglmm(phen~spec_mean_cf,random=~phylo+species,
  family="gaussian",ginverse=list(phylo=phylo.inv$Ainv),
  prior=prior2,data=data,nitt=5000000,burnin=1000,thin=500)
```

Since the columns `phylo` and `species` have the same content and are called in the same fashion in the model, one would wonder if they are not accounting for the same thing. Actually, a careful inspection of the `ginverse` argument would show that we are providing the phylogenetic variance-covariance for the `phylo` effect, but not for the `species` one. This means that `species` is here to account for any specific effect that would be independent from the phylogenetic relationship between species (e.g. environmental/niche effects). This is also visible in Eqn. 4 in the main text. Here, because we distinguish these effects from the residual variance, the residual variance `units` now actually corresponds to the intra-specific variance (which is assumed equal across species). The output summary follows:

```
> summary(model_repeat1)

Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998

DIC: 7187.331

G-structure: ~phylo

      post.mean 1-95% CI u-95% CI eff.samp
phylo    277.7    152.7    403.4     8804

      ~species

      post.mean 1-95% CI u-95% CI eff.samp
species     24.38    8.477    41.66     9444

R-structure: ~units

      post.mean 1-95% CI u-95% CI eff.samp
units     65.78    59.49    72.28     9343
```

Location effects: phen ~ spec_mean_cf

	post.mean	l-95% CI	u-95% CI	eff.samp	pMCMC
(Intercept)	38.320	23.023	54.504	9998	<1e-04 ***
spec_mean_cf	5.097	4.888	5.289	9253	<1e-04 ***

Because our sampling of individuals within species was totally unbiased, the results are similar, except we now have an estimate of the intra-specific variance, which is represented by the residual variance units.

We did not, however, use the whole dataset in the previous model. We totally ignored the intra-specific variability of the cofactor. In order to get an estimate for the “between-species” and “within-species” (see Eqn. 7 in the main text), we need to use the *within-group centring* technique:

```
> data$within_spec_cf<-data$cofactor-data$spec_mean_cf
```

We can now use a slightly more elaborate model:

```
> model_repeat2<-MCMCglmm(phen~spec_mean_cf+within_spec_cf,
  random=~phylo+species,family="gaussian",
  ginverse=list(phylo=phylo.inv$Ainv),prior=prior2,data=data,
  nitt=5000000,burnin=1000,thin=500)
```

The fixed effect `spec_mean_cf` corresponds to the between-species slope (just as the previous model) and the fixed effect `within_spec_cf` corresponds to the within-species slope. As usual, we can have a look at the results:

```
> summary(model_repeat2)
```

```
Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998
```

```
DIC: 7189.433
```

```
G-structure: ~phylo
```

	post.mean	l-95% CI	u-95% CI	eff.samp
phylo	278.3	158.7	408.9	9998

```
~species
```

	post.mean	l-95% CI	u-95% CI	eff.samp
species	24.39	8.303	41.67	9998

```
R-structure: ~units
```

```
      post.mean 1-95% CI u-95% CI eff.samp
units      65.78   59.31   72.42   9998
```

```
Location effects: phen ~ spec_mean_cf + within_spec_cf
```

```
      post.mean 1-95% CI u-95% CI eff.samp pMCMC
(Intercept)   38.30006 22.54138 53.80957   9998 <1e-04 ***
spec_mean_cf   5.09879  4.89180  5.29816   9998 <1e-04 ***
within_spec_cf -0.05757 -0.42837  0.30606   9998  0.747
```

The results are almost unchanged, with apparently no relationship between the phenotype `phen` and `cofactor` on the intra-specific level. Finally, we can again calculate λ using:

```
> lambda <- model_repeated2$VCV[, 'phylo'] /
  (model_repeated2$VCV[, 'phylo'] + model_repeated2$VCV[, 'species'] +
  model_repeated2$VCV[, 'units'])
```

11.3 An example of meta-analysis with Fisher's transformation of r (Zr)

Sources

`ape`{R} (Paradis et al, 2004)
`MCMCglmm`{R} (Hadfield, 2010a)

Data

`data_effect.txt` Data frame containing the effects for meta-analysis
`phylo.nex` Phylogeny file (NEXUS file)

Let's use the same phylogeny as in the OPM section 11.2. We have an effect size in Fisher's z-transformation of correlation coefficient Zr per species along with corresponding sample sizes (e.g. correlations between male coloration and reproductive success):

```
> phylo <- read.nexus("phylo.nex")
> data <- read.table("data_effect.txt", header=TRUE)
> head(data)
      Zr  N  phylo
1 0.28917549 13 sp_1
2 0.02415579 40 sp_2
3 0.19513651 39 sp_3
4 0.09831239 40 sp_4
5 0.13780152 66 sp_5
6 0.13710587 41 sp_6
```


It is straightforward to fit a meta-analytic model in MCMCglmm. Let's see how the syntax looks like:

```
> inv.phylo<-inverseA(phylo,nodes="ALL",scale=TRUE)
> prior<-list(G=list(G1=list(V=1,nu=0.02)),R=list(V=1,nu=0.02))
> model_effect<-MCMCglmm(Zr~1,random=~phylo,family="gaussian",
  mev=1/(data$N-3),ginverse=list(phylo=phylo.inv$Ainv),
  prior=prior,data=data,nitt=5000000,burnin=1000,thin=500)
```

As you may have noticed the syntax is pretty much the same as in the OPM section 11.1, including the same `prior` specification. A notable difference is that we now use the argument `mev`, which stands for measurement error variance. We can pass vector of sampling variances to `mev`. We do not need a prior for sampling error variances because we assume they are known (remember for Zr , it is $\frac{1}{n-3}$), thus not to be estimated. Also, for a quicker computation, we used `nodes="ALL"` in the `inverseA` function this time⁶ in the main text. The result looks like this:

```
> summary(model_effect)

Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998

DIC: -320.2919

G-structure: ~phylo

      post.mean 1-95% CI u-95% CI eff.samp
phylo  0.009016 0.001747  0.01931   10423

R-structure: ~units

      post.mean 1-95% CI u-95% CI eff.samp
units  0.006078 0.001671  0.01113    9998

Location effects: Zr ~ 1

      post.mean 1-95% CI u-95% CI eff.samp pMCMC
(Intercept)  0.15887  0.06005  0.25381   9147 0.0038 **
```

⁶ In most analyses, this default option will be fine. As mentioned before, "TIPS" option could reduce auto-correlation and will improve the chain "mixing". Note that you will get a warning message, saying some missing records are generated. This is because MCMCglmm is using ancestral nodes and trait values of such nodes are treated as missing values. See more on this in section 3.2

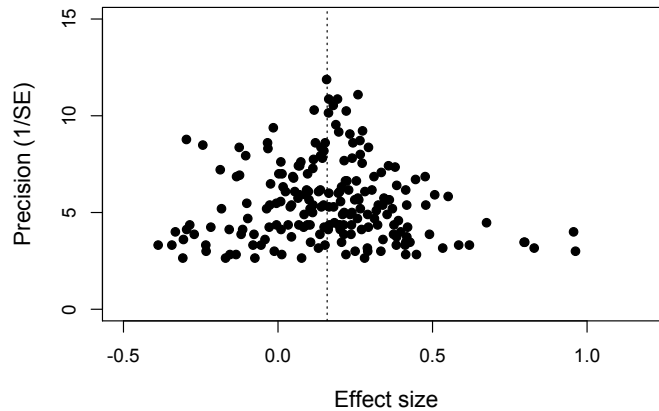


Fig. 11.2 Funnel plot for effect sizes (Zr) with the meta analytic mean (dashed line).

The meta-analytic mean is $Zr = 0.15887$ and is significantly larger than zero. In meta-analysis, it is common to plot what is called a funnel plot where effect sizes are plotted with the inverse of the squared root of sampling error variance, called ‘precision’ (Fig. 11.2). As you see, effect sizes funnel down around the meta-analytic mean. This is what we exactly expect because effect sizes with low precisions (low sample sizes) should have larger sampling errors. Here, we do not go any further with phylogenetic meta-analysis. But to follow up on this topic, you may want to see recent examples of phylogenetic meta-regression models using `MCMCglmm` in Horvathova et al (2012) and Prokop et al (2012). Other important issues in meta-analysis include statistical heterogeneity and publication bias (for further information, see Nakagawa and Santos, 2012; Koricheva et al, 2013).

11.4 An example of non-Gaussian trait using count data

Sources

`ape`{R} (Paradis et al, 2004)
`MCMCglmm`{R} (Hadfield, 2010a)

Data

`data_pois.txt` Data frame containing the count phenotypic data
`phylo.nex` Phylogeny file (NEXUS file)

Suppose we have to analyse a dataset alike the one in the OPM section 11.1, but we are now interested in count data without multiple measurement:

```

> data<-read.table("data_pois.txt",header=TRUE)
> head(data)
  phen_pois  cofactor  phylo
1         1  7.8702830  sp_1
2         0  3.4690529  sp_2
3         1  2.5478774  sp_3
4        14 18.2286628  sp_4
5         1  2.5302806  sp_5
6         1  0.5145559  sp_6

```

Because we don't have multiple measurement, we can use the same prior and the same model as in our first example in the OPM section 11.1:

```

> prior<-list(G=list(G1=list(V=1,nu=0.02)),R=list(V=1,nu=0.02))
> model_pois<-MCMCglmm(phen_pois~cofactor,random=~phylo,
  family="poisson",ginverse=list(phylo=phylo.inv$Ainv),
  prior=prior,data=data,nitt=5000000,burnin=1000,thin=500)

```

Note that we are now using `family="poisson"`, which automatically assumes the canonical logarithmic link function. We can now print the summary of the results:

```

> summary(model_pois)
Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998

DIC: 690.3027

G-structure: ~phylo

      post.mean 1-95% CI u-95% CI eff.samp
phylo  0.04032 0.002295  0.1094    9998

R-structure: ~units

      post.mean 1-95% CI u-95% CI eff.samp
units  0.04206 0.002638  0.09635    9482

Location effects: phen_pois ~ cofactor

      post.mean 1-95% CI u-95% CI eff.samp pMCMC
(Intercept) -2.0849 -2.5007 -1.6930    9667 <1e-04 ***
cofactor     0.2508  0.2287  0.2732    9998 <1e-04 ***

```

As we can observe, random effects variances and fixed effects values are low. This is partly due to the assumed logarithmic link function which "impose"

low values for the latent trait \mathbf{I} (see Eqs. 14 and 15 in the main text). However, the fixed effects are significantly different from zero ($\text{pMCMC} < 10^{-4}$).

Generally, fitting the generalised phylogenetic mixed model using an MCMC algorithm is not much harder than fitting the Gaussian one. However one can encounter several issues. First, the algorithm will be slower for non-Gaussian traits. Second, issues due to auto-correlation might arise, so that one will be forced to run the algorithm for longer. Third, as noted above, the overall expected variances can be much smaller than for Gaussian traits (e.g. for binary and Poisson traits). In this case, issues related to the choice of the prior can arise, especially for small datasets. This is due to the fact that most variance priors (including those available in `MCMCglmm`) are a bit informative for small variances.

11.5 An example with MAR comparative data

Sources

`ape`{R} (Paradis et al, 2004)
`MCMCglmm`{R} (Hadfield, 2010a)

Data

`data_pois_missing.txt` Data frame containing the count phenotypic data with missing values
`phylo.nex` Phylogeny file (NEXUS file)

We will use the same data set as in the OPM section 11.4, but this time, we are missing one half of trait data `phen_pois` (100 out of 200 species). However, we have a complete data set for the cofactor, which, in this case, was related to missingness; we deleted the 50 trait values that had the 50 lowest values for the cofactor. Therefore, missing data in this data set is MAR. Let's look at the data:

```
> data<-read.table("data_pois_missing.txt",header=TRUE)
> head(data)
  phen_pois  cofactor  phylo
1      NA  7.8702830  sp_1
2      NA  3.4690529  sp_2
3      NA  2.5478774  sp_3
4      14 18.2286628  sp_4
5      NA  2.5302806  sp_5
6      NA  0.5145559  sp_6
>
```

NA indicates missing values. It is almost too easy to run a model dealing with missing data in the response variable. We just run an identical model as in

the OPM section 11.4. The only difference is that we will use the default `node="ALL"` rather than `node="TIPS"`:

```
> inv.phylo<-inverseA(phylo,nodes="ALL",scale=TRUE)
> prior<-list(G=list(G1=list(V=1,nu=0.02)),R=list(V=1,nu=0.02))
> model_missing<-MCMCglmm(phen_pois~cofactor,random=~phylo,
  family="poisson",ginverse=list(phylo=phylo.inv$Ainv),
  prior=prior,data=all,nitt=5000000,burnin=1000,thin=500)
> summary(model_missing)
```

If you finish running this model, you will get this warning message:

Warning message:

```
In MCMCglmm(phen_pois ~ cofactor, random = ~phylo, family = "poisson", :
some combinations in phylo do not exist and 198 missing records have been generated
```

This is *not* the data augmentation we are intending to show, but MCMCglmm is using its data augmentation technique to estimate all ancestral values at the internal nodes of the tree (`phylo`) apart from the root⁷, assuming MAR⁸. We now look at the output from this analysis:

```
> summary(model_missing)

Iterations = 1001:4999501
Thinning interval = 500
Sample size = 9998

DIC: 490.3962

G-structure: ~phylo

      post.mean 1-95% CI u-95% CI eff.samp
phylo  0.04045 0.002245  0.109      8700

R-structure: ~units

      post.mean 1-95% CI u-95% CI eff.samp
units  0.03824 0.002411  0.08832      8938

Location effects: phen_pois ~ cofactor

      post.mean 1-95% CI u-95% CI eff.samp pMCMC
```

⁷ This tree `phylo` has 200 tips and so it has 199 nodes ($n_{node} = n_{tip} - 1$). MCMCglmm augmented 198 missing values, which corresponds to $199 - 1$ nodes (the number of nodes without the root).

⁸ MCMCglmm uses the MAR assumption in this case, by assuming missingness depends on phylogeny (see section 3.2 in the main text).

(Intercept)	-2.0548	-2.6662	-1.4252	9126	<1e-04	***
cofactor	0.2495	0.2123	0.2836	9142	<1e-04	***

Compare the estimate of `cofactor` between this and one in the OPM section 11.4 (0.2508, 95% CI = 0.2287 to 0.2732). Notably, in this MAR missingness in the bivariate data context, complete-case analysis will not result in biased estimates; this is shown in Nakagawa and Freckleton (2008). Thus, we will not run such analysis.

References

- Gelman A (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian analysis* 1(3):515–533, DOI 10.1214/06-BA117A
- Hadfield JD (2010a) MCMC methods for multi-response generalised linear mixed models: The MCMCglmm R package. *Journal of Statistical Software* 33(2):1–22, URL <http://www.jstatsoft.org/v33/i02>
- Hadfield JD (2010b) MCMCglmm course notes. URL <http://cran.r-project.org/web/packages/MCMCglmm/index.html>
- Horvathova T, Nakagawa S, Uller T (2012) Strategic female reproductive investment in response to male attractiveness in birds. *Proceedings of the Royal Society B-Biological Sciences* 279(1726):163–170, 856WD Times Cited:2 Cited References Count:79
- Koricheva J, Gurevitch J, Mengersen K (2013) *The handbook of meta-analysis in ecology and evolution*. Princeton University Press, Princeton
- Nakagawa S, Freckleton RP (2008) Missing inaction: The dangers of ignoring missing data. *Trends in Ecology & Evolution* 23(11):592–596, 375RF Times Cited:0 Cited References Count:25
- Nakagawa S, Santos ESA (2012) Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* 26(5):1253–1274, sp. Iss. SI 989IC Times Cited:4 Cited References Count:115
- Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in r language. *Bioinformatics* 20(2):289–290, DOI 10.1093/bioinformatics/btg412, URL <http://bioinformatics.oxfordjournals.org/content/20/2/289.abstract>
- Prokop ZM, Michalczyk L, Drobniak SM, Herdegen M, Radwan J (2012) Meta-analysis suggests choosy females get sexy sons more than “good genes”. *Evolution* 66(9):2665–73, prokop, Zofia M Michalczyk, Lukasz Drobniak, Szymon M Herdegen, Magdalena Radwan, Jacek Evolution. 2012 Sep;66(9):2665-73. doi: 10.1111/j.1558-5646.2012.01654.x. Epub 2012 May 14.