Atelier sur les méthodes comparées

Simon Joly

2024-11-09

Contents

Chapter 1

À propos

Ce document consiste en une introduction aux méthodes comparatives. Il contient de la théorie ainsi que des exemples pratiques en R sur les moindres carrés phylogénétiques généralisés (PGLS). Il a été développé pour un atelier d'une demi-journée composé de courtes présentations suivies d'exercices R. Notez que le présent document devrait être autonome car la plupart de la théorie donnée dans les présentations est incorporée dans les sections théoriques. Par conséquent, ce document doit contenir toutes les informations nécessaires pour comprendre les exemples.

Je suppose que les lecteurs sont « raisonnablement » familiers avec R ainsi qu'avec la régression linéaire et ses hypothèses. Il existe de nombreux bons didacticiels d'introduction à R sur le Web et pour les modèles linéaires. Zuur et coll. (Zuur et al., 2007) fournit une bonne introduction aux modèles linéaires, aux modèles à effets mixtes et à la comparaison de modèles. De bonnes introductions à l'ajustement de modèles dans R peuvent également être trouvées sur la [page Web] de Dolph Schluter (https://www.zoology.ubc.ca/~schluter/R/fit-model/) et parmi les [ateliers QCBS] (http : //qcbs.ca/wiki/r_workshop4).

1.1 Resources utiles

Ces liens contiennent des information complémentaires pertinentes.

Le livre de Luke Harmon - Phylogenetic Comparative Methods

Le blog de Liam Revell

Le livre de Liam Revell et Luke Harmon book on Phylogenetic comparative methods in R

La liste des packages R pour les phylogénies

Mes tutoriels sur les méthodes comparées

Le package R V.PhyloMaker2 qui peut générer de larges phylogénies pour les plantes vasculaires, et le package U.PhyloMaker qui peut générer des phylogénies pour les plantes et les animaux.

1.2 Source

Ce tutoriel est disponible publiquement et est hébergé sur github dans de dépôt github.com/simjoly/AtelierPGLS

1.3 Disclaimer

Ce tutoriel est distribué tel quel, avec aucun garantie qu'il va fonctionner ou que les analyses seront à jour.

Chapter 2

Avant l'atelier

Voici quelques éléments que vous devriez connaître et faire avant l'atelier.

2.1 Installer R et les packages requis

Pour réaliser les exemples de ce document, vous aurez besoin d'avoir le logiciel R installé sur votre ordinateur. Je vous recommande fortement d'installer RStudio. Bien que R Studio ne soit pas requis, il facilite les interactions entre les scripts et la console R et offre de nombreux outils utiles.

Après avoir installé R, vous devrez installer certains packages. Pour ce tutoriel spécifique, nous devrons charger les packages R suivants.

```
library(nlme)
library(ape)
library(RColorBrewer)
library(ggplot2)
```

Pour exécuter le code de ce tutoriel dans R, je vous suggère de créer un nouveau script (Fichier>Nouveau Fichier>Script R) où vous collerez le code copié depuis les encadrés. Dans R Studio, vous pouvez ensuite exécuter ce code en sélectionnant les lignes que vous souhaitez exécuter puis en appuyant sur "Run" (ou le raccourci associé). Cela reproduira les analyses présentées dans le tutoriel. Vous devriez enregistrer le fichier script dans un répertoire dédié pour l'atelier où vous placerez également les fichiers de données requis (voir section 2.2). Ensuite, vous devez vous assurer que votre script (et les données) sont dans le répertoire de travail R. Dans R Studio, cela peut être défini via le menu : 'Session > Set Working Directory'.

Si certains des packages ci-dessus ne sont pas encore installés sur votre ordinateur, vous recevrez des messages d'erreur lors de leur chargement. Dans ce cas,

vous devrez les installer en utilisant la fonction `install.packages()`. Vous n'avez besoin de les installer qu'une seule fois.

```
install.packages('nlme')
install.packages('ape')
install.packages('RColorBrewer')
install.packages('ggplot2')
```

Une fois les packages installés, vous pouvez les charger en utilisant la fonction `library()`. Notez également que si vous utilisez à la fois les packages `nlme` et `ape`, `nlme` doit être chargé en premier. Sinon, vous pourriez rencontrer des erreurs; dans ce cas, vous pouvez redémarrer R et recommencer.

2.2 Téléchargement des données

Les données dont vous aurez besoin pour ce tutoriel peuvent être téléchargées depuis ce dépôt : data.zip.

Je vous suggère de télécharger le dossier, de le décompresser et de le placer dans un dossier dédié où vous enregistrerez également le script avec toutes les commandes que vous utiliserez.

2.3 Familiarisez-vous avec les arbres phylogénétiques dans R

Si vous n'avez jamais utilisé d'arbres phylogénétiques dans R, vous pouvez apprendre quelques techniques de base pour les manipuler et simuler des arbres et des caractères en lisant le chapitre 14.

Chapter 3

Une introduction aux méthodes comparatives phylogénétiques

Les méthodes comparatives phylogénétiques ont été introduites par Joseph Felsenstein en 1985. L'idée des méthodes comparatives phylogénétiques était de corriger la non-indépendance des espèces dans les tests statistiques en raison de leurs histoires évolutives partagées. En effet, deux espèces peuvent se ressembler non pas parce qu'elles vivent dans le même environnement mais parce qu'elles sont étroitement liées. Considérez la phylogénie des angiospermes suivante.

Il est clair que *Fagus* (hêtre) et *Pisum* (pois) sont plus susceptibles de partager des caractéristiques similaires par rapport à *Asplenium* (une fougère), car ils partagent un ancêtre commun plus récent. En d'autres termes, leurs histoires évolutives sont partagées sur une période plus longue qu'avec *Asplenium*. Ainsi, ils ont plus de chances d'avoir des traits plus similaires (et en fait, ils en ont). Par exemple, prenez deux caractères, l'ovule et le type de fécondation, au sein de ce groupe.

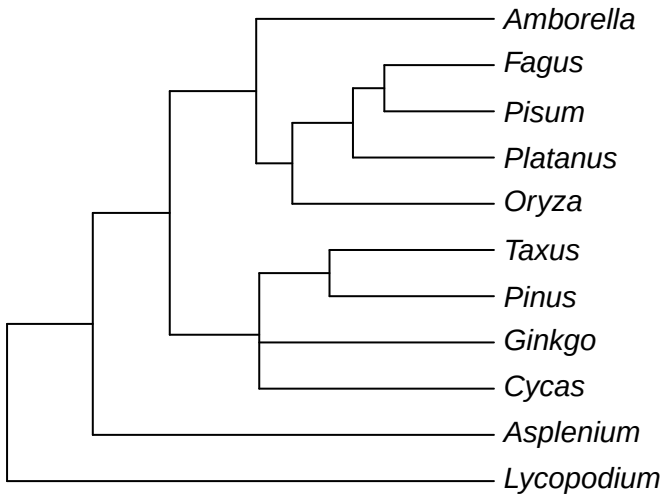
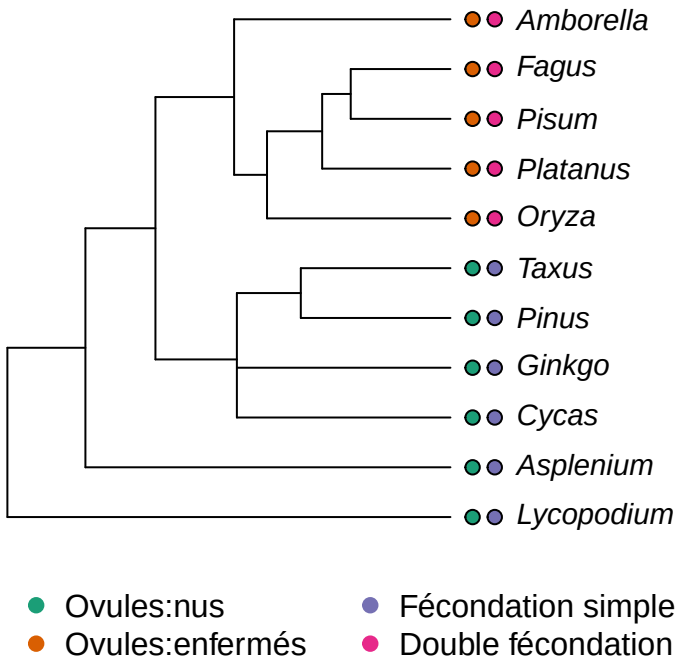


Figure 3.1: land plant phylogeny



En ignorant la phylogénie, nous pourrions être tentés de voir une forte corrélation entre ces deux caractères. En effet, les états entre les deux caractères montrent une correspondance parfaite. En utilisant les statistiques de tableau de contingence standard, nous pourrions faire un test exact de Fisher :

```
fisher.test(matrix(c(5,0,0,6),ncol=2))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  matrix(c(5, 0, 0, 6), ncol = 2)
## p-value = 0.002165
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  2.842809      Inf
## sample estimates:
## odds ratio
##        Inf
```

Le test suggère que l'association est hautement significative. Cependant, nous savons que les comparaisons faites ne sont pas complètement indépendantes. En réalité, les deux caractères n'ont évolué qu'une seule fois, et ce, le long de la même branche.

Une façon plus appropriée de poser la question serait "quelle est la probabilité que deux caractères aient évolué le long de la même branche ?". Cela peut également être calculé en utilisant un tableau de contingence, mais cette fois en prenant les branches de la phylogénie comme unités d'observation.

Dans cet exemple, il y a 18 branches et les deux caractères n'ont évolué qu'une fois et sur la même branche. Le tableau de contingence en considérant les changements le long des branches ressemble à ceci :

	Changement dans le trait 2	Pas de changement dans le trait 2
Changement dans le trait 1	1	0
Pas de changement dans le trait 1	0	17

Avec ce tableau, le test exact de Fisher donnera le résultat suivant :

```
fisher.test(matrix(c(1,0,0,17),ncol=2))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  matrix(c(1, 0, 0, 17), ncol = 2)
## p-value = 0.05556
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4358974       Inf
## sample estimates:
## odds ratio
##        Inf
```

Vous pouvez voir que le résultat n'est plus significatif.

Bien que cette approche pour prendre en compte les relations phylogénétiques soit correcte, des méthodes comparatives plus puissantes ont été développées. Une approche utile et puissante est le modèle de moindres carrés généralisés phylogénétiques (PGLS). Mais avant d'introduire PGLS, nous allons faire une révision et examiner brièvement la régression standard.

Chapter 4

Le modèle de régression linéaire

4.1 Théorie

Le modèle linéaire a la forme suivante :

$$\mathbf{y} = \alpha + \beta\mathbf{x} + \mathbf{e}$$

y est la variable de réponse (ou dépendante), **x** est la variable explicative (ou indépendante), et **e** représente les résidus ou en d'autres termes la variation non expliquée par le modèle. Pour un modèle de régression linéaire simple, cela représente la distance entre les observations (c'est-à-dire les données réelles) et la ligne de régression (c'est-à-dire la prédiction du modèle) le long de l'axe y. Les paramètres α représentent l'ordonnée à l'origine, qui es…

Obtenir des estimations fiables avec une régression linéaire implique que les données respectent plusieurs hypothèses, parmi lesquelles la normalité, l'homogénéité, X fixe, l'indépendance et la spécification correcte du modèle. Nous ne passerons pas en revue toutes ces hypothèses ici, mais nous nous concentrerons sur l'une d'entre elles, souvent violée lorsque les données sont structurées phylogénétiquement, qui est **l'indépendance**. Cette hypothèse est importante car un manque d'indépendance inval…

Vous obtenez une violation de l'indépendance lorsque la valeur $\mathbf{y}_i$ à $\mathbf{x}_i$ est influencée par d'autres $\mathbf{x}_i$. Évidemment, cela peut se produire avec des données structurées phylogénétiquement, car une variable de réponse est plus susceptible de réagir de manière similaire chez des espèces étroitement apparentées, car elles partagent de nombreux caractères par filiation. En d'autres termes, la valeur y d'une espèce n'est pas complètement indépendante de la valeur y d'un…

4.2 Pratique

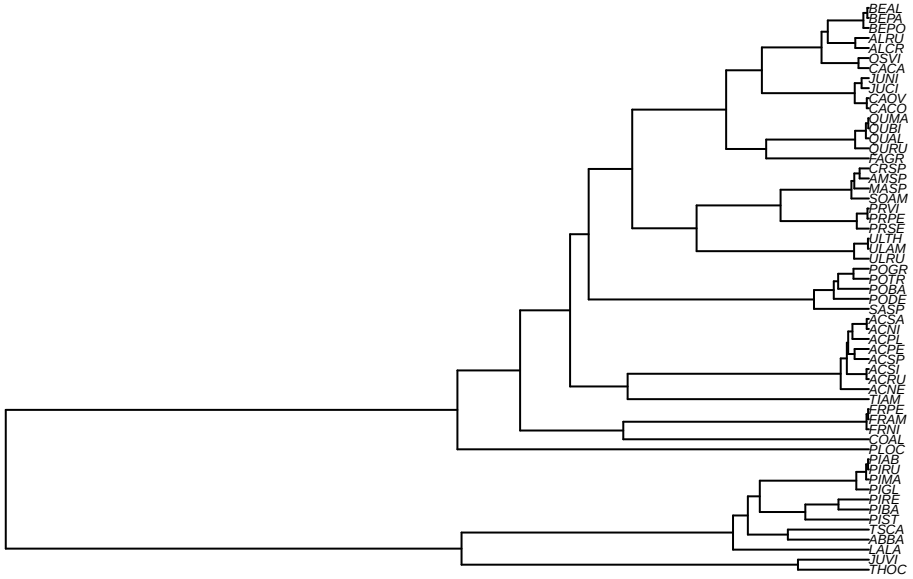
Pour fournir des exemples pratiques dans cet atelier, nous utiliserons un jeu de données de traits fonctionnels d'arbres de la province de Québec (Paquette et al., 2015). Le jeu de données comprend un certain nombre de traits fonctionnels des plantes et une phylogénie moléculaire construite à l'aide des marqueurs *rbc*L et *mat*K. Le jeu de données dont vous avez besoin pour exécuter les exemples se trouve déjà dans le dossier /data/ du dépôt github. Cependant, vous pouvez également les télécharger en …

seedplants.tre

seedplants.csv

Avant d'analyser les données, nous commencerons par ouvrir les données et l'arbre phylogénétique et les nettoyer pour ne conserver que les espèces présentes à la fois dans l'arbre et dans le tableau des traits. Cela est nécessaire car certaines espèces supplémentaires ont été incluses dans l'analyse phylogénétique.

```
require(ape)
# Ouvrir les documents ; cela suppose que vous êtes dans le répertoire principal du do
seedplantstree <- read.nexus("./data/seedplants.tre")
seedplantsdata <- read.csv2("./data/seedplants.csv")
# Supprimer les espèces pour lesquelles nous n'avons pas de données complètes
seedplantsdata <- na.omit(seedplantsdata)
# Supprimer les espèces de l'arbre qui ne sont pas dans la matrice de données
species.to.exclude <- seedplantstree$tip.label[!(seedplantstree$tip.label %in% seedplan
seedplantstree <- drop.tip(seedplantstree,species.to.exclude)
# Supprimer l'objet inutile
rm(species.to.exclude)
# Ordonner l'arbre pour qu'il soit plus esthétique lors du tracé
seedplantstree <- ladderize(seedplantstree, right = FALSE)
# Maintenant, regardons l'arbre
plot(seedplantstree,cex=0.4)
```



Maintenant, nous pouvons jeter un œil aux données, puis ordonner le trait des plantes pour qu'il soit dans le même ordre que les espèces de l'arbre.

```
# Voici à quoi ressemblent les données chargées
head(seedplantsdata)
```

```
##   Code       Species.name Occurrence maxH   Wd    Sm Shade    N
## 1 ABBA     Abies balsamea       7759   25 0.34   7.6   5.0 1.66
## 2 ACNE       Acer negundo          0   20 0.44  34.0   3.5 2.50
## 3 ACNI        Acer nigrum          1   30 0.52  65.0   3.0 1.83
## 4 ACPE Acer pensylvanicum        665   10 0.44  41.0   3.5 2.22
## 5 ACPL   Acer platanoides          0   15 0.51 172.0   4.2 1.99
## 6 ACRU        Acer rubrum       3669   25 0.49  20.0   3.4 1.91
```

```
# Nommer les lignes du data.frame avec les codes des espèces utilisés comme étiquettes d'arbre
rownames(seedplantsdata) <- seedplantsdata$Code
# Ordonner les données dans le même ordre que les étiquettes de l'arbre. Dans cet exemple
#  précis, c'était déjà le cas, mais c'est une étape importante pour toute analyse.
seedplantsdata <- seedplantsdata[seedplantstree$tip.label,]
```

Maintenant que les données sont prêtes, ajustons un modèle linéaire et essayons d'expliquer la tolérance à l'ombre (Shade) des arbres en utilisant la densité du bois (Wd). En R, une manière très simple de faire une régression est d'utiliser la fonction 'lm', qui signifie modèle linéaire. Pour ajuster un modèle linéaire, vous devez dire à la fonction `lm` quelle variable est la variable de réponse et laquelle est la variable explicative. Cela se fait en utilisant des formules de la forme `Shade ~ Wd`. La variab…

```r
# Ajuster un modèle linéaire en utilisant les moindres carrés ordinaires (MCO)
shade.lm <- lm(Shade ~ Wd, data = seedplantsdata)
# Imprimer les résultats
summary(shade.lm)
```

```
##
## Call:
## lm(formula = Shade ~ Wd, data = seedplantsdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.87120 -1.02501  0.05628  0.70132  2.38261
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.0010     0.7501   2.668    0.010 *
## Wd            1.8130     1.5676   1.157    0.252
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.146 on 55 degrees of freedom
## Multiple R-squared:  0.02374,    Adjusted R-squared:  0.005992
## F-statistic: 1.338 on 1 and 55 DF,  p-value: 0.2525
```

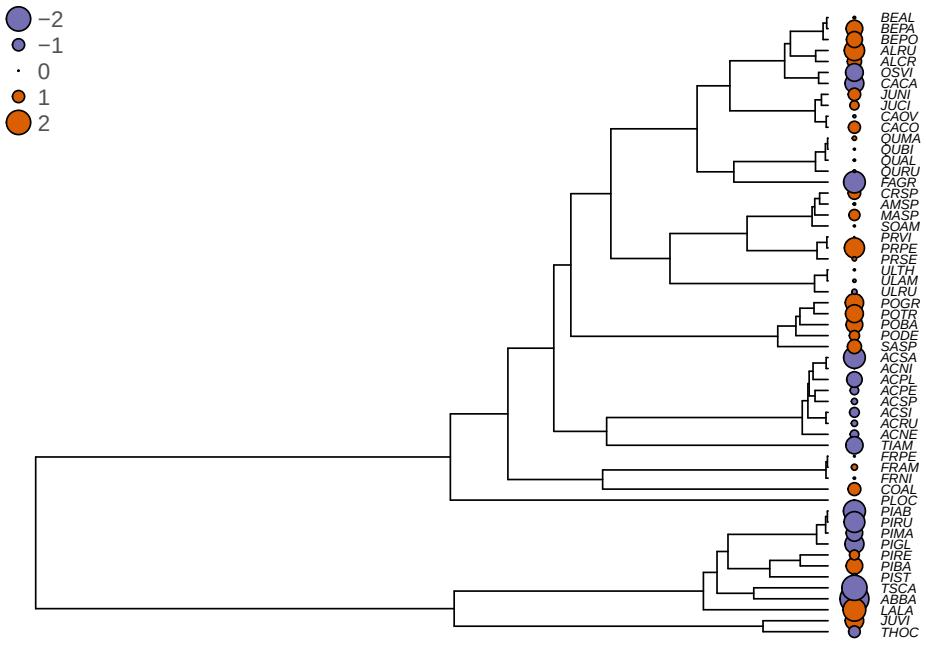
Vous pouvez voir que l'estimation de la pente (ici le paramètre `Wd`) est 1.81 et qu'elle n'est pas significative (p=0.252). Les graphiques descriptifs standards obtenus avec `plot(shade.lm)` montrent qu'il y a une légère variation plus grande dans les résidus pour les faibles valeurs ajustées, mais elles ne sont pas extrêmes. Cependant, une autre façon de violer l'hypothèse d'indépendance est si les résidus sont corrélés phylog…

```r
# Extraire les résidus
shade.res <- residuals(shade.lm)

#
# Tracer les résidus à côté de la phylogénie

# La commande suivante modifie les paramètres graphiques pour un meilleur rendu de l'a
op <- par(mar=c(1,1,1,1))
# Vecteur de couleurs pour le tracé de l'arbre
cols <- c("#7570b3","#d95f02")
# Les trois commandes suivantes vont tracer l'arbre, puis des cercles reflétant
#  les valeurs résiduelles aux extrémités de l'arbre, et enfin
#  ajouter une légende.
# La commande plot trace l'arbre et laisse un espace pour tracer les
#  résidus aux extrémités avec l'option 'label.offset=0.01'
```

```
plot(seedplantstree,type="p",TRUE,label.offset=0.01,cex=0.5,no.margin=FALSE)
# La commande suivante trace les résidus. l'option 'bg' est pour la couleur de fond.
#  Si les résidus sont supérieurs à 0 (shade.res>0), il affichera la première couleur
#  (1) du tableau 'cols' et s'il est inférieur à zéro, il affiche la deuxième couleur (2).
#  La taille du cercle (l'option 'cex') est relative à la valeur absolue
#  des résidus (abs(shade.res). Pour tracer d'autres valeurs, remplacez simplement le
#  vecteur 'shade.res' par un autre.
tiplabels(pch=21,bg=cols[ifelse(shade.res>0,1,2)],col="black",cex=abs(shade.res),adj=0.505)
# Imprimer la légende
legend("topleft",legend=c("-2","-1","0","1","2"),pch=21,
       pt.bg=cols[c(1,1,1,2,2)],bty="n",
       text.col="gray32",cex=0.8,pt.cex=c(2,1,0.1,1,2))
```




```
# Réinitialiser les paramètres graphiques par défaut
par(op)
```

Vous pouvez voir que dans plusieurs cas, des espèces étroitement apparentées ont tendance à avoir des résidus similaires (elles sont de la même couleur, ce qui signifie qu'elles sont du même côté de la pente de régression). Cela est problématique. En effet, cela montre que l'hypothèse d'indépendance de la régression des moindres carrés ordinaires (MCO) ne tient plus et les tests statistiques pour les hypothèses nulles ne sont plus valides. Nous verrons ensuite comment les moindres carrés généralisés phylo…

4.3 Défi no. 1

Dans le data frame `seedplantsdata`, il y avait plusieurs traits différents. Essayez d'ajuster une régression de la tolérance à l'ombre des arbres (`Shade`) en fonction de la masse des graines (`Sm`). En d'autres termes, testez si la tolérance à l'ombre peut être expliquée par la masse des graines des arbres. Ensuite, essayez de voir si les résidus sont corrélés phylogénétiquement.

Chapter 5

Phylogenetic generalized least squares (PGLS)

5.1 Théorie

Les moindres carrés généralisés phylogénétiques (PGLS) sont juste une application spécifique de la méthode plus générale appelée moindres carrés généralisés (GLS). Les moindres carrés généralisés relâchent l'hypothèse selon laquelle l'erreur du modèle linéaire doit être non corrélée. Ils permettent à l'utilisateur de spécifier la structure de cette corrélation résiduelle. Cela est utilisé, par exemple, pour corriger la corrélation spatiale, les séries temporelles ou la corrélation phylogénétique.

Les GLS ont la même structure que les Moindres Carrés Ordinaires (OLS) :

$$\mathbf{y} = \alpha + \beta\mathbf{x} + \mathbf{e}$$

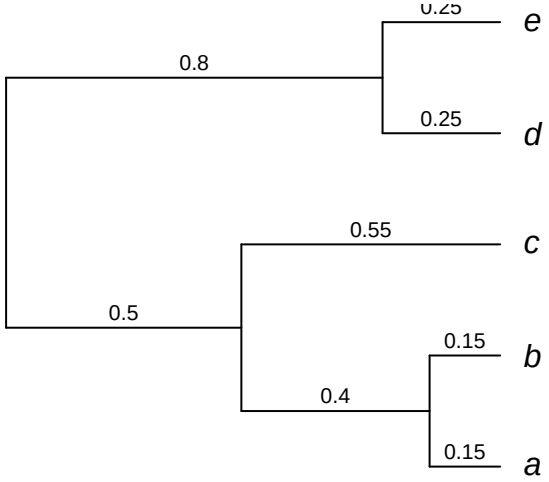
La seule différence est que les résidus sont corrélés entre eux selon une structure de corrélation $\mathbf{C}$:

$$\mathbf{e} \sim N(0, \sigma^2\mathbf{C})$$

Ici, $\mathbf{C}$ est une matrice de corrélation qui décrit comment les résidus sont corrélés entre eux. Pour pouvoir tenir compte des relations phylogénétiques dans un PGLS, nous devons donc être capables d'exprimer les relations phylogénétiques sous la forme d'une matrice de corrélation.

5.1.1 Structure de corrélation phylogénétique

Les relations phylogénétiques peuvent être décrites en utilisant une structure de corrélation. Ci-dessous, vous avez un arbre phylogénétique avec les longueurs de branches indiquées au-dessus des branches.



Maintenant, cet arbre peut être parfaitement représenté par une matrice de variance-covariance.

```
##      a    b    c    d    e
## a 1.05 0.90 0.50 0.00 0.00
## b 0.90 1.05 0.50 0.00 0.00
## c 0.50 0.50 1.05 0.00 0.00
## d 0.00 0.00 0.00 1.05 0.80
## e 0.00 0.00 0.00 0.80 1.05
```

Les éléments diagonaux de la matrice sont les variances des espèces ; ces nombres représentent la distance totale de la racine de l'arbre aux extrémités. Cela détermine dans quelle mesure les extrémités ont évolué par rapport à la racine. Les éléments hors diagonale sont les covariances entre les espèces. Ils indiquent la proportion du temps pendant laquelle les espèces ont évolué ensemble. Cela correspond à la longueur des branches que deux espèces partagent, à partir de la racine de l'arbre. Par exemple, les espèces a et c ont partagé une histoire commune pendant 0,5 unité de temps ; elles ont donc une covariance de 0,5. Plus la covariance est grande, plus les deux espèces ont partagé la même histoire évolutive.

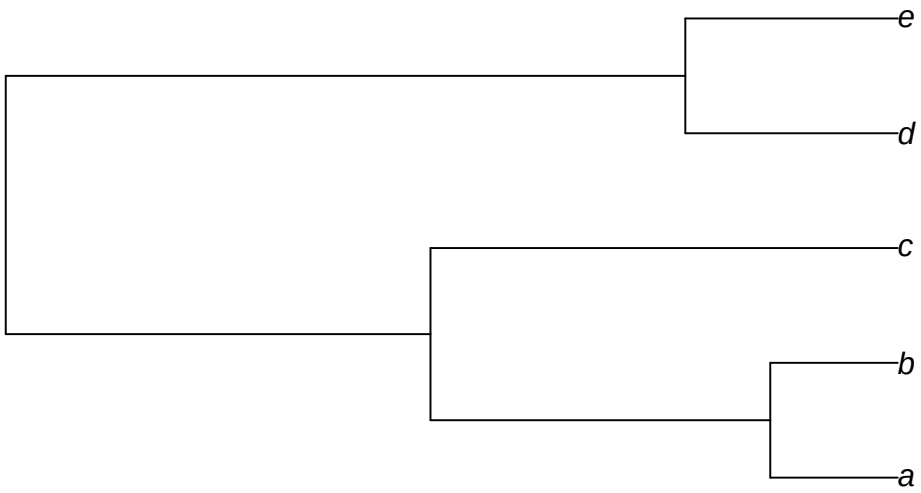
Si toute la variation entre les espèces était due à la phylogénie et non à la sélection, alors cette matrice de variance-covariance représenterait l'attente de la similitude entre toutes les espèces.

> Notez que toutes les extrémités sont équidistantes de la racine. Lorsque les arbres ont cette propriété, on dit qu'ils sont ***ultra-***

métriques. La plupart des méthodes comparatives phylogénétiques nécessitent que les arbres soient ultramétriques, bien qu'il existe parfois des moyens de relâcher cette hypothèse. Si vous n'avez pas un arbre ultramétrique, il est possible de le rendre ultramétrique en utilisant la fonction chronopl du package ape. Mais idéalement, il est préférable d'utiliser une méthode phylogénétique qui reconstruit directement des arbres ultramétriques.

La matrice de variance-covariance d'un arbre phylogénétique peut être obtenue à partir d'un arbre en utilisant la fonction vcv du package ape.

```
# 'atree' correspond à l'arbre phylogénétique montré ci-dessus au format newick
atree <- "(((a:0.15,b:0.15):0.4,c:0.55):0.5,(d:0.25,e:0.25):0.8);"
# Lisons maintenant cet arbre et stockons-le comme un objet d'arbre phylogénétique dans R
atree <- read.tree(text=atree)
# Afficher l'arbre
plot(atree)
```




```
# Extraire la matrice de variance-covariance
varcovar <- vcv(atree)
# Imprimer la matrice de variance-covariance
varcovar
```

```
##      a    b    c    d    e
## a 1.05 0.90 0.50 0.00 0.00
## b 0.90 1.05 0.50 0.00 0.00
## c 0.50 0.50 1.05 0.00 0.00
## d 0.00 0.00 0.00 1.05 0.80
## e 0.00 0.00 0.00 0.80 1.05
```

C'est excellent, mais nous avons mentionné ci-dessus qu'il s'agit d'une matrice de corrélation dont nous avons besoin dans un GLS pour tenir compte de la

corrélation dans les résidus. Pour obtenir une matrice de corrélation à partir de la matrice de variance-covariance montrée ci-dessus, il suffit de diviser la matrice de variance-covariance par la longueur de l'arbre, ou la distance de la racine aux extrémités. Elle peut également être obtenue en utilisant la fonction R `cov2cor`.

```
# Convertir la matrice de covariance en une matrice de corrélation
corrmat <- cov2cor(varcovar)
# Imprimer la matrice, arrondie à trois décimales
round(corrmat,3)
```

```
##       a     b     c     d     e
## a 1.000 0.857 0.476 0.000 0.000
## b 0.857 1.000 0.476 0.000 0.000
## c 0.476 0.476 1.000 0.000 0.000
## d 0.000 0.000 0.000 1.000 0.762
## e 0.000 0.000 0.000 0.762 1.000
```

Maintenant, les éléments diagonaux sont égaux à 1, indiquant que les espèces sont parfaitement corrélées avec elles-mêmes. Notez qu'il est également possible d'obtenir directement la matrice de corrélation à partir de la fonction `vcv` en utilisant l'option `corr=TRUE`.

```
# Obtention d'une matrice de corrélation en utilisant la fonction 'vcv'
corrmat <- vcv(atree,corr=TRUE)
round(corrmat,3)
```

```
##       a     b     c     d     e
## a 1.000 0.857 0.476 0.000 0.000
## b 0.857 1.000 0.476 0.000 0.000
## c 0.476 0.476 1.000 0.000 0.000
## d 0.000 0.000 0.000 1.000 0.762
## e 0.000 0.000 0.000 0.762 1.000
```

Maintenant que nous savons comment obtenir une matrice de corrélation à partir d'un arbre phylogénétique, nous sommes prêts à exécuter un PGLS.

5.2 Défi no. 2

Pouvez-vous obtenir la matrice de covariance et la matrice de corrélation pour l'arbre phylogénétique des plantes à graines de l'exemple ci-dessus (`seedplantstree`)?

5.3 Exercices pratiques

Il existe plusieurs façons d'exécuter un PGLS en R. Par exemple, le package `caper` est un package bien connu pour PGLS. Cependant, nous allons utiliser

ici la fonction gls du package `nlme`. Cette fonction est robuste et a l'avantage d'être très flexible. En effet, elle permet d'utiliser facilement des modèles plus complexes tels que les modèles à effets mixtes, bien que cela ne soit pas abordé ici.

Avant d'exécuter le PGLS, exécutons le modèle de base avec la fonction `gls` comme référence. Exécuter le modèle linéaire standard avec le package nlme permettra d'exécuter des fonctions de comparaison de modèles dans R (voir ci-dessous), ce qui ne serait pas possible si différents modèles étaient ajustés en utilisant différents packages.

```
require(nlme)
shade.pgls0 <- gls(Shade ~ Wd, data = seedplantsdata)
summary(shade.pgls0)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Wd
##   Data: seedplantsdata
##       AIC     BIC    logLik
##   180.472 186.494 -87.23602
##
## Coefficients:
##               Value Std.Error  t-value p-value
## (Intercept) 2.00098 0.7500707 2.667722  0.0100
## Wd          1.81296 1.5675668 1.156544  0.2525
##
##  Correlation:
##    (Intr)
## Wd -0.979
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -1.63307700 -0.89457443  0.04911902  0.61207032  2.07940955
##
## Residual standard error: 1.145813
## Degrees of freedom: 57 total; 55 residual
```

Vous pouvez voir que la sortie est essentiellement identique à celle de la fonction lm. Cependant, il y a quelques différences. L'une est la présence de l'élément « Correlation: » qui donne la corrélation entre les paramètres estimés. De plus, les « résidus standardisés » sont les résidus bruts divisés par l'erreur standard des résidus (les résidus bruts peuvent être affichés avec `residuals(shade.gls,"response")`).

Maintenant, exécutons un modèle PGLS. Pour attribuer la matrice de corrélation à la fonction `gls`, il suffit d'utiliser l'option `corr` de la fonction `gls`. Cependant, vous devez utiliser une fonction de corrélation spécifique pour que R comprenne qu'il s'agit d'une matrice de corrélation et estime correctement le

modèle.

Il existe plusieurs types de structures de corrélation disponibles dans R. Nous commencerons par utiliser l'une des plus simples, appelée `corSymm`, qui suppose que la matrice de corrélation est symétrique. C'est le cas avec les arbres phylogénétiques; la corrélation entre les espèces a et b est la même qu'entre b et a. Seule la partie triangulaire inférieure de la matrice doit être transmise à la structure corSymm. Si `mat` est la matrice de corrélation, cela se fait avec la commande `mat[lower.tri(mat)]`. Ensuite, vous passez la matrice de corrélation à la function `gls` en utilisant l'argument `correlation`.

```
# Calculer la matrice de corrélation à partir de l'arbre
mat <- vcv(seedplantstree,corr=TRUE)
# Créer la structure de corrélation pour gls
corr.struct <- corSymm(mat[lower.tri(mat)],fixed=TRUE)
# Exécuter le pgls
shade.pgls1 <- gls(Shade ~ Wd, data = seedplantsdata, correlation=corr.struct)
summary(shade.pgls1)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Wd
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   214.3762 220.3982 -104.1881
##
## Correlation Structure: General
##  Formula: ~1
##  Parameter estimate(s):
##  Correlation:
##    1     2     3     4     5     6     7     8     9     10    11    12
## 2  0.000
## 3  0.000 0.967
## 4  0.000 0.967 0.976
## 5  0.000 0.967 0.981 0.976
## 6  0.000 0.967 0.974 0.974 0.974
## 7  0.000 0.967 0.997 0.976 0.981 0.974
## 8  0.000 0.967 0.974 0.974 0.974 0.997 0.974
## 9  0.000 0.967 0.976 0.983 0.976 0.974 0.976 0.974
## 10 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## 11 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.984
## 12 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726
## 13 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.952 0.952 0.726
## 14 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.952 0.952 0.726
## 15 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.952 0.952 0.726
## 16 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.945 0.945 0.726
## 17 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.876 0.876 0.726
## 18 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.876 0.876 0.726
```

```
## 19 0.000 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596
## 20 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.989
## 21 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.835 0.835 0.726
## 22 0.000 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596
## 23 0.000 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596
## 24 0.000 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596 0.596
## 25 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.876 0.876 0.726
## 26 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.876 0.876 0.726
## 27 0.528 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 28 0.843 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 29 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.983
## 30 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.945 0.945 0.726
## 31 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 32 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 33 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 34 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 35 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 36 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 37 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 38 0.000 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523
## 39 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.675 0.675 0.675
## 40 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.675 0.675 0.675
## 41 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.675 0.675 0.675
## 42 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.675 0.675 0.675
## 43 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.898
## 44 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.898
## 45 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.898
## 46 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.835 0.835 0.726
## 47 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.835 0.835 0.726
## 48 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.835 0.835 0.726
## 49 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.835 0.835 0.726
## 50 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.675 0.675 0.675
## 51 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.980
## 52 0.528 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 53 0.000 0.720 0.720 0.720 0.720 0.720 0.720 0.720 0.720 0.654 0.654 0.654
## 54 0.906 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 55 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.800
## 56 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.800
## 57 0.000 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.726 0.726 0.800
##    13    14    15    16    17    18    19    20    21    22    23    24
## 2
## 3
## 4
## 5
## 6
## 7
```

```
## 8
## 9
## 10
## 11
## 12
## 13
## 14 0.998
## 15 0.994 0.994
## 16 0.945 0.945 0.945
## 17 0.876 0.876 0.876 0.876
## 18 0.876 0.876 0.876 0.876 0.998
## 19 0.596 0.596 0.596 0.596 0.596 0.596
## 20 0.726 0.726 0.726 0.726 0.726 0.726 0.596
## 21 0.835 0.835 0.835 0.835 0.835 0.835 0.596 0.726
## 22 0.596 0.596 0.596 0.596 0.596 0.596 0.715 0.596 0.596
## 23 0.596 0.596 0.596 0.596 0.596 0.596 0.715 0.596 0.596 0.997
## 24 0.596 0.596 0.596 0.596 0.596 0.596 0.715 0.596 0.596 0.999 0.997
## 25 0.876 0.876 0.876 0.876 0.984 0.984 0.596 0.726 0.835 0.596 0.596 0.596
## 26 0.876 0.876 0.876 0.876 0.984 0.984 0.596 0.726 0.835 0.596 0.596 0.596
## 27 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 28 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 29 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.983 0.726 0.596 0.596 0.596
## 30 0.945 0.945 0.945 0.988 0.876 0.876 0.596 0.726 0.835 0.596 0.596 0.596
## 31 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 32 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 33 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 34 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 35 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 36 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 37 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 38 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523 0.523
## 39 0.675 0.675 0.675 0.675 0.675 0.675 0.596 0.675 0.675 0.596 0.596 0.596
## 40 0.675 0.675 0.675 0.675 0.675 0.675 0.596 0.675 0.675 0.596 0.596 0.596
## 41 0.675 0.675 0.675 0.675 0.675 0.675 0.596 0.675 0.675 0.596 0.596 0.596
## 42 0.675 0.675 0.675 0.675 0.675 0.675 0.596 0.675 0.675 0.596 0.596 0.596
## 43 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.898 0.726 0.596 0.596 0.596
## 44 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.898 0.726 0.596 0.596 0.596
## 45 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.898 0.726 0.596 0.596 0.596
## 46 0.835 0.835 0.835 0.835 0.835 0.835 0.596 0.726 0.881 0.596 0.596 0.596
## 47 0.835 0.835 0.835 0.835 0.835 0.835 0.596 0.726 0.881 0.596 0.596 0.596
## 48 0.835 0.835 0.835 0.835 0.835 0.835 0.596 0.726 0.881 0.596 0.596 0.596
## 49 0.835 0.835 0.835 0.835 0.835 0.835 0.596 0.726 0.881 0.596 0.596 0.596
## 50 0.675 0.675 0.675 0.675 0.675 0.675 0.596 0.675 0.675 0.596 0.596 0.596
## 51 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.980 0.726 0.596 0.596 0.596
## 52 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 53 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
```

```
## 54 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 55 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.800 0.726 0.596 0.596 0.596
## 56 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.800 0.726 0.596 0.596 0.596
## 57 0.726 0.726 0.726 0.726 0.726 0.726 0.596 0.800 0.726 0.596 0.596 0.596
##    25    26    27    28    29    30    31    32    33    34    35    36
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26 0.992
## 27 0.000 0.000
## 28 0.000 0.000 0.528
## 29 0.726 0.726 0.000 0.000
## 30 0.876 0.876 0.000 0.000 0.726
## 31 0.000 0.000 0.528 0.843 0.000 0.000
## 32 0.000 0.000 0.528 0.843 0.000 0.000 0.874
## 33 0.000 0.000 0.528 0.843 0.000 0.000 0.985 0.874
## 34 0.000 0.000 0.528 0.843 0.000 0.000 0.997 0.874 0.985
## 35 0.000 0.000 0.528 0.843 0.000 0.000 0.874 0.965 0.874 0.874
## 36 0.000 0.000 0.528 0.843 0.000 0.000 0.999 0.874 0.985 0.997 0.874
## 37 0.000 0.000 0.528 0.843 0.000 0.000 0.874 0.926 0.874 0.874 0.926 0.874
## 38 0.523 0.523 0.000 0.000 0.523 0.523 0.000 0.000 0.000 0.000 0.000 0.000
## 39 0.675 0.675 0.000 0.000 0.675 0.675 0.000 0.000 0.000 0.000 0.000 0.000
## 40 0.675 0.675 0.000 0.000 0.675 0.675 0.000 0.000 0.000 0.000 0.000 0.000
## 41 0.675 0.675 0.000 0.000 0.675 0.675 0.000 0.000 0.000 0.000 0.000 0.000
## 42 0.675 0.675 0.000 0.000 0.675 0.675 0.000 0.000 0.000 0.000 0.000 0.000
```

```
## 43 0.726 0.726 0.000 0.000 0.898 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 44 0.726 0.726 0.000 0.000 0.898 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 45 0.726 0.726 0.000 0.000 0.898 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 46 0.835 0.835 0.000 0.000 0.726 0.835 0.000 0.000 0.000 0.000 0.000 0.000
## 47 0.835 0.835 0.000 0.000 0.726 0.835 0.000 0.000 0.000 0.000 0.000 0.000
## 48 0.835 0.835 0.000 0.000 0.726 0.835 0.000 0.000 0.000 0.000 0.000 0.000
## 49 0.835 0.835 0.000 0.000 0.726 0.835 0.000 0.000 0.000 0.000 0.000 0.000
## 50 0.675 0.675 0.000 0.000 0.675 0.675 0.000 0.000 0.000 0.000 0.000 0.000
## 51 0.726 0.726 0.000 0.000 0.980 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 52 0.000 0.000 0.918 0.528 0.000 0.000 0.528 0.528 0.528 0.528 0.528 0.528
## 53 0.654 0.654 0.000 0.000 0.654 0.654 0.000 0.000 0.000 0.000 0.000 0.000
## 54 0.000 0.000 0.528 0.843 0.000 0.000 0.860 0.860 0.860 0.860 0.860 0.860
## 55 0.726 0.726 0.000 0.000 0.800 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 56 0.726 0.726 0.000 0.000 0.800 0.726 0.000 0.000 0.000 0.000 0.000 0.000
## 57 0.726 0.726 0.000 0.000 0.800 0.726 0.000 0.000 0.000 0.000 0.000 0.000
##    37    38    39    40    41    42    43    44    45    46    47    48
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## 31
```

```
## 32
## 33
## 34
## 35
## 36
## 37
## 38 0.000
## 39 0.000 0.523
## 40 0.000 0.523 0.959
## 41 0.000 0.523 0.964 0.959
## 42 0.000 0.523 0.964 0.959 0.982
## 43 0.000 0.523 0.675 0.675 0.675 0.675
## 44 0.000 0.523 0.675 0.675 0.675 0.675 0.986
## 45 0.000 0.523 0.675 0.675 0.675 0.675 0.998 0.986
## 46 0.000 0.523 0.675 0.675 0.675 0.675 0.726 0.726 0.726
## 47 0.000 0.523 0.675 0.675 0.675 0.675 0.726 0.726 0.726 0.997
## 48 0.000 0.523 0.675 0.675 0.675 0.675 0.726 0.726 0.726 0.997 0.999
## 49 0.000 0.523 0.675 0.675 0.675 0.675 0.726 0.726 0.726 0.984 0.984 0.984
## 50 0.000 0.523 0.936 0.936 0.936 0.936 0.675 0.675 0.675 0.675 0.675 0.675
## 51 0.000 0.523 0.675 0.675 0.675 0.675 0.898 0.898 0.898 0.726 0.726 0.726
## 52 0.528 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 53 0.000 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## 54 0.860 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## 55 0.000 0.523 0.675 0.675 0.675 0.675 0.800 0.800 0.800 0.726 0.726 0.726
## 56 0.000 0.523 0.675 0.675 0.675 0.675 0.800 0.800 0.800 0.726 0.726 0.726
## 57 0.000 0.523 0.675 0.675 0.675 0.675 0.800 0.800 0.800 0.726 0.726 0.726
##    49    50    51    52    53    54    55    56
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
```

```
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## 31
## 32
## 33
## 34
## 35
## 36
## 37
## 38
## 39
## 40
## 41
## 42
## 43
## 44
## 45
## 46
## 47
## 48
## 49
## 50 0.675
## 51 0.726 0.675
## 52 0.000 0.000 0.000
## 53 0.654 0.654 0.654 0.000
## 54 0.000 0.000 0.000 0.528 0.000
## 55 0.726 0.675 0.800 0.000 0.654 0.000
## 56 0.726 0.675 0.800 0.000 0.654 0.000 0.983
## 57 0.726 0.675 0.800 0.000 0.654 0.000 0.999 0.983
##
## Coefficients:
##                Value Std.Error   t-value p-value
## (Intercept) 0.911433  4.409058 0.2067184  0.8370
## Wd          4.361028  1.693349 2.5753865  0.0127
##
##  Correlation:
##    (Intr)
## Wd -0.166
```

```
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.26890642 -0.16431866 -0.02645422  0.09638984  0.34953444
##
## Residual standard error: 7.455109
## Degrees of freedom: 57 total; 55 residual
```

Notez que le terme `fixed=TRUE` dans la structure corSymm indique que la structure de corrélation est fixée pendant l'optimisation des paramètres.

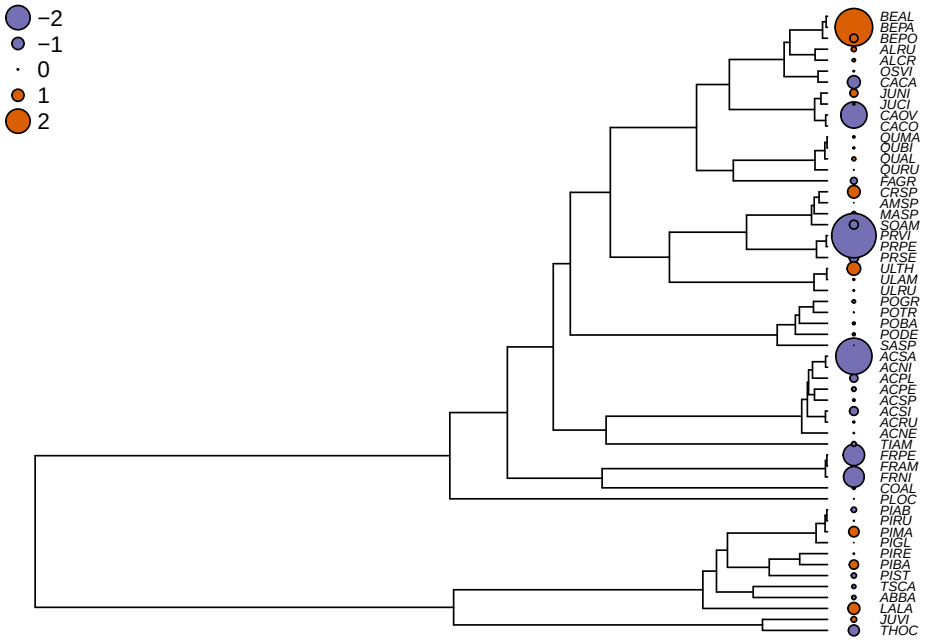
La sortie est similaire à celle du modèle sans corrélation, sauf pour la sortie de la matrice de corrélation.

Fait intéressant, vous pouvez voir que l'estimation du coefficient pour la pente est plus grande (4.361) qu'avec la régression standard et également significative (p=0.0127). C'est un exemple positif de PGLS. En effet, la relation entre la tolérance à l'ombre et la densité du bois était obscurcie par la corrélation phylogénétique des résidus. Une fois cette corrélation prise en compte, la relation significative est révélée.

Une relation significative entre la tolérance à l'ombre et la densité du bois a en fait un sens, bien que cette relation ne soit probablement pas causale. En effet, les arbres tolérants à l'ombre sont généralement des espèces de succession et poussent souvent plus lentement, en partie à cause de la disponibilité limitée de la lumière, et ont donc tendance à développer des bois de densité plus élevée.

Maintenant, regardons les résidus du modèle. Pour extraire les résidus corrigés par la structure de corrélation, vous devez demander les résidus normalisés.

```
# Extraire les résidus corrigés par la structure de corrélation
pgls1.res <- residuals(shade.pgls1,type="normalized")
# Modifier les paramètres graphiques
op <- par(mar=c(1,1,1,1))
# Même tracé que ci-dessus sauf pour utiliser pgls1.res comme résidus
plot(seedplantstree,type="p",TRUE,label.offset=0.01,cex=0.5,no.margin=FALSE)
tiplabels(pch=21,bg=cols[ifelse(pgls1.res>0,1,2)],col="black",
          cex=abs(pgls1.res),adj=0.505)
legend("topleft",legend=c("-2","-1","0","1","2"),pch=21,
       pt.bg=cols[c(1,1,1,2,2)],bty="n",
       text.col="black",cex=0.8,pt.cex=c(2,1,0.1,1,2))
```




```
# Réinitialiser les paramètres graphiques par défaut
par(op)
```

Si vous comparez avec l'optimisation des moindres carrés ordinaires, les résidus sont beaucoup moins corrélés phylogénétiquement.

5.3.1 Autres structures de corrélation

Dans le PGLS précédent, nous avons utilisé la structure corSymm pour transmettre la structure de corrélation phylogénétique à gls. Cela fonctionne parfaitement, mais il existe des moyens plus simples. Julien Dutheil a développé des structures phylogénétiques à utiliser spécialement dans les PGLS.

Celui que nous avons utilisé ci-dessus est équivalent à la structure `corBrownian` de `ape`. Cette approche est plus simple et il suffit de transmettre l'arbre à la structure de corrélation. Voici le même exemple en utilisant la structure `corBrownian`.

```
# Obtenir la structure de corrélation
bm.corr <- corBrownian(phy=seedplantstree, form=~1)
# PGLS
shade.pgls1b <- gls(Shade ~ Wd, data = seedplantsdata, correlation=bm.corr)
```

```
## Warning in Initialize.corPhyl(X[[i]], ...): No covariate specified, species
## will be taken as ordered in the data frame. To avoid this message, specify a
## covariate containing the species names with the 'form' argument.
```

```
summary(shade.pgls1b)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Wd
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   214.3762 220.3982 -104.1881
##
## Correlation Structure: corBrownian
##  Formula: ~1
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##                Value Std.Error   t-value p-value
## (Intercept) 0.911433  4.409058 0.2067184  0.8370
## Wd          4.361028  1.693349 2.5753865  0.0127
##
##  Correlation:
##    (Intr)
## Wd -0.166
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.26890642 -0.16431866 -0.02645422  0.09638984  0.34953444
##
## Residual standard error: 7.455109
## Degrees of freedom: 57 total; 55 residual
```

Vous pouvez voir que les résultats sont identiques. La seule différence est que la structure de corrélation n'est pas affichée dans le résumé. Le `numeric(0)` signifie qu'aucun paramètre n'a été estimé pendant l'optimisation (il et fixe).

Maintenant, vous vous demandez peut-être pourquoi la structure de corrélation ets appelée corBrownian. C'est parce qu'elle utilise le mouvement Brownien pour modéliser l'évolution le long des branches de l'arbre. Ce processus est souvent référé comme un modèle neutre. Si vous voulez en savoir davantage du le modèle Brownien, vous pouvez lire la section 12 à propos de ce modèle.

5.4 Défi no. 3

Ajuster un modèle PGLS pour voir si le poids des graines (`Sm`) explique la tolérance à l'ombre (`Shade`) à l'aide du jeu de données seedplantdataset. Comment est-ce que ces résultats se comparent avec les résultats d'une régression standard.

Chapter 6

Contrastes Indépendants Phylogénétiques

Faisons une digression pour examiner les Contrastes Indépendants Phylogénétiques (PIC). Les PIC ont été la première approche comparative proposée pour traiter la non-indépendance phylogénétique (Felsenstein, 1985). Bien qu'ils soient moins flexibles que les PGLS, ils donnent les mêmes résultats. Voyons comment ils peuvent être utilisés.

Les contrastes indépendants phylogénétiques sont estimés un trait à la fois. Ils transforment essentiellement le trait observé en contrastes qui ne sont pas corrélés avec la phylogénie. Cela peut être fait dans R en utilisant la fonction `pic` du package `ape`.

```
# Estimer le PIC pour la tolérance à l'ombre
Shade.pic <- pic(seedplantsdata$Shade, phy=seedplantstree)
# Estimer le PIC pour la densité du bois
Wd.pic <- pic(seedplantsdata$Wd, phy=seedplantstree)
```

Une fois cela fait, il suffit d'ajuster une régression entre ces contrastes. Notez qu'il est important que l'ordonnée à l'origine soit fixée à 0 dans le modèle. Cela se fait en ajoutant `- 1` à droite de la formule.

```
# Estimer le PIC pour la tolérance à l'ombre
pic.results <- lm(Shade.pic ~ Wd.pic - 1)
summary(pic.results)
```

```
##
## Call:
## lm(formula = Shade.pic ~ Wd.pic - 1)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -71.943  -4.106   1.013   5.679  21.614
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## Wd.pic    4.361      1.693   2.575   0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.21 on 55 degrees of freedom
## Multiple R-squared:  0.1076, Adjusted R-squared:  0.09139
## F-statistic: 6.633 on 1 and 55 DF,  p-value: 0.01273
```

Vous pouvez voir que l'estimation de la pente, `4.361`, est identique à celle obtenue avec les PGLS. Même chose pour la valeur p. La principale limitation des PIC est que vous êtes limité à toujours comparer deux variables. Avec les PGLS, une flexibilité beaucoup plus grande est possible.

Chapter 7

Assouplir l'hypothèse selon laquelle les résidus doivent être parfaitement corrélés phylogénétiquement

Les moindres carrés généralisés phylogénétiques supposent que les résidus sont parfaitement corrélés phylogénétiquement. Cela est relativement contraignant car cela signifie que d'autres sources d'erreurs non corrélées phylogénétiquement ne sont pas autorisées par le modèle. De plus, si elles existent, elles peuvent biaiser les résultats des PGLS (Revell, 2010).

Il existe des moyens d'assouplir cette hypothèse, et l'un d'eux consiste à utiliser un type de structure de corrélation qui permet cet assouplissement.

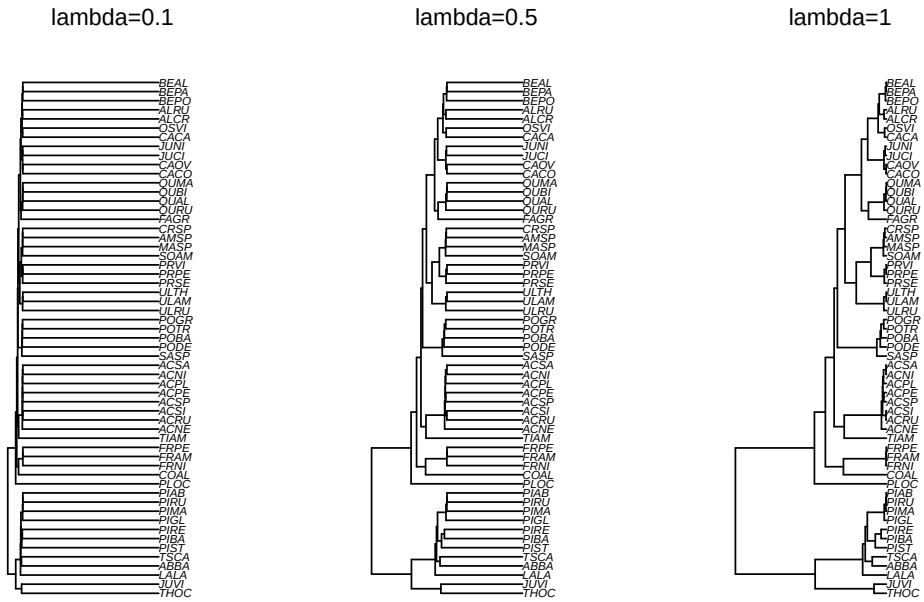
7.1 Théorie : Structure de corrélation de Pagel

Lors du contrôle des relations phylogénétiques avec les moindres carrés généralisés phylogénétiques, nous supposons que les résidus sont parfaitement corrélés en fonction de la structure de corrélation. En pratique, ce n'est pas toujours le cas, et il est difficile de vraiment savoir à quel point il est important de contrôler la relation phylogénétique dans un cas spécifique. Par exemple, pour une étude donnée, la corrélation dans les résidus peut ne pas être fortement corrélée phylogénétiquement.

Il est possible de prendre cela en compte en utilisant le modèle λ de Pagel (Pagel, 1999). L'idée est de multiplier les éléments hors diagonale de la matrice de corrélation (essentiellement les longueurs de branches de la phylogénie) par

un paramètre λ, mais pas les valeurs diagonales. Cela entraîne essentiellement une modification des longueurs de branches de la phylogénie. Une valeur de λ proche de zéro donne des branches internes très courtes et de longues branches terminales. Cela réduit, en effet, les corrélations phylogénétiques (l'effet de la phylogénie est réduit). À l'opposé, si λ est proche de 1, alors la phylogénie modifiée ressemble à la phylogénie réelle. En effet, le paramètre λ est souvent interprété comme un paramètre de signal phylogénétique ; ainsi, une valeur de λ plus élevée implique un signal phylogénétique plus fort.

La figure suivante montre comment différentes valeurs de lambda affectent la forme de la phylogénie des arbres du Québec.



Vous pouvez voir qu'avec des valeurs de lambda faibles, le poids accordé à l'histoire partagée (la phylogénie) est considérablement réduit. Les longues branches terminales indiquent en quelque sorte qu'il pourrait y avoir beaucoup plus de variation dans les résidus indépendants des autres espèces. Cette variation pourrait être due à d'autres facteurs inclus dans les estimations de chaque espèce mais indépendants de la phylogénie (comme les erreurs de mesure, par exemple).

7.2 Exercices pratiques

Le modèle λ de Pagel peut être utilisé dans les PGLS en utilisant la structure de corrélation `corPagel`. L'utilisation de cette structure de corrélation est similaire à celle de la structure `corBrownian`, sauf que vous devez fournir une valeur de paramètre initiale pour λ.

```
# Obtenir la structure de corrélation
pagel.corr <- corPagel(0.3, phy=seedplantstree, fixed=FALSE, form=~Code)
```

La valeur donnée à corPagel est la valeur de départ pour le paramètre λ. Notez également que l'option `fixed=` est définie sur `FALSE`. Cela signifie que le paramètre λ sera optimisé en utilisant les moindres carrés généralisés. S'il était défini sur `TRUE`, alors le modèle serait ajusté avec le paramètre de départ, ici `0.3`. Le terme `form~Code` indique à la fonction d'utiliser l'ordre de la variable `Code` pour ordonner les noms d'espèces dans l'arbre.

Ajustons maintenant le PGLS avec cette structure de corrélation.

```
# PGLS avec corPagel
shade.pgls2 <- gls(Shade ~ Wd, data = seedplantsdata, correlation=pagel.corr)
summary(shade.pgls2)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Wd
##   Data: seedplantsdata
##        AIC     BIC    logLik
##   163.3967 171.426 -77.69833
##
## Correlation Structure: corPagel
##  Formula: ~Code
##  Parameter estimate(s):
##    lambda
## 0.9581665
##
## Coefficients:
##                Value Std.Error   t-value p-value
## (Intercept) 1.254987  1.636575 0.7668377  0.4465
## Wd          3.573527  1.497808 2.3858381  0.0205
##
##  Correlation:
##    (Intr)
## Wd -0.397
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.75145692 -0.44908843 -0.05417524  0.25655008  0.96493685
##
## Residual standard error: 2.621947
## Degrees of freedom: 57 total; 55 residual
```

Vous pouvez voir que gls a estimé le paramètre λ, qui est ici de 0,958. Étant donné que le λ estimé est très proche de 1, nous pouvons conclure que les résidus du modèle étaient fortement corrélés phylogénétiquement. Cela confirme donc

l'importance d'utiliser un PGLS avec ce modèle. Si le λ estimé avait été proche de 0, cela aurait suggéré que le PGLS n'était pas nécessaire. Notez cependant qu'en utilisant cette approche, vous êtes assuré de ne jamais obtenir un résultat statistiquement biaisé. En fait, je vous **recommande fortement** d'utiliser toujours cette structure de corrélation dans vos analyses statistiques.

7.3 Défi no. 4

Essayez d'ajuster un PGLS avec une structure de corrélation de Pagel en régressant la tolérance à l'ombre sur la masse des graines. Les résidus sont-ils aussi corrélés phylogénétiquement que dans la régression précédente avec la densité du bois?

7.4 Autres structures de corrélation (ou modèles évolutifs)

Les structures de corrélation disponibles dans le package `ape` offrent d'autres alternatives pour le modèle d'évolution des caractères supposé. Par exemple, la structure de corrélation `corMartins` modélise la sélection en utilisant le modèle d'Ornstein-Uhlenbeck (ou Hansen) avec le paramètre α qui détermine la force de la sélection. De plus, `corBlomberg` modélise une évolution brownienne accélérée ou décélérée, c'est-à-dire que le taux d'évolution du mouvement brownien s'accélère ou ralentit avec le temps avec ce modèle. Il est possible de faire des comparaisons de modèles pour décider quel modèle correspond le mieux à la variation résiduelle.

Chapter 8

ANOVA phylogénétique

Jusqu'à présent, nous n'avons analysé que des caractères quantitatifs continus. Mais il est également possible de réaliser une ANOVA avec PGLS.

L'avantage de PGLS, tel qu'il est implémenté avec la fonction `gls`, est qu'il peut être facilement adapté pour tester de nombreux types de modèles différents. Pour donner un exemple ici, il est facile de mettre en œuvre une ANOVA phylogénétique en R. En effet, il vous suffit de fournir à `gls` un trait catégoriel en tant que variable indépendante.

Comme il n'y a pas de variable catégorielle dans le jeu de données des traits fonctionnels des plantes, nous allons en créer une en divisant la catégorie de densité du bois en deux catégories : bois léger et bois dense.

```
# Créer une variable catégorielle
seedplantsdata$Wd.cat<-cut(seedplantsdata$Wd,breaks=2,labels=c("light","dense"))
# Regarder le résultat
seedplantsdata$Wd.cat
```

```
##  [1] light light dense light dense dense dense light light light light dense
## [13] dense light light dense dense dense dense dense dense dense light dense
## [25] light dense light light dense dense light light light light light light
## [37] light light light light light light light light light light dense dense dense
## [49] dense light light light light light light light dense
## Levels: light dense
```

Nous pouvons maintenant ajuster une ANOVA phylogénétique.

```
# ANOVA phylogénétique
shade.pgls3 <- gls(Shade ~ Wd.cat, data = seedplantsdata, correlation=pagel.corr)
summary(shade.pgls3)
```

```
## Generalized least squares fit by REML
```

```
##   Model: Shade ~ Wd.cat
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   166.7352 174.7646 -79.36762
##
## Correlation Structure: corPagel
##  Formula: ~Code
##  Parameter estimate(s):
##    lambda
## 0.9439646
##
## Coefficients:
##                 Value Std.Error  t-value p-value
## (Intercept) 2.6826723 1.3844404 1.937730  0.0578
## Wd.catdense 0.6179855 0.2526902 2.445626  0.0177
##
##  Correlation:
##             (Intr)
## Wd.catdense -0.037
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.69257567 -0.48677930 -0.04143001  0.33640615  0.95379525
##
## Residual standard error: 2.429586
## Degrees of freedom: 57 total; 55 residual
```

Vous pouvez voir que la densité du bois, même transformée en variable catégorielle, a un effet significatif sur la tolérance à l'ombre.

Chapter 9

Comparaison de modèles

Vous pourriez être intéressé à comparer différents modèles, ce qui est une approche courante en biologie pour la modélisation. Cependant, il y a une petite subtilité à prendre en compte avec les PGLS.

La méthode par défaut pour l'ajustement des modèles avec `gls` est l'estimation par maximum de vraisemblance restreinte (REML), obtenue avec `method="REML"`. Cela diffère de l'estimation standard par maximum de vraisemblance (ML), qui peut être obtenue avec `method="ML"`. La différence entre ces deux méthodes est complexe, mais il suffit de dire qu'elles diffèrent dans la manière dont les paramètres de variance sont estimés. REML fournit des estimations de paramètres moins biaisées et est la méthode privilégiée pour rapporter les coefficients des paramètres dans une publication. C'est également la méthode de choix si vous souhaitez comparer des modèles avec des structures de corrélation (ou de variance) différentes (Zuur et al., 2009). Par exemple, si vous voulez tester si un modèle PGLS avec un λ de Pagel optimisé s'ajuste mieux aux données qu'un modèle sans corrélation phylogénétique (c'est-à-dire avec λ de Pagel = 0) :

```
pagel.0 <- gls(Shade ~ Wd, data = seedplantsdata,
               correlation=corPagel(0,phy=seedplantstree,
                                    fixed=TRUE, form=~Code),
               method="REML")
pagel.fit <- gls(Shade ~ Wd, data = seedplantsdata,
                 correlation=corPagel(0.8,phy=seedplantstree,
                                      fixed=FALSE, form=~Code),
                 method="REML")
anova(pagel.0,pagel.fit)
```

```
##           Model df      AIC     BIC    logLik   Test  L.Ratio p-value
## pagel.0       1  3 180.4720 186.494 -87.23602
```

```
## pagel.fit     2  4 163.3967 171.426 -77.69833 1 vs 2 19.07537  <.0001
```

Vous pouvez utiliser l'AIC ou le BIC pour comparer le modèle, ou le test du rapport de vraisemblance. Vous pouvez voir ici que le modèle PGLS avec un λ de Pagel ajusté offre un meilleur ajustement que celui avec un $\lambda = 0$ (AIC plus faible). D'ailleurs, il s'agit également d'un test pour déterminer si un modèle PGLS est meilleur qu'un modèle de régression standard, car une structure `corPagel` avec $\lambda = 0$ est un modèle standard (= pas de corrélation phylogénétique).

Maintenant, si vous êtes intéressé par le test des paramètres fixes dans le modèle, vous devez utiliser l'ajustement par maximum de vraisemblance (Zuur et al., 2009). Par exemple, si vous souhaitez utiliser un test du rapport de vraisemblance pour tester le modèle avec la densité du bois comme variable indépendante par rapport à un modèle nul avec seulement l'ordonnée à l'origine, vous pouvez procéder comme suit.

```
wd <- gls(Shade ~ Wd, data = seedplantsdata,
          correlation=corBrownian(phy=seedplantstree, form=~Code),
          method="ML")
null <- gls(Shade ~ 1, data = seedplantsdata,
            correlation=corBrownian(phy=seedplantstree, form=~Code),
            method="ML")
anova(wd,null)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## wd       1  3 222.0088 228.1380 -108.0044
## null     2  2 226.4988 230.5848 -111.2494 1 vs 2 6.489907  0.0108
```

Vous pouvez voir que le modèle avec la variable de densité du bois est meilleur que le modèle avec seulement l'ordonnée à l'origine. Cependant, comme mentionné ci-dessus, étant donné que l'ajustement REML fournit de meilleures estimations des paramètres, vous devrez réajuster le modèle en utilisant REML pour présenter les résultats.

```
wd.final <- gls(Shade ~ Wd, data = seedplantsdata,
                correlation=corBrownian(phy=seedplantstree, form=~Code),
                method="REML")
summary(wd.final)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Wd
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   214.3762 220.3982 -104.1881
##
## Correlation Structure: corBrownian
##  Formula: ~Code
##  Parameter estimate(s):
```

```
## numeric(0)
##
## Coefficients:
##                Value Std.Error   t-value p-value
## (Intercept) 0.911433  4.409058 0.2067184  0.8370
## Wd          4.361028  1.693349 2.5753865  0.0127
##
##  Correlation:
##    (Intr)
## Wd -0.166
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.26890642 -0.16431866 -0.02645422  0.09638984  0.34953444
##
## Residual standard error: 7.455109
## Degrees of freedom: 57 total; 55 residual
```

Chapter 10

Quand devrions-nous utiliser les méthodes comparatives ?

Les méthodes comparatives devraient toujours être utilisées lorsqu'on travaille avec des ensembles de données qui comprennent plusieurs espèces. Un bon conseil est d'utiliser une méthode permettant aux résidus du modèle de ne pas être tous corrélés phylogénétiquement, comme lors de l'utilisation des PGLS avec la structure `corPagel` ou en utilisant le modèle mixte phylogénétique. Des études antérieures ont montré que l'utilisation de telles méthodes comparatives donne lieu à des estimations plus précises et exactes de l'effet fixe, un taux d'erreur de type I plus faible et une puissance statistique plus grande (Revell, 2010). Par conséquent, il est toujours avantageux d'utiliser ces méthodes.

Une erreur courante consiste à utiliser les PGLS pour tester le signal phylogénétique dans Y ou X en utilisant soit le λ de Pagel ou le K de Blomberg, et si un signal phylogénétique est présent, utiliser un PGLS pour analyser les données et sinon utiliser une régression standard. C'est une **grosse erreur**. Comme nous l'avons vu précédemment, les PGLS corrigent la corrélation phylogénétique dans les résidus et non dans les variables. Par conséquent, la présence d'un signal phylogénétique dans les variables ne signifie pas nécessairement que les résidus sont corrélés phylogénétiquement. Et l'inverse est également vrai : les variables peuvent ne pas être corrélées phylogénétiquement, mais les résidus pourraient l'être !

Une autre idée fausse courante concernant les méthodes comparatives est qu'elles éliminent toute variation dans les données liée à la phylogénie et que cela pourrait affecter l'interprétation de la variable d'intérêt. Cela était vrai pour les anciennes méthodes telles que l'autorégression phylogénétique qui, d'abord, élimi-

nait le signal phylogénétique des données avant de les analyser. Ces approches posaient effectivement problème. Mais les méthodes présentées ici ne souffrent pas de ces problèmes. Elles tiennent compte de la structure phylogénétique et la quantifient, mais elles n'éliminent pas la variation du modèle.

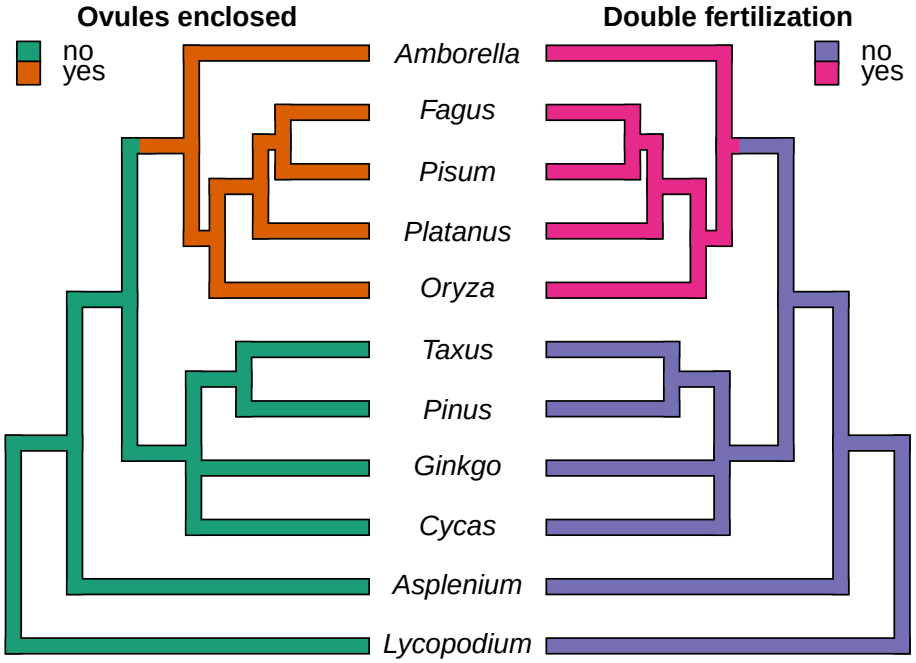
Chapter 11

Un dernier mot : le problème de la réplication

Tout biologiste est bien conscient de l'importance de répliquer ses expériences afin d'avoir confiance en ses conclusions. Cela est beaucoup plus complexe lorsque nous considérons l'évolution. Pour tester nos hypothèses sur l'évolution, l'approche idéale serait de rembobiner la "bande de l'évolution" (S. J. Gould) et de laisser l'histoire se répéter plusieurs fois pour voir ce qui se passe. Cela n'est malheureusement pas possible, bien que certaines études d'évolution expérimentale parviennent à reproduire des expériences évolutives.

La méthode comparative phylogénétique introduite dans ce tutoriel est une approche appropriée pour nous protéger d'arriver à des conclusions qui ne sont pas fortement soutenues dans un contexte évolutif. Cependant, même cette approche peut parfois échouer. C'est pourquoi une attention supplémentaire est nécessaire dans de telles études.

En interprétant leurs résultats, les biologistes devraient d'abord se demander s'ils ont suffisamment de réplicats dans leurs données pour tirer des conclusions solides. Et par réplicat, j'entends réplicat évolutif. Considérons l'exemple des plantes à graines présenté ci-dessus.



S'il existe plusieurs espèces avec des ovules enfermés ou non et qui effectuent ou non une double fécondation, le scénario le plus parcimonieux pour les deux caractères est que chacun a évolué une fois le long de la branche de l'arbre qui mène aux plantes à fleurs. En d'autres termes, il n'y a eu qu'une seule transition entre les états de chaque caractère dans l'évolution de ce groupe.

Ainsi, même s'il semble y avoir une réplication lorsque nous regardons les espèces (plusieurs espèces avec chaque état de caractère ont été échantillonnées), il n'y a pas de réplication évolutive ! Donc, même si la probabilité que ces deux événements se produisent sur la même branche est très faible et même si un test de contingence pour calculer la probabilité d'un tel événement est significatif, c'est un peu comme une expérience avec un seul réplicat. Par conséquent, même lorsqu'un test qui prend en compte la phylogénie est significatif, une grande prudence est nécessaire lors de l'interprétation de ces résultats. Idéalement, une étude devrait avoir un nombre décent de réplicats évolutifs pour que les résultats soient significatifs sur le plan biologique. Je vous encourage à lire le très bon article de Maddison et Fitzjohn sur le sujet (Maddison and FitzJohn, 2015).

Idéalement, avant de planifier une expérience, on devrait s'assurer qu'il existe une réplication suffisante dans l'évolution des traits étudiés parmi les espèces considérées pour avoir plus de confiance dans les résultats. Par exemple, il serait beaucoup mieux si chaque caractère avait évolué 5 à 6 fois chacun dans l'exemple précédent, en particulier si les deux caractères évoluaient toujours simultanément !

Chapter 12

Le modèle de Mouvement Brownien (BM)

Lorsque nous souhaitons prendre en compte la non-indépendance des espèces en raison de leurs histoires évolutives dans les analyses statistiques, un modèle d'évolution est nécessairement impliqué. En effet, nous supposons que les caractères ont évolué au fil du temps (le long de la phylogénie) et que les espèces étroitement apparentées sont plus susceptibles d'être en moyenne plus similaires pour un trait donné que des espèces éloignées. En biologie évolutive, le modèle de base (souvent utilisé comme modèle nul dans de nombreuses analyses) est le modèle de mouvement brownien. Ce modèle d'évolution porte le nom de Robert Brown, un botaniste célèbre qui a publié une importante *Flora of Australia* en 1810. Il fut aussi le premier à distinguer les gymnospermes des angiospermes. Sa découverte du mouvement brownien est due à l'observation que de petites particules en solution ont tendance à se déplacer dans toutes les directions, une observation faite pour la première fois en observant du pollen de *Clarkia* au microscope. L'explication viendrait plus tard, en termes d'impacts moléculaires aléatoires.

Les mathématiciens ont construit un processus stochastique destiné à approcher le mouvement brownien. Dans ce modèle, chaque étape est indépendante des autres et peut aller dans n'importe quelle direction. Le déplacement moyen est nul et la variance est uniforme dans tout l'espace paramétrique. Les déplacements peuvent être additionnés, ce qui signifie que les variances des déplacements indépendants peuvent s'ajouter. Si σ^2 est la variance d'un seul déplacement, la variance après un temps t sera $\sigma^2 t$. Lorsque le nombre d'étapes est grand, comme dans un contexte phylogénétique, le résultat est distribué normalement.

Lorsqu'il est appliqué aux phylogénies, le modèle de mouvement brownien est en quelque sorte appliqué indépendamment à chaque branche de la phylogénie.

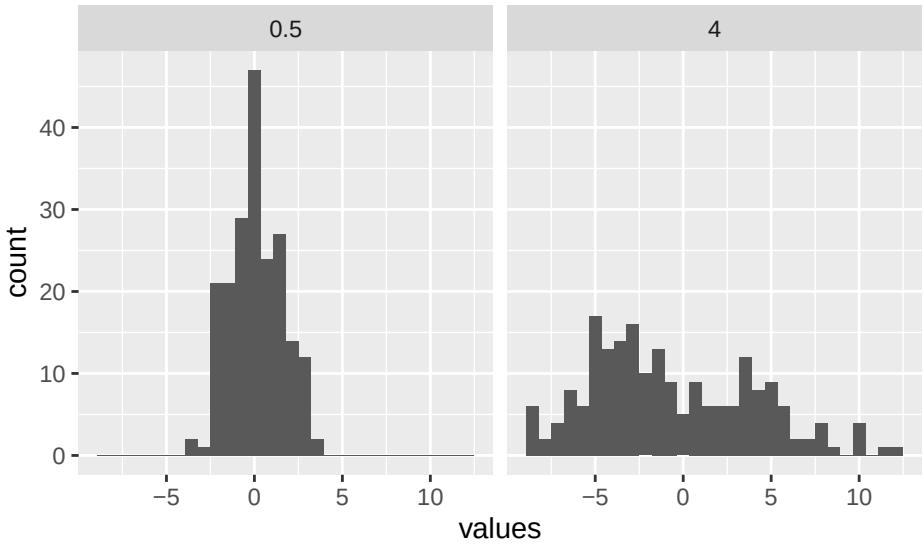
Cela permet de modéliser la quantité de changement qui s'est produite le long d'une branche donnée. Si la variance du modèle de mouvement brownien est σ^2 par unité de temps t, alors le changement net le long d'une branche de temps t est tiré d'une distribution normale de moyenne 0 et de variance $\sigma^2 t$. Ce modèle peut également être représenté mathématiquement de la manière suivante, où la quantité de changement pour le caractère X sur le temps infinitésimal dans l'intervalle entre le temps t et $t + dt$ est :

$$dX(t) = \sigma^2 dB(t),$$

où $dB(t)$ est la distribution gaussienne. Il est important de noter que ce modèle suppose que :

1. L'évolution se produisant dans chaque branche de la phylogénie est indépendante de celle se produisant dans les autres branches.
2. L'évolution est complètement aléatoire (c'est-à-dire sans sélection).

Le paramètre σ^2 dans le modèle donne la variance, ou en d'autres termes, la vitesse d'évolution. Plus la variance est élevée, plus le caractère évoluera rapidement. Voici deux exemples de caractères simulés sur un arbre de 200 espèces avec $\sigma^2 = 0.5$ et $\sigma^2 = 4$.

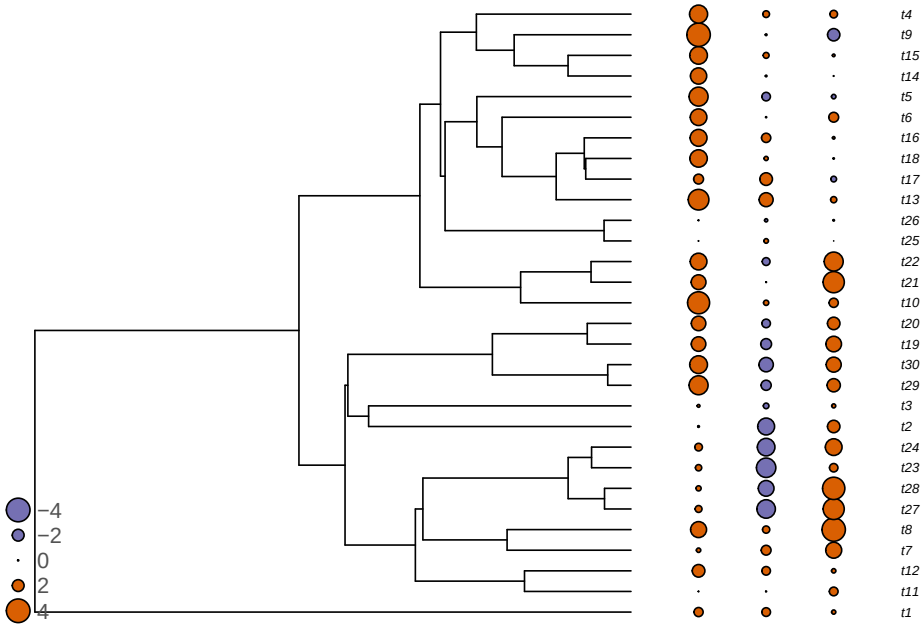


Une introduction plus approfondie au modèle de Mouvement Brownien se trouve au chapitre 23 du livre de Joe Felsenstein (Felsenstein and Felenstein, 2004).

Le modèle de mouvement brownien est souvent dit modéliser la dérive neutre, bien qu'un bon ajustement à ce modèle ne signifie pas nécessairement que les données ont évolué via des dérives aléatoires, car d'autres processus peuvent

également donner des motifs similaires à ceux du mouvement brownien (Hansen and Martins, 1996).

Notez également que le modèle est stochastique. C'est-à-dire que même si deux espèces étroitement apparentées sont plus susceptibles de partager des états de caractère similaires qu'une espèce éloignée, cela n'est vrai qu'en moyenne. Pour un caractère donné simulé, des espèces étroitement apparentées peuvent parfois être plus différentes qu'une espèce éloignée. Regardez la figure suivante, qui montre trois caractères simulés selon le mouvement brownien.



Chapter 13

Lectures supplémentaires

Pour bien comprendre un nouveau domaine de recherche, il est toujours conseillé de lire beaucoup à ce sujet. Voici quelques références que vous pourriez trouver utiles. Les différentes sources expliquent parfois la théorie de différentes manières ou utilisent des exemples différents, ce qui peut vous aider à mieux comprendre.

- Felsenstein, J. (1985) Phylogenies and the comparative method. *The American Naturalist* 125, 1-15. **The classic initial paper that launched the field of comparative analyses. The phylogenetic independent contrasts are introduced here**
- Felsenstein, J. (2004) *Inferring phylogenies*. Sinauer Associates, Inc. Sunderland, MA. **A thorough reference on phylogenies, from reconstruction to phylogenetic methods**
- Hadfield, J. D., S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* 23:494–508. **This paper describes the phylogenetic mixed model and its implementation in MCMCglmm. It is a very important paper**
- Housworth, E.A., E.P. Martins, M. Lynch. 2004. The phylogenetic mixed model. *The American Naturalist* 163:84–96. **Excellent paper on the Phylogenetic Mixed Model**
- Paradis, E. (2012). *Analysis of phylogenetics and evolution with R*. New York, USA: Springer. **This is the book that explains the analyses available in the R package APE. It is also a great reference on many phylogenetic analyses, including the comparative method. This is a classic and a must for users of phylogenies in R.**
- Revell, L J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution* 1: 319-329. **A great paper on PGLS. It uses simulations to show when it is important to use**

PGLS.

- Villemereuil, P., S. Nakagawa. 2014. General quantitative genetic methods for comparative biology. Pp. 287–303 in L. Z. Garamszegi, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology.* Springer-Verlag, Berlin, Heidelberg. **Nice book chapter explaining the phylogenetic mixed model**
- Zuur, A.F., E.N. Ieno, N. Walker, A. A. Saveliev, G.M. Smith. (2009). *Mixed effects models and extensions in ecology with R.* New York, NY: Springer New York. **This is not a book on phylogenetic methods, but it is a great book on the analysis of ecological data with examples in R. Its chapter 6 and 7 discuss correlation structures and although they are not about phylogenies, they are very instructive on how to deal with them and how to compare models and analyse complex data. It also has tons of information on how to deal with more complex data, along with correlation structure. A very good read!**

Chapter 14

Introduction aux phylogénies dans R

Il existe de nombreux packages pour les analyses phylogénétiques dans R. Je ne vais pas tous les énumérer ici, mais vous pouvez avoir une bonne idée des options disponibles en consultant la vignette phylogénétique de R maintenue par Brian O'Meara. Elle est principalement orientée vers les méthodes comparatives phylogénétiques, mais c'est un bon point de départ.

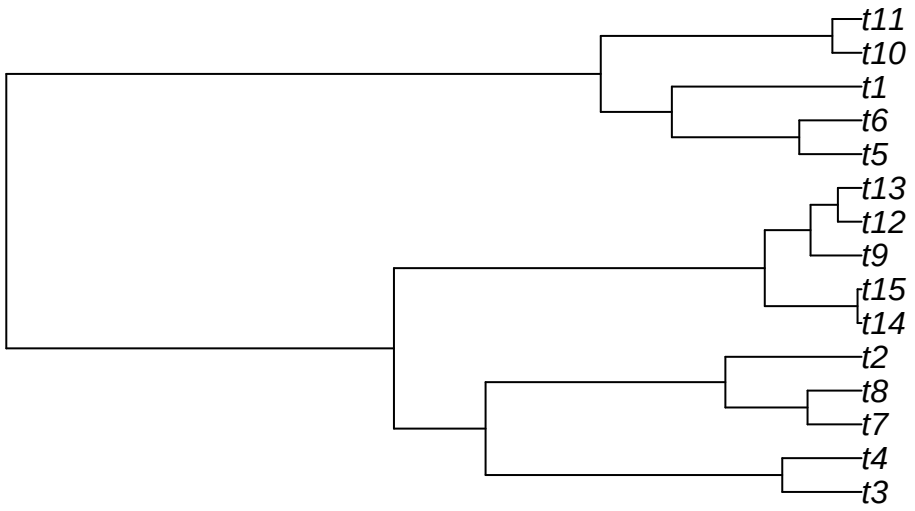
Le package le plus basique pour utiliser des arbres dans R est *ape*, qui vous permet de lire et de tracer des arbres.

14.1 Importer et tracer des arbres

14.1.1 Simuler un arbre

Tout au long de ces exercices, nous utiliserons souvent des arbres simulés, qui sont très utiles à des fins pédagogiques. Les arbres peuvent être simulés en utilisant plusieurs fonctions, mais voici un exemple pour simuler un arbre avec 15 espèces.

```
require(phytools)
tree <- pbtree(n=15,nsim=1)
plot(tree)
```



Vous enregistrez l'arbre au format nexus dans un fichier. Mais avant de le faire, il est recommandé de définir le répertoire de travail dans le même dossier où votre script est enregistré. Vous pouvez le faire dans RStudio dans le menu Session>Set Working Directory>To Source File Location.

```
require(ape)
write.nexus(tree, file="My_first_tree.tre")
```

14.1.2 Simulation de caractères

Les caractères peuvent également être facilement simulés dans R. Par exemple, vous pourriez simuler un caractère en utilisant un modèle de Mouvement Brownien (BM) avec le code suivant.

```
trait1 <- fastBM(tree, sig2=0.01, nsim=1, internal=FALSE)
# To get trait values for tree tips:
trait1
```

```
##           t3           t4           t7           t8           t2          t14
## -0.007306853  0.043251831  0.004858391 -0.114795593 -0.043737891  0.113964926
##          t15           t9          t12          t13           t5           t6
##  0.097243984  0.096978766  0.063069656  0.053040643  0.072476832  0.034675083
##           t1          t10          t11
## -0.084659118  0.118596501  0.059071786
```

Ensuite, enregistrons ce trait dans un fichier en faisant comme s'il s'agissait de nos données d'origine.

```
write.table(matrix(trait1,ncol=1,dimnames=list(names(trait1),"trait1")), file="mytrait
```

Maintenant que nous avons simulé un arbre et un caractère, effaçons ce que nous avons fait jusqu'à présent de l'environnement R et faisons comme si c'étaient

nos données pour les prochaines sections.

```
rm(tree, trait1)
```

14.2 Importer des données dans R

Voici comment vous devez importer vos données dans R.

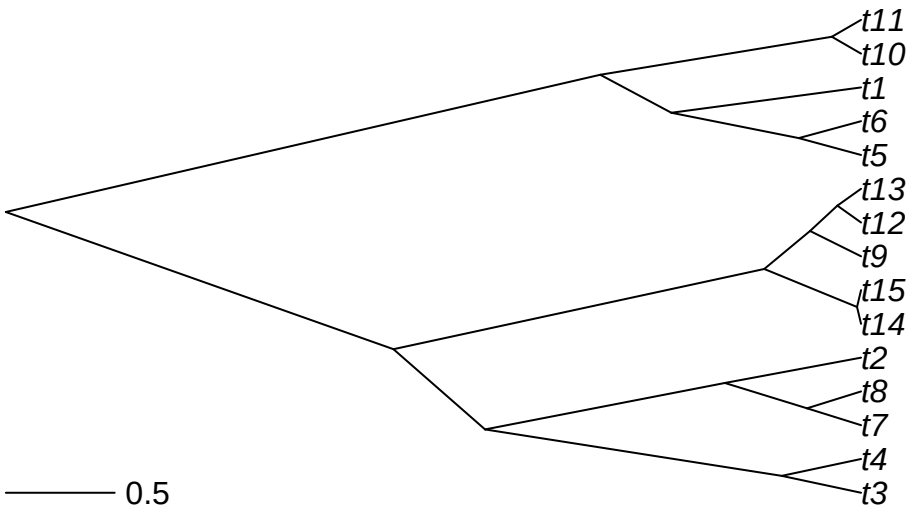
```
tree <- read.nexus(file="My_first_tree.tre")
trait1 <- read.csv2(file="mytrait.csv",dec=".")
```

Le format d'arbre dans ape contient plusieurs informations, et il est utile de savoir comment y accéder. Par exemple, les étiquettes des extrémités peuvent être consultées avec `tree$tip.label` et les longueurs des branches avec `tree$edge.length`. Nous verrons d'autres options dans d'autres exercices, mais si vous voulez des informations plus détaillées sur la façon dont les objets "phylo" sont organisés, vous pouvez consulter le fichier d'aide `?read.tree` ou ce document préparé par Emmanuel Paradis, l'auteur de `ape`.

14.3 Représenter des arbres

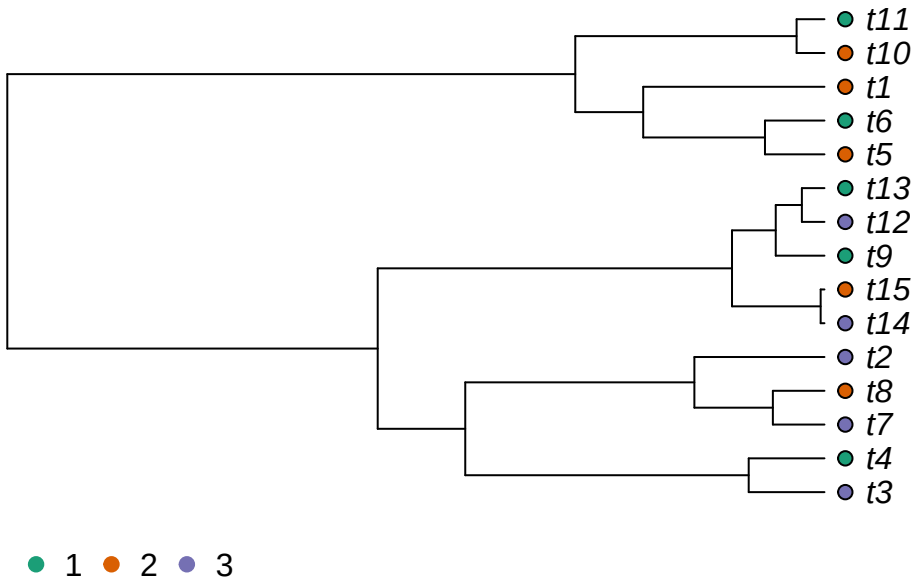
Tracer des arbres est l'un des aspects les plus intéressants de l'utilisation de R. Les options sont nombreuses et les possibilités larges. La fonction la plus courante est `plot.phylo` du package ape, qui propose de nombreuses options différentes. Je vous conseille vivement de regarder de près les différentes options de la fonction `?plot.phylo`. Voici un exemple de base.

```
plot(tree, type="c")
add.scale.bar()
```



Mais R est aussi intéressant pour représenter des caractères à côté des arbres. Si vous avez un caractère catégoriel, vous pouvez l'utiliser pour colorer les extrémités de la phylogénie.

```
# Générer un caractère discret
trait2 <- as.factor(sample(c(1,2,3),size=length(tree$tip.label),replace=TRUE))
# Créer une palette de couleur
library(RColorBrewer)
ColorPalette1 <- brewer.pal(n = length(levels(trait2)), name = "Dark2")
plot(tree, type="p", use.edge.length = TRUE, label.offset=0.2,cex=1)
tiplabels(pch=21,bg=ColorPalette1[trait2],col="black",cex=1,adj=0.6)
op<-par(xpd=TRUE)
legend(0,0,legend=levels(trait2),col=ColorPalette1,
       pch=20,bty="n",cex=1,pt.cex=1.5,ncol=length(levels(trait2)))
```

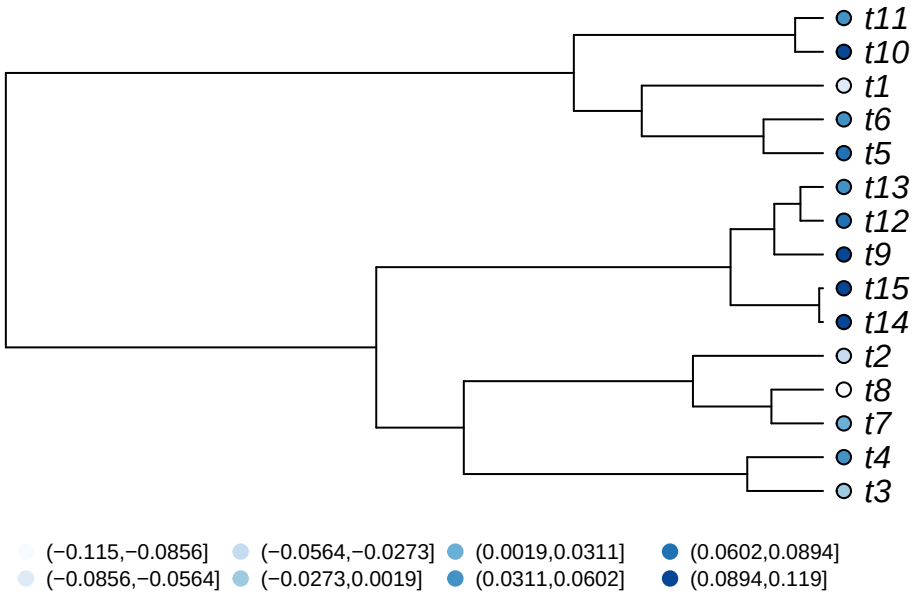



```
par(op) #Remettre les paramètres graphiques par défaut
```

Un résultat similaire pourrait être obtenu avec une variable continue. Ici, nous utiliserons le modèle de Mouvement Brownien, que nous étudierons dans une prochaine leçon, pour simuler le caractère continu.

```
# Diviser un trait continu en catégories
trait1.cat <- cut(trait1[,1],breaks=8,labels=FALSE)
# Créer une palette de couleur
ColorPalette2 <- brewer.pal(n = 8, name = "Blues")
# Représenter l'arbre
plot(tree, type="p", use.edge.length = TRUE, label.offset=0.2,cex=1)
tiplabels(pch=21,bg=ColorPalette2[trait1.cat],col="black",cex=1,adj=0.6)
```

```
op<-par(xpd=TRUE)
legend(0,0,legend=levels(cut(trait1[,1],breaks=8)),
       col=ColorPalette2,pch=20,bty="n",cex=0.7,pt.cex=1.5,ncol=4)
```



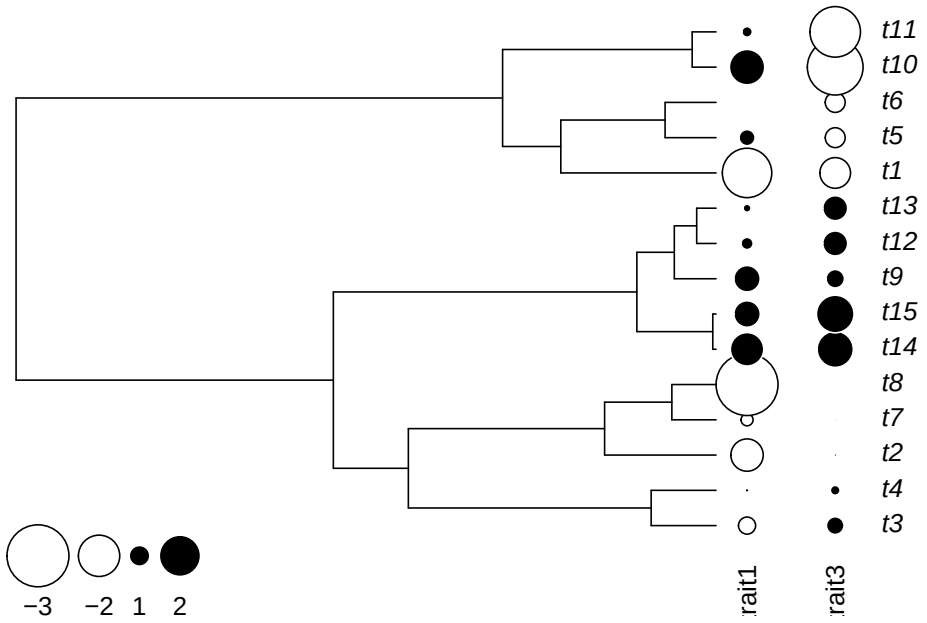

```
par(op)
```

Comme prévu pour un caractère simulé avec un mouvement brownien, vous pouvez voir que les espèces étroitement apparentées ont tendance à avoir des valeurs de caractère plus similaires.

Une autre option pour représenter un paramètre continu est d'utiliser la fonction `table.phylo4d` du package `adephylo` pour représenter le trait, où ses valeurs sont représentées par des tailles et des couleurs différentes. Il est également possible de tracer plusieurs caractères en même temps.

Notez que vous devrez installer les packages `phylobase` et `adephylo` pour exécuter ces fonctions s'ils ne sont pas installés.

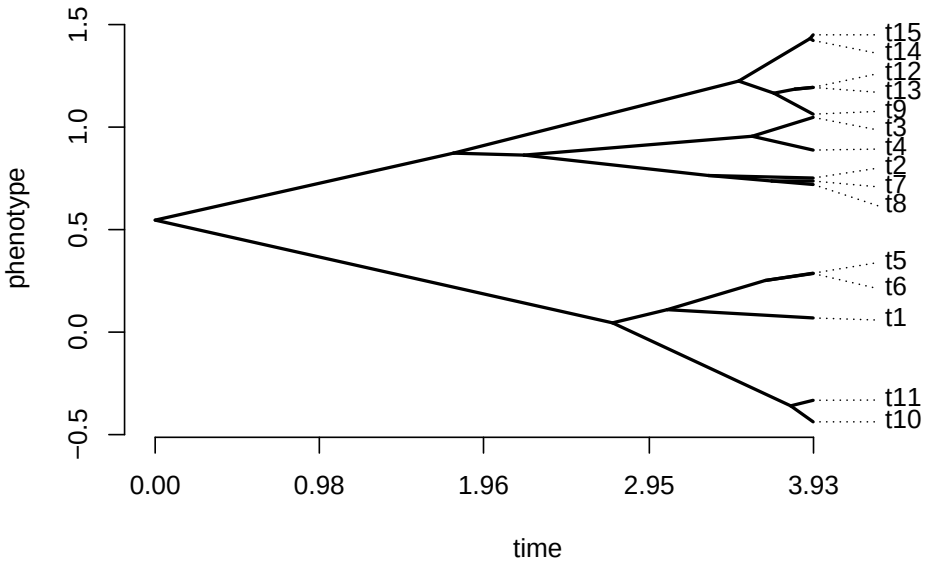
```
library(phylobase)
library(adephylo)
trait3 <- fastBM(tree, sig2=0.1, nsim=1, internal=FALSE) #simuler un caractère
trait.table <- data.frame(trait1=trait1[,1], trait3)
obj <- phylo4d(tree, trait.table) # construire un objet phylo4d
op <- par(mar=c(1,1,1,1))
table.phylo4d(obj,cex.label=1,cex.symbol=1,ratio.tree=0.8,grid=FALSE,box=FALSE)
```




```
par(op)
```

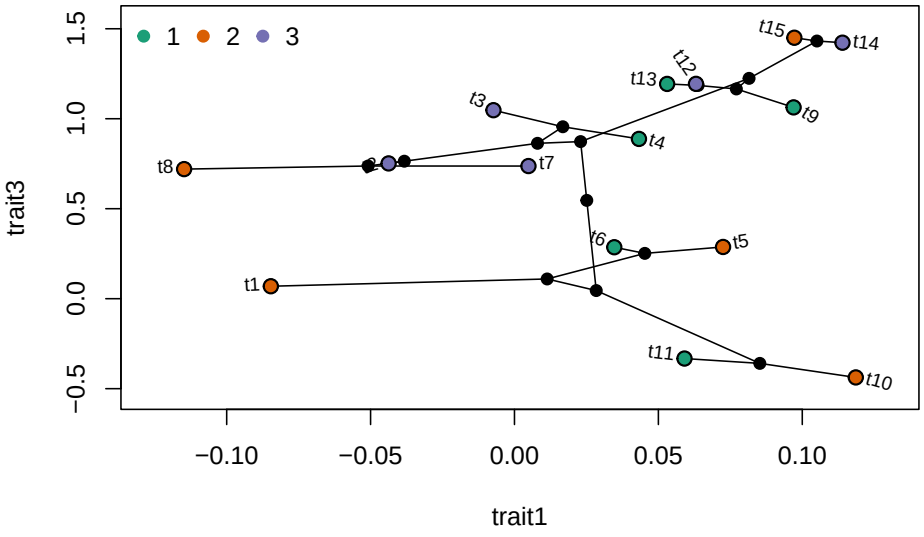
On peut aussi représenter avec un traitgram:

```
require(phytools)
phenogram(tree,trait3,spread.labels=TRUE)
```



Enfin, il est également possible de représenter un arbre sur un graphique en deux dimensions, en colorant les points avec la variable catégorielle.

```
phylomorphospace(tree,trait.table)
points(trait.table,pch=21,bg=ColorPalette1[trait2],col="black",cex=1.2,adj=1)
legend("topleft",legend=levels(trait2),
       col=ColorPalette1,pch=20,bty="n",cex=1,pt.cex=1.5,ncol=length(levels(trait2)))
```



14.4 Gérer plusieurs arbres

Dans plusieurs cas, il est important de savoir comment gérer plusieurs arbres dans R. Ceux-ci sont normalement stockés dans un objet `multiPhylo`. Voyons un exemple.

```
trees <- pbtree(n=15,nsim=10)
trees
```

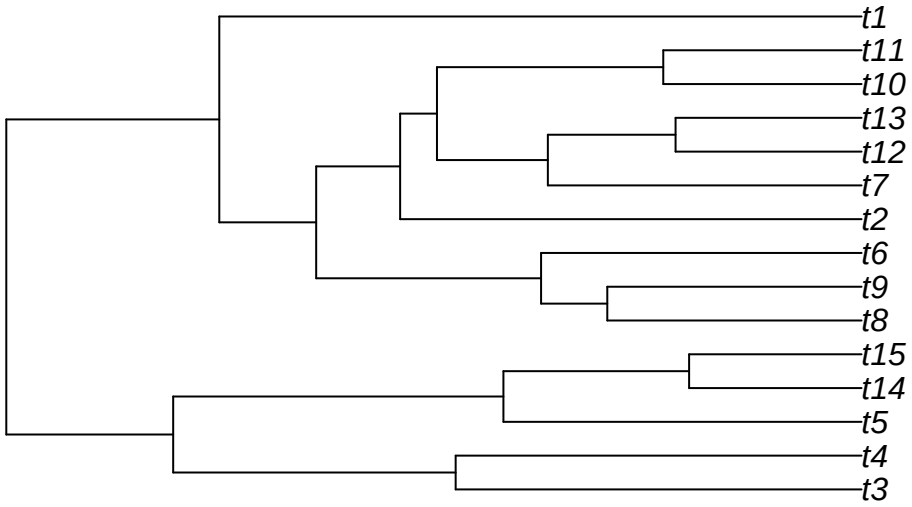
```
## 10 phylogenetic trees
```

Vous pouvez voir que l'objet n'est pas le même qu'un objet phylo. Par exemple, si vous utilisez le code `plot(trees)`, vous serez invité à appuyer sur Entrée pour passer d'un arbre à l'autre. Pour accéder aux arbres individuels, vous devez utiliser la technique suivante.

```
trees[[1]]
```

```
##
## Phylogenetic tree with 15 tips and 14 internal nodes.
##
## Tip labels:
##   t3, t4, t5, t14, t15, t8, ...
##
## Rooted; includes branch lengths.
```

```
plot(trees[[1]])
```

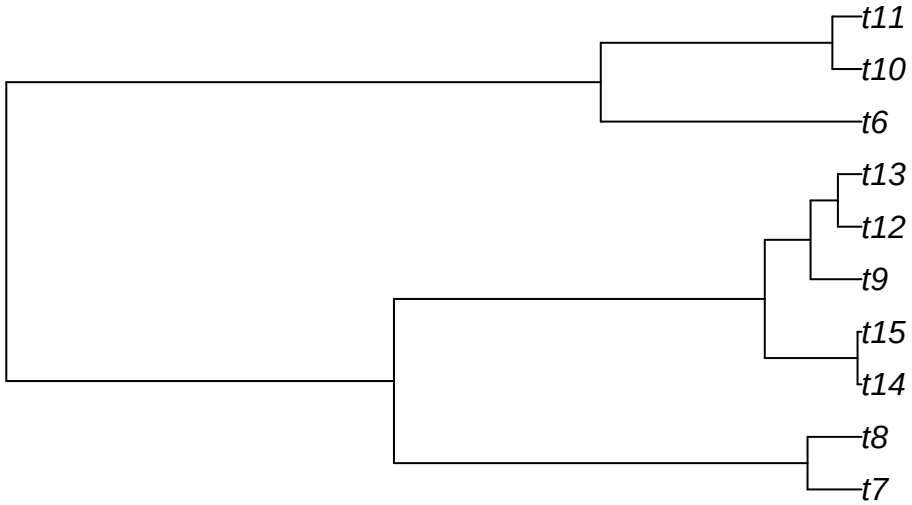


14.5 Manipuler les arbres

Il existe plusieurs manipulations qui peuvent être effectuées sur les arbres. Voici quelques exemples.

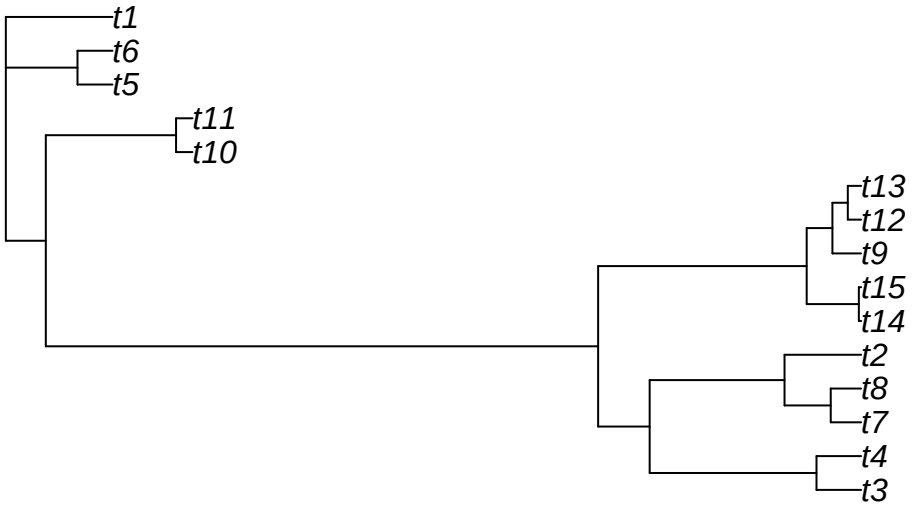
14.5.1 Supprimer des feuilles de l'arbre

```
plot(drop.tip(tree,c("t1","t2","t3","t4","t5")))
```



14.5.2 Réenraciner les arbres

```
plot(root(tree,"t1"))
```



14.5.3 Obtenir les distances cophénétiques

```
cophenetic.phylo(tree)
```

```
##            t3        t4        t7        t8       t2        t14        t15
## t3  0.0000000 0.7285289 3.4533793 3.4533793 3.453379 4.29496600 4.29496600
## t4  0.7285289 0.0000000 3.4533793 3.4533793 3.453379 4.29496600 4.29496600
## t7  3.4533793 3.4533793 0.0000000 0.4962115 1.250834 4.29496600 4.29496600
## t8  3.4533793 3.4533793 0.4962115 0.0000000 1.250834 4.29496600 4.29496600
## t2  3.4533793 3.4533793 1.2508336 1.2508336 0.000000 4.29496600 4.29496600
## t14 4.2949660 4.2949660 4.2949660 4.2949660 4.294966 0.00000000 0.03691962
## t15 4.2949660 4.2949660 4.2949660 4.2949660 4.294966 0.03691962 0.00000000
## t9  4.2949660 4.2949660 4.2949660 4.2949660 4.294966 0.88909442 0.88909442
## t12 4.2949660 4.2949660 4.2949660 4.2949660 4.294966 0.88909442 0.88909442
## t13 4.2949660 4.2949660 4.2949660 4.2949660 4.294966 0.88909442 0.88909442
## t5  7.8556290 7.8556290 7.8556290 7.8556290 7.855629 7.85562897 7.85562897
## t6  7.8556290 7.8556290 7.8556290 7.8556290 7.855629 7.85562897 7.85562897
## t1  7.8556290 7.8556290 7.8556290 7.8556290 7.855629 7.85562897 7.85562897
## t10 7.8556290 7.8556290 7.8556290 7.8556290 7.855629 7.85562897 7.85562897
## t11 7.8556290 7.8556290 7.8556290 7.8556290 7.855629 7.85562897 7.85562897
##            t9       t12       t13        t5        t6       t1       t10
## t3  4.2949660 4.2949660 4.2949660 7.8556290 7.8556290 7.855629 7.855629
## t4  4.2949660 4.2949660 4.2949660 7.8556290 7.8556290 7.855629 7.855629
## t7  4.2949660 4.2949660 4.2949660 7.8556290 7.8556290 7.855629 7.855629
## t8  4.2949660 4.2949660 4.2949660 7.8556290 7.8556290 7.855629 7.855629
```

```
## t2  4.2949660 4.2949660 4.2949660 7.8556290 7.8556290 7.855629 7.855629
## t14 0.8890944 0.8890944 0.8890944 7.8556290 7.8556290 7.855629 7.855629
## t15 0.8890944 0.8890944 0.8890944 7.8556290 7.8556290 7.855629 7.855629
## t9  0.0000000 0.4685481 0.4685481 7.8556290 7.8556290 7.855629 7.855629
## t12 0.4685481 0.0000000 0.2174902 7.8556290 7.8556290 7.855629 7.855629
## t13 0.4685481 0.2174902 0.0000000 7.8556290 7.8556290 7.855629 7.855629
## t5  7.8556290 7.8556290 7.8556290 0.0000000 0.5720997 1.742611 2.395633
## t6  7.8556290 7.8556290 7.8556290 0.5720997 0.0000000 1.742611 2.395633
## t1  7.8556290 7.8556290 7.8556290 1.7426110 1.7426110 0.000000 2.395633
## t10 7.8556290 7.8556290 7.8556290 2.3956328 2.3956328 2.395633 0.000000
## t11 7.8556290 7.8556290 7.8556290 2.3956328 2.3956328 2.395633 0.268201
##          t11
## t3  7.855629
## t4  7.855629
## t7  7.855629
## t8  7.855629
## t2  7.855629
## t14 7.855629
## t15 7.855629
## t9  7.855629
## t12 7.855629
## t13 7.855629
## t5  2.395633
## t6  2.395633
## t1  2.395633
## t10 0.268201
## t11 0.000000
```

Chapter 15

Solutions aux défis

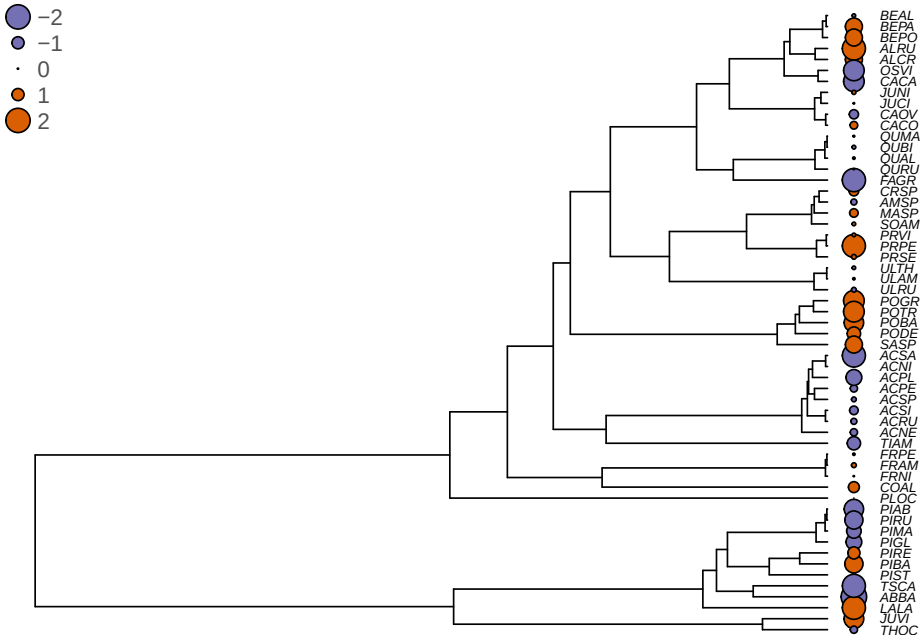
15.1 Défi 1

Dans la trame de données `seedplantsdata`, il y avait de nombreux traits différents. Essayez d'ajuster une régression de la tolérance à l'ombre des arbres (`Shade`) en fonction de la masse des graines (`Sm`). En d'autres termes, testez si la tolérance à l'ombre peut être expliquée par la masse des graines des arbres. Ensuite, essayez de voir si les résidus sont corrélés phylogénétiquement.

```
# Ajuster un modèle linéaire en utilisant les moindres carrés (Ordinary Least Squares; OLS)
Sm.lm <- lm(Shade ~ Sm, data = seedplantsdata)
summary(Sm.lm)
```

```
##
## Call:
## lm(formula = Shade ~ Sm, data = seedplantsdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9042 -0.9009  0.1481  0.5982  2.0962
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.904e+00  1.608e-01  18.064   <2e-16 ***
## Sm          -5.824e-05  5.640e-05  -1.033    0.306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.149 on 55 degrees of freedom
## Multiple R-squared:  0.01902,    Adjusted R-squared:  0.001184
## F-statistic: 1.066 on 1 and 55 DF,  p-value: 0.3063
```

```
# Extraire les résidus
Sm.res <- residuals(Sm.lm)
# Représenter les résidus à côté de la phylogénie
op <- par(mar=c(1,1,1,1))
plot(seedplantstree,type="p",TRUE,label.offset=0.01,cex=0.5,no.margin=FALSE)
tiplabels(pch=21,bg=cols[ifelse(Sm.res>0,1,2)],col="black",cex=abs(Sm.res),adj=0.505)
legend("topleft",legend=c("-2","-1","0","1","2"),pch=21,
       pt.bg=cols[c(1,1,1,2,2)],bty="n",
       text.col="gray32",cex=0.8,pt.cex=c(2,1,0.1,1,2))
```




```
par(op)
```

15.2 Défi 2

Pouvez-vous obtenir la matrice de covariance et la matrice de corrélation pour l'arbre phylogénétique des plantes à graines de l'exemple ci-dessus (`seedplantstree`)?

```
# Covariance matrix
seedplants.cov <- vcv(seedplantstree,corr=FALSE)
# Regarder les premières lignes de la matrice
head(round(seedplants.cov,3))
```

```
##       ABBA  ACNE  ACNI  ACPE  ACPL  ACRU  ACSA  ACSI  ACSP  ALCR  ALRU  AMSP
## ABBA 0.151 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```

```
## ACNE 0.000 0.151 0.146 0.146 0.146 0.146 0.146 0.146 0.146 0.099 0.099 0.099
## ACNI 0.000 0.146 0.151 0.147 0.148 0.147 0.150 0.147 0.147 0.099 0.099 0.099
## ACPE 0.000 0.146 0.147 0.151 0.147 0.147 0.147 0.147 0.148 0.099 0.099 0.099
## ACPL 0.000 0.146 0.148 0.147 0.151 0.147 0.148 0.147 0.147 0.099 0.099 0.099
## ACRU 0.000 0.146 0.147 0.147 0.147 0.151 0.147 0.150 0.147 0.099 0.099 0.099
##       BEAL  BEPA  BEPO  CACA  CACO  CAOV COAL  CRSP  FAGR FRAM FRNI FRPE  JUCI
## ABBA 0.000 0.000 0.000 0.000 0.000 0.000 0.00 0.000 0.000 0.00 0.00 0.00 0.000
## ACNE 0.099 0.099 0.099 0.099 0.099 0.099 0.09 0.099 0.099 0.09 0.09 0.09 0.099
## ACNI 0.099 0.099 0.099 0.099 0.099 0.099 0.09 0.099 0.099 0.09 0.09 0.09 0.099
## ACPE 0.099 0.099 0.099 0.099 0.099 0.099 0.09 0.099 0.099 0.09 0.09 0.09 0.099
## ACPL 0.099 0.099 0.099 0.099 0.099 0.099 0.09 0.099 0.099 0.09 0.09 0.09 0.099
## ACRU 0.099 0.099 0.099 0.099 0.099 0.099 0.09 0.099 0.099 0.09 0.09 0.09 0.099
##       JUNI JUVI  LALA  MASP  OSVI PIAB PIBA PIGL PIMA PIRE PIRU PIST  PLOC
## ABBA 0.000 0.08 0.127 0.000 0.000 0.13 0.13 0.13 0.13 0.13 0.13 0.13 0.000
## ACNE 0.099 0.00 0.000 0.099 0.099 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.079
## ACNI 0.099 0.00 0.000 0.099 0.099 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.079
## ACPE 0.099 0.00 0.000 0.099 0.099 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.079
## ACPL 0.099 0.00 0.000 0.099 0.099 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.079
## ACRU 0.099 0.00 0.000 0.099 0.099 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.079
##       POBA  PODE  POGR  POTR  PRPE  PRSE  PRVI  QUAL  QUBI  QUMA  QURU  SASP
## ABBA 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## ACNE 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099
## ACNI 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099
## ACPE 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099
## ACPL 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099
## ACRU 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099 0.099
##       SOAM THOC  TIAM  TSCA  ULAM  ULRU  ULTH
## ABBA 0.000 0.08 0.000 0.137 0.000 0.000 0.000
## ACNE 0.099 0.00 0.109 0.000 0.099 0.099 0.099
## ACNI 0.099 0.00 0.109 0.000 0.099 0.099 0.099
## ACPE 0.099 0.00 0.109 0.000 0.099 0.099 0.099
## ACPL 0.099 0.00 0.109 0.000 0.099 0.099 0.099
## ACRU 0.099 0.00 0.109 0.000 0.099 0.099 0.099
```

```
# Matrice de corrélation
seedplants.cor <- vcv(seedplantstree,corr=TRUE)
# Regarder les premières lignes de la matrice
head(round(seedplants.cor,3))
```

```
##      ABBA  ACNE  ACNI  ACPE  ACPL  ACRU  ACSA  ACSI  ACSP  ALCR  ALRU  AMSP
## ABBA    1 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## ACNE    0 1.000 0.967 0.967 0.967 0.967 0.967 0.967 0.967 0.654 0.654 0.654
## ACNI    0 0.967 1.000 0.976 0.981 0.974 0.997 0.974 0.976 0.654 0.654 0.654
## ACPE    0 0.967 0.976 1.000 0.976 0.974 0.976 0.974 0.983 0.654 0.654 0.654
## ACPL    0 0.967 0.981 0.976 1.000 0.974 0.981 0.974 0.976 0.654 0.654 0.654
## ACRU    0 0.967 0.974 0.974 0.974 1.000 0.974 0.997 0.974 0.654 0.654 0.654
```

```
##       BEAL  BEPA  BEPO  CACA  CACO  CAOV  COAL  CRSP  FAGR  FRAM  FRNI  FRPE
## ABBA 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## ACNE 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
## ACNI 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
## ACPE 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
## ACPL 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
## ACRU 0.654 0.654 0.654 0.654 0.654 0.654 0.596 0.654 0.654 0.596 0.596 0.596
##       JUCI  JUNI  JUVI  LALA  MASP  OSVI PIAB PIBA PIGL PIMA PIRE PIRU PIST
## ABBA 0.000 0.000 0.528 0.843 0.000 0.000 0.86 0.86 0.86 0.86 0.86 0.86 0.86
## ACNE 0.654 0.654 0.000 0.000 0.654 0.654 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## ACNI 0.654 0.654 0.000 0.000 0.654 0.654 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## ACPE 0.654 0.654 0.000 0.000 0.654 0.654 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## ACPL 0.654 0.654 0.000 0.000 0.654 0.654 0.00 0.00 0.00 0.00 0.00 0.00 0.00
## ACRU 0.654 0.654 0.000 0.000 0.654 0.654 0.00 0.00 0.00 0.00 0.00 0.00 0.00
##       PLOC  POBA  PODE  POGR  POTR  PRPE  PRSE  PRVI  QUAL  QUBI  QUMA  QURU
## ABBA 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## ACNE 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## ACNI 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## ACPE 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## ACPL 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
## ACRU 0.523 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654 0.654
##       SASP  SOAM  THOC TIAM  TSCA  ULAM  ULRU  ULTH
## ABBA 0.000 0.000 0.528 0.00 0.906 0.000 0.000 0.000
## ACNE 0.654 0.654 0.000 0.72 0.000 0.654 0.654 0.654
## ACNI 0.654 0.654 0.000 0.72 0.000 0.654 0.654 0.654
## ACPE 0.654 0.654 0.000 0.72 0.000 0.654 0.654 0.654
## ACPL 0.654 0.654 0.000 0.72 0.000 0.654 0.654 0.654
## ACRU 0.654 0.654 0.000 0.72 0.000 0.654 0.654 0.654
```

15.3 Défi 3

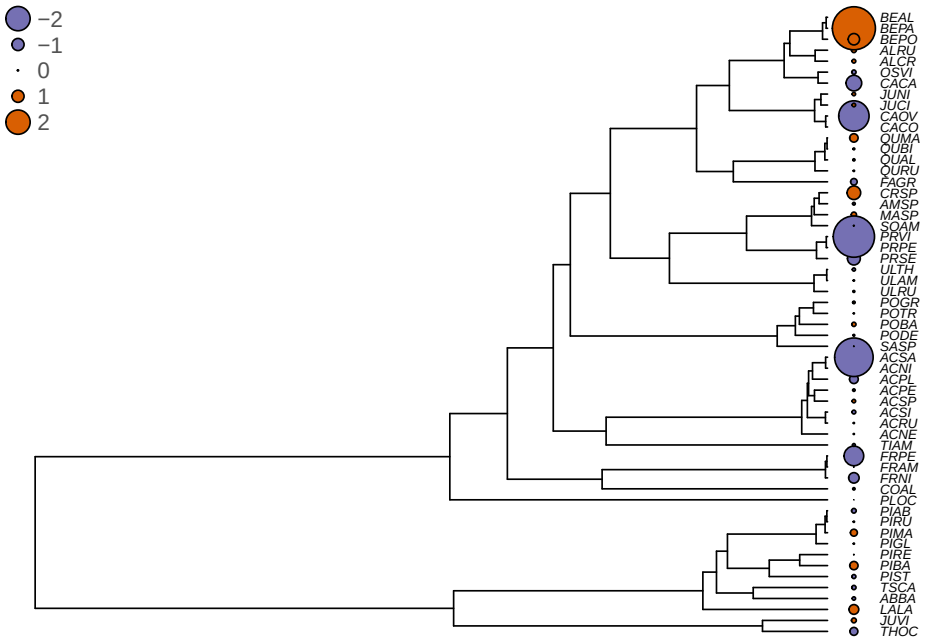
Ajustez un modèle PGLS pour voir si la masse des graines (`Sm`) explique la tolérance à l'ombre (`Shade`) avec le jeu de données `seedplantsdata`. Comment cela se compare-t-il aux résultats de la régression standard?

```
# Ajuster un PGLS
Sm.pgls <- gls(Shade ~ Sm, data = seedplantsdata, correlation=bm.corr)
summary(Sm.pgls)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Sm
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   240.3701 246.3921 -117.1851
##
## Correlation Structure: corBrownian
```

```
##  Formula: ~1
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##                   Value Std.Error    t-value p-value
## (Intercept)  2.8031105  4.591805  0.6104594  0.5441
## Sm          -0.0000417  0.000081 -0.5117076  0.6109
##
##  Correlation:
##    (Intr)
## Sm -0.004
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -0.22901901 -0.10170487  0.02535202  0.08873220  0.27907713
##
## Residual standard error: 7.873115
## Degrees of freedom: 57 total; 55 residual
```

```
# Extraire les résidus corrigés de la structure de corrélation
Sm.pgls.res <- residuals(Sm.pgls,type="normalized")
# Représenter les résidus à côté de la phylogénie
op <- par(mar=c(1,1,1,1))
plot(seedplantstree,type="p",TRUE,label.offset=0.01,cex=0.5,no.margin=FALSE)
tiplabels(pch=21,bg=cols[ifelse(Sm.pgls.res>0,1,2)],col="black",cex=abs(Sm.pgls.res),adj=0.505)
legend("topleft",legend=c("-2","-1","0","1","2"),pch=21,
       pt.bg=cols[c(1,1,1,2,2)],bty="n",
       text.col="gray32",cex=0.8,pt.cex=c(2,1,0.1,1,2))
```




```
par(op)
```

15.4 Défi 4

Essayez d'ajuster un PGLS avec une structure de corrélation de Pagel en régressant la tolérance à l'ombre sur la masse des graines. Les résidus sont-ils aussi corrélés phylogénétiquement que dans la régression précédente avec la densité du bois ?

```
# Ajuster un PGLS
Sm.pgls2 <- gls(Shade ~ Sm, data = seedplantsdata, correlation=pagel.corr)
# Résultats
summary(Sm.pgls2)
```

```
## Generalized least squares fit by REML
##   Model: Shade ~ Sm
##   Data: seedplantsdata
##        AIC      BIC    logLik
##   187.6889 195.7183 -89.84447
##
## Correlation Structure: corPagel
##  Formula: ~Code
##  Parameter estimate(s):
##   lambda
## 0.951553
```

```
##
## Coefficients:
##                   Value Std.Error   t-value p-value
## (Intercept)  2.8204268  1.497276  1.883705  0.0649
## Sm          -0.0000716  0.000060 -1.193604  0.2378
##
##  Correlation:
##    (Intr)
## Sm -0.009
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -0.6946682 -0.3115198  0.1068607  0.2604470  0.8319385
##
## Residual standard error: 2.620527
## Degrees of freedom: 57 total; 55 residual
```

Bibliography

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15.

Felsenstein, J. and Felenstein, J. (2004). *Inferring phylogenies*, volume 2. Sinauer associates Sunderland, MA.

Hansen, T. F. and Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4):1404–1417.

Maddison, W. P. and FitzJohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic biology*, 64(1):127–136.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756):877–884.

Paquette, A., Joly, S., and Messier, C. (2015). Explaining forest productivity using tree functional traits and phylogenetic information: two sides of the same coin over evolutionary scale? *Ecology and Evolution*, 5(9):1774–1783.

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution*, 1(4):319–329.

Zuur, A. F., Ieno, E. N., Smith, G. M., et al. (2007). *Analysing ecological data*, volume 680. Springer.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M., et al. (2009). *Mixed effects models and extensions in ecology with R*, volume 574. Springer.