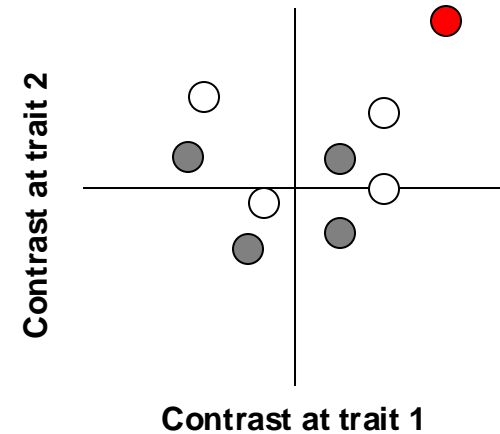
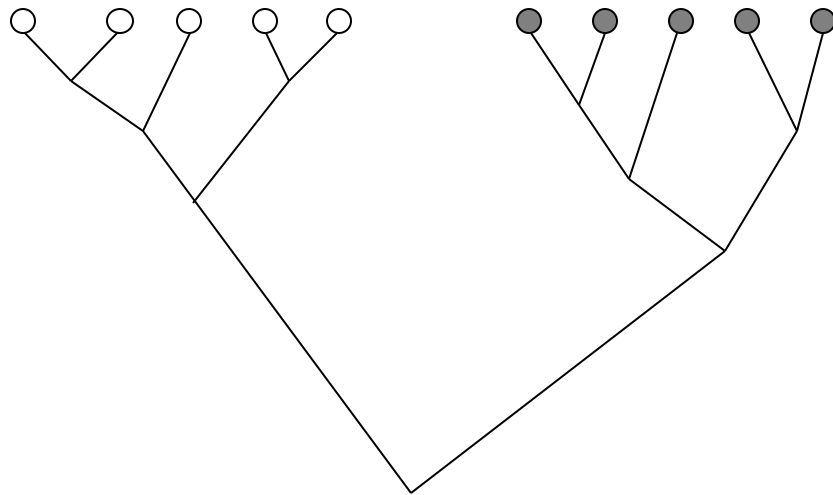
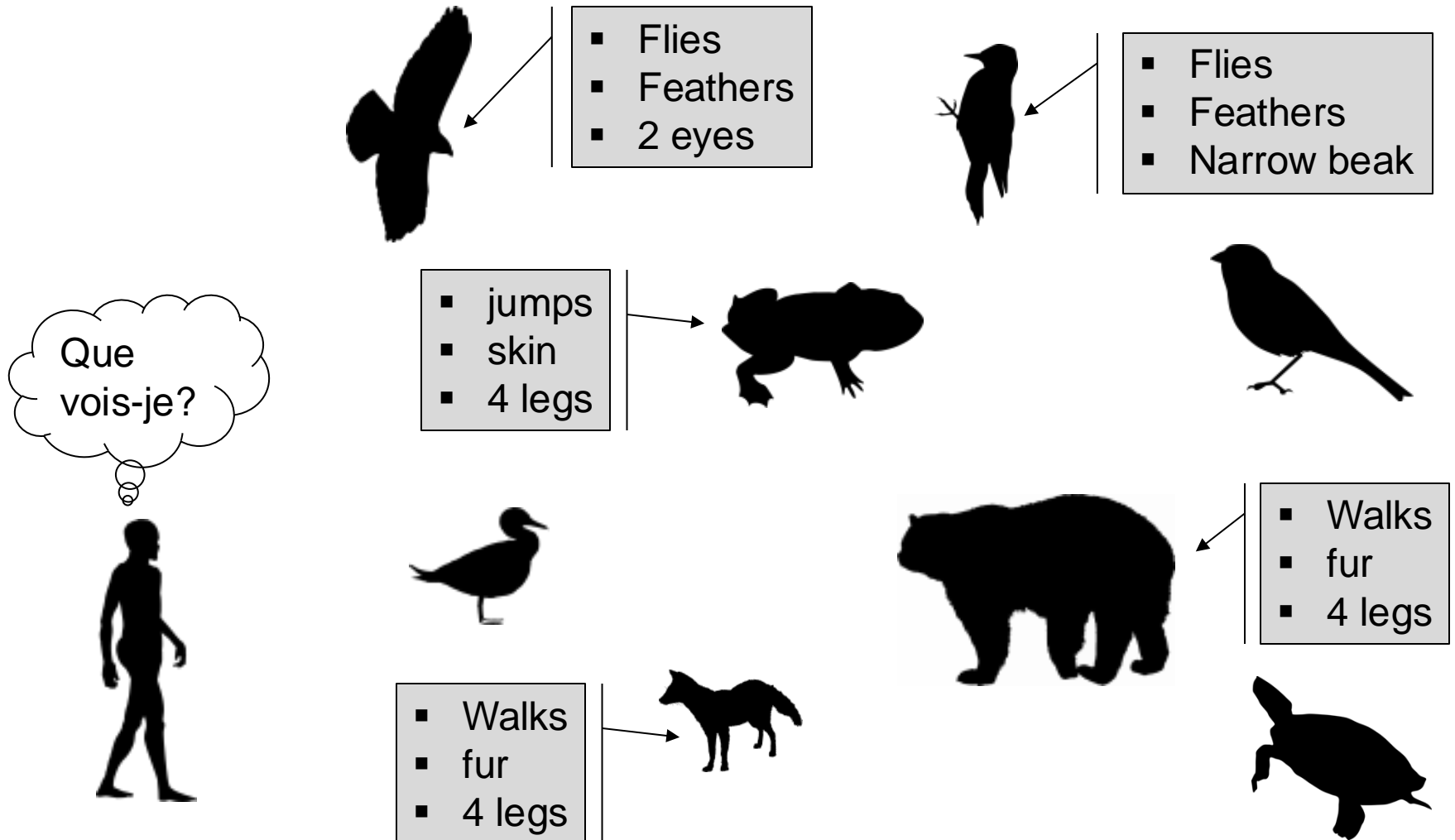


La méthode comparée



Introduction

Introduction



Introduction



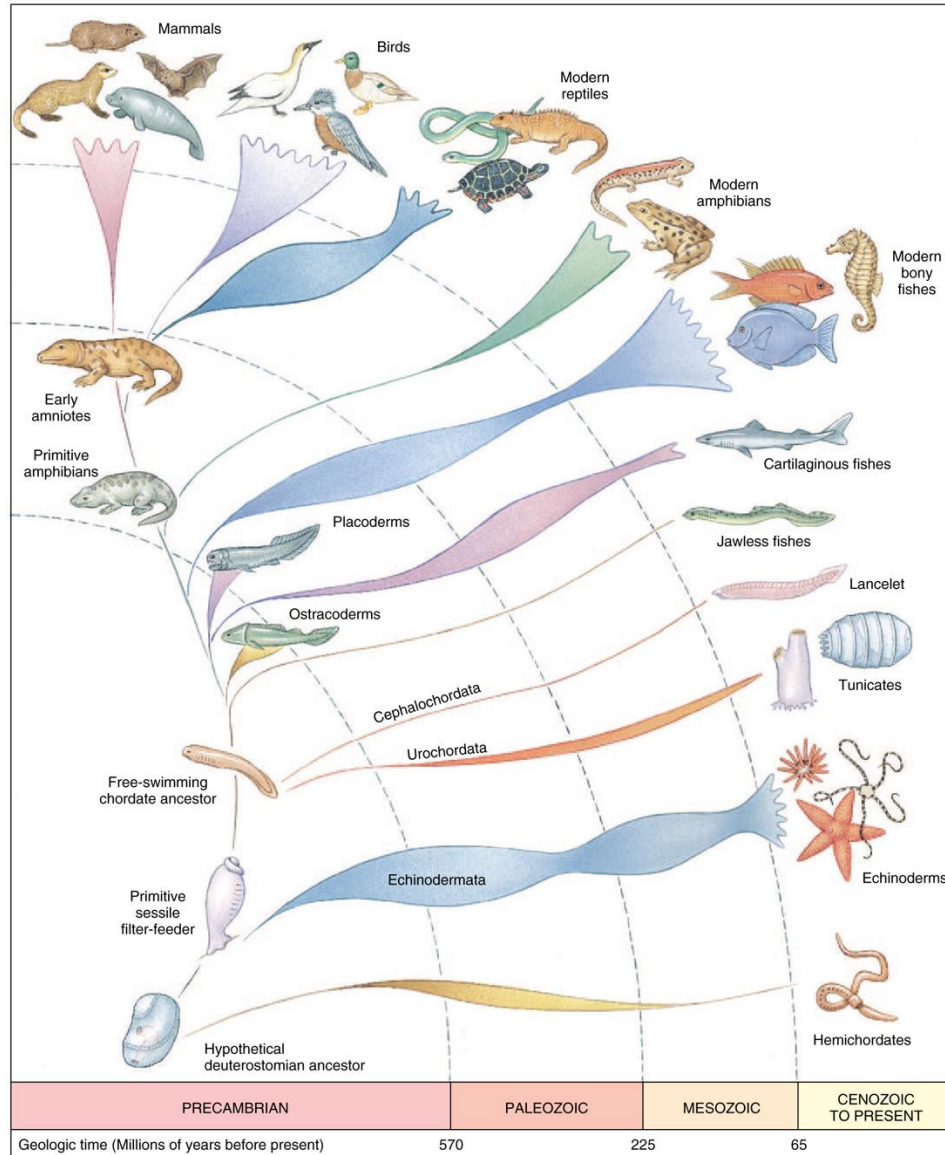
	feathers	other
Animal flies	9	0
Animal doesn't fly	0	7

```
> fisher.test(matrix(c(9,0,0,7),2))
```

Fisher's Exact Test for Count Data

```
data:  matrix(c(9, 0, 0, 7), 2)  
p-value = 8.741e-05
```

Evolution de la capacité de voler



La méthode comparée

La méthode comparée

- Definition
 - C'est le nom donné à une famille de méthodes qui vise à corriger les analyses statistiques pour la non-indépendance des espèces
- Pourquoi ?
 - Les espèces ne peuvent pas être considérées sans le contexte de l'évolution; les observations ne sont pas indépendantes.

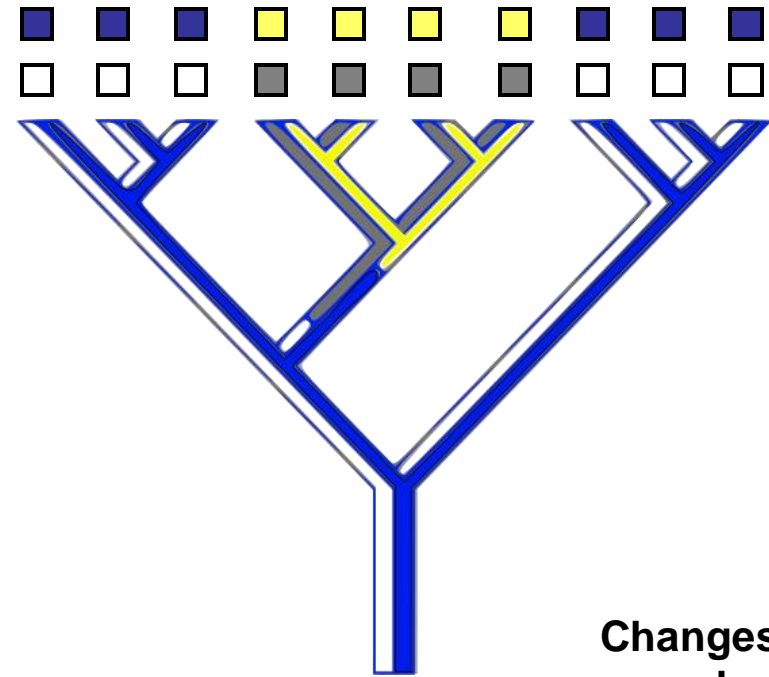






*Nothing in Biology Makes Sense
Except in the Light of Evolution*

Theodosius Dobzhansky, 1973

Un exemple qualitatif

species	flower	stem
1	blue	spiny
2	blue	spiny
3	blue	spiny
4	blue	spiny
5	yellow	smooth
6	yellow	smooth
7	yellow	smooth
8	yellow	smooth
9	blue	spiny
10	blue	spiny



		
	blue	yellow
 spiny	6	0
 smooth	0	4

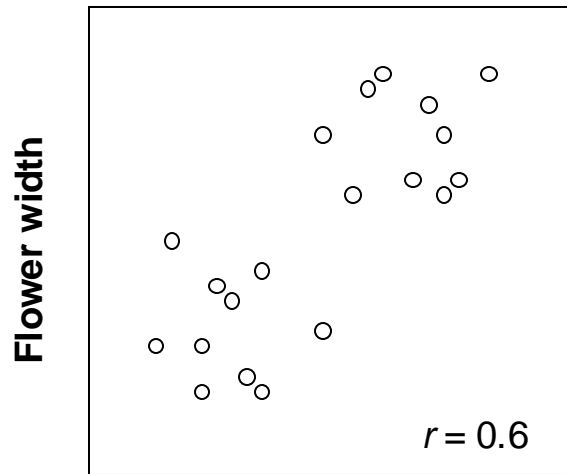
Fisher exact test
 $p = 0.00476$

Fisher exact test
 $p = 0,056$

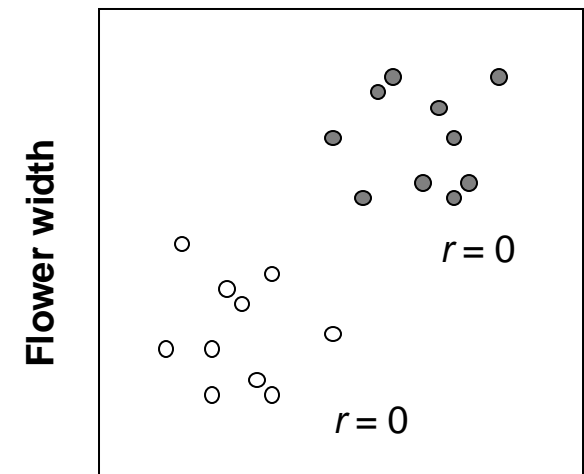
Changes along
branches

		flower	
		yes	no
stem	yes	1	0
	no	0	17

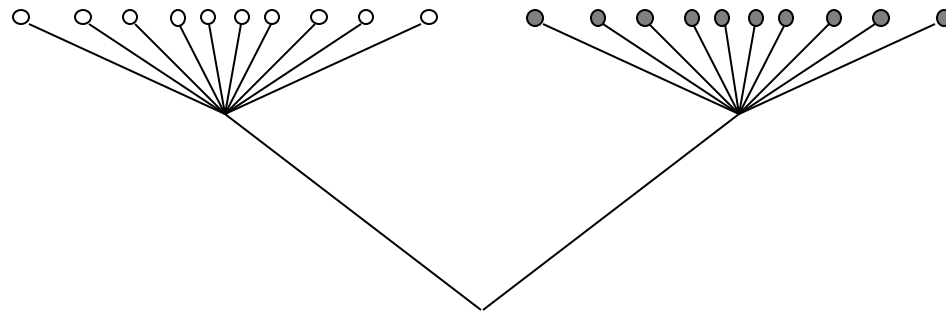
Un exemple quantitatif



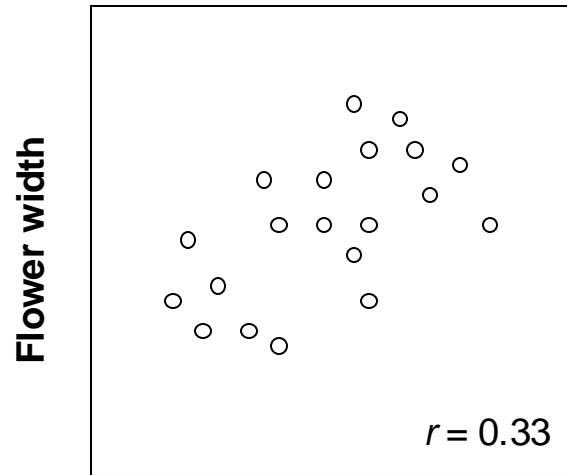
Trichome number
on the leaves



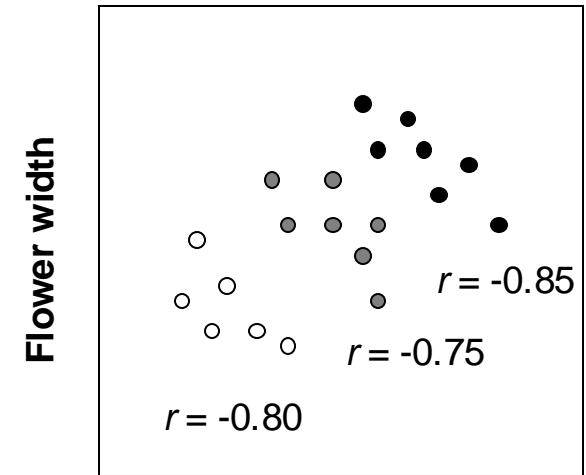
Trichome number
on the leaves



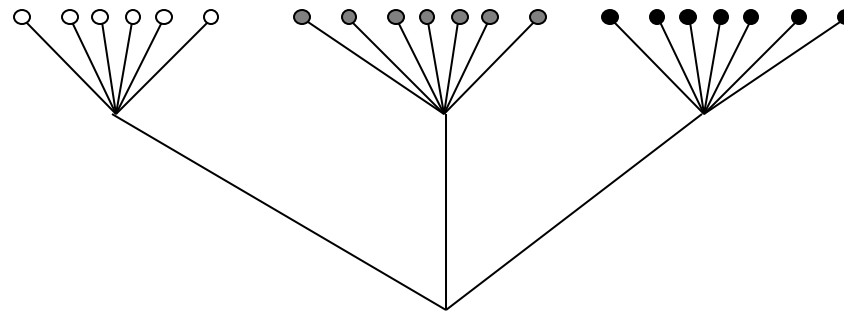
Un autre exemple quantitatif



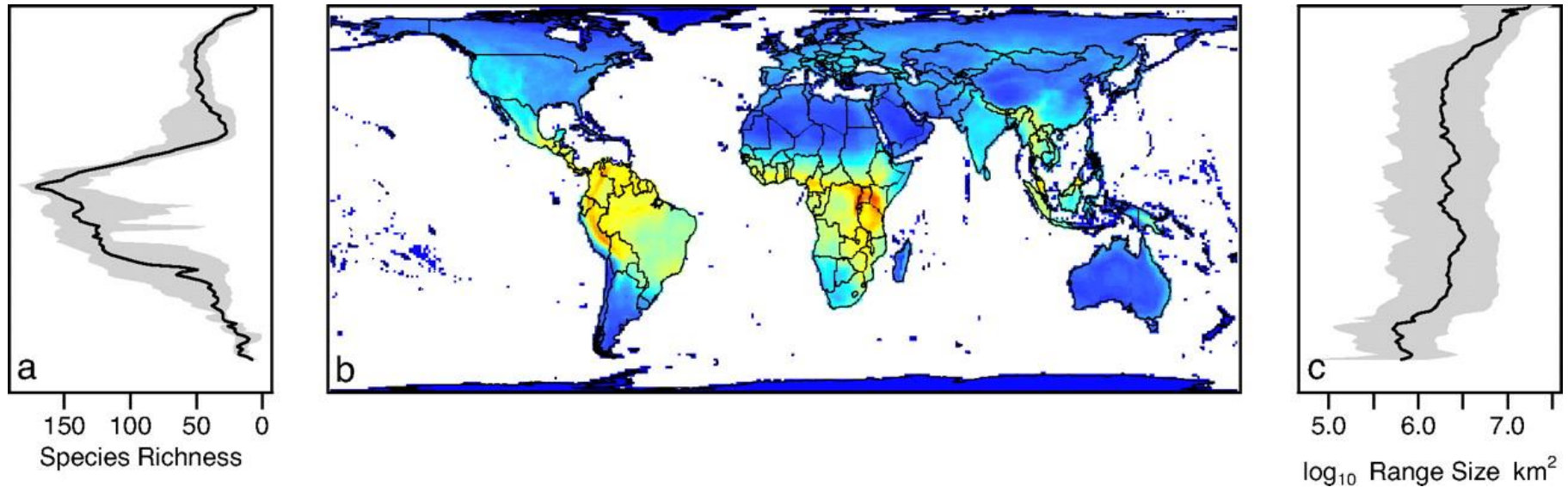
Trichome number
on the leaves



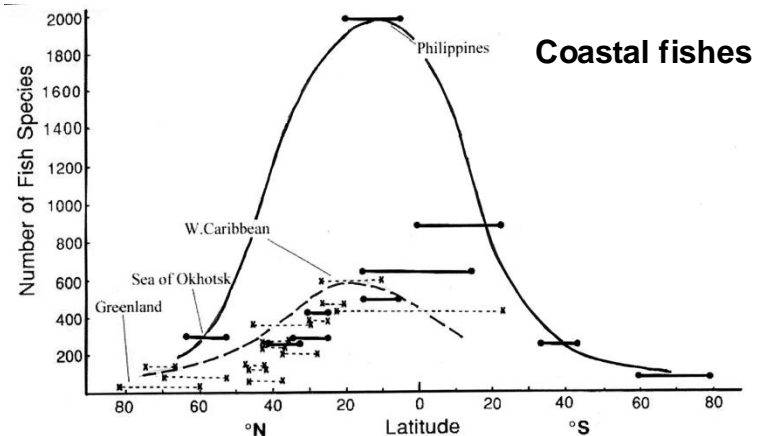
Trichome number
on the leaves



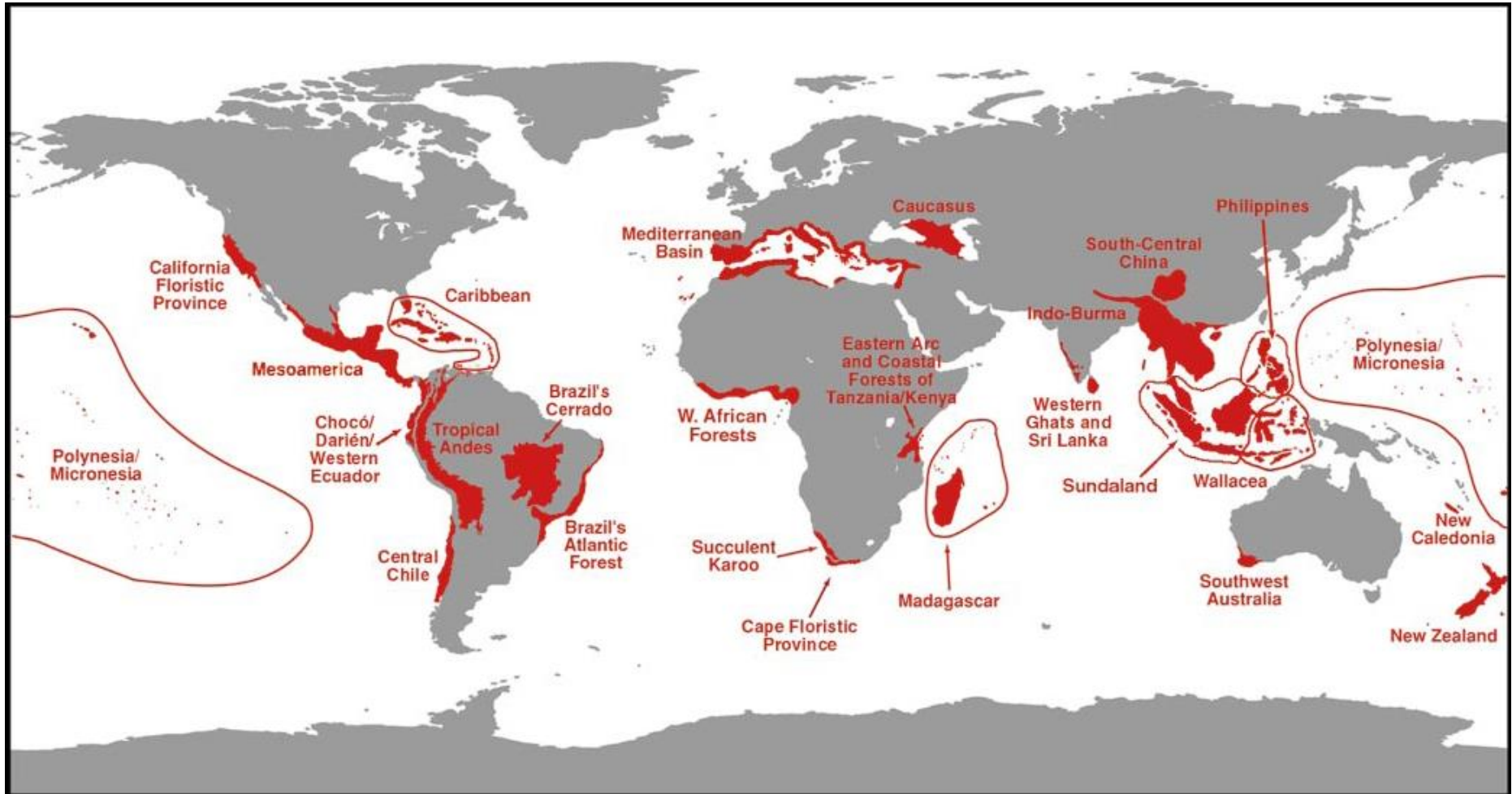
Le gradient latitudinal d'espèces



Mamals - Davies T J et al. PNAS 2008;105:11556-11563

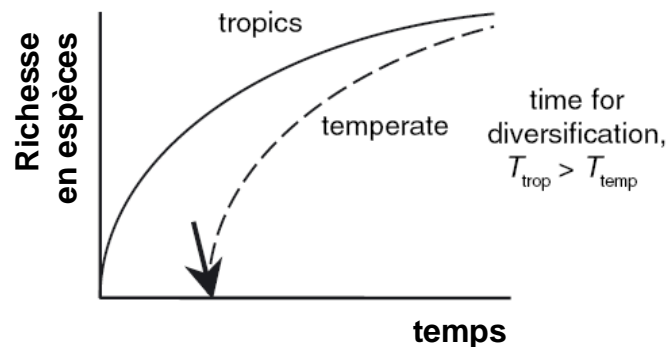


25 points chauds de biodiversité

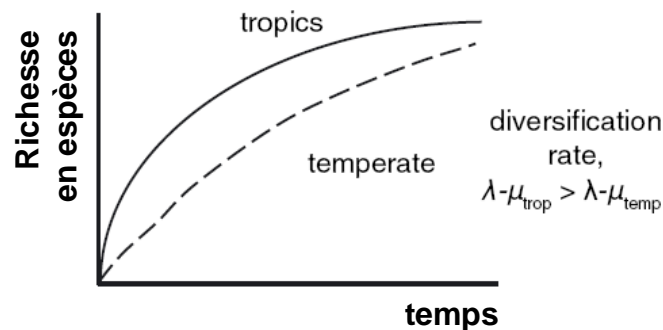


Quelques hypotheses sur le gradient d'espèces

- Plus de temps pour la speciation dans les tropiques



- Taux de diversification plus élevés dans les tropiques:
 - Taux de speciation plus élevés dans les tropiques (théorie du berceau)
 - Taux d'extinction plus bas dans les tropiques (Théorie du musée)

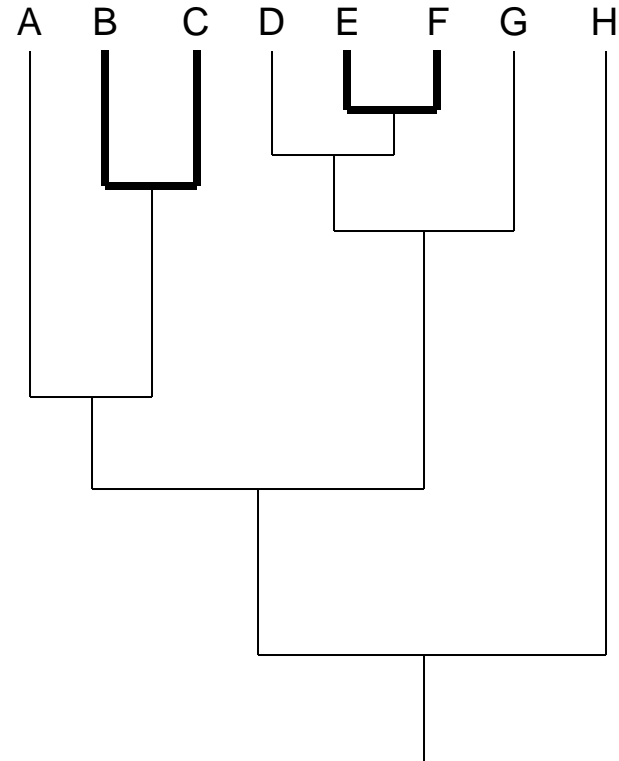
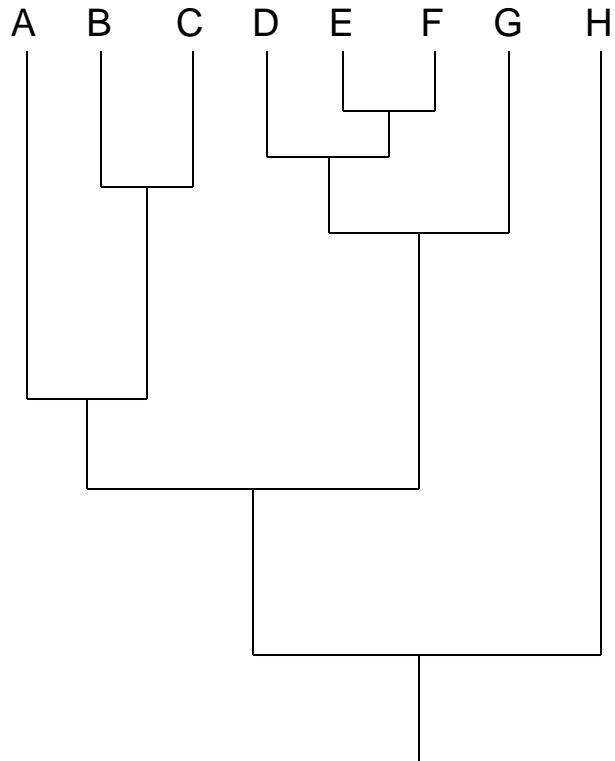


Méthodes comparées

Différentes approches

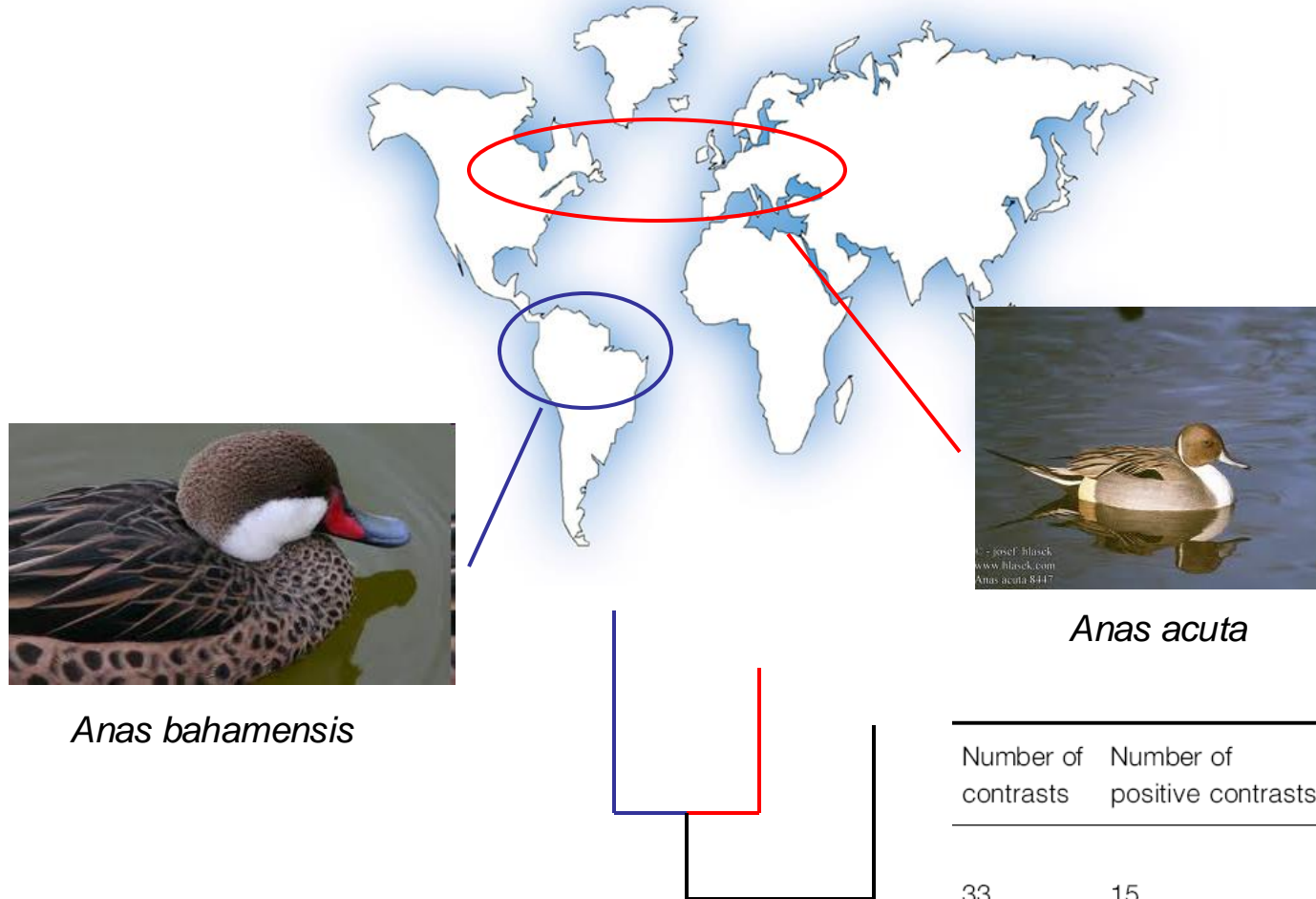
- Étudier des paires d'espèces
- Contrastes indépendants
- Moindres carrés généralisés phylognétiques (Phylogenetic generalized least squares: PGLS)

Paires d'espèces



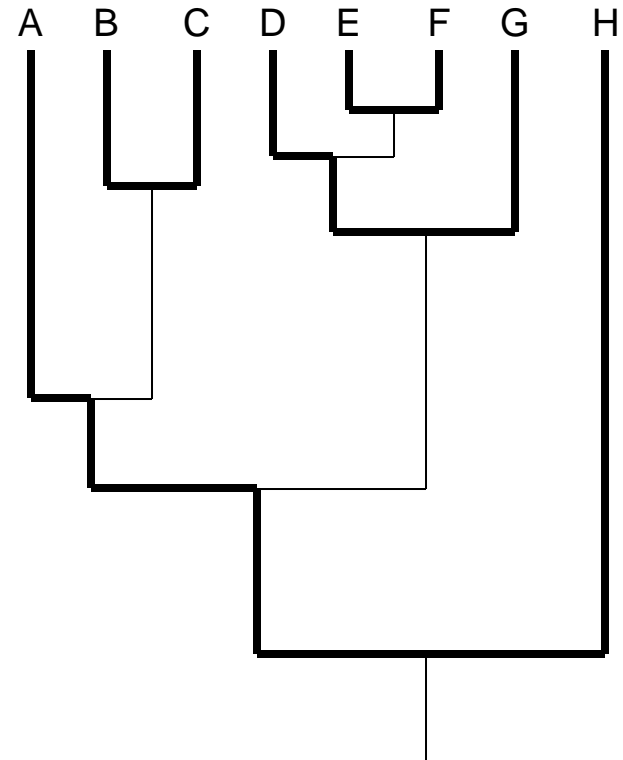
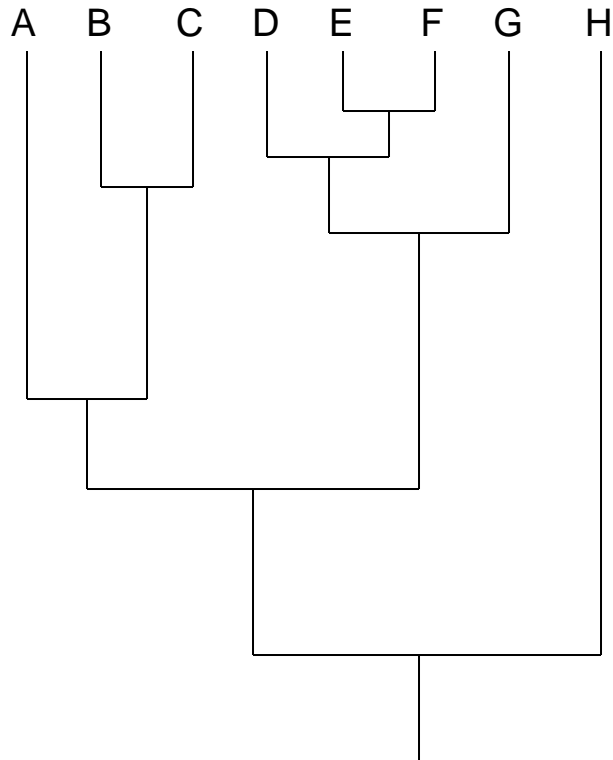
Ces comparaisons sont indépendantes

Tester un haut taux de diversification dans les tropiques



Pas significatif

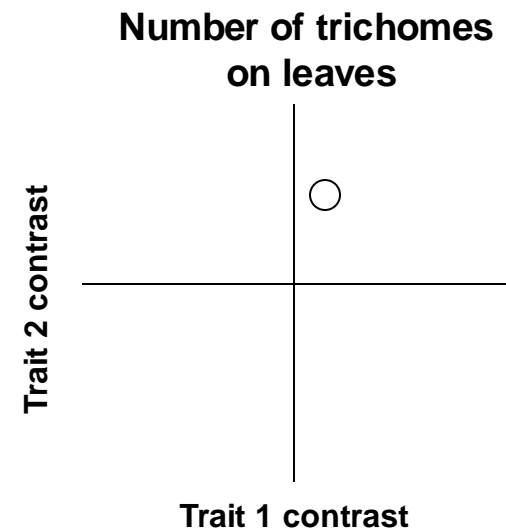
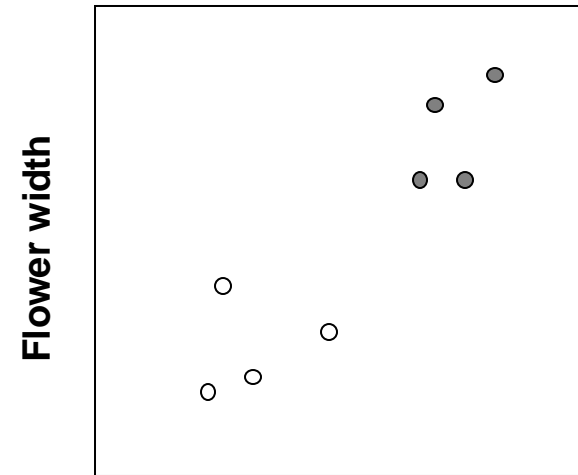
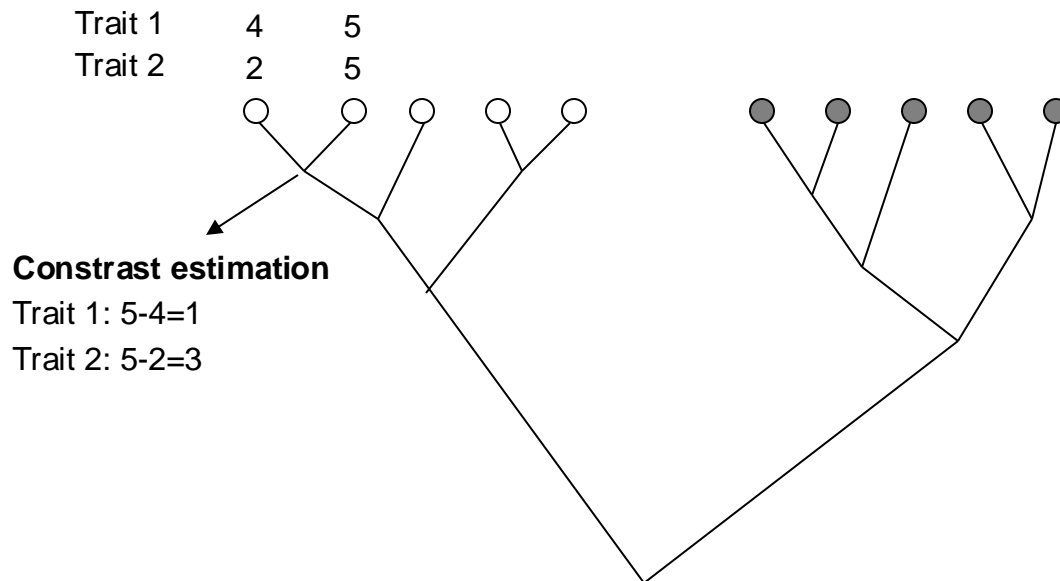
Paires d'espèces (plus complexe!)



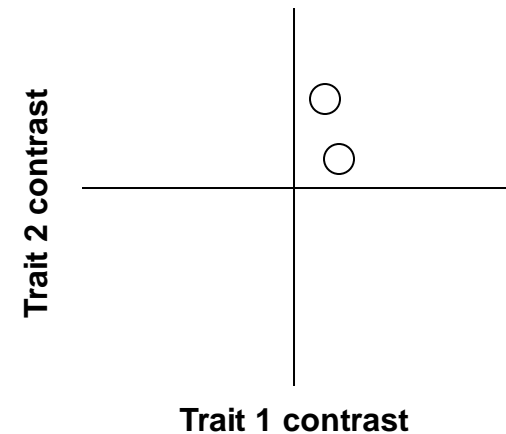
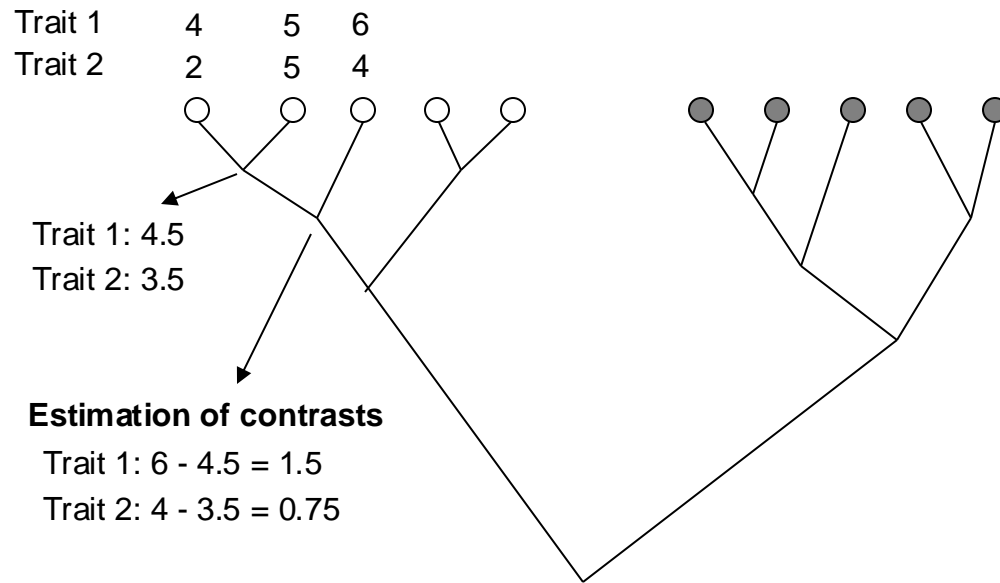
Ces observations sont aussi indépendantes les unes des autres

Contrastes indépendants

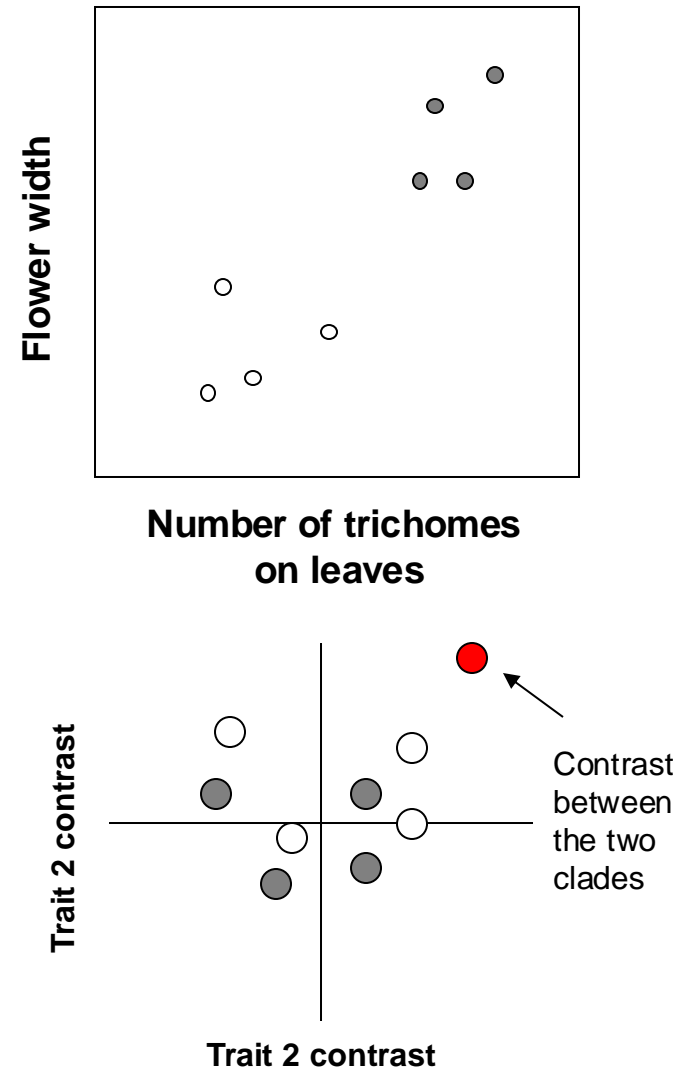
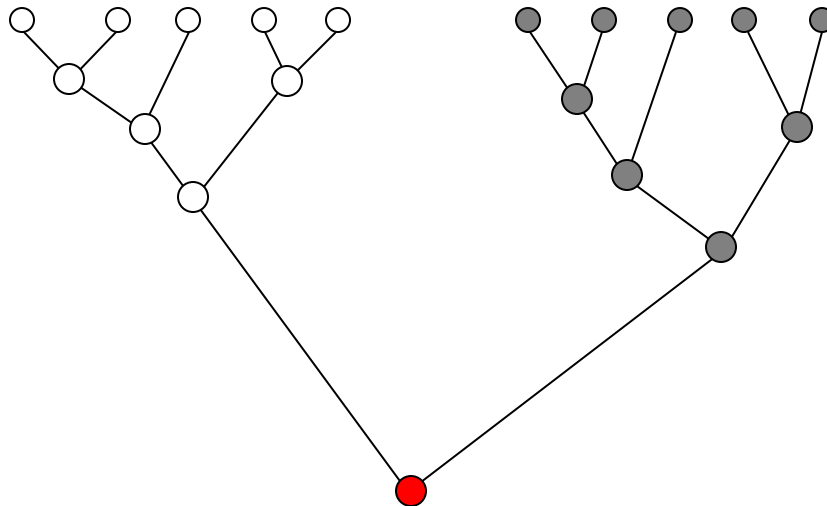
Objective: Tester les corrélations entre caractères tout en tenant comptes de l'histoire évolutive des espèces.



Contrastes indépendants

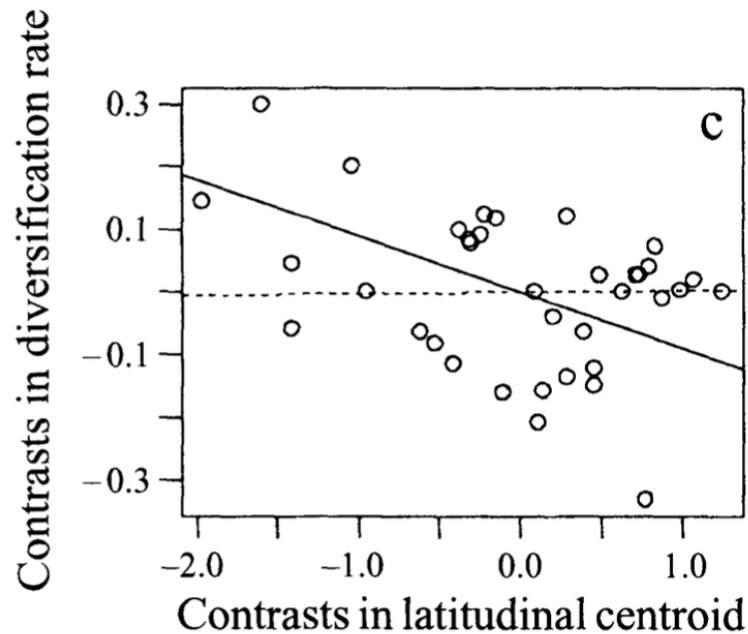


Contrastes indépendants



Tester un haut taux de diversification dans les tropiques

Test: En utilisant une phylogénie de genres d'oiseaux, Cardillo et al. (2005) on estimé le taux de diversification de chaque genre et sa latitude moyenne. Ils ont ensuite estimé les contrastes indépendants et testé s'il y a une corrélation entre la latitude et le taux de diversification.



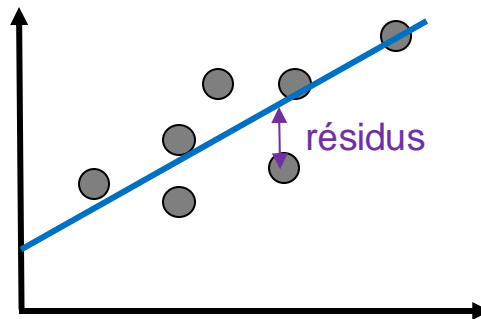
P-value = 0.023

Plus haut taux de
diversification dans les
tropiques

PGLS

Molèle linéaire (ordinary least squares – OLS)

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



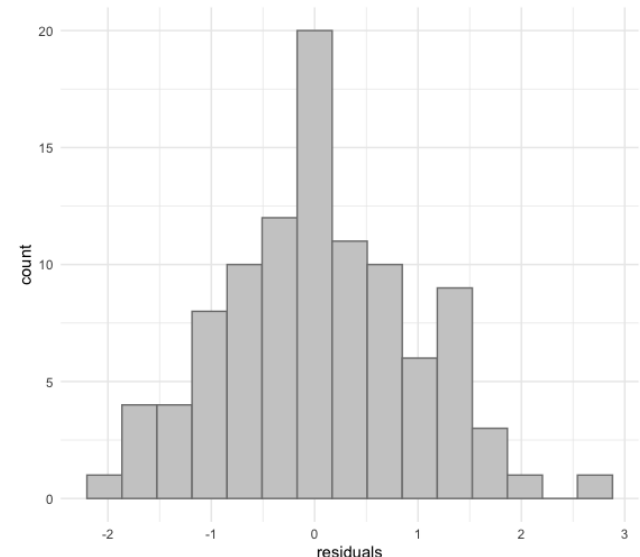
Molèle linéaire (ordinary least squares – OLS)

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

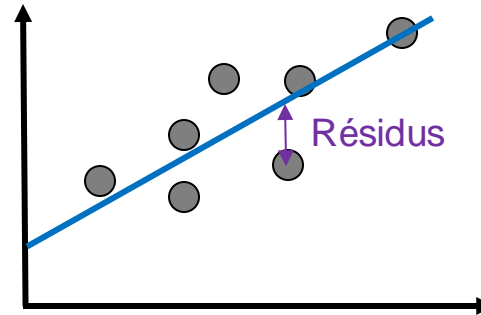
Plusieurs prémisses

- Normalité, indépendance, modèle approprié
- Normalité:

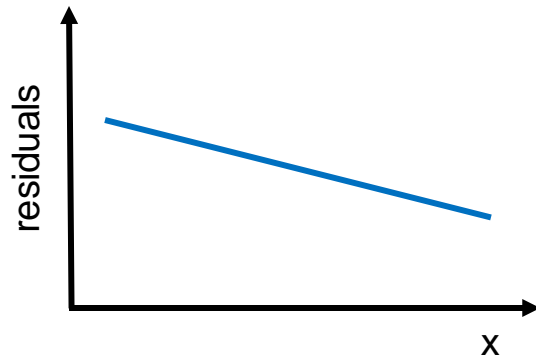
$$\mathbf{e} \sim \mathcal{N}(0, \sigma^2)$$



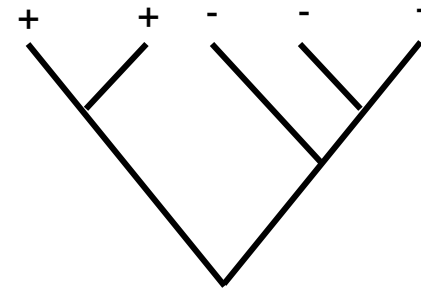
Exemples de violations



Résidus corrélés avec la
variables dépendante



Résidus corrélés avec la
phylogénie



Travaux pratiques dans R



www.github.com/simjoly/ComparativeMethods-HalfDayWorkshop

Allez jusqu'au défi #1

PGLS

PGLS

- Un cas special des moindres carrés généralisés (generalized least squares: GLS)
 - Permet aux résidus du modèle d'être corrélé d'une façon spécifique
 - Utilisé pour les correlations spatiales, les series temporelles, les relations phylogénétiques, etc.

PGLS

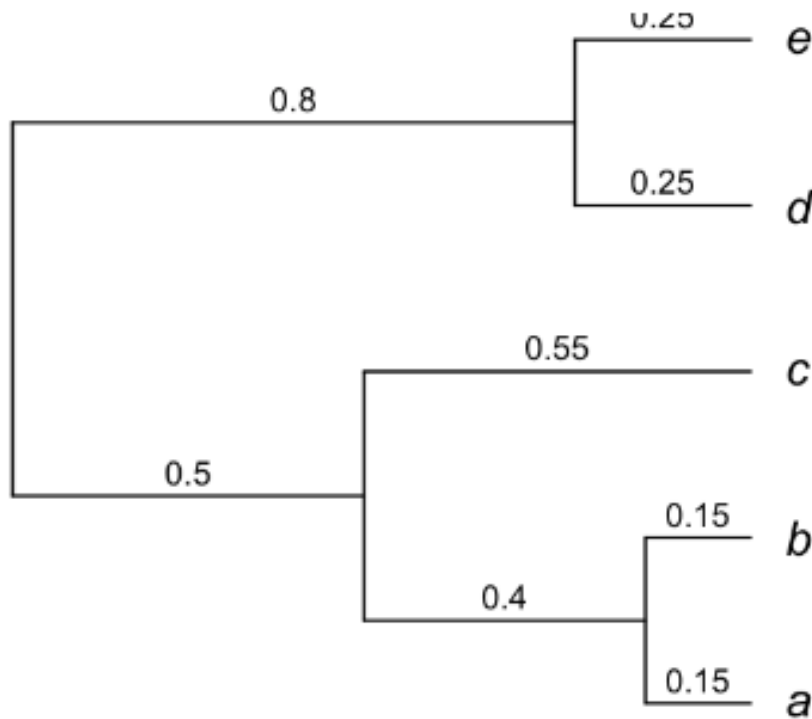
$$\mathbf{y} = \alpha + \beta \mathbf{x} + \mathbf{e}$$

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

$$\sigma^2 \mathbf{C} = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1i} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2i} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{i1} & \rho_{i2} & \rho_{i3} & \dots & 1 \end{pmatrix}$$

Matrice de corrélation

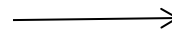
D'un arbre à une matrice de variance - covariance



	a	b	c	c	d
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
c	0.00	0.00	0.00	1.05	0.80
d	0.00	0.00	0.00	0.80	1.05

... à une matrice de corrélation

	a	b	c	c	d
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
c	0.00	0.00	0.00	1.05	0.80
d	0.00	0.00	0.00	0.80	1.05



	a	b	c	c	d
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
c	0.000	0.000	0.000	1.000	0.762
d	0.000	0.000	0.000	0.762	1.000

Matrice de variance - covariance

Matrice de corrélation

De retour au PGLS

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

$$\sigma^2 \mathbf{C} = \sigma^2 \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1i} \\ \rho_{21} & 1 & \rho_{23} & \dots & \rho_{2i} \\ \rho_{31} & \rho_{32} & 1 & \dots & \rho_{3i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{i1} & \rho_{i2} & \rho_{i3} & \dots & 1 \end{pmatrix}$$

Matrice de corrélation

Travaux pratiques dans R

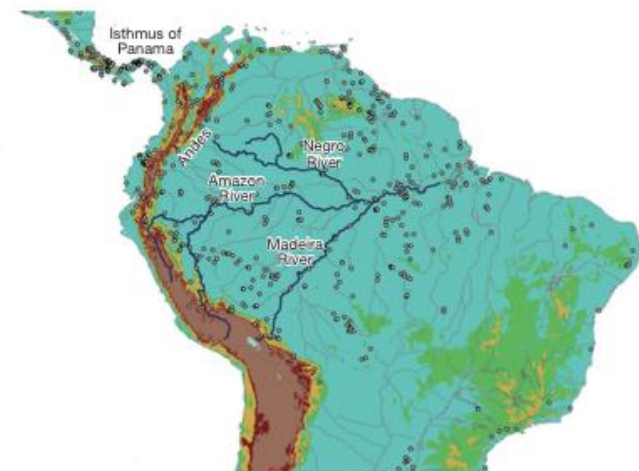
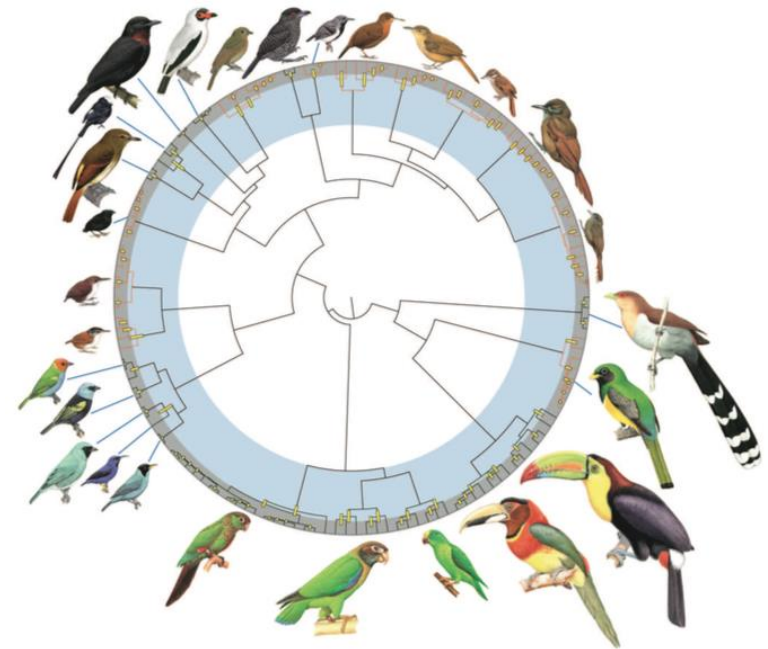


Jusqu'au défi #3

(Vous pouvez regarder la section suivante sur les contrastes indépendants)

Example

- Qu'est-ce qui est responsable de la diversification des oiseaux dans les néotropiques?
 - Age des lignées
 - Comportement de recherche de nourriture (sous-bois ou canopée)
 - Aire ancestrale
 - Taille de la niche
- 27 lignées d'oiseaux ont été étudiées



Example

- Qu'est-ce qui est responsable de la diversification des oiseaux dans les néotropiques?

Table 1 | Phylogenetic generalized least-squares regression showing the effects of historical and ecological variables on species diversity

Effect	Estimate	Standard error	<i>t</i> value	<i>P</i>	ΔAICc
Lineage age	0.1187	0.0283	4.1907	0.0004	6.9586
Foraging stratum	0.5188	0.2025	2.5623	0.0178	4.0122
Ancestral origin	−0.1921	0.2023	−0.9495	0.3527	−1.9546
Niche breadth	1.0097	1.0658	0.9473	0.3538	−1.9595

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

Le problème

- Que se passe-t-il s'il y a d'autres types d'erreurs dans les données que les corrélations phylogénétiques?
- Supposons par exemple qu'il y a des erreurs de mesures qui sont indépendantes de la phylogénie

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

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

La solution: utiliser le modèle λ

- Estimer le parameter λ dans le GLS. λ détermine la force du signal phylogénétique dans les résidus.
 - $\lambda = 0$: Aucun signal phylogénétique (= OLS)
 - $\lambda = 1$: Signal phylogénétique parfait

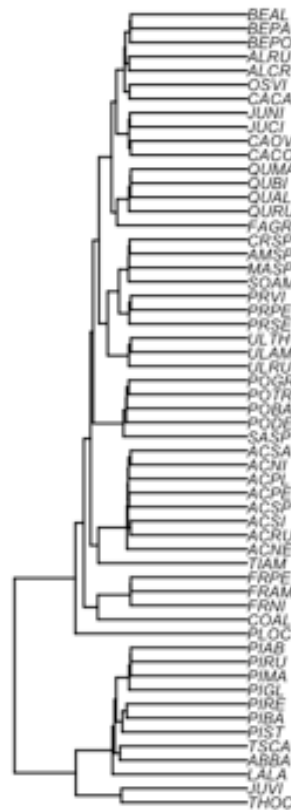
$$\sigma^2 \mathbf{C} = \sigma^2 \begin{pmatrix} 1 & \lambda\rho_{12} & \lambda\rho_{13} & \dots & \lambda\rho_{1i} \\ \lambda\rho_{21} & 1 & \lambda\rho_{23} & \dots & \lambda\rho_{2i} \\ \lambda\rho_{31} & \lambda\rho_{32} & 1 & \dots & \lambda\rho_{3i} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda\rho_{i1} & \lambda\rho_{i2} & \lambda\rho_{i3} & \dots & 1 \end{pmatrix}$$

Effet du paramètre λ sur la phylogénie

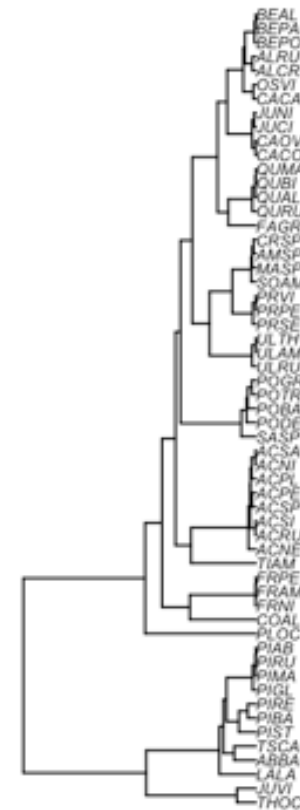
lambda=0.1



lambda=0.5



lambda=1



Exercices dans R



Jusqu'au défi #4

Optionel: regarder la prochaine section sur l'ANOVA phylogénétique

Sélection (ou test) de modèles

Concept du choix de modèle

Un modèle plus complexe aura toujours une meilleure vraisemblance (meilleure probabilité), mais l'estimation de ces paramètres auront une plus grande variance (et donc une plus grande incertitude).

Comment alors décider combien de paramètres ajouter?

Le choix d'un modèle repose généralement sur la notion de parcimonie, qui stipule qu'il faut choisir un modèle plus complexe seulement si c'est nécessaire.

Likelihood Ratio Test

$$D = -2 \ln \frac{L(D|M_0)}{L(D|M_1)}$$

$$D = -2(\ln L(D|M_0) - \ln L(D|M_1))$$

La distribution de D peut-être évalué à l'aide d'une distribution χ^2 avec x degrés de libertés, où le nombre de degré de liberté correspond à la difference en nombre de paramètres entre les deux modèles.

M_0 est le modèle null (plus simple) et M_1 est le modèle plus complexe.

Aikake Information Criterion (AIC)

$$AIC = -2\ln L(D|M) + 2k$$

où k est le nombre de paramètres dans le modèle.

Le meilleur modèle est celui avec le plus petit AIC.

Important de savoir!

- Pour comparer des modèles avec différentes structures de corrélation (e.g., OLS vs. PGLS), il faut ajuster les modèles avec le restricted maximum likelihood (REML) plutôt que le maximum likelihood (ML).
- Pour comparer les variables, il faut ajuster le modèles avec ML (e.g., $\text{shade} \sim \text{Wd}$ vs. $\text{Shade} \sim 1$).
- Finalement, avec les PGLS, il faut rapporter les résultats avec le modèle ajusté par REML

Exercices dans R

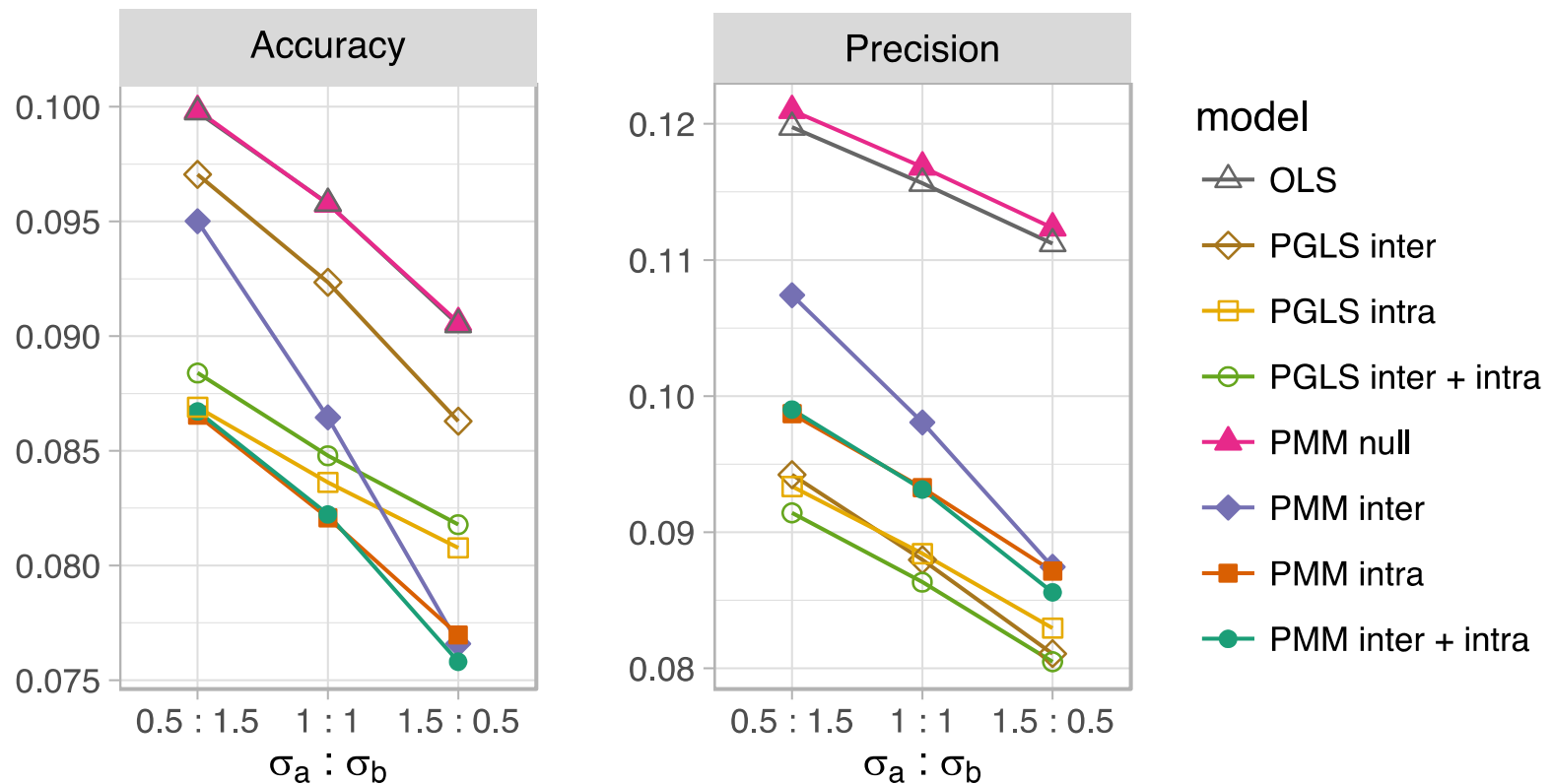


Voir les exercices sur le choix de modèle

Pourquoi utiliser les PGLS?

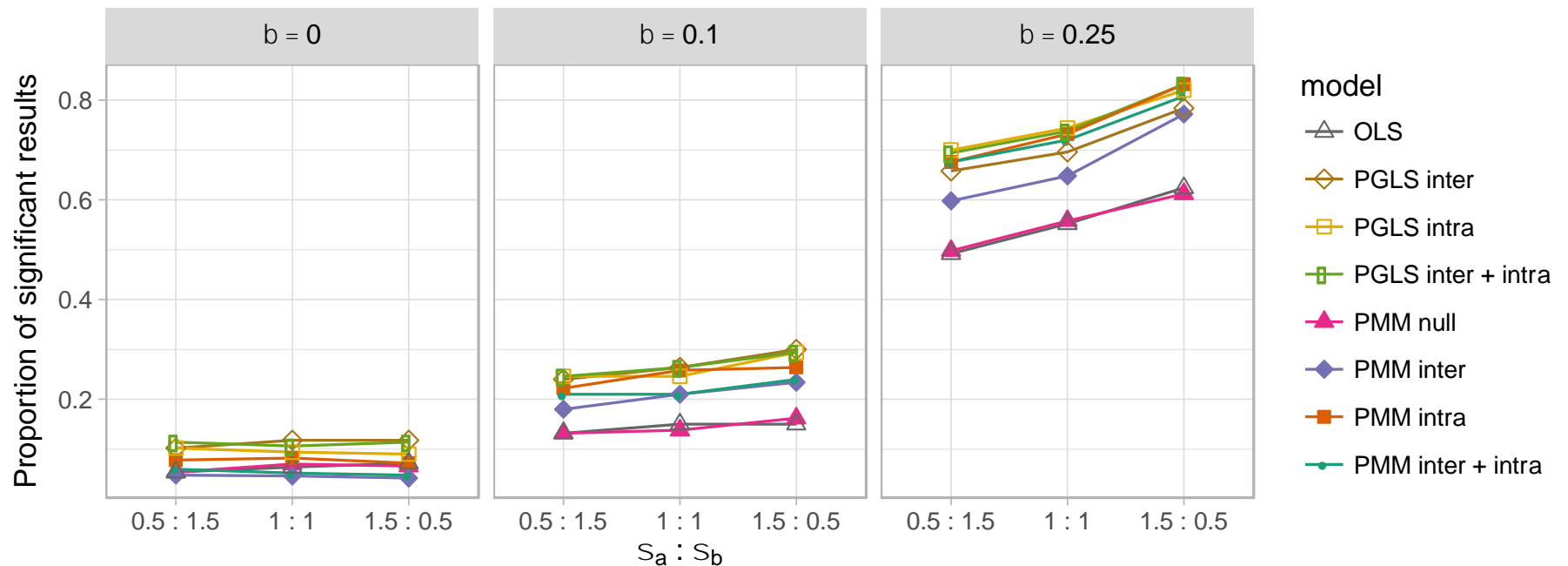
Pourquoi utiliser les PGLS?

- Meilleure précision et justesse



Pourquoi utiliser les PGLS?

- Erreur de type I plus basse et plus grande puissance



Quand utiliser les méthodes comparées?

Erreur commune

On devrait utiliser les PGLS seulement quand les caractères ont un signal phylogénétique

- Deux traits peuvent n'avoir aucun signal phylogénétique, mais leur corrélation pourrait être corrélée à la phylogénie !
- PGLS corrige les corrélations phylogénétiques dans les résidus, pas dans les traits.

Meilleure approche

- Toujours utiliser un PGLS avec le modèle λ .
 - Plus de puissance
 - Moins d'erreur de type I
 - Estimés plus précis

Quand utiliser les méthodes comparées?

Toujours!

(Quand vous étudiez plusieurs espèces)

Éviter l'approche des groupes frères

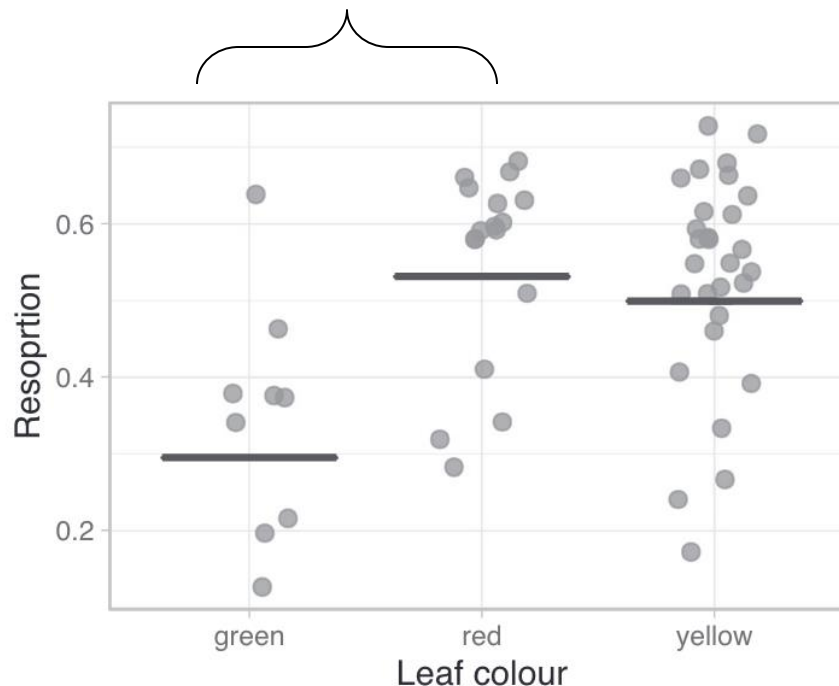
- Ils réduisent la puissance du test en supprimant beaucoup de données !

L'hypothèse de photoprotection

Les pigments rouges des feuilles à l'automne protègent du soleil et permettent une meilleure résorption

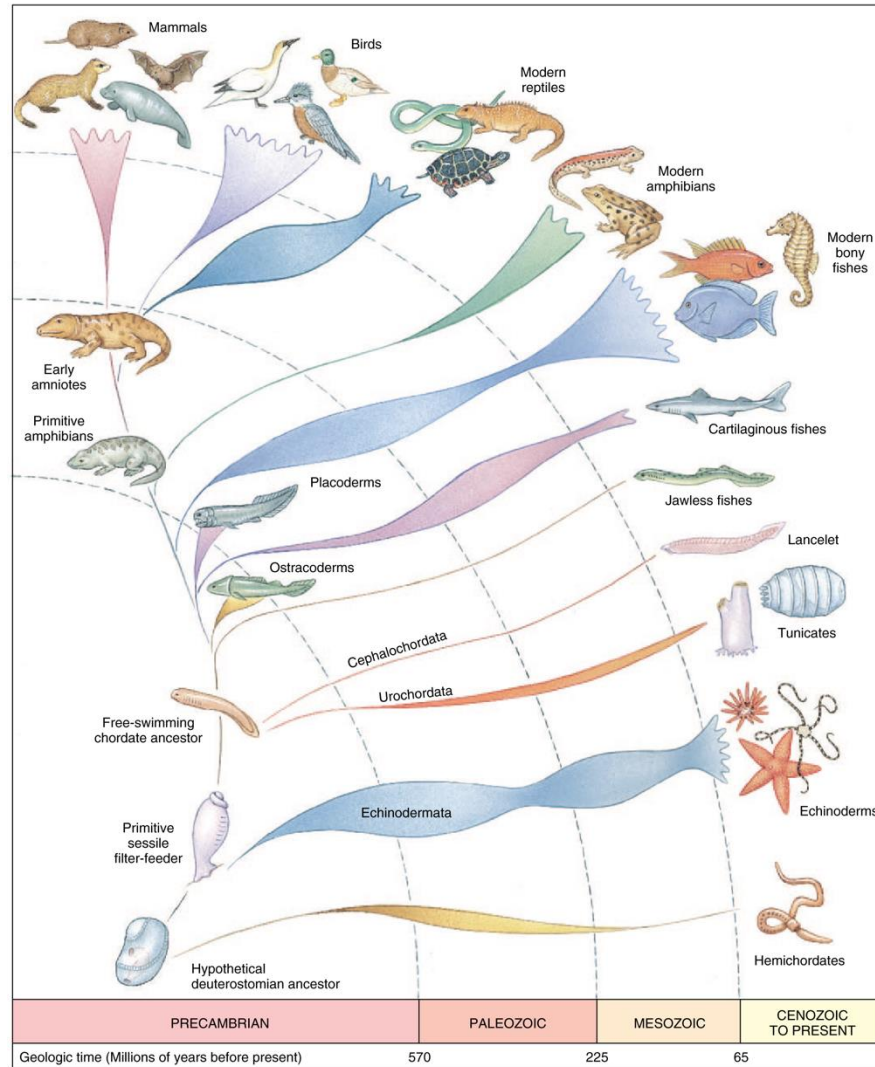
Sister species approach : $p > 0.05$

PGLS : $p = 0.0009$

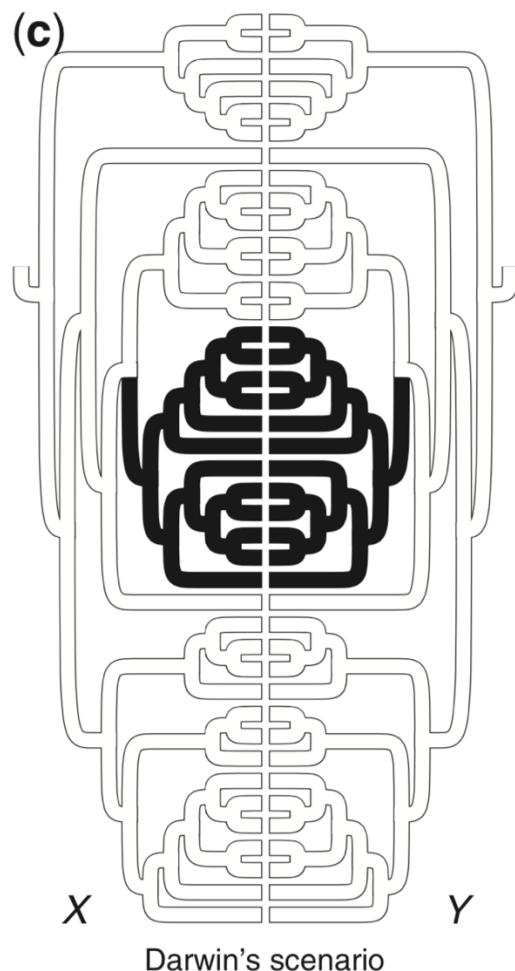


Le problème de la réplication

Replication dans les études évolutives



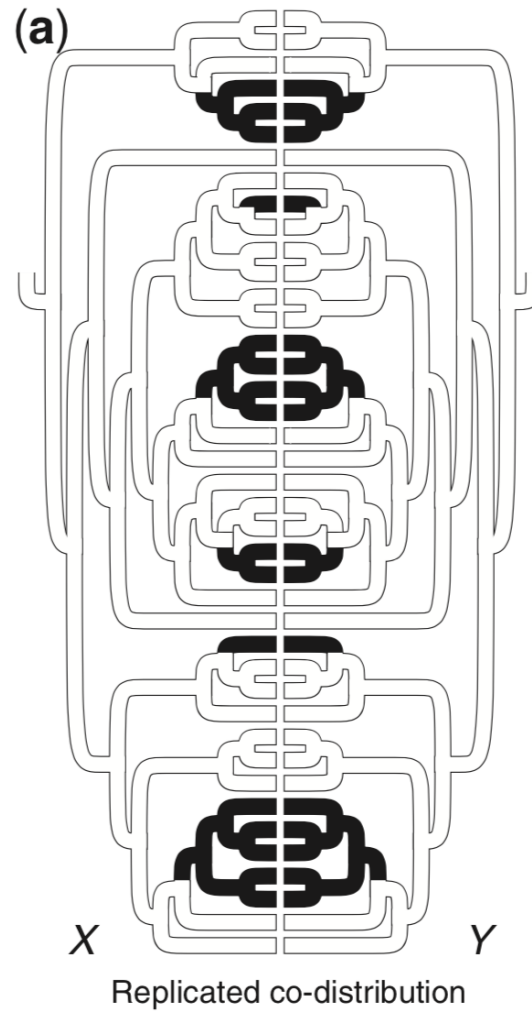
Aucune répllication (Darwin's scenario)



“We may often falsely attribute to correlated variation structures which are common to whole groups of species, and which in truth are simply due to inheritance; for an ancient progenitor may have acquired through natural selection some one modification in structure, and, after thousands of generations, some other and independent modification; and these two modifications, having been transmitted to a whole group of descendants with diverse habits, would naturally be thought to be in some necessary manner correlated.”

Darwin (1872)

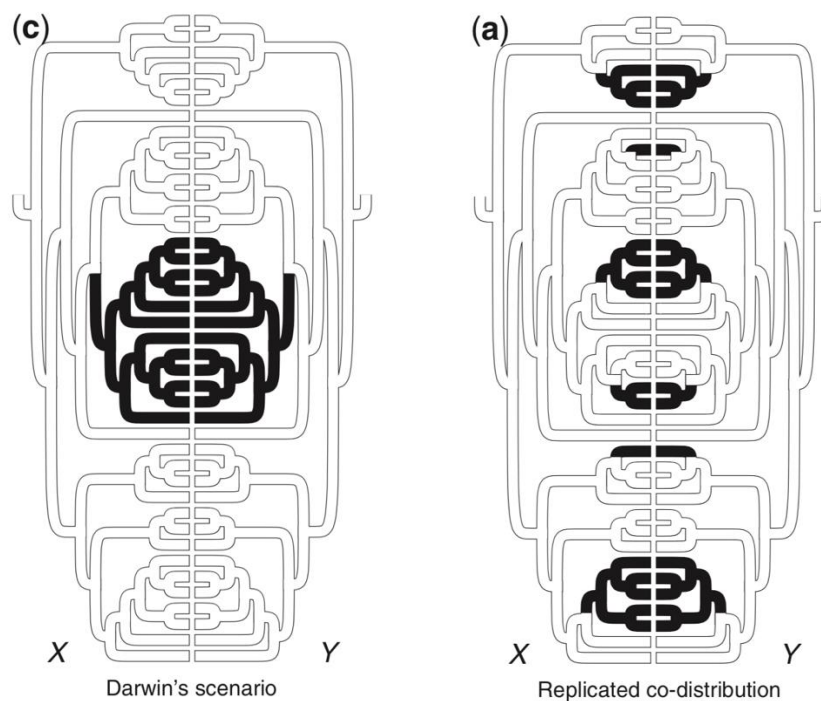
True replication



Attention ...

Les méthodes comparées sont actuellement très mauvaises pour expliquer un tel manque de réplication... Elles peuvent parfois donner des résultats très significatifs même si un caractère n'est pas répliqué.

C'est le travail de l'utilisateur de s'assurer que l'interprétation est correcte !



Nous devons changer nos perspectives!

