

Méthodes Phylogénétiques Comparatives

Mahendra Mariadassou (et Paul Bastide)

INRAE - MAIAGE

Journée de rentrée de l'EDMH
IHES, 18 Octobre 2021

1

Motivation

- Un exemple simple
- Un autre exemple
- Un dernier exemple (thèse de P. Bastide)
- Message Principal

2

Traits Continus

- Mouvement Brownien (BM)
- Mouvement brownien multivarié
- Corrélation / Régression Phylogénétique
- Vers le mouvement brownien et au delà

3

Traits Discrets

- Modèle univarié
- Traits multivariés

4

Exemples d'application

5

Résumé

- 1 **Motivation**
 - **Un exemple simple**
 - Un autre exemple
 - Un dernier exemple (thèse de P. Bastide)
 - Message Principal

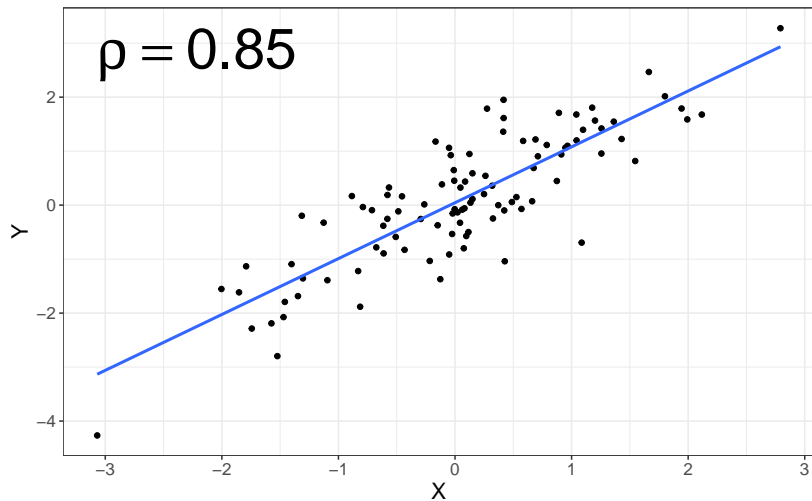
- 2 Traits Continus

- 3 Traits Discrets

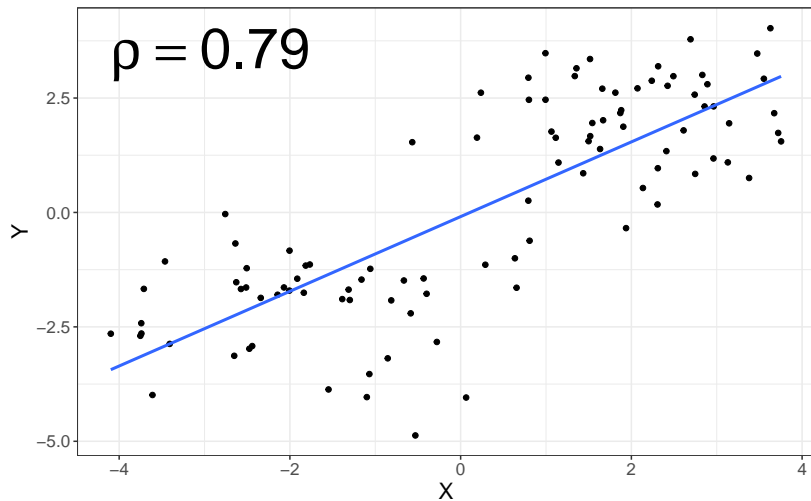
- 4 Exemples d'application

- 5 Résumé

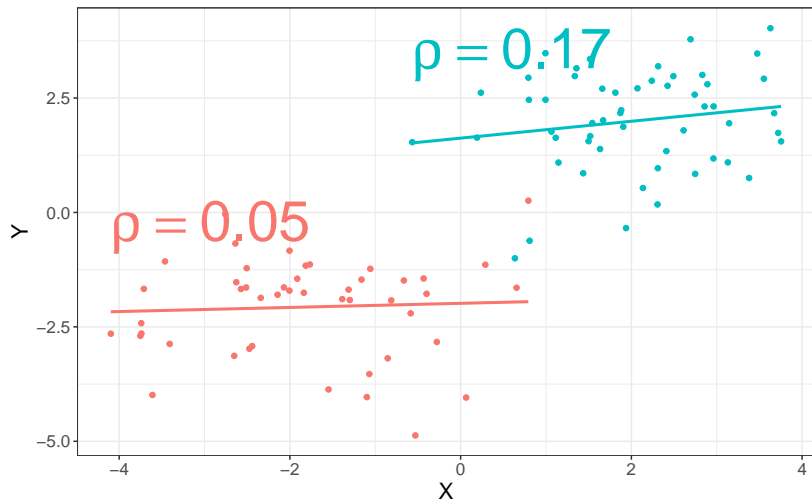
Comparer des quantités



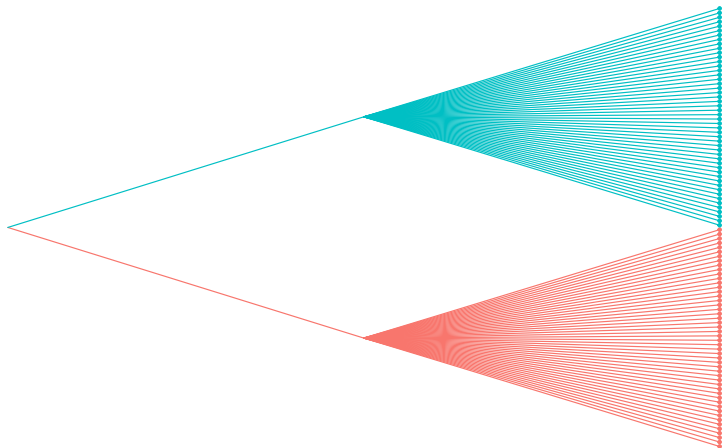
Comparer des quantités (suite)



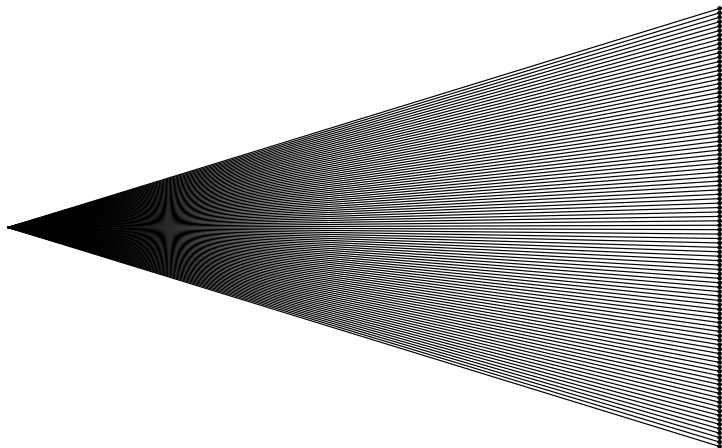
Comparer des quantités (suite)



Le "pire cas" de Felsenstein



Ce que vous aviez en tête



- 1 Motivation
 - Un exemple simple
 - **Un autre exemple**
 - Un dernier exemple (thèse de P. Bastide)
 - Message Principal

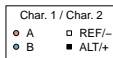
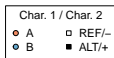
2 Traits Continus

3 Traits Discrets

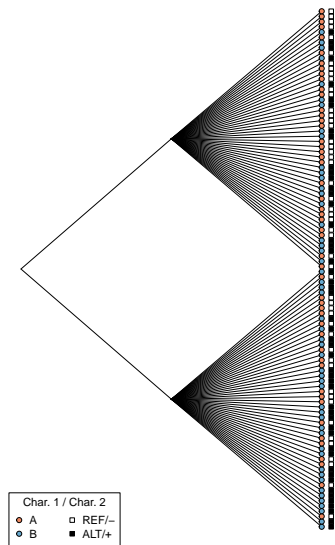
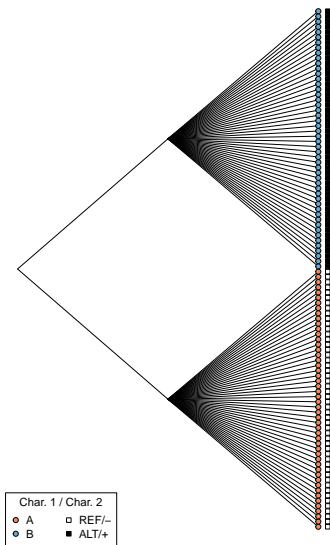
4 Exemples d'application

5 Résumé

Comparaison de variables binaires



Comparaison de variables binaires (suite)



- 1 Motivation
 - Un exemple simple
 - Un autre exemple
 - **Un dernier exemple (thèse de P. Bastide)**
 - Message Principal

2 Traits Continus

3 Traits Discrets

4 Exemples d'application

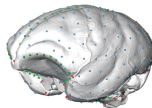
5 Résumé



Callithrix penicillata

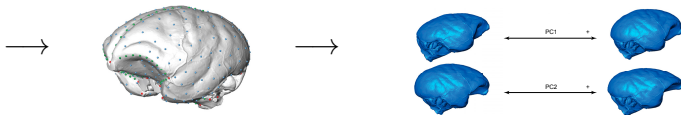


Callithrix penicillata



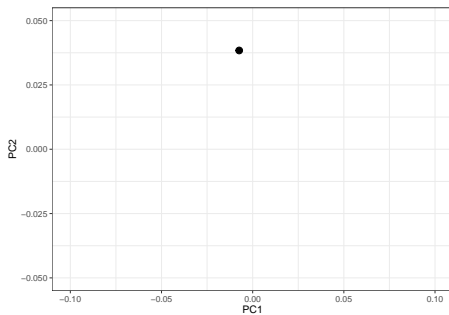
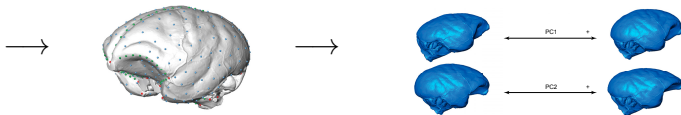


Callithrix penicillata



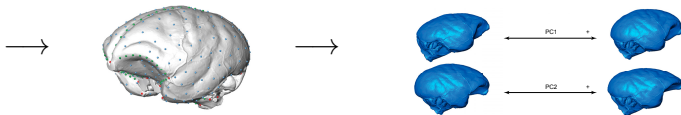


Callithrix penicillata

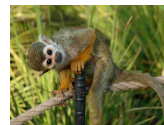
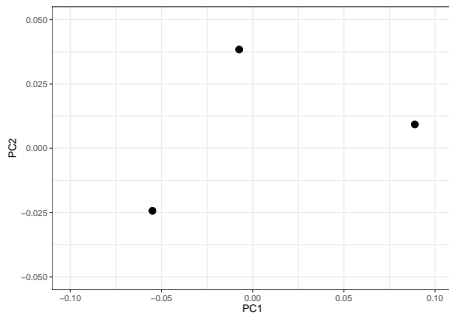




Callithrix penicillata



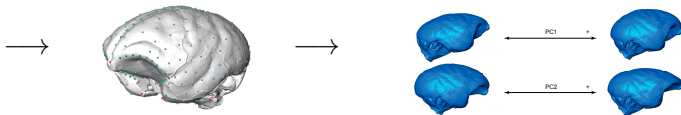
Alouatta palliata



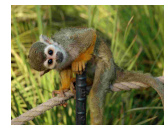
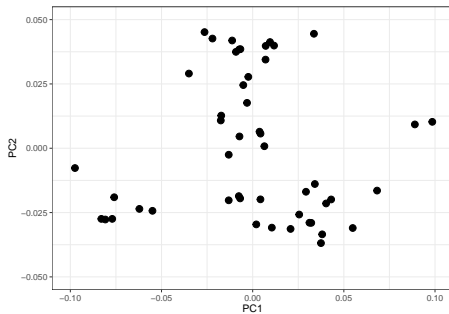
Saimiri sciureus



Callithrix penicillata



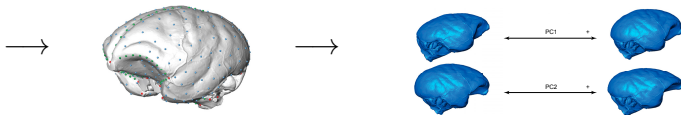
Alouatta palliata



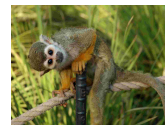
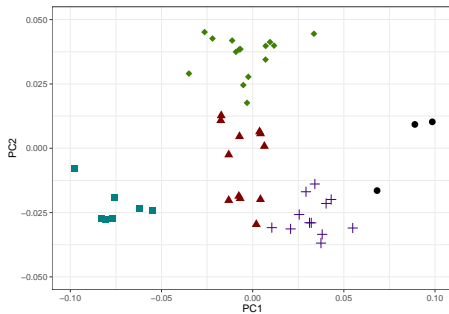
Saimiri sciureus



Callithrix penicillata



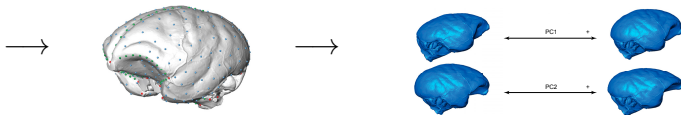
Alouatta palliata



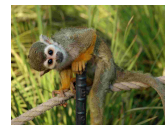
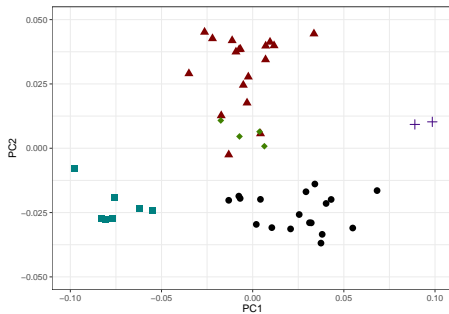
Saimiri sciureus



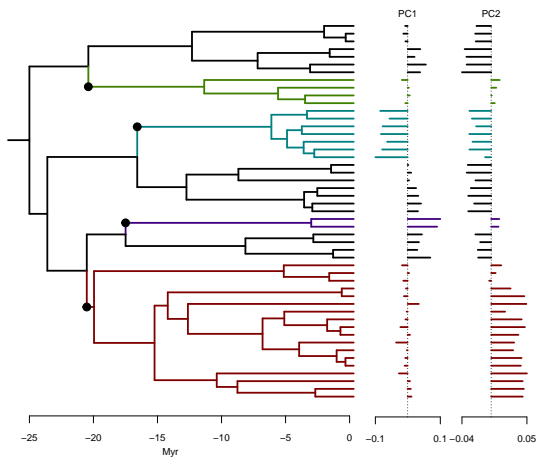
Callithrix penicillata



Alouatta palliata



Saimiri sciureus



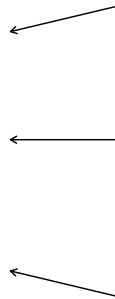
Alouatta palliata



Saimiri sciureus



Callithrix penicillata



- 1 Motivation
 - Un exemple simple
 - Un autre exemple
 - Un dernier exemple (thèse de P. Bastide)
 - **Message Principal**

2 Traits Continus

3 Traits Discrets

4 Exemples d'application

5 Résumé

L'évolution compte!!

L'évolution compte!!

Modéliser l'évolution

- Attention lorsqu'on compare des traits mesurés sur des **organismes apparentés**
- Si les traits ne répondent pas instantanément à la sélection naturelle, il y a une **inertie phylogénétique**
- Nécessité de **modéliser** l'évolution du trait le long de l'arbre

1

Motivation

- Un exemple simple
- Un autre exemple
- Un dernier exemple (thèse de P. Bastide)
- Message Principal

2

Traits Continus

- **Mouvement Brownien (BM)**
- **Mouvement brownien multivarié**
- **Corrélation / Régression Phylogénétique**
- **Vers le mouvement brownien et au delà**

3

Traits Discrets

- Modèle univarié
- Traits multivariés

4

Exemples d'application

5

Résumé

1 Motivation

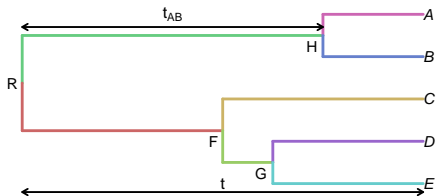
2 **Traits Continus**

- **Mouvement Brownien (BM)**
- Mouvement brownien multivarié
- Corrélacion / Régression Phylogénétique
- Vers le mouvement brownien et au delà

3 Traits Discrets

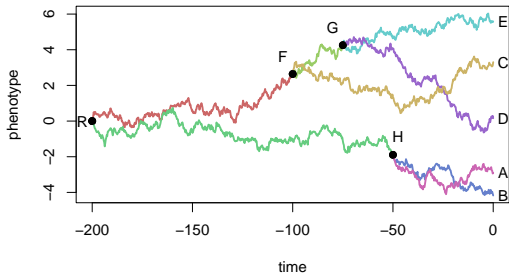
4 Exemples d'application

5 Résumé



L'arbre est **connu**.

Seules les **feuilles** sont observées.



Processus décrit sur **une** branche et **dupliqué** à chaque noeud.

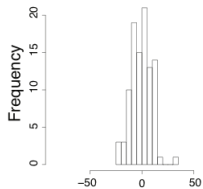
Le MB est la solution de l'équation différentielle stochastique:

$$dX(t) = \sigma dB(t)$$

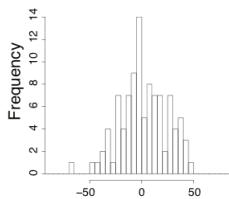
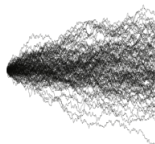
et vérifie:

- $E[X(0)] = \mu$
- Les incréments de X sont indépendents
- $X(t) - X(0) \sim \mathcal{N}(0, \sigma^2 t)$

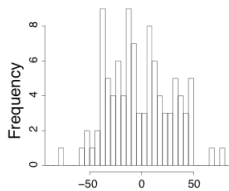
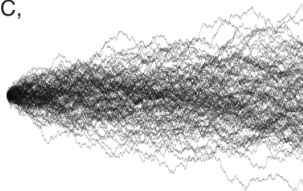
A,



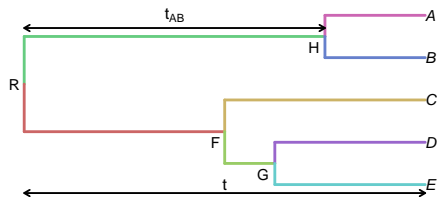
B,



C,



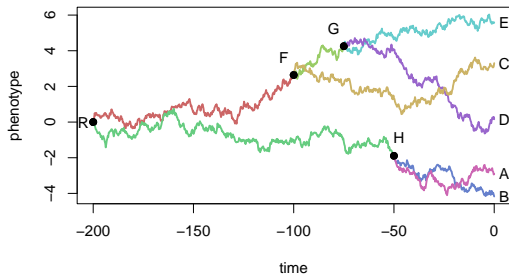
Corrélation Phylogénétique



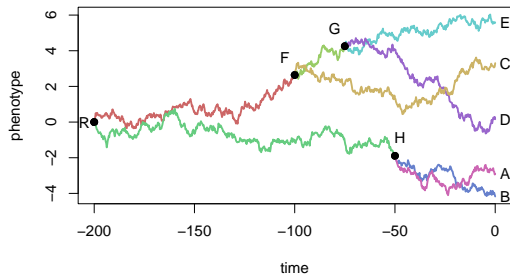
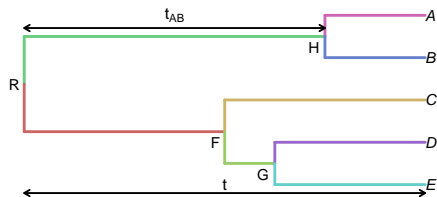
$$\text{Var}[A | R] = \sigma^2 t$$

$$A - R = (A - H) + (H - R)$$

$$B - R = (B - H) + (H - R)$$



Corrélation Phylogénétique



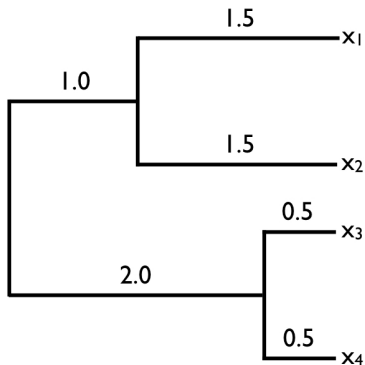
$$\text{Var}[A | R] = \sigma^2 t$$

$$A - R = (A - H) + (H - R)$$

$$B - R = (B - H) + (H - R)$$

$$\text{Cov}[A; B | R] = \sigma^2 t_{AB}$$

Corrélation Phylogénétique



$$C = \begin{bmatrix} 2.5 & 1.0 & 0 & 0 \\ 1.0 & 2.5 & 0 & 0 \\ 0 & 0 & 2.5 & 2.0 \\ 0 & 0 & 2.0 & 2.5 \end{bmatrix}$$

Model

En notant \mathbf{Y} le vecteur de taille n des observations

$$\mathbf{Y} \sim \mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathbf{C})$$

avec \mathbf{C} complètement déterminée par l'arbre.

Model

En notant \mathbf{Y} le vecteur de taille n des observations

$$\mathbf{Y} \sim \mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathbf{C})$$

avec \mathbf{C} complètement déterminée par l'arbre.

Vraisemblance

$$L(\mathbf{Y}; \mu, \sigma^2, \mathbf{C}) = -\frac{(\mathbf{Y} - \mu \mathbf{1}_n)^\top \mathbf{C}^{-1} (\mathbf{Y} - \mu \mathbf{1}_n)}{2\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\mathbf{C}|$$

Model

En notant \mathbf{Y} le vecteur de taille n des observations

$$\mathbf{Y} \sim \mathcal{N}(\mu \mathbf{1}_n, \sigma^2 \mathbf{C})$$

avec \mathbf{C} complètement déterminée par l'arbre.

Vraisemblance

$$L(\mathbf{Y}; \mu, \sigma^2, \mathbf{C}) = -\frac{(\mathbf{Y} - \mu \mathbf{1}_n)^\top \mathbf{C}^{-1} (\mathbf{Y} - \mu \mathbf{1}_n)}{2\sigma^2} - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2} \log |\mathbf{C}|$$

Estimateurs

$$\hat{\mu} = (\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1})^{-1} (\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{Y})$$
$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \hat{\mu} \mathbf{1}_n)^\top \mathbf{C}^{-1} (\mathbf{Y} - \hat{\mu} \mathbf{1}_n)}{n}$$

- $\mathbf{1}^\top \mathbf{C}^{-1}$ agit comme un vecteur de **poids**
- $\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1}$ agit comme une **taille d'échantillon** efficace
- Donc
 - $\hat{\mu}$ est une moyenne pondérée des Y_i
 - $\hat{\sigma}^2$ est la norme de de $(\mathbf{Y} - \hat{\mathbf{Y}})$ sous la métrique induite par \mathbf{C} .

- On sait modéliser un trait sur un arbre :)

- On sait modéliser un trait sur un arbre :)
- Mais on ne sait pas encore comparer deux traits :(

1 Motivation

2 Traits Continus

- Mouvement Brownien (BM)
- **Mouvement brownien multivarié**
- Corrélation / Régression Phylogénétique
- Vers le mouvement brownien et au delà

3 Traits Discrets

4 Exemples d'application

5 Résumé

Principe

- Modéliser l'évolution **jointe** de plusieurs traits sur **une** branche...
- En utilisant une gaussienne multivariée avec matrice de taux \mathbf{R} .
- Avant d'ajouter une structure phylogénétique.

Principe

- Modéliser l'évolution **jointe** de plusieurs traits sur **une** branche...
- En utilisant une gaussienne multivariée avec matrice de taux \mathbf{R}
- Avant d'ajouter une structure phylogénétique.

Principe

- \mathbf{C} capture la variance entre *organismes* due à leur évolution partagée
- \mathbf{R} capture la variance entre *traits* due à la coévolution

Le mvBM est solution de l'équation différentielle stochastique:

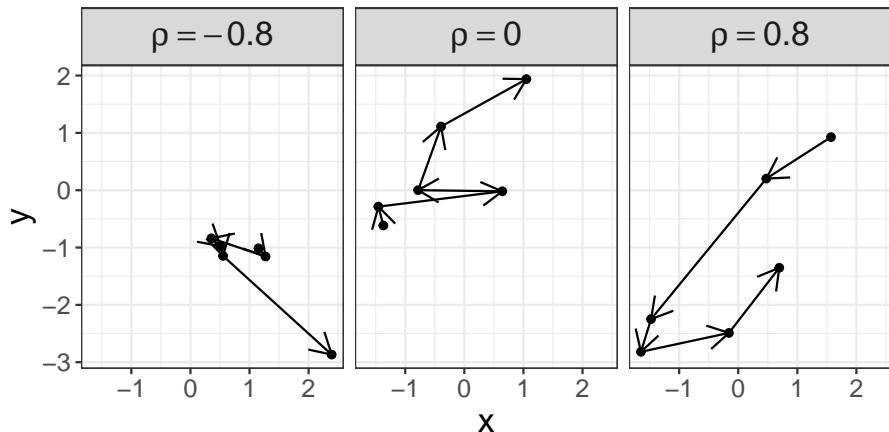
$$d\mathbf{X}(t) = \mathbf{R}^{1/2}\sigma d\mathbf{B}(t)$$

et vérifie:

- $E[\mathbf{X}(0)] = \boldsymbol{\mu}$
- Les incréments de \mathbf{X} sont indépendants
- $\mathbf{X}(t) - \mathbf{X}(0) \sim \mathcal{N}(\mathbf{0}, t\mathbf{R})$

Exemple: un brownien bivarié

Considérons $\mathbf{R} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$



Y est désormais une matrice $n \times p$ (n organismes $\times p$ traits) vérifiant

$$\text{Cov}(Y_{ik}, Y_{jl}) = R_{kl} \times C_{ij}$$

\mathbf{Y} est désormais une matrice $n \times p$ (n organismes $\times p$ traits) vérifiant

$$\text{Cov}(Y_{ik}, Y_{jl}) = R_{kl} \times C_{ij}$$

La covariance se factorise en un produit d'une composante **phylogénétique** (C_{ij}) et d'une composante **phénotypique** (R_{kl}).

\mathbf{Y} est désormais une matrice $n \times p$ (n organismes $\times p$ traits) vérifiant

$$\text{Cov}(Y_{ik}, Y_{jl}) = R_{kl} \times C_{ij}$$

La covariance se factorise en un produit d'une composante **phylogénétique** (C_{ij}) et d'une composante **phénotypique** (R_{kl}).

En particulier

$$\mathbf{V} = \text{Var}(\mathbf{Y}) = \mathbf{R} \otimes \mathbf{C}$$

\mathbf{Y} est désormais une matrice $n \times p$ (n organismes $\times p$ traits) vérifiant

$$\text{Cov}(Y_{ik}, Y_{jl}) = R_{kl} \times C_{ij}$$

La covariance se factorise en un produit d'une composante **phylogénétique** (C_{ij}) et d'une composante **phénotypique** (R_{kl}).

En particulier

$$\mathbf{V} = \text{Var}(\mathbf{Y}) = \mathbf{R} \otimes \mathbf{C}$$

\mathbf{V} est de taille $np \times np$ et capture la covariance de tous les traits au sein de toutes les espèces.

Modèle

Notons \mathbf{Y} le vecteur de taille $n \times p$ des observations (obtenu en empilant les colonnes de \mathbf{Y})

$$\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{R} \otimes \mathbf{C})$$

où \mathbf{C} est complètement déterminée par l'arbre.

Modèle

Notons \mathbf{Y} le vecteur de taille $n \times p$ des observations (obtenu en empilant les colonnes de \mathbf{Y})

$$\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{R} \otimes \mathbf{C})$$

où \mathbf{C} est complètement déterminée par l'arbre.

Vraisemblance

$$L(\mathbf{Y}; \boldsymbol{\mu}, \mathbf{R}, \mathbf{C}) = -\frac{(\mathbf{Y} - \boldsymbol{\mu} \otimes \mathbf{1}_n)^\top \mathbf{R}^{-1} \otimes \mathbf{C}^{-1} (\mathbf{Y} - \boldsymbol{\mu} \otimes \mathbf{1}_n)}{2} - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{R} \otimes \mathbf{C}|$$

Modèle

Notons \mathbf{Y} le vecteur de taille $n \times p$ des observations (obtenu en empilant les colonnes de \mathbf{Y})

$$\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{R} \otimes \mathbf{C})$$

où \mathbf{C} est complètement déterminée par l'arbre.

Vraisemblance

$$L(\mathbf{Y}; \boldsymbol{\mu}, \mathbf{R}, \mathbf{C}) = -\frac{(\mathbf{Y} - \boldsymbol{\mu} \otimes \mathbf{1}_n)^\top \mathbf{R}^{-1} \otimes \mathbf{C}^{-1} (\mathbf{Y} - \boldsymbol{\mu} \otimes \mathbf{1}_n)}{2} - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{R} \otimes \mathbf{C}|$$

Estimateurs

$$\hat{\boldsymbol{\mu}} = (\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{1})^{-1} (\mathbf{1}^\top \mathbf{C}^{-1} \mathbf{Y})^\top$$
$$\hat{\mathbf{R}} = \frac{(\mathbf{Y} - \mathbf{1}_n \hat{\boldsymbol{\mu}}^\top)^\top \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{1}_n \hat{\boldsymbol{\mu}}^\top)}{n}$$

1 Motivation

2 Traits Continus

- Mouvement Brownien (BM)
- Mouvement brownien multivarié
- **Corrélation / Régression Phylogénétique**
- Vers le mouvement brownien et au delà

3 Traits Discrets

4 Exemples d'application

5 Résumé

\hat{R}_{kl} est la **corrélation évolutive** estimée entre les traits k et l

\hat{R}_{kl} est la **corrélation évolutive** estimée entre les traits k et l

On peut donc revenir aux statistiques classiques et tester

- $R_{kl} = 0$ contre
- $R_{kl} \neq 0$

en utilisant par exemple un test de rapport de vraisemblance ou en construisant un intervalle de confiance autour de \hat{R}_{kl}

Pourquoi s'arrêter à la corrélation?

- La corrélation est **symétrique**

Pourquoi s'arrêter à la corrélation?

- La corrélation est **symétrique**
- Pour calculer l'effet d'un trait sur un autre, on utilise plutôt une **régression**

Pourquoi s'arrêter à la corrélation?

- La corrélation est **symétrique**
- Pour calculer l'effet d'un trait sur un autre, on utilise plutôt une **régression**
- Construisons une régression phylogénétique!!

Régression Phylogénétique

Pourquoi s'arrêter à la corrélation?

- La corrélation est **symétrique**
- Pour calculer l'effet d'un trait sur un autre, on utilise plutôt une **régression**
- Construisons une régression phylogénétique!!

Régression Phylogénétique

On considère le modèle de régression suivant:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{E} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\theta}, \sigma^2\mathbf{C})$$

où on suppose que:

- \mathbf{Y} a une structure phylogénétique;
- \mathbf{E} a une structure phylogénétique;
- \mathbf{X} peut avoir ou non une structure phylogénétique.

Estimateurs habituels

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{Y})^T$$
$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}})^T \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}})}{n}$$

Estimateurs habituels

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{C}^{-1} \mathbf{Y})^T$$
$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T \mathbf{C}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})}{n}$$

Remarques

- La régression phylogénétique est un cas particuliers des moindres carrés généralisés;
- Beaucoup de travaux sur les modèles linéaires peuvent être importés dans ce cadre.

1 Motivation

2 Traits Continus

- Mouvement Brownien (BM)
- Mouvement brownien multivarié
- Corrélacion / Régression Phylogénétique
- **Vers le mouvement brownien et au delà**

3 Traits Discrets

4 Exemples d'application

5 Résumé

Extensions

- Remplacer BM par un processus de Ornstein-Uhlenbeck (rappel vers une valeur optimale)
- Remplacer BM par des processus de Lévy (évolution Simpsonian, par à coups)
- Ajouter des sauts discrets dans le processus (changement de niche)
- Remplacer σ^2 / \mathbf{R} par une fonction du temps (pour accélérer/ralentir l'évolution)
- ...

Extensions

- Remplacer BM par un processus de Ornstein-Uhlenbeck (rappel vers une valeur optimale)
- Remplacer BM par des processus de Lévy (évolution Simpsonian, par à coups)
- Ajouter des sauts discrets dans le processus (changement de niche)
- Remplacer σ^2 / \mathbf{R} par une fonction du temps (pour accélérer/ralentir l'évolution)
- ...

Difficultés

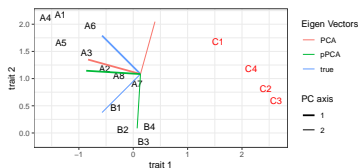
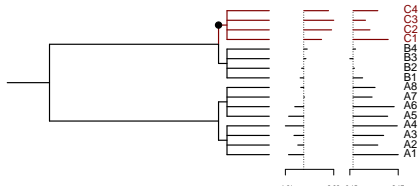
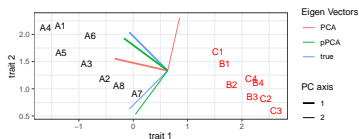
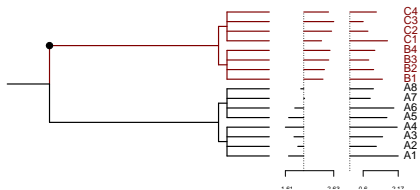
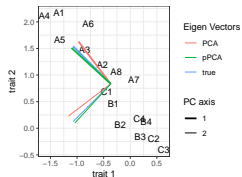
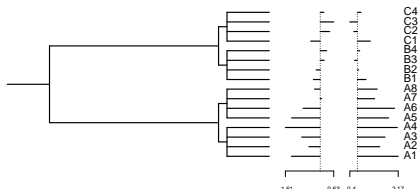
- Toutes ces extensions entraînent des calculs (bien) plus **difficiles**.

Extensions

- Remplacer BM par un processus de Ornstein-Uhlenbeck (rappel vers une valeur optimale)
- Remplacer BM par des processus de Lévy (évolution Simpsonian, par à coups)
- Ajouter des sauts discrets dans le processus (changement de niche)
- Remplacer σ^2 / \mathbf{R} par une fonction du temps (pour accélérer/ralentir l'évolution)
- ...

Difficultés

- Toutes ces extensions entraînent des calculs (bien) plus **difficiles**.
- Mais les négliger peut causer de **sérieuses** erreurs...



1

Motivation

- Un exemple simple
- Un autre exemple
- Un dernier exemple (thèse de P. Bastide)
- Message Principal

2

Traits Continus

- Mouvement Brownien (BM)
- Mouvement brownien multivarié
- Corrélation / Régression Phylogénétique
- Vers le mouvement brownien et au delà

3

Traits Discrets

- **Modèle univarié**
- **Traits multivariés**

4

Exemples d'application

5

Résumé

- 1 Motivation
- 2 Traits Continus
- 3 Traits Discrets**
 - **Modèle univarié**
 - Traits multivariés
- 4 Exemples d'application
- 5 Résumé

Modèle de Markov

- Un trait discret évolue selon un modèle de **Markov**
- La transition vers un état j dépend seulement de l'état **courant** i
- Si le taux de transition de i vers j est q_{ij} alors

$$\mathbb{P}[X(t + dt) = j \mid X(t) = i] \simeq q_{ij}dt$$

Modèle de Markov

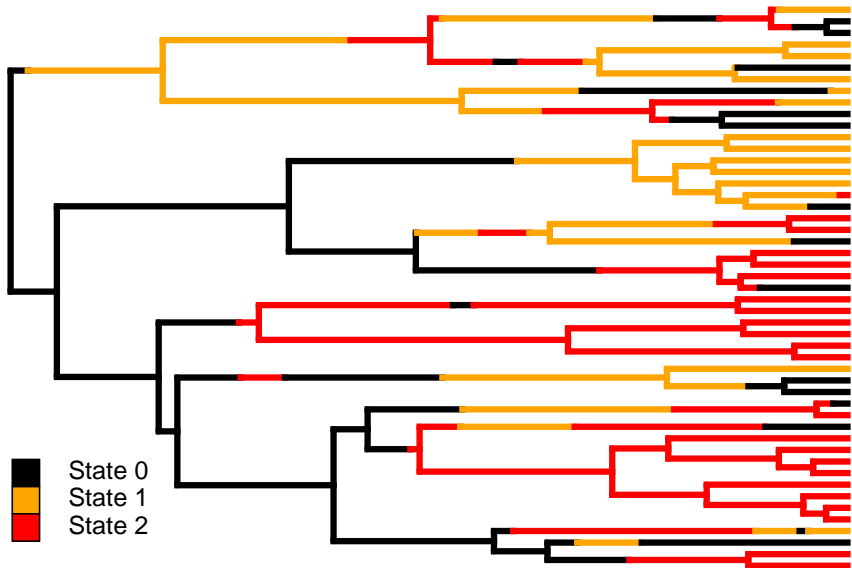
- Un trait discret évolue selon un modèle de **Markov**
- La transition vers un état j dépend seulement de l'état **courant** i
- Si le taux de transition de i vers j est q_{ij} alors

$$\mathbb{P}[X(t + dt) = j \mid X(t) = i] \simeq q_{ij}dt$$

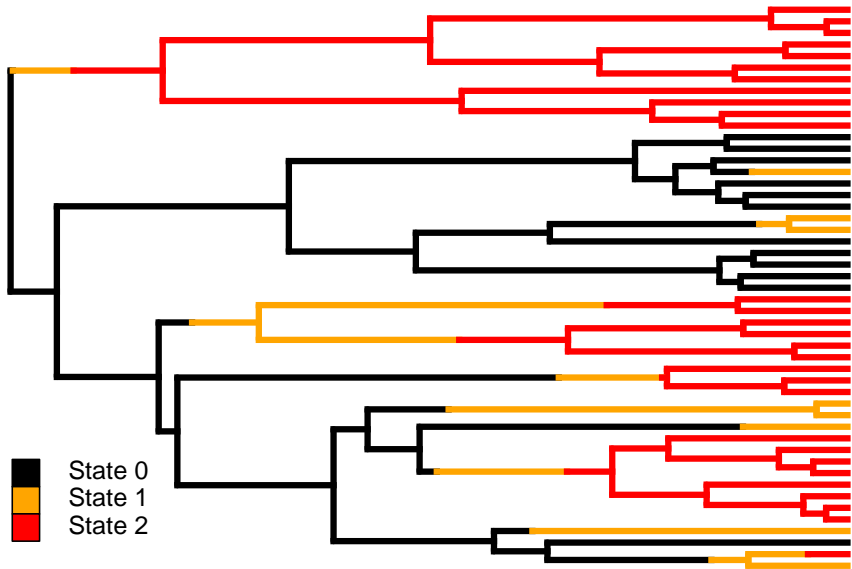
Matrice de taux

$$Q = \begin{bmatrix} \bullet & q_{12} & \dots & q_{1k} \\ q_{21} & \bullet & \dots & q_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & \dots & \bullet \end{bmatrix}$$

Un trait sur un arbre (I)



Un trait sur un arbre (II)



Quand le trait est connu aux deux extrémités

$$P_{ij}(t) = \mathbb{P}[X(t) = j \mid X(0) = i] = (e^{t\mathbf{Q}})_{ij}$$

Calculer $e^{t\mathbf{Q}}$ est coûteux!!

Quand le trait est connu aux deux extrémités

$$P_{ij}(t) = \mathbb{P}[X(t) = j \mid X(0) = i] = (e^{t\mathbf{Q}})_{ij}$$

Calculer $e^{t\mathbf{Q}}$ est coûteux!!

Autrement

$$P_{\bullet j}(t) = \mathbb{P}[X(t) = j] = \sum_{i=1}^k \mathbb{P}[X(0) = i] (e^{t\mathbf{Q}})_{ij}$$

Algorithme d'élagage

- Un exemple de programmation dynamique
- Basé sur une formule de récursion pour les **vraisemblances conditionnelles** sur sous-arbres: $L_N(i)$: la probabilité d'obtenir les observations sachant que le sous-arbre raciné en N est dans l'état i

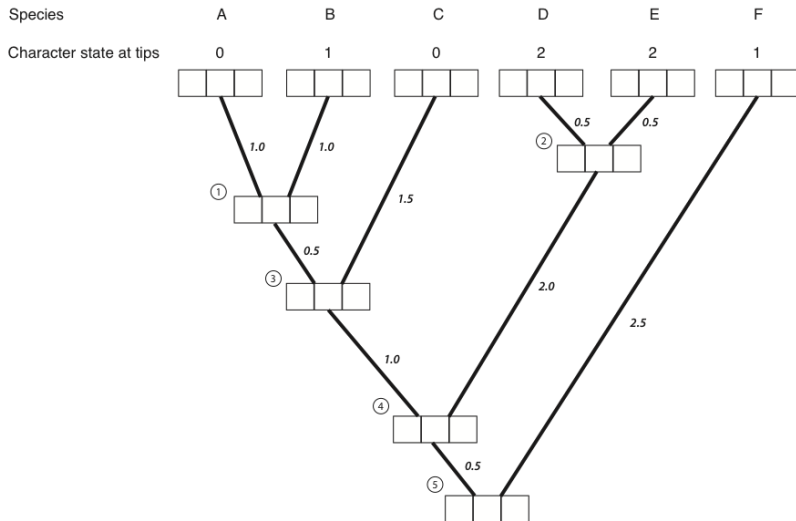
Algorithme d'élagage

- Un exemple de programmation dynamique
- Basé sur une formule de récursion pour les **vraisemblances conditionnelles** sur sous-arbres: $L_N(i)$: la probabilité d'obtenir les observations sachant que le sous-arbre raciné en N est dans l'état i

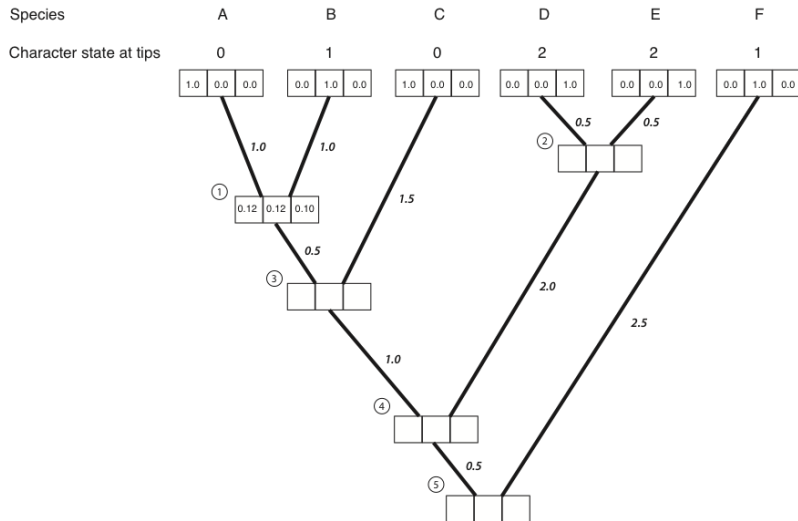
Formule de récurrence

$$L_N(i) = \left(\sum_x \mathbb{P}[X_L = x \mid X_P = i] L_L(x) \right) \times \left(\sum_y \mathbb{P}[X_R = y \mid X_P = i] L_R(y) \right)$$

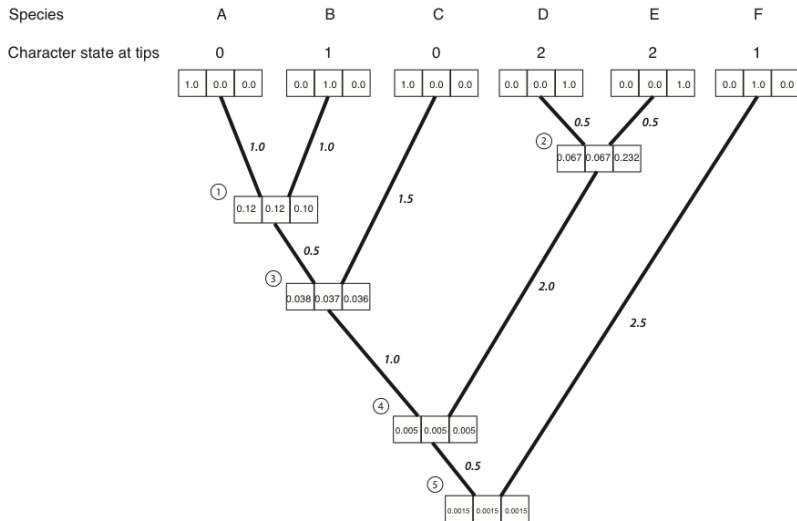
Vraisemblance sur un arbre



Vraisemblance sur un arbre



Vraisemblance sur un arbre



On peut calculer la vraisemblance et estimer (numériquement) Q

- 1 Motivation
- 2 Traits Continus
- 3 Traits Discrets**
 - Modèle univarié
 - Traits multivariés**
- 4 Exemples d'application
- 5 Résumé

Pour des traits **indépendants** X_1 et X_2 :

$$\mathbb{P}[X_1(t) = j, X_2(t) = l \mid X_1(0) = i, X_2(0) = k] = \\ \mathbb{P}[X_1(t) = j \mid X_1(0) = i] \times \mathbb{P}[X_2(t) = l \mid X_2(0) = k]$$

Traits indépendants

Soit

$$Q_A = \begin{matrix} & \begin{matrix} a & A \end{matrix} \\ \begin{bmatrix} \bullet & q_A \\ q_a & \bullet \end{bmatrix} & \begin{matrix} a \\ A \end{matrix} \end{matrix} \quad \text{et} \quad Q_B = \begin{matrix} & \begin{matrix} b & B \end{matrix} \\ \begin{bmatrix} \bullet & q_B \\ q_b & \bullet \end{bmatrix} & \begin{matrix} b \\ B \end{matrix} \end{matrix}$$

Alors

$$Q_{AB} = \begin{matrix} & \begin{matrix} ab & aB & Ab & AB \end{matrix} \\ \begin{bmatrix} \bullet & q_B & q_A & \cdot \\ q_b & \bullet & \cdot & q_A \\ q_a & \cdot & \bullet & q_B \\ \cdot & q_a & q_b & \bullet \end{bmatrix} & \begin{matrix} ab \\ aB \\ Ab \\ AB \end{matrix} \end{matrix}$$

Traits dépendants

En général, pour des traits dépendants, on n'a pas:

$$q_{Ab,AB} = q_{ab,aB} = q_B \quad \text{and} \quad q_{AB,Ab} = q_{aB,ab} = q_b$$

$$q_{ab,Ab} = q_{aB,AB} = q_A \quad \text{and} \quad q_{AB,aB} = q_{Ab,ab} = q_a$$

Traits dépendants

En général, pour des traits dépendants, on n'a pas:

$$q_{Ab,AB} = q_{ab,aB} = q_B \quad \text{and} \quad q_{AB,Ab} = q_{aB,ab} = q_b$$

$$q_{ab,Ab} = q_{aB,AB} = q_A \quad \text{and} \quad q_{AB,aB} = q_{Ab,ab} = q_a$$

Cas particulier: b/B dépend de a/A

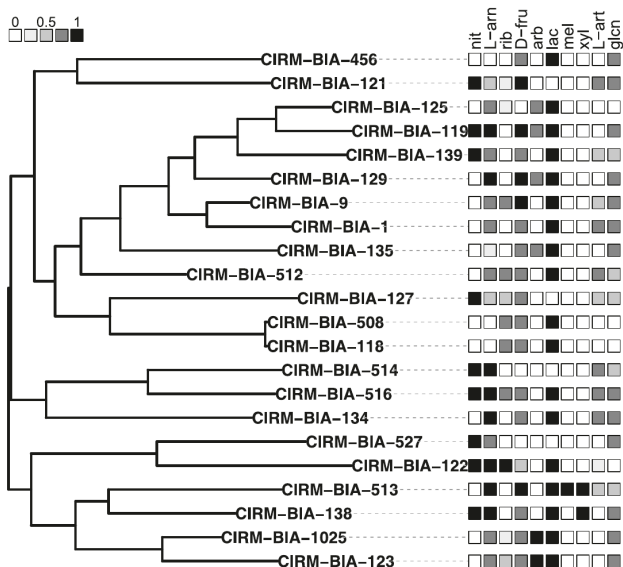
Si le trait b/B dépend du trait a/A

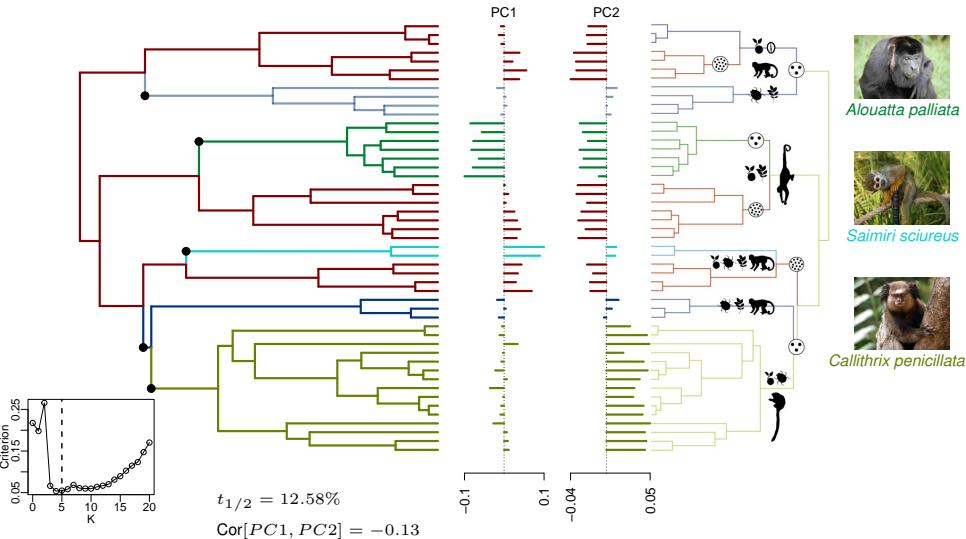
$$q_{Ab,AB} \neq q_{ab,aB} \quad \text{and/or} \quad q_{AB,Ab} \neq q_{aB,ab}$$

On peut tester la corrélation évolutive en testant

- H_0 : les taux vérifient les 4 égalités précédentes (indépendance)
- H_1 : Ils en violent certaines (dépendance)

En utilisant des tests de rapport de vraisemblance (4 paramètres sous H_0 , 8 en général sous H_1).





L'évolution est importante lorsqu'on compare des choses!!

L'évolution est importante lorsqu'on compare des choses!!

Modéliser l'évolution

- Beaucoup (beaucoup) de modèles existent pour la coévolution de traits discrets et/ou continus.
- Séparer la corrélation évolutive de la corrélation phylogénétique est possible (et souhaitable)...
- et permet d'éviter des erreurs d'interprétation.

References I

- Aristide, L., dos Reis, S. F., Machado, A. C., Lima, I., Lopes, R. T., and Perez, S. I. (2016). Brain shape convergence in the adaptive radiation of New World monkeys. *Proceedings of the National Academy of Sciences*, 113(8):2158–2163.
- Bastide, P., Ané, C., Robin, S., and Mariadassou, M. (2018). Inference of adaptive shifts for multivariate correlated traits. *Systematic Biology*, page syy005.
- Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist*, 125(1):1–15.
- Felsenstein, J. (1983). Statistical inference of phylogenies. *J.R. Statist. Soc.*, 146(3):246–272.
- Loux, V., Mariadassou, M., Almeida, S., Chiapello, H., Hammani, A., Buratti, J., Gendrault, A., Barbe, V., Aury, J.-M., Deutsch, S.-M., Parayre, S., Madec, M.-N., Chuat, V., Jan, G., Peterlongo, P., Azevedo, V., Le Loir, Y., and Falentin, H. (2015). Mutations and genomic islands can explain the strain dependency of sugar utilization in 21 strains of *propionibacterium freudenreichii*. *BMC Genomics*, 16(1):296.
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings: Biological Sciences*, 255(1342):37–45.

Photo Credits:

- Miguelrangeljr - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=28294644>
- Steven G. Johnson - Own work, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=4372777>
- Braboowi at the English language Wikipedia, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=7069103>
- Xiphophorus Genetic Stock Center, Texas State University, <http://www.xiphophorus.txstate.edu/resources/galleries/comprehensive.html>
- "Lonesome George in profile" by Mike Weston - Flickr: Lonesome George 2. Licensed under CC BY 2.0 via Wikimedia Commons