Modèles d'évolution de caractères continus

Paul Bastide, Mahendra Mariadassou, Stéphane Robin

Modèles d'évolution de caractères continus

Paul Bastide[1], Mahendra Mariadassou[2], and Stéphane Robin[3]

[1] *IMAG, Université de Montpellier, CNRS, Montpellier, France.*
[2] *MaIAGE, INRAE, Université Paris-Saclay, Jouy-en-Josas, France.*
[3] *LPSM, Sorbonne Université, Paris, France.*
Contacts : paul.bastide@umontpellier.fr ; mahendra.mariadassou@inra.fr ; stephane.robin@sorbonne-universite.fr.

Résumé

Ce document est un chapitre d'ouvrage, présentant les modèles d'évolution de caractères continus, dans le cadre général des méthodes phylogénétiques comparatives. Partant du mouvement brownien, univarié puis multivarié, nous introduisons les modèles gaussiens et leurs extensions utilisés dans la littérature, en écologie évolutive ou en phylodynamique. Le chapitre se conclue par un tableau récapitulatif des logiciels R permettant de mettre en œuvre ces modèles.

Il est paru sous les références suivantes :
Bastide, Mariadassou, Robin. 2022. Modèles d'évolution de caractères continu. In : Didier, Guindon, éditeurs. *Modèles et méthodes pour l'évolution biologique*, pages 47–85. ISTE Group. https://doi.org/10.51926/ISTE.9069.ch3.

Mots clefs— modèle d'évolution, traits quantitatifs, méthodes phylogénétiques comparatives, mouvement brownien, processus d'Ornstein-Uhlenbeck, calcul de vraisemblance

Table des matières

1 Motivations

Les modèles d'évolution présentés dans le chapitre précédent sont principalement motivés par l'étude de macromolécules biologiques : séquences d'ADN, d'ARN ou de protéines. Ces modèles ont tous en commun d'utiliser un espace d'état, ou alphabet, de faible cardinal : 4 nucléotides pour l'ADN et l'ARN et 20 acides aminés pour les protéines. Il est cependant fréquent d'observer ou de mesurer des traits *quantitatifs* sur des organismes apparentés : taille, poids, taux de croissance, espérance de vie, longueur d'une partie du corps, etc. Ces traits peuvent être discrétisés afin d'être analysés comme des traits discrets, par exemple en utilisant le modèle Mk présenté dans le chapitre 2, mais il est plus naturel de les considérer comme intrinsèquement *continus*.

De même que pour les séquences d'ADN, on s'attend à ce que des espèces fortement apparentées (par exemple le chimpanzé et le bonobo) possèdent des valeurs de traits plus proches que deux espèces plus distantes (par exemple le chimpanzé et le ouistiti). Ce lien entre similarité de séquences (ou de traits) et apparentement est à la base de nombreuses méthodes de reconstruction d'arbres phylogénétique (cf chapitres 7 et 8). Les modèles d'évolution de traits continus ont pour objet de décrire cette similarité afin de la comprendre ou de la corriger.

1.1 Méthodes comparatives

Lorsqu'on mesure plusieurs traits, il est naturel d'étudier leurs relations. La taille est-elle corrélée à la longévité chez les animaux ? La longueur des sépales est-elle corrélée à celle des pétales chez les fleurs ? Une façon naïve de répondre à ces questions consiste à calculer le coefficient de corrélation entre les traits.

Malheureusement, et comme l'illustre la figure 1, cette vision naïve peut mettre en évidence des corrélations apparentes (panel de gauche) là où il n'existe qu'un effet de composition (panel de droite) dû à la phylogénie des espèces (panel du milieu). Dans cet exemple caricatural, la structure induite par la phylogénie peut être corrigée simplement en considérant un effet *groupe* (en *gris* ou *noir* sur la figure 1, droite) mais la correction optimale est moins évidente pour des phylogénies plus réalistes.

Dans son travail fondateur sur les méthodes comparatives en phylogénie, Felsenstein (1985) propose une correction systématique basée sur la notion de contrastes phylogénétiques indépendants. L'idée particulièrement astucieuse consiste à comparer non plus des *valeurs de traits* mesurées aux feuilles mais des *différences de valeurs de traits* observées entre des portions d'arbre disjointes. Cette idée est illustrée dans la figure 2. Le contraste C_{ij} entre un nœud i et un nœud j est donné par $C_{ij} = (Y_j - Y_i)/\sqrt{d_{ij}}$ où Y_i (respectivement, Y_j) est la valeur du trait pour le nœud i (respectivement, j) et d_{ij} est la distance dans l'arbre entre les nœuds i et j. Cette méthode nécessite de calculer préalablement la valeur du trait pour les nœuds ancestraux. On peut montrer que, sous certaines hypothèses, elle transforme les n valeurs corrélées de trait (1 par espèce) en $n-1$ contrastes indépendants et identiquement distribués (voir section 3.2).

Cette méthode est surtout intéressante pour ce qu'elle suppose sans l'expliciter : (i) l'intérêt porte sur la variation conjointe des traits plus que sur leurs valeurs conjointes, (ii) les variations du trait sur des portions disjointes de l'arbre sont indépendantes, (iii) l'échelle caractéristique de ces variations augmente avec la distance évolutive (ici comme la racine carrée). Ces 3 aspects évoquent tous une vision *dynamique* du trait qui peut être formalisée par un processus d'évolution, décrivant la façon dont le trait varie au cours du temps.

Nous présentons en section 2 un exemple simple de modèle d'évolution pour un trait avant de le généraliser en section 3 à plusieurs traits. Le modèle multivarié est particulièrement utile pour calculer la corrélation entre traits en corrigeant (i) l'effet de la phylogénie, à l'instar des

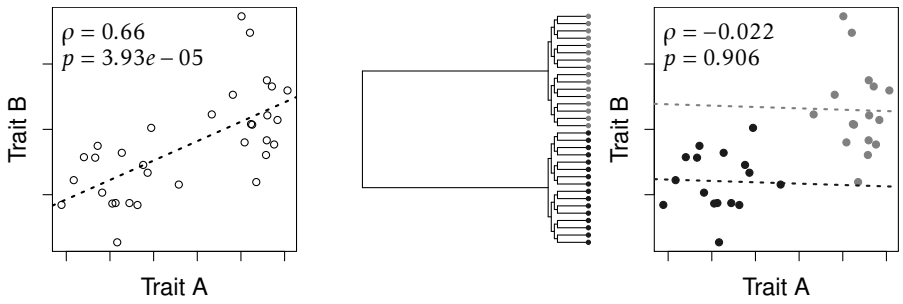
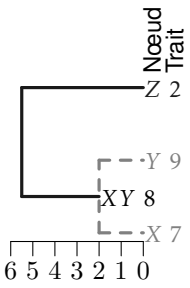


Fig. 1: Deux traits (A et B) sont mesurés sur 32 espèces. Une analyse naïve (gauche) révèle une corrélation forte ($\rho = 0.66$) et significative ($p = 3.93 \times 10^{-5}$) mais uniquement induite par l'histoire évolutive des espèces (panel du milieu) : les 32 espèces sont réparties en 2 groupes (en gris ou noirs) de 16 espèces fortement apparentées. Une fois cette structure prise en compte, il n'y a plus de corrélation entre les traits A et B (droite).



Nœuds	Différence	d_{ij}	Contraste
$X - Y$	2	4	1
$XY - Z$	6	9	2

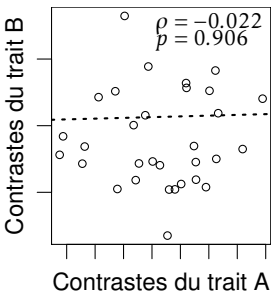


Fig. 2: Gauche : Illustration du principe de calcul des contrastes phylogénétiques. La différence de trait entre les nœuds X et Y est de 2 et ces deux nœuds sont à distance 4 dans l'arbre. Leur contraste vaut donc $2/\sqrt{4} = 1$. On note que les contrastes $X - Y$ (gris, trait pointillé) et $XY - Z$ (noir, trait plein) impliquent des portions d'arbre disjointes. Droite : Contrastes des traits A et B de l'exemple précédent. Aucune corrélation n'est apparente entre les traits.

contrastes, mais aussi (ii) l'effet d'autres variables, dans le cadre plus général des méthodes phylogénétiques comparatives (MPC).

1.2 Études des phénomènes évolutifs

Dans la description précédente, l'arbre est vu comme un paramètre de nuisance, qui induit des corrélations phylogénétiques sans intérêt que le formalisme des processus d'évolution sur arbre permet de corriger. L'arbre est cependant une source d'informations précieuse pour identifier des phénomènes adaptatifs, comme par exemple le gigantisme insulaire des tortues (Jaffe et al., 2011), ou l'adaptation de la forme du cerveau à des changements de régime alimentaire (Aristide et al., 2018). Dans ces situations, l'objet de l'étude est moins la comparaison de traits que les paramètres intrinsèques du modèle qui peuvent capturer la vitesse d'évolution du trait, sa valeur ancestrale dans certaines branches de l'arbre, sa valeur optimale, des changements adaptatifs qui peuvent coïncider avec des changements de niche, des mécanismes de compétition entre espèces, des interactions entre valeur du trait et fitness des individus, entre valeur du trait et géographie, etc.

Notre capacité à comprendre la dynamique passée des traits est évidemment contrainte par la complexité et le réalisme des modèles : il est peu commode d'étudier la valeur optimale d'un trait quand cette notion n'est pas modélisée par le processus d'évolution ! Mais elle l'est également par notre (in)capacité à estimer les modèles : les modèles les plus réalistes sont à la fois les plus complexes et ceux nécessitant le plus de données pour être ajustés. Nous verrons

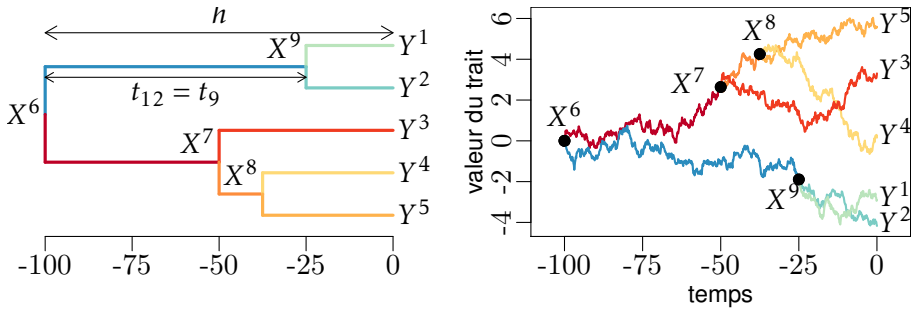


Fig. 3: Réalisation d'un MB (avec $\mu = 0$ et $\sigma^2 = 0.04$) sur un arbre phylogénétique calibré en temps. Les couleurs des branches de l'arbre (à gauche) correspondent aux couleurs des processus stochastiques (à droite). Seules les feuilles de l'arbre sont observées (au temps $t = 0$).

en détails dans la section 4 des modèles permettant d'étudier la valeur optimale d'un trait, de détecter des évènements adaptatifs modifiant cette dernière (typiquement un changement de niche), et de prendre en compte la dispersion du trait au sein d'une espèce. Ces modèles sont tous gaussiens et se prêtent bien à l'estimation. Nous évoquerons pour finir dans la section 5 quelques modèles non-gaussiens, capturant des dynamiques plus sophistiquées au prix d'une estimation souvent bien plus complexe.

2 Le mouvement brownien

Un modèle d'évolution de trait se doit, à tout le moins, de décrire des fluctuations de la valeur du trait le long d'une branche de l'arbre phylogénétique. Ces fluctuations sont généralement décrites par un processus aléatoire à temps continu. Pour présenter un intérêt évolutif, le modèle doit également décrire comment ces fluctuations sont partagées entre les différentes branches du même arbre. Si les modèles d'évolution diffèrent dans la définition du processus courant le long d'une branche (gaussien ou non, univarié ou non, *etc.*), la plupart font l'hypothèse de "branchement" qui stipule que le processus donne naissance à deux copies indépendantes, partant de l'état courant du processus, à chaque nœud de l'arbre (voir la figure 3). Cette représentation permet de modéliser la corrélation observée entre les traits d'espèces apparentées au travers de l'intersection de leurs trajectoires évolutives respectives.

2.1 Description

Le processus le plus simple et le plus classique pour décrire l'évolution d'un trait univarié est le mouvement brownien (MB), régi par l'équation :

$$W_0 = \mu, \qquad \mathrm{d}W_t = \sigma \mathrm{d}B_t, \ \forall t \in [0, h], \tag{1}$$

où B_t est le mouvement brownien de variance 1, défini comme l'unique processus à incréments stationnaires et indépendants, presque sûrement continu, et tel que $B_t \sim \mathcal{N}(0,\ t)$ pour tout temps t, $0 \leq t \leq h$ (voir par exemple Méléard, 2016 pour une introduction à ce processus dans un contexte de modélisation écologique). Le fait de supposer la valeur du trait fixe à la racine ($W_0 = \mu$) revient à travailler conditionnellement à cette valeur. Le modèle résultant de ce processus combiné avec l'hypothèse de branchement permet de préciser le lien entre les valeurs du trait Y mesuré chez deux espèces i et j observées respectivement aux temps t_i et t_j et s'étant séparées l'une de l'autre au temps t_{ij} ($0 \leq t_{ij} \leq \min(t_i, t_j)$, voir figure 3, gauche) :

$$\mathbb{E}\left[Y^i\right] = \mu, \qquad \mathbb{V}\mathrm{ar}\left[Y^i\right] = \sigma^2 t_i, \qquad \mathbb{C}\mathrm{ov}\left[Y^i; Y^j\right] = \sigma^2 t_{ij}. \tag{2}$$

On renvoie à Felsenstein (2004, Chap. 23) pour une dérivation formelle et intuitive de ces équations. On parle d'arbre ultramétrique quand toutes les espèces sont contemporaines et observées à un même temps $t_i \equiv h$ appelée « hauteur de l'arbre ». Dans ce cas, la corrélation entre les traits observés chez deux espèces dépend uniquement de leur temps d'évolution partagé, *via* la formule $\mathbb{C}\text{or}\left[Y^i; Y^j\right] = t_{ij}/h$.

2.2 Régression phylogénétique et transformations statistiques

2.2.1 Régression phylogénétique

Comme les autres modèles gaussiens uni- ou multi-variés qui seront présentés dans les sections 3 et 4, le modèle MB induit une distribution marginale gaussienne aux feuilles de l'arbre, avec une structure de variance-covariance qui lui est spécifique. La *régression phylogénétique* (Grafen, 1989, 1992) consiste à décrire les relations statistiques entre le trait observé $\mathbf{Y}$ et une série de prédicteurs (par exemple, des variables environnementales) rangés dans une matrice $\mathbf{X}$ de taille $n \times k$ comme un modèle linéaire, dont la loi des résidus est précisément celle donnée par le processus considéré :

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{E} \quad \text{avec} \quad \mathbf{E} \sim \mathcal{N}\left(\mathbf{0},\ \sigma^2 \mathbf{C}_n\right) \tag{3}$$

où $\mathbf{C}_n = [C_{ij}]_{1 \leq i,j \leq n}$ est la matrice représentant la structure de variance aux feuilles, soit pour le modèle MB : $C_{ij} = t_{ij}$ (voir l'équation (2)).

Cette représentation présente un double intérêt. Elle permet tout d'abord d'estimer les effets des prédicteurs sur la valeurs du trait, sans les confondre avec ceux de la structure phylogénétique liant les différentes espèces. Elle donne également un cadre pour mesurer l'intensité de l'effet de la structure phylogénétique sur la diversité du trait. En effet, celle-ci étant entièrement capturée par la matrice $\mathbf{C}_n$, relâcher la structure revient à « atténuer » les coefficients hors diagonaux de cette matrice, pour la rapprocher d'une matrice diagonale. Cela revient à changer la structure de l'arbre sous-jacent, pour le rapprocher d'un arbre étoile, dans lequel toutes les espèces sont indépendantes. En effet, la matrice $\mathbf{C}_n$ dépendant uniquement des longueurs des branches de l'arbre considéré, changer la structure de variance revient à imposer une transformation sur les branches de l'arbre. C'est l'objectif des transformations proposées par Pagel (1999).

2.2.2 Les transformations de Pagel

On présente ici une revue des transformations des branches de l'arbre dues à Pagel (1999) en commençant par la plus utilisée : la transformation λ. Il se trouve que les vraisemblances résultant de ces différentes transformations sont explicites et ne dépendent que d'un paramètre, elles peuvent donc être optimisées numériquement (Revell, 2012; Pennell et al., 2014). En lien avec la discussion précédente, ce paramètre est souvent vu comme mesurant un « signal phylogénétique », défini comme « la non-indépendance statistique entre les valeurs des traits de plusieurs espèces du fait de leurs relations phylogénétiques » (Revell et al., 2008). Comme souligné dans la référence précédente à laquelle on renvoie le lecteur intéressé, il existe d'autres manières de mesurer ce signal, l'une des plus populaires étant la statistique K (Blomberg et al., 2003), non discutée ici.

λ de Pagel. Une première idée est d'atténuer les corrélations entre espèces en multipliant tous les coefficients hors diagonaux de la matrice $\mathbf{C}_n$ par un paramètre λ entre 0 et 1 :

$$\begin{cases} C(\lambda)_{ij} = \lambda C_{ij} & \forall i \neq j \\ C(\lambda)_{ii} = C_{ii} & \forall 1 \leq i \leq n \end{cases} \tag{4}$$

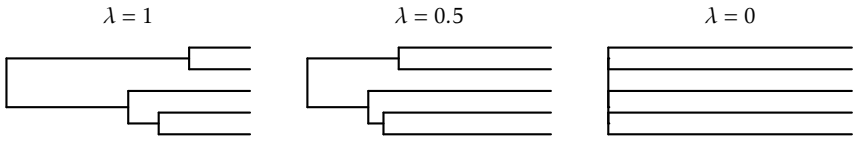


Fig. 4: Transformation λ de Pagel. Gauche : $\lambda = 1$, arbre original. Milieu : $\lambda < 1$, corrélations phylogénétiques atténuées. Droite : $\lambda = 0$, observations indépendantes.

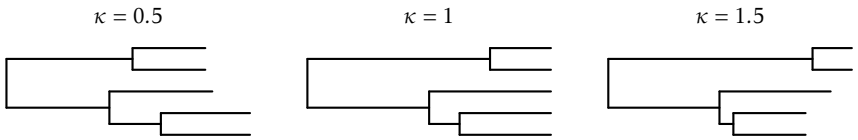


Fig. 5: Transformation κ de Pagel. Gauche : $\kappa < 1$, longueurs de branches homogènes. Milieu : $\kappa = 1$, arbre original. Droite : $\kappa > 1$, longueurs de branches hétérogènes.

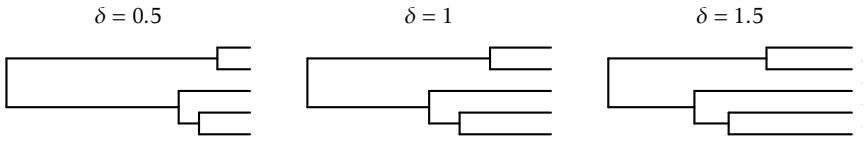


Fig. 6: Transformation δ de Pagel. Gauche : $\delta < 1$, évolution concentrée dans les branches ancestrales. Milieu : $\delta = 1$, arbre original. Droite : $\delta > 1$, évolution concentrée dans les branches récentes.

On peut vérifier que cette opération revient à faire courir un MB sur un arbre dont les branches ont subi la transformation suivante :

$$\ell_i(\lambda) = \begin{cases} \lambda \ell_i & \text{si } i \text{ est un nœud interne} \\ \ell_i + (1-\lambda) t_{\mathsf{pa}(i)} = \lambda \ell_i + (1-\lambda) t_i & \text{si } i \text{ est une feuille,} \end{cases} \tag{5}$$

où ℓ_i est la longueur de la branche allant de l'unique parent $\mathsf{pa}(i)$ de i à i : $\ell_i = t_i - t_{\mathsf{pa}(i)}$. Cette transformation est représentée figure 4. Elle revient à multiplier toutes les branches internes par λ, puis à allonger les branches externes au diapason pour que l'arbre reste ultramétrique et de même hauteur totale h. Le paramètre λ peut être vu comme mesurant le « signal phylogénétique » : si $\lambda = 1$, l'arbre n'est pas modifié, et la distribution des données aux feuilles est inchangée (signal phylogénétique fort, l'arbre a une grande influence sur la structure des données), tandis que si $\lambda = 0$, l'arbre se ramène à un arbre étoile, et les données aux feuilles sont indépendantes les unes des autres (signal phylogénétique faible, l'arbre n'a pas d'influence sur les données observées). Comme on le verra section 4.6, ce paramètre λ peut être interprété comme un rapport de variances intra et inter spécifiques, et est pour cette raison parfois interprété comme un paramètre d'« héritabilité phylogénétique » (Lynch, 1991; Leventhal and Bonhoeffer, 2016).

Cette transformation est très utilisée dans la littérature, avec plus de 2300 citations[1] pour l'article original de Pagel (1999). Le paramètre λ est souvent vu comme un paramètre de nuisance, permettant de contrôler de manière *ad hoc* l'importance de la phylogénie dans une régression phylogénétique (voir Revell, 2010 et Ceccarelli et al., 2018; Law et al., 2018 pour des exemples d'applications récents en écologie évolutive, pour l'étude de groupes d'araignées ou de mustéloïdes). Il peut également être un paramètre d'intérêt en lui même, pour quantifier l'importance relative de la phylogénie par rapport aux variations environnementales indépendantes, par exemple en virologie (Vrancken et al., 2015).

[1] D'après le site « *Web of Science* », dernière consultation le 14/02/2020.

κ de Pagel. Cette transformation consiste à élever toutes les longueurs à une puissance $\kappa > 0$:

$$\ell_i(\kappa) = (\ell_i)^\kappa \quad \forall 1 \leq i \leq n. \tag{6}$$

Elle ne permet pas de conserver un arbre ultramétrique (voir figure 5). Si $\kappa > 1$, l'hétérogénéité entre les longueurs de branches est accrue, et l'on s'attend à encore plus de diversification sur les longues branches, et encore moins sur les branches courtes. À l'inverse, si $\kappa < 1$, la transformation a tendance à homogénéiser la longueur de toutes les branches.

δ de Pagel. Dans cette dernière transformation, on élève toutes les hauteurs de nœuds à une puissance δ, tout en compensant par un facteur $h^{1-\delta}$ de sorte à ce que l'arbre reste ultramétrique et de hauteur totale h (voir figure 6) :

$$t_i(\delta) = (t_i)^\delta \cdot h^{1-\delta} \quad \forall 1 \leq i \leq n. \tag{7}$$

Cette transformation a une interprétation similaire à celle du modèle ACDC (voir section 4.4.1). Si $\delta < 1$, la plus grande partie de l'évolution a lieu dans les branches ancestrales (« *decelerating* », ou « *early burst* »), tandis que si $\delta > 1$, ce sont les branches terminales qui sont responsables de toute la variance (« *accelerating* »).

2.3 Algorithmes récursifs pour l'inférence

La formulation en terme de régression phylogénétique de l'équation (3) revient à un simple modèle linéaire avec corrélation connues (conditionnellement à l'arbre phylogénétique, les temps évolutifs partagés, et donc la matrice $\mathbf{C}_n$, sont connus), si bien que les estimateurs du maximum de vraisemblance des paramètres $\boldsymbol{\theta}$ et σ^2 peuvent s'obtenir comme simples solutions des moindres carrés généralisés (voir par exemple Mardia et al., 1979) :

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{X}^T \mathbf{C}_n^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{C}_n^{-1} \mathbf{Y} \\ \hat{\sigma^2} &= (n-1)^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}})^T \mathbf{C}_n^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}). \end{aligned} \tag{8}$$

Ces formules (et nombre de celles qui apparaîtront par la suite) font intervenir $\mathbf{C}_n^{-1}$. L'approche naïve qui consiste à inverser cette matrice de taille $n \times n$ a une complexité algorithmique d'ordre $\mathcal{O}(n^{2.373})$ (Raz, 2003) et devient donc prohibitive pour des arbres de grandes tailles. Une approche plus sophistiquée tire partie de la structure d'arbre et passe par un algorithme récursif d'*élagage* (dit en anglais « *pruning algorithm* ») pour calculer ces estimateurs sans calculer explicitement $\mathbf{C}_n^{-1}$ et avec une complexité d'ordre $\mathcal{O}(n)$ en temps et en espace.

Cet algorithme récursif, semblable à celui utilisé pour le calcul de la vraisemblance pour des caractères discrets (voir chapitre 7), est un algorithme de type « propagation de croyance » (« *belief propagation* » : Kim and Perl, 1983). Depuis sa première description dans ce contexte par Felsenstein (1973) pour le MB, de nombreuses adaptations (parfois redondantes) en ont été proposées (voir par exemple Hadfield and Nakagawa, 2010; FitzJohn, 2012; Freckleton, 2012; Lartillot, 2014; Pybus et al., 2012; Cybis et al., 2015; Bastide et al., 2018a) pour des processus gaussiens plus complexes (présentés ci dessous, voir section 4). Mitov et al. (2020) proposent une implémentation très générale de l'algorithme qui englobe les adaptations précédentes. L'algorithme récursif peut même dans certains cas (Landis et al., 2013; Duchen et al., 2017; Hiscott et al., 2016) s'étendre à des processus non-gaussiens (voir section 5). Enfin, Ho and Ané (2014a) décrivent un autre algorithme linéaire, basé sur une formulation un peu différente (structure en « 3 points »), qui leur permet de traiter les régressions phylogénétiques généralisées (*i.e.* logistique, poissonnienne, etc).

3 Analyse multivariée

Dans la section précédente, on s'est limité au cas où l'on n'observait qu'un seul trait aux feuilles de l'arbre. Les techniques décrites plus haut peuvent s'étendre de manière immédiate au cas où l'on observe plusieurs traits qui évoluent de manière *indépendante* sur l'arbre. Cependant, comme on l'a vu dans l'exemple introductif (voir section 1.1), on s'intéresse précisément dans les jeux de données observés aux relations de *corrélations* entre les traits. Dans cette section, on montre comment le MB multivarié permet de modéliser l'évolution corrélée de multiples traits sur un arbre phylogénétique.

3.1 Description

3.1.1 Définition

Le mouvement brownien multivarié décrit l'évolution d'un vecteur $\mathbf{W}_t$ de p traits au cours du temps t, avec une moyenne $\boldsymbol{\mu}$ et une variance $\mathbf{R} = \mathbf{R}^{1/2}(\mathbf{R}^{1/2})^T$ par l'EDS suivante :

$$\mathbf{W}_0 = \boldsymbol{\mu}, \qquad \mathrm{d}\mathbf{W}_t = \mathbf{R}^{1/2}\mathrm{d}\mathbf{B}_t, \ \forall 0 \leq t \leq h, \tag{9}$$

où $\mathbf{B}_t$ est le mouvement brownien de variance $\mathbf{I}_p$, défini comme l'unique processus à incréments stationnaires et indépendants, presque sûrement continu, et tel que $\mathbf{B}_t \sim \mathcal{N}\left(\mathbf{0}_p, \ t\mathbf{I}_p\right)$ pour tout temps t, $0 \leq t \leq h$. Là encore, supposer le vecteur des traits à la racine fixe revient à travailler conditionnellement aux valeurs initiales des traits.

3.1.2 Loi aux feuilles

Les données consistent à présent en une *matrice* de traits $\mathbf{Y}$ de taille $n \times p$: pour chaque feuille i de l'arbre, $1 \leq i \leq n$, on observe un vecteur de traits $\mathbf{Y}^i$ de dimension p, tel que $\mathbf{Y}^T = (\mathbf{Y}^1, \ldots, \mathbf{Y}^n)$. De même que pour le mouvement univarié, on peut écrire la distribution marginale observée aux feuilles comme suit (voir également Felsenstein 2004, Chap. 23) :

$$\mathbb{E}\left[\mathbf{Y}^i\right] = \boldsymbol{\mu}, \qquad \mathbb{C}\mathrm{ov}\left[\mathbf{Y}^i; \mathbf{Y}^j\right] = t_{ij}\mathbf{R}. \tag{10}$$

La moyenne aux feuilles est uniforme et égale à la moyenne à la racine $\boldsymbol{\mu}$, et la covariance entre le trait q de la feuille i et le trait r de la feuille j est simplement le produit entre la variance R_{qr} entre les deux traits, et le temps d'évolution partagé t_{ij} entre les deux feuilles $\mathbb{C}\mathrm{ov}\left[Y_{iq}; Y_{jr}\right] = t_{ij}R_{qr}$.

Cette structure multiplicative simple implique que l'on peut naturellement *factoriser* la contribution de l'arbre et celle du processus. Cela signifie que l'on peut écrire la loi de $\mathbf{Y}$ directement comme une gaussienne matricielle :

$$\mathrm{vec}(\mathbf{Y}) \sim \mathcal{N}\left(\mathrm{vec}(\mathbf{1}_n\boldsymbol{\mu}^T), \ \mathbf{R} \otimes \mathbf{C}_n\right) \tag{11}$$

où vec est l'opérateur de vectorisation, qui à une matrice de taille $n \times p$ associe un vecteur de dimension np en « empilant » les vecteurs colonnes de la matrice les uns au dessus des autres, $\otimes$ est le produit de Kronecker, et $\mathbf{C}_n = [t_{ij}]_{1 \leq i,j \leq n}$ est la matrice des temps d'évolutions partagés (voir 2.2). La vraisemblance du modèle se calcule simplement et permet d'écrire les estimateurs du maximum de vraisemblance de $\boldsymbol{\mu}$ et $\mathbf{R}$. Ces derniers sont détaillés dans la section 3.3.

3.2 Contrastes phylogénétiques

Les contrastes introduits dans la section 1.1 trouvent leur pleine justification dans ce cadre gaussien. En effet (i) le contraste du trait r entre les nœuds i et j défini par $C_{ij}^r = (Y_{ir} - Y_{jr})/\sqrt{d_{ij}}$ suit une loi gaussienne $\mathcal{N}(0,\ R_{rr})$. De plus, (ii) par la propriété d'incréments indépendants du MB, les contrastes C_{ij}^r et C_{kl}^q sont indépendants dès que les chemins $i \leftrightarrow j$, reliant i à j dans l'arbre, et $k \leftrightarrow l$, reliant k à l n'ont aucune branche en commun. Enfin, l'équation (10) montre que (iii) la covariance entre les mêmes contrastes calculés sur des traits différents vaut $\mathbb{C}\mathrm{ov}\left[C_{ij}^r; C_{ij}^q\right] = R_{rq}$. On renvoie le lecteur intéressé au livre de Felsenstein (2004), Chapitres 23 et 24, pour une dérivation plus détaillée de ces contrastes par leur inventeur.

Les contrastes permettent donc (i) de remplacer n valeurs corrélées par $n-1$ différences indépendantes et (ii) de neutraliser simultanément la matrice de covariance phylogénétique $\mathbf{C}_n$ et la valeur ancestrale du trait $\boldsymbol{\mu}$ pour étudier simplement la matrice de covariance des traits $\mathbf{R}$. Cette simplification se fait au prix d'une légère perte de puissance : passage de n à $n-1$ observations lors du calcul des contrastes. La méthode des contrastes n'est présentée que pour son intérêt historique. Elle a aujourd'hui largement cédé la place aux MPC, notamment les modèles phylogénétiques gaussiens dont elle est un cas particulier, et qui se prêtent beaucoup mieux à des généralisations comme l'ACP phylogénétique.

3.3 ACP phylogénétique

L'équation (11) qui décrit la loi marginale des traits aux feuilles se ramène à une simple gaussienne matricielle, de moyenne $\mathbf{1}_n\boldsymbol{\mu}^T$, de variance en colonnes $\mathbf{R}$, et de variance en lignes $\mathbf{C}_n$. Si l'on suppose que l'on raisonne conditionnellement à la phylogénie, cette matrice $\mathbf{C}_n$, étant entièrement caractérisée par l'arbre, est connue. Il est alors possible de décorréler les observations en lignes en multipliant les observations à droite par la matrice de Cholesky de $\mathbf{C}_n$. Si $\mathbf{L}_n$ est telle que $\mathbf{C}_n = \mathbf{L}_n\mathbf{L}_n^T$, alors :

$$\mathrm{vec}(\mathbf{L}_n^{-1}\mathbf{Y}) \sim \mathcal{N}\left(\mathrm{vec}(\mathbf{L}_n^{-1}\mathbf{1}_n\boldsymbol{\mu}^T),\ \mathbf{R}\otimes\mathbf{I}_n\right), \tag{12}$$

et l'on se ramène au cas classique en analyse multivariée où l'on observe n vecteurs indépendants. On dispose alors d'estimateurs explicites pour $\boldsymbol{\mu}$ et $\mathbf{R}$ (voir par exemple Mardia et al., 1979) :

$$\begin{aligned} \hat{\boldsymbol{\mu}}^T &= (\mathbf{1}_n^T\mathbf{C}_n^{-1}\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{C}_n^{-1}\mathbf{Y} \\ \hat{\mathbf{R}} &= (n-1)^{-1}(\mathbf{Y}-\mathbf{1}_n\hat{\boldsymbol{\mu}}^T)^T\mathbf{C}_n^{-1}(\mathbf{Y}-\mathbf{1}_n\hat{\boldsymbol{\mu}}^T). \end{aligned} \tag{13}$$

En utilisant ces notations, il est possible, comme en analyse multivariée classique, de tenter de réduire la dimension des observations par le biais d'une analyse en composante principale, dite « phylogénétique » (ACPp, Revell, 2009). Cette transformation revient à essayer de supprimer les corrélations en colonnes des observations, en utilisant les vecteurs propres de l'estimateur de la variance $\hat{\mathbf{R}} = \hat{\mathbf{V}}\hat{\mathbf{D}}^2\hat{\mathbf{V}}^T$ pour calculer des scores empiriquement indépendants $\mathbf{S} = (\mathbf{Y}-\mathbf{1}_n\hat{\boldsymbol{\mu}}^T)\hat{\mathbf{V}}$. Cette ACPp ne diffère de l'ACP classique que par la prise en compte des corrélations induites par l'arbre, telle que représentée par $\mathbf{C}_n$. Comme présenté figure 7, si les traits sont effectivement issus d'un MB multivarié, elle décorrèle mieux les traits que l'ACP classique, et permet de corriger les effets de parentés entre les individus.

L'ACPp a été utilisée dans la littérature comme moyen de résumer et représenter des données morphologiques de grande dimension (voir par exemple Ceccarelli et al., 2018 pour une application sur des morphologies d'araignées). Cependant, il est important de garder à l'esprit que les performances de cette ACPp dépendent fortement de l'hypothèse sous-jacente d'évolution brownienne des traits. Si cette hypothèse n'est pas vérifiée, il est facile de montrer

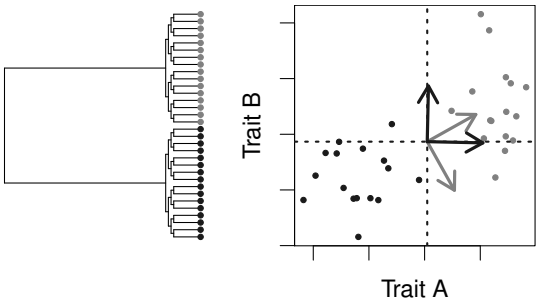


Fig. 7: Comparaison entre l'ACP et l'ACPp sur l'exemple introductif présenté section 1. L'arbre (à gauche) induit deux groupes fortement apparentés, ce qui biaise la structure de variance des traits observés (à droite). L'ACP classique (axes gris) est biaisée, tandis que l'ACP phylogénétique (axes noirs) approxime les vrais vecteurs propres (lignes pointillées, aucune corrélation).

que cette procédure induit des biais structurels importants (Uyeda et al., 2015; Khabbazian et al., 2016; Bastide et al., 2018a), et il est déconseillé de l'utiliser comme pré-traitement des données avant une analyse plus poussée (Uyeda et al., 2015). Le caractère brownien de l'évolution peut se justifier dans certains cas, mais n'est pas toujours pertinent pour analyser des données phylogénétiques. Dans la suite de ce chapitre, on présente d'autres types de modèles pour représenter l'évolution de traits de manière un peu plus réaliste.

4 Modèles gaussiens

4.1 Quelques limites du mouvement brownien

On a vu que toutes les méthodes phylogénétiques comparatives étudiées dans les sections précédentes reposaient sur le même modèle d'évolution d'un (ou plusieurs) traits suivant un mouvement brownien branchant sur un arbre phylogénétique. Ce modèle, très simple, nous a permis d'incorporer explicitement les relations de parentés phylogénétiques dans nos analyses statistiques. Si l'on se place dans ce cadre, le MB représente l'évolution des traits phénotypiques au cours du temps. Cette hypothèse d'évolution peut être justifiée pour certains types de traits, dans certaines conditions environnementales (voir ci-après la section 4.3). Cependant, le MB peut être vu comme représentant un simple « bruit blanc », qui ne favorise aucune valeur de trait par rapport à une autre, et dont la variance, qui augmente linéairement avec le temps, n'est pas bornée. De telles propriétés peuvent entrer en contradiction avec l'idée intuitive que l'on se fait d'un trait soumis à une sélection environnementale. Plus généralement, ce qui a fait la force et le succès du mouvement brownien, sa simplicité, fait aussi sa limite : un tel processus ne permet pas de modéliser ou d'explorer les différents mécanismes sous-jacents à l'évolution phénotypique.

D'un point de vue théorique, cette limitation est facile à lever. En gardant la même structure branchante sous-jacente liée à l'arbre phylogénétique, il suffit de remplacer le mouvement brownien par un autre processus stochastique bien choisi. Ce choix repose en général sur deux contraintes : il faut que le processus utilisé soit, premièrement, cohérent et à même de modéliser de manière utile certains des mécanismes biologiques à l'œuvre, et, deuxièmement, assez simple pour en autoriser l'analyse statistique. On explore dans la suite plusieurs options qui ont été proposées dans la littérature.

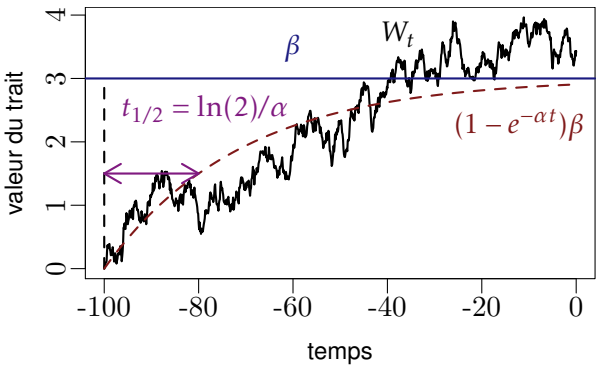


Fig. 8: Réalisation d'un OU univarié W_t, avec une valeur à la racine fixée $\mu = 0$, une variance $\sigma^2 = 0.05$, un valeur optimale $\beta = 3$ (ligne horizontale bleue), et une force de sélection α telle que $t_{1/2} = \ln(2)/\alpha = 20\%$ du temps total alloué ($h = 100$, double flèche violette). L'évolution exponentielle de l'espérance de W_t vers β est représentée par la courbe pointillée violette.

4.2 Le processus d'Ornstein-Uhlenbeck

4.2.1 Description univariée

Le processus d'Ornstein-Uhlenbeck (OU) a été proposé pour modéliser une *évolution stabilisatrice* autour d'un optimum (Hansen and Martins, 1996; Hansen, 1997). Comparé au mouvement brownien, ce processus présente, en plus d'un bruit gaussien, un mécanisme de rappel élastique vers une valeur centrale β, interprétée comme la valeur optimale du trait dans un environnement donné (voir la section 4.3 pour plus de précisions). L'équation stochastique définissant ce processus est donnée ci-après, pour un trait univarié W_t évoluant sur un temps t entre 0 et h. Un exemple de réalisation de ce processus est présenté figure 8.

$$W_0 = \mu, \qquad \mathrm{d}W_t = -\alpha(W_t - \beta)\mathrm{d}t + \sigma \mathrm{d}B_t, \ \forall 0 \leq t \leq h. \tag{14}$$

La vitesse avec laquelle le trait est attiré par son optimum est contrôlée par la *force de sélection* α. Pour interpréter ce paramètre, on le remplace souvent par le *temps de demi-vie phylogénétique* $t_{1/2} = \ln(2)/\alpha$ associé (Hansen, 1997). Comme on peut le voir figure 8, $t_{1/2}$ représente le temps nécessaire pour que l'espérance du trait fasse la moitié du chemin entre sa valeur actuelle et la valeur de l'optimum. Ce temps $t_{1/2}$ peut être comparé au temps total h d'évolution accordé au processus, qui est souvent la hauteur totale de l'arbre phylogénétique dans le cadre des MPC. Pour rendre cette relation explicite, on exprime souvent $t_{1/2}$ en pourcentage du temps total h :

- Si $t_{1/2}$ est *petit* comparé à h, cela signifie que le processus dispose du temps nécessaire pour aller vers son optimum, et donc que la force de sélection α peut-être considérée comme *forte*. La partie déterministe de rappel élastique tend à l'emporter sur la partie stochastique.
- Dans le cas contraire, si $t_{1/2}$ est *grand* par rapport à h, le processus est « coupé » avant d'avoir pu atteindre son optimum, et la force de sélection est *faible*. La partie stochastique brownienne tend à dominer.

La loi du trait W_t à l'instant t est toujours gaussienne :

$$W_t \sim \mathcal{N}\left(e^{-\alpha t}\mu + (1 - e^{-\alpha t})\beta, \ (1 - e^{-2\alpha t})s^2\right), \tag{15}$$

avec $s^2 = \sigma^2/(2\alpha)$ la *variance stationnaire*, qui représente la variance limite d'un processus après un temps long. L'existence d'une loi stationnaire, ainsi que le caractère borné de la variance, sont deux propriétés qui rendent l'OU intéressant à utiliser par rapport au MB.

4.2.2 Loi aux feuilles et identifiabilité

En reprenant les notations des sections précédentes, on obtient (voir par exemple Butler and King 2004 pour une dérivation détaillée dans un cas simple) :

$$\begin{aligned}\mathbb{E}\left[Y^i\right] &= e^{-\alpha t_i}\mu + (1-e^{-\alpha t_i})\beta \\ \mathbb{C}\mathrm{ov}\left[Y^i; Y^j\right] &= e^{-\alpha(t_i-t_{ij})}s^2 e^{-\alpha(t_j-t_{ij})} - e^{-\alpha t_i}s^2 e^{-\alpha t_j} \\ &= \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t_{ij}})e^{-\alpha(t_i+t_j-2t_{ij})}\end{aligned} \tag{16}$$

La loi marginale du trait aux feuilles est ainsi complètement explicite, ce qui permet d'en faire une inférence directe (Butler and King, 2004; Beaulieu et al., 2012; Pennell et al., 2014).

Remarque 4.1. Lorsque α tend vers 0, on a :

$$e^{-\alpha t_i}\mu + (1-e^{-\alpha t_i})\beta \to \mu \quad \text{et} \quad \frac{\sigma^2}{2\alpha}(1-e^{-2\alpha t_{ij}})e^{-\alpha(t_i+t_j-2t_{ij})} \to \sigma^2 t_{ij},$$

si bien que les moments donnés ci-dessus convergent vers ceux obtenus par simple mouvement brownien (voir équation (2)). Ceci confirme l'intuition donnée précédemment : lorsque $t_{1/2}$ tend vers l'infini (i.e. α tend vers 0), un OU homogène sur l'arbre se comporte comme un MB de paramètres μ et σ^2.

Pour un arbre ultramétrique, on a $t_i = h$ pour toutes les feuilles de l'arbre, si bien que le trait a la même espérance $\lambda = e^{-\alpha h}\mu + (1-e^{-\alpha h})\beta$ à toutes les feuilles, et que seul ce paramètre λ est identifiable. Il est donc impossible dans le cas d'un arbre ultramétrique d'inférer la moyenne à la racine de l'arbre et l'optimum de manière indépendante.

Comme présenté figure 9, ceci a des conséquences pratiques du point de vue de l'interprétation biologique. En effet, dans le scénario A de la figure 9, le trait part de sa valeur optimale ($\mu = \beta = 1$), et y reste jusqu'aux feuilles, si bien que le trait est vu comme étant à l'équilibre dans la population d'espèces observées aux feuilles. En revanche, dans le scénario B, le trait part d'une valeur élevée $\mu = 10$, et décroit vers son optimum $\beta = -2$ de manière exponentielle, mais le temps total alloué à son évolution (la hauteur de l'arbre h) est trop court pour lui permettre de l'atteindre, et tous les traits sont « coupés » à une valeur moyenne $\lambda = 1$. Dans ce scénario, la population d'espèces observée est hors équilibre, et l'on s'attend à ce que le trait continue à décroître dans le futur. Cependant, comme $\lambda = 1$ dans les deux scénarios, ils sont indistinguables à partir des seules données aux feuilles. Seul un signal temporel, obtenu par exemple par l'échantillonnage de fossiles lorsqu'ils sont disponibles, pourrait nous permettre de discriminer entre ces deux scénarios pourtant très différents.

Pour régler ce problème d'identifiabilité, on suppose souvent que $\lambda = \mu = \beta$, c'est-à-dire que la racine de l'arbre est déjà à l'optimum (scénario A). Cette hypothèse peut se justifier si l'on suppose par exemple que le trait a évolué longtemps avant d'arriver à la racine du clade qui nous intéresse. Cette difficulté, ainsi que d'autres problèmes relatifs à l'inférence de l'OU, sont présentés dans Ho and Ané (2014b).

4.2.3 L'OU multivarié

Lorsque le trait est multivarié, la dynamique de rappel du trait vers son optimum n'est plus décrite par un scalaire α, mais par une matrice $\mathbf{A}$ de taille $p \times p$:

$$\mathbf{W}_0 = \boldsymbol{\mu}, \qquad \mathrm{d}\mathbf{W}_t = -\mathbf{A}(\mathbf{W}_t - \boldsymbol{\beta})\mathrm{d}t + \mathbf{R}^{1/2}\mathrm{d}\mathbf{B}_t, \ \forall 0 \le t \le h. \tag{17}$$

En fonction de la structure imposée à $\mathbf{A}$, cette dynamique peut-être plus ou moins complexe (voir figure 10). Pour que le trait soit bien attiré vers son optimum, il suffit que toutes les

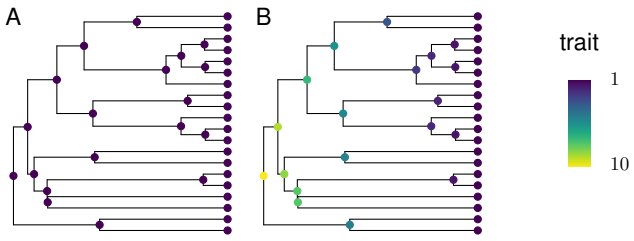


Fig. 9: Deux processus OU équivalents, tels que $\lambda = 1$ et $t_{1/2} = 50\%$ de la hauteur totale h de l'arbre. L'espérance du trait aux nœuds est représentée par l'échelle de couleur. A : $\lambda = \mu = \beta = 1$. B : $\lambda = 1$, $\mu = 10$ et $\beta = -2$.

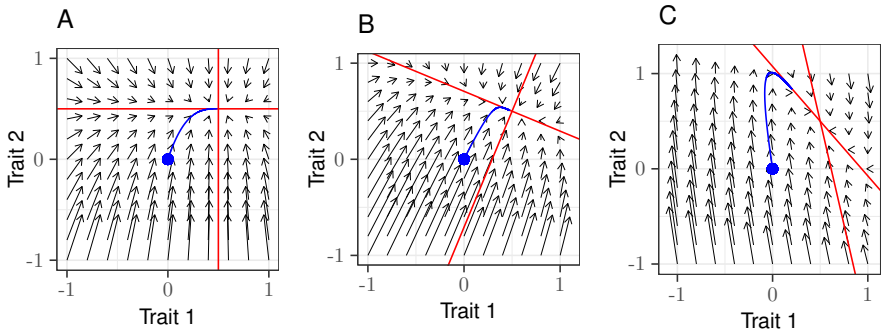


Fig. 10: Dynamiques de retour à l'optimum pour trois OU bivariés avec une matrice $\mathbf{A}$ diagonale (A), symétrique (B) ou de valeurs propres strictement positives (C). La trajectoire en bleu représente l'évolution du trait en espérance, de $\boldsymbol{\mu} = (0,0)^T$ vers l'optimum $\boldsymbol{\beta} = (0.5, 0.5)^T$. Les axes en rouge suivent la direction des vecteurs propres de la matrice, et sont centrés en $\boldsymbol{\beta}$. Les flèches noires représentent le champ de vecteur associé à la partie déterministe de l'équation (17).

valeurs propres de $\mathbf{A}$ aient des parties réelles strictement positives (Bartoszek et al., 2012). En pratique, l'inférence de cette matrice nécessite un signal temporel fort, et il est souvent préférable de lui imposer une structure a priori (Clavel et al., 2015).

La loi marginale du trait aux feuilles s'écrit de manière similaire au cas univarié (voir par exemple Bartoszek et al. 2012, Annexe B, pour une dérivation détaillée) :

$$\begin{aligned} \mathbb{E}\left[\mathbf{Y}^i\right] &= e^{-\mathbf{A}t_i}\boldsymbol{\mu} + (1 - e^{-\mathbf{A}t_i})\boldsymbol{\beta}, \\ \mathbb{C}\mathrm{ov}\left[\mathbf{Y}^i; \mathbf{Y}^j\right] &= e^{-\mathbf{A}(t_i - t_{ij})}\mathbf{S}e^{-\mathbf{A}^T(t_j - t_{ij})} - e^{-\mathbf{A}t_i}\mathbf{S}e^{-\mathbf{A}^T t_j}, \end{aligned} \tag{18}$$

où $\mathbf{S}$ est la variance stationnaire multivariée du processus, donnée par la formule $\mathrm{vec}(\mathbf{S}) = (\mathbf{A} \oplus \mathbf{A})^{-1}\,\mathrm{vec}(\mathbf{R})$, avec $\oplus$ la somme de Kronecker (Meucci, 2009).

De même que dans le cas univarié, il est possible d'utiliser directement cette distribution pour faire l'inférence des paramètres de l'OU multivarié (Bartoszek et al., 2012; Clavel et al., 2015). Comparé au cas brownien (voir équations (10) et (11)), on remarque que l'on ne peut plus factoriser de manière élémentaire les contributions de l'arbre et du processus en lui même à la matrice de variance. Cela rend l'inférence de ce processus plus complexe, et sujette à des instabilités numériques. En particulier, on ne dispose pas d'estimateur explicite pour $\mathbf{A}$ dans le cas général, et son estimation dépend de manière critique de la paramétrisation choisie pour cette matrice (Clavel et al., 2015). L'estimation est relativement aisée quand $\mathbf{A}$ et $\mathbf{R}^{1/2}$ commutent : Khabbazian et al. (2016) ont considéré le cas $\mathbf{A}$ et $\mathbf{R}^{1/2}$ diagonales et Bastide et al. (2018a) le cas $\mathbf{A}$ scalaire et $\mathbf{R}^{1/2}$ quelconque.

Depuis son introduction en phylogénie (Hansen and Martins, 1996), l'OU a connu un

grand succès dans le domaine, le nombre de publications utilisant ce modèle croissant exponentiellement sur une période de dix ans entre 2005 et 2014 (Cooper et al., 2016). Ce succès est en partie dû à son interprétation intuitive et séduisante d'un trait soumis à la sélection naturelle. Cependant, cette interprétation biologique n'est pas toujours justifiée, et certaines subtilités de modélisation, présentées dans la section suivante, sont importantes à garder en tête lors de l'utilisation de ce processus.

4.3 Interprétations biologiques et mises en gardes

L'interprétation des processus décrits ci-dessus dépend de manière cruciale de l'*échelle de temps* considérée. Pour des échelles de temps allant de quelques milliers à quelques centaines de milliers d'années, ces processus jouissent de garanties théoriques bien établies, et peuvent être liés à des mécanismes biologiques bien précis. En utilisant des outils de génétique quantitative, il est en effet possible de montrer (Lande, 1976) que l'évolution d'un phénotype soumis à une dérive génétique additive dans un paysage adaptatif constant, une fois remise à l'échelle, peut converger vers un mouvement brownien (pour un paysage adaptatif plat) ou un processus d'Ornstein-Uhlenbeck (pour un paysage comportant un unique pic adaptatif). Ce type de phénomène peut être observé lorsque l'on dispose d'un registre fossile assez dense, porteur d'un signal temporel suffisamment fort. C'est par exemple le cas pour la famille des poissons épinoches (Gasterosteidae), dont on voit l'évolution de la forme de l'épine dorsale d'un optimum à un autre en seulement quelques milliers d'années, et ce malgré une force de sélection relativement faible (Hunt et al., 2008; Hunt and Rabosky, 2014). Lorsque les pressions de sélection sont plus fortes, quelques années peuvent même être suffisantes pour observer une telle adaptation. Dans la famille des sauriens (Anolis) insulaires, certaines études suggèrent par exemple que l'introduction d'une espèce invasive (*Anolis sagrei*) a pu conduire à l'adaptation de l'espèce endémique (*Anolis carolinensis*) vers une niche de type arboricole en une quinzaine d'années seulement (1995–2010, Stuart et al., 2014).

Ces échelles de temps très courtes n'ont rien de commun avec les durées typiques obtenues lors de la calibration d'arbres phylogénétiques de type écologiques, qui se comptent en plusieurs dizaines voir centaines de millions d'années. Pour prendre quelques exemples, les ancêtres commun aux clades des singes du nouveau monde (Platyrrhini), aux oiseaux (Aves) et aux tortues (Testudines) remontent, respectivement, à environ 25, 110 et 210 millions d'années avant le présent (Aristide et al., 2016; Jetz et al., 2012; Jaffe et al., 2011). Sur de telles durées, les hypothèses utilisées par Lande (1976) ne sont plus valides. En particulier, on a pu observer que le taux de changement phénotypique inféré à partir du registre fossile était en général bien plus bas que celui qui serait attendu sous de tels modèles, phénomène connu sous le nom de « paradoxe de la stabilité » (« *paradox of stasis* », voir par exemple Hansen and Houle, 2004). Dans un tel contexte, l'interprétation des processus MB et OU doit donc être revue.

Une idée souvent invoquée est que, pour ce type de phénomènes macro-évolutifs, les processus stochastiques utilisés reflètent, plutôt que les conséquences de la dérive génétique, celles de l'évolution aléatoire des niches écologiques successives. L'hypothèse sous-jacente est que, pour des conditions environnementales données, le trait considéré atteint son optimum de manière presque instantanée (par rapport à l'échelle de temps macro-évolutive), si bien que c'est l'optimum du trait (ou « *optimum secondaire* », Hansen, 1997) que l'on suit au cours du temps, plutôt que sa valeur en elle même. Le MB peut alors être utilisé pour modéliser, dans un environnement stochastique, l'évolution de traits adaptatifs sur des phylogénies (Felsenstein, 2004, Chap. 24). Cette interprétation permettrait en particulier d'expliquer les divergences constatées entre les taux d'évolutions observés au niveaux micro et macro évolutifs. Dans le cadre d'un OU, l'optimum secondaire, soumis aux variations environnementales locales, évolue lui-même vers un « *optimum primaire* » (Hansen, 1997) censé représenter

un équilibre adaptatif macroscopique. Cet équilibre est lui-même susceptible de subir des changements brusques, ou *sauts*, reflétant des événements écologiques majeurs, comme une migration, ou une rupture climatique (Jaffe et al., 2011, voir aussi section 4.5).

Si l'OU est préféré au MB pour modéliser l'évolution d'un trait stationnaire (Hansen and Orzack, 2005; Hansen et al., 2008), dans le cadre de données écologiques sur un temps long, il est important de garder en tête que l'interprétation « naïve » d'un trait allant à son optimum est au mieux très simplifiée, et au pire peut induire en erreur (Cooper et al., 2016). Ces difficultés d'interprétation peuvent conduire à adopter une approche plus « pragmatique » de l'OU, qui peut être vu comme une simple *transformation statistique* de la phylogénie, conduisant à relâcher la structure arborescente des données imposée par un MB, avec un degré de liberté supplémentaire représenté par α (voir ci dessous, section 4.4.3).

4.4 Autres processus gaussiens

4.4.1 Modèles ACDC et EB

Le modèle ACDC pour (« *Accelerating / Decelerating model* ») a été introduit par Blomberg et al. (2003) pour modéliser un trait évoluant avec une variance qui croît ou décroît exponentiellement en temps. Il s'agit d'une variation simple sur le MB, défini par l'équation :

$$W_0 = \mu, \qquad \mathrm{d}W_t = \sigma_0 e^{rt/2} \mathrm{d}B_t, \ \forall 0 \leq t \leq h, \tag{19}$$

avec r un réel, qui règle le mode d'évolution. Si r est strictement négatif, la variance décroît au cours du temps, et ce modèle a été utilisé pour décrire une *radiation évolutive*, sous le nom de « *Early Burst* » (EB, Harmon et al., 2010). Il est particulièrement utilisé en paléontologie, pour des traits morphologiques pour lesquels on disposes de données fossiles, essentielles pour se faire une idée de la dynamique temporelle d'évolution du trait. Bien que séduisant, ce modèle est peu souvent sélectionné de manière statistiquement significative par rapport au MB simple (Harmon et al., 2010; Slater and Pennell, 2014).

Comme pour le MB, l'espérance du trait sous ce modèle reste constante égale à μ, tandis que la covariance induite entre deux feuilles est donnée par (voir Blomberg et al. 2003, Annexe 3, pour une dérivation détaillée) :

$$\mathbb{C}\mathrm{ov}\left[Y^i; Y^j\right] = \sigma_0^2 \frac{e^{rt_{ij}} - 1}{r} \tag{20}$$

Ce processus peut-être vu comme une transformation des branches de l'arbre, et est équivalent, sous certaines conditions, à un OU univarié (section 4.4.3).

4.4.2 Modèles OUBM and OUOU

Il est aussi possible d'introduire une « couche » supplémentaire au modèle, en supposant par exemple que la valeur optimale $\boldsymbol{\beta}(\mathbf{E}_t)$ d'un OU dépend de variables explicatives $\mathbf{E}_t$ (par exemple, des données climatiques) qui évoluent elles-mêmes dans le temps comme un processus stochastique, comme un MB ou un OU. C'est le principe des modèles OUBM et OUOU introduits respectivement par Hansen et al. (2008) et Bartoszek et al. (2012). Dans le cas le plus simple d'un OUBM univarié, le processus est défini comme :

$$\begin{cases} W_0 = \mu \\ \mathrm{d}W_t = -\alpha(W_t - \beta(E_t))\mathrm{d}t + \sigma_w \mathrm{d}B_t^w, \quad \forall 0 \leq t \leq h \\ \beta(E_t) = b_0 + b_1 E_t \\ \mathrm{d}E_t = \sigma_e \mathrm{d}B_t^e. \end{cases} \tag{21}$$

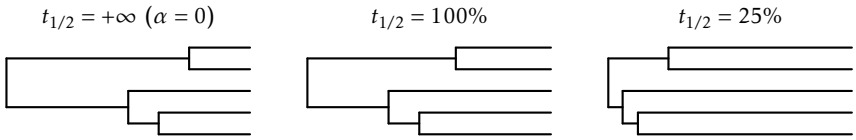


Fig. 11: Transformations de l'arbre équivalentes à un OU. Pour $\alpha = 0$, l'arbre est inchangé, et l'OU se ramène à un MB (gauche). Lorsque la sélection croît (c'est-à-dire lorsque le temps de demie vie décroît, en pourcentage de la hauteur totale de l'arbre), l'arbre se déforme pour ressembler de plus en plus à un arbre étoile (milieu et droite).

La structure de covariance obtenue aux feuilles peut être dérivée de manière explicite, et est donnée dans les deux références précédentes. Ce modèle OUBM a par exemple permis de mettre en évidence un lien statistiquement significatif entre les préférences thermiques de certaines espèces de sauriens et la température ambiante de leur environnement (Labra et al., 2009).

4.4.3 OU et ACDC comme transformations statistiques

Si l'on a à faire à un arbre ultramétrique, toutes les feuilles sont à la même distance de la racine ($t_i = h,\ \forall 1 \leq i \leq n$), et la covariance entre deux feuilles induite par un OU donnée par l'équation (16) se réduit à :

$$C_{ij}(\alpha) = e^{-2\alpha h}\frac{e^{2\alpha t_{ij}} - 1}{2\alpha}. \tag{22}$$

Il est alors facile de voir que l'OU se ramène à un simple MB courant sur les branches d'un arbre dont les hauteurs des nœuds internes ont été transformées par la relation (voir figure 11) :

$$t_i(\alpha) = e^{-2\alpha h}\frac{e^{2\alpha t_i} - 1}{2\alpha} \quad \forall 1 \leq i \leq n. \tag{23}$$

Cette relation a été observée de nombreuses fois dans la littérature, et peut être à la base de procédures d'inférence efficaces d'un OU sur un arbre ultramétrique (voir par exemple Blomberg et al., 2003; Ho and Ané, 2014a; Pennell et al., 2014; Bastide et al., 2018a). On remarque également que, lorsque α tend vers 0, $t_i(\alpha)$ tend vers t_i, et l'arbre n'est pas modifié. Ceci rejoint l'interprétation donnée précédemment (voir note 4.1) : lorsque α tends vers 0, l'OU univarié homogène converge vers un MB.

De la même manière, la structure de covariance induite par l'ACDC (équation (20)) peut être obtenue par un MB de variance σ_0^2 courant sur un arbre dont les longueurs de branches ont été transformées par la relation :

$$t_i(\alpha) = \frac{e^{rt_i} - 1}{r} \quad \forall 1 \leq i \leq n. \tag{24}$$

On constate que cette transformation est très similaire à celle de l'OU. Il est en effet possible de montrer que, sous certaines conditions, ces deux processus sont équivalents.

Proposition 4.1 (Équivalence de l'OU et de l'ACDC, Uyeda et al., 2015)**.** *Sur un arbre ultramétrique de hauteur h, un processus OU de variance à la racine $\gamma^2 = 0$, de variance σ^2 et de force de sélection α induit la même distribution marginale du trait aux feuilles de l'arbre qu'un processus ACDC de paramètres :*

$$\begin{cases} r = 2\alpha \\ \sigma_0^2 = \sigma^2 e^{-2\alpha h}. \end{cases} \tag{25}$$

On remarque que si l'on prend la définition stricte de l'OU, on a $\alpha > 0$, et l'on ne peut reproduire que le processus AC ($r > 0$). Si l'on s'autorise un « OU généralisé » où α peut être négative (c'est à dire, un processus tel que le trait est *repoussé* par la valeur β), on peut également reproduire le processus DC ou EB ($r < 0$). Cette observation, qui permet de voir l'ACDC comme un cas particulier de l'OU, peut être utile lors de l'analyse statistique d'un jeu de données (Aristide et al., 2018).

4.5 Évolution hétérogène

Dans tous les modèles décrits précédemment, on suppose qu'un unique processus règle l'évolution des traits continus sur l'ensemble de l'arbre phylogénétique. Cette hypothèse peut être raisonnable pour de petits arbres, mais devient difficile à justifier pour des phylogénies plus grandes, couvrant des classes entières du vivant, comme les mammifères ou les oiseaux (Meredith et al., 2011; Jetz et al., 2012). Il devient alors nécessaire de la relâcher, en autorisant par exemple certains paramètres du processus à changer au cours du temps. On définit alors des *régimes* différents sur certains clades de l'arbre, chacun ayant son propre mode d'évolution.

La figure 12 présente un exemple d'une telle étude, portant sur le groupe des singes du nouveau monde. Les traits étudiés ici correspondent à deux scores statistiques, choisis pour représenter au mieux les variations de forme du cerveau de ces singes (Aristide et al., 2016). On s'attend à ce que de tels traits fonctionnels soient influencés par la niche écologique dans laquelle évoluent les espèces étudiées (Butler and King, 2004). Dans le cadre de l'étude citée, ces niches évolutives sont définies par une combinaison de paramètres écologiques, dont le régime alimentaire (composé de feuilles, de fruits, de graines ou d'insectes), le mode de locomotion (quadrupédie arboricole, brachiation, etc.) et la taille typique des communautés (Aristide et al., 2016). Les différentes couleurs représentées sur l'arbre de la figure 12 recoupent largement ces différentes niches évolutives (Bastide et al., 2018a) : toutes les espèces d'une même couleur partagent des modes de vie similaires.

Ces régimes évolutifs peuvent être définis *a priori*, en fonction d'hypothèses biologiques adéquates, ou, ce qui est plus complexe d'un point de vue statistique, être inférés directement à partir des données.

4.5.1 Régimes préétablis

Si les régimes sont fixés à l'avance, et que l'on se donne un processus dont les paramètres sont constants sur chaque régime, il est possible d'écrire et de maximiser la vraisemblance du modèle complet (Clavel et al., 2015). On peut alors comparer la performance de plusieurs hypothèses biologiques bien choisies de positions des régimes sur l'arbre (voir la Section 4.7 sur la sélection de modèle).

L'outil le plus complet à notre connaissance est le progiciel R mvMORPH (Clavel et al., 2015). Ce progiciel permet d'utiliser tous les processus et transformations citées, dans un cadre multivarié, et avec la possibilité de définir des régimes a priori sur l'arbre. Cet outil est l'aboutissement d'une longue lignée de modèles similaires, dont on cite les principales étapes ci-après : Butler and King (2004) font uniquement varier l'optimum β d'un OU univarié ; O'Meara et al. (2006) la variance σ^2 d'un MB univarié ; et Beaulieu et al. (2012) l'optimum et la variance d'un OU univarié.

Le jeu de données présenté dans le paragraphe précédent et illustré figure 12 a par exemple été étudié dans un tel cadre par Aristide et al. (2016). Une étude comparative leur permet ainsi de privilégier un modèle d'OU avec sauts, c'est-à-dire un scénario dans lequel chaque groupe d'espèces, tels que définis figure 12, a sa propre forme optimale de cerveau, adaptée à sa niche écologique. Ce modèle s'avère plus en accord avec les données collectées qu'un

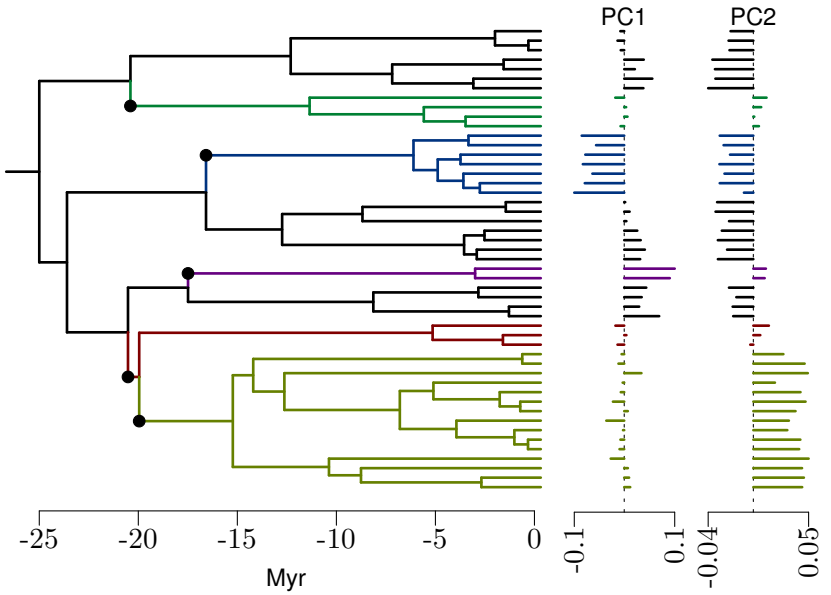


Fig. 12: Arbre phylogénétique et traits représentant la forme des cerveaux (composantes principales 1 et 2) pour le groupe des singes du nouveau monde (Aristide et al., 2016). Les clades en couleur représentent des régimes d'évolution différents, tels que inférés par PhylogeneticEM (Bastide et al., 2018a).

simple modèle stochastique de type EB (voir la section 4.4.1), où le trait évolue avec une grande variance au début de l'arbre, pour se stabiliser ensuite, et ainsi donner naissance à des clades d'espèces aux traits similaires mais pas nécessairement particulièrement bien adaptés à leur environnement spécifique.

Dans une optique un peu différente, Lemey et al. (2010) autorisent chaque branche b de l'arbre à avoir sa propre variance $\sigma_b^2 = \phi_b \sigma^2$ dans un MB univarié. Dans un cadre d'inférence bayésienne, les paramètres d'hétérogénéité ϕ_b sont alors tirés dans une loi d'émission donnée bien choisie (log-normale ou Cauchy, par exemple).

4.5.2 Détection automatique

La détection automatique de la position des régimes sur l'arbre est un problème plus complexe, en raison du grand nombre de modèles différents à considérer (toutes les partitions identifiables de l'arbre en k régimes différents, pour $1 \leq k \leq n + m$, où m est le nombre de nœuds internes de l'arbre, voir Bastide et al. 2017).

Mahler et al. (2013) ont proposé une première méthode pour détecter la position de sauts sur l'arbre dans le paramètre β. Cette approche est basée sur une heuristique, et permet également l'étude de phénomènes de convergence évolutive. Elle est limitée à des traits univariés ou indépendants. Eastman et al. (2011, 2013); Uyeda and Harmon (2014) ont proposé une méthode bayésienne, reposant sur un algorithme de Monte Carlo par chaînes de Markov à sauts réversibles (RJMCMC, Green, 1995), pour la détection de sauts, respectivement, dans la variance σ^2 d'un MB, la moyenne μ d'un MB et l'optimum β d'un OU univariés. En exploitant une formulation sous forme de modèle linéaire, et en utilisant la transformation décrite section 4.4.3, Khabbazian et al. (2016) ont proposé une méthode basée sur le LASSO (Tibshirani, 1996), pour des traits indépendants. En utilisant les mêmes outils, Bastide et al. (2018a) ont proposé une méthode de maximum de vraisemblance pour la détection de sauts dans ce même paramètre β pour un OU multivarié corrélé, mais tel que la matrice de force

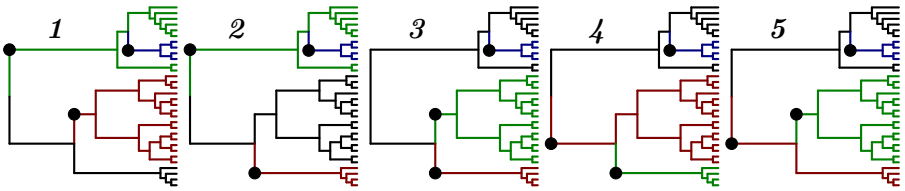


Fig. 13: Cinq allocations équivalentes de trois sauts (points noirs) sur un arbre de 32 espèces. La partition induite aux feuilles et effectivement observée est toujours la même, bien que les scénarios évolutifs représentés par la position des sauts peuvent être très différents.

de sélection **A** soit proportionnelle à l'identité.

C'est cette dernière méthode qui a été utilisée pour produire la classification présentée figure 12 (Bastide et al., 2018a). On constate, *a posteriori*, que cette classification est en accord avec la répartition en niches écologiques tels que décrite précédemment (Aristide et al., 2016). Cependant, cette relation n'est pas prise en compte par le modèle, qui n'utilise que la valeur des traits continus comme données, si bien que l'on ne peut tirer de conclusion sur l'impact réel de ces niches sur l'évolution du trait. Un modèle intégratif plus précis permettant d'étudier ce genre de questions reste à développer.

4.5.3 Identifiabilité

Dans la plupart des études comparatives, comme celle portant sur les singes du nouveau monde présentée ci-dessus, on n'a en général accès qu'aux traits mesurés aux feuilles de l'arbre, pour les espèces actuelles. Pour un arbre ultramétrique, on essaye ainsi d'obtenir des informations sur un processus d'évolution dynamique (tel que le processus stochastique avec sauts de la section précédente) à l'aide de mesures effectuées à un seul et unique point de temps. Une telle situation est propre à faire naître des problèmes d'identifiabilité. La figure 13 présente un exemple d'une telle situation. Si l'on modélise l'évolution d'un trait comme un OU pour lequel chaque clade d'une couleur différente sur l'arbre a un optimum β distinct, alors il est possible de montrer que les cinq coloriages présentés figure 13 induisent exactement la même distribution marginale du trait observée aux feuilles. En d'autres termes, ces cinq allocations de sauts auront la même vraisemblance, et ne seront donc pas distinguables à partir des seules observations.

Ces scénarios peuvent pourtant être bien distincts d'un point de vue biologique. Par exemple, dans le scénario 1, les quatre espèces en bas de l'arbre (en noir) sont dans l'état ancestral : au cours de toute la durée de l'évolution, leur niche écologique n'a jamais changé. En revanche, dans le scénario 4, ces mêmes espèces (en vert) ont subi deux changements successifs dans la valeur de leur optimum, pouvant représenter des événements écologiques tels que des migrations, ou des changements climatiques. Il est cependant impossible de distinguer ces deux scénarios à la seule vue des traits observés aujourd'hui. Seules des données fossiles, si elles sont disponibles, pourraient, en introduisant des mesures asynchrones, nous permettre de choisir entre ces scénarios. Ce problème, et ses implications pour l'inférence statistique de la position des sauts sur l'arbre telle que présentée dans la section précédente, a été étudié dans Bastide et al. (2017).

4.6 Modèles d'observation

Dans toutes les sections précédentes, on a supposé que la seule source de variabilité entre les observations étaient dues au processus stochastique sur l'arbre. L'interprétation qu'on a donnée des MB et OU (section 4.3) suppose même explicitement que, dans le cas macro

évolutif, on observe directement l'optimum secondaire du trait de chaque espèce, qui peut être approché par la moyenne du trait au sein d'une espèce. Cette hypothèse est en pratique souvent loin d'être vérifiée, et peut entraîner des biais importants dans les résultats d'une MPC (Silvestro et al., 2015; Cooper et al., 2016). Ce biais est particulièrement marqué pour des jeux de données hétérogènes, issus de collectes larges et faisant usage de plusieurs sources. Par exemple, Rose et al. (2016) fait une étude morphologique de la famille des mousses (Bryophyta) en se basant sur des spécimens collectés sur site, mais aussi sur une étude extensive de la littérature scientifique, de laquelle plus de milles planches manuscrites d'illustrations botaniques sont extraites. Une telle disparité dans la provenance des données doit être prise en compte dans les analyses statistiques.

Une première manière simple de prendre en compte cette variabilité est d'introduire, en plus de la variance phylogénétique $\sigma_p^2 \mathbf{C}_n$ due au processus stochastique sur l'arbre, une variance résiduelle $\sigma_e^2 \mathbf{I}_n$, qui représente l'effet indépendant de l'environnement sur chaque mesure. La matrice de variance totale aux feuilles $\mathbf{V}_n$ s'écrit alors :

$$\mathbf{V}_n = \sigma_p^2 \mathbf{C}_n + \sigma_e^2 \mathbf{I}_n. \tag{26}$$

Par analogie avec les modèles de génétique quantitative, la première variance est vue comme résultante d'effets dits « héritables », tandis que la seconde représente la variabilité dite « non héritable ». Ce modèle se retrouve dans la littérature sous le nom de *modèle mixte phylogénétique* (PMM, Lynch, 1991; Housworth et al., 2004).

Dans le cas d'un mouvement brownien sur un arbre ultramétrique de hauteur h, le ratio h_p^2 de la variance phylogénétique sur la variance totale est ainsi parfois désigné sous le nom d'« héritabilité phylogénétique » (Lynch, 1991) :

$$h_p^2 = \frac{\sigma_p^2 h}{\sigma_p^2 h + \sigma_e^2}. \tag{27}$$

Du fait de son interprétation biologique séduisante au premier abord, cette quantité a reçu beaucoup d'attention, notamment et de manière récente en virologie (voir par exemple Alizon et al., 2010; Leventhal and Bonhoeffer, 2016; Blanquart et al., 2017; Mitov and Stadler, 2018). Dans ce cadre simple, on peut constater que ce modèle est équivalent à une transformation λ de Pagel (voir section 2.2.2). Sous ce modèle, la variance totale aux feuilles $\mathbf{V}_n^\lambda$ s'écrit en effet (Leventhal and Bonhoeffer, 2016) :

$$\mathbf{V}_n^\lambda = \sigma^2 [\lambda \mathbf{C}_n + (1-\lambda) h \mathbf{I}_n]. \tag{28}$$

Ce qui revient au modèle (26), avec $\sigma^2 = \sigma_p^2 + \sigma_e^2/h$ et $\lambda = h_p^2$. Cette équivalence permet de voir que le paramètre λ, souvent interprété comme mesurant un « signal phylogénétique », peut aussi être interprété comme un rapport de variances représentant une « héritabilité phylogénétique ». Notons que cette équivalence formelle entre ces deux notions ne tient que pour un mouvement brownien simple, et n'est valable que pour la définition « statistique » du signal phylogénétique, tel que donnée par Revell et al. (2008) et rappelée dans la section 2.2.2.

Dans un cadre multivarié, un modèle similaire a été proposé par Ives et al. (2007), avec une variance environnementale supposée connue. Lorsque plusieurs observations sont disponibles pour chaque espèce, Felsenstein (2008) propose de voir ces observations individuelles comme des feuilles additionnelles sur l'arbre, connectées au nœud de leur espèce par des branches de longueur nulle. Ceci lui permet de proposer une extension des contrastes phylogénétiques (voir section 3.2) à cette situation plus complexe. Dans un cadre similaire, Goolsby et al. (2017) propose une procédure d'inférence rapide de la variance intra-spécifique se basant sur un algorithme développé par Ho and Ané (2014a).

De manière plus générale, Hadfield and Nakagawa (2010) reprennent et étendent l'analogie entre les MPC et les modèles de génétique quantitative basés sur l'étude de pédigrées. Plus spécifiquement, dans le cas d'un MB, ils montrent l'équivalence mathématique entre le calcul de la matrice de corrélations phylogénétiques $\mathbf{C}_n$ et la matrice de parenté $\mathbf{A}$ (Crow and Kimura, 1970, Chap. 4) calculée avec des coefficients de consanguinité égaux aux longueurs de branches (voir aussi Bastide et al. 2018b pour une explicitation de cette relation dans le cas plus général d'un réseau phylogénétique). Le modèle général décrit dans Hadfield and Nakagawa (2010) s'écrit alors

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{P} + \mathbf{E} + \mathbf{M}, \tag{29}$$

où :

- $\mathbf{Y}$ est la matrice d'observation de taille $n \times p$, où n est le nombre d'observations, et p la dimension du trait ;
- $\boldsymbol{\mu}$ est la matrice des effets fixes $(n \times p)$;
- $\mathbf{P}$ est la matrice des effets aléatoires, dont la covariance est dictée par le modèle stochastique d'évolution sur l'arbre comme vu précédemment ;
- $\mathbf{E}$ est la matrice des variations résiduelles, de variance σ_e^2 indépendantes pour toutes les observations et tous les traits ;
- $\mathbf{M}$ est la matrice d'hétérogénéité, représentant les erreurs spécifiques à chaque groupe de mesures, liées à leur provenance. Les observations peuvent être groupées par espèces, mais pas nécessairement. Elle permet de prendre en compte l'effet de « méta-analyse », en homogénéisant des données issues de sources variées.

L'inférence d'un tel modèle peut se faire en empruntant des techniques issues de la génétique quantitative (voir par exemple Henderson, 1976 et Thompson, 2000 pour une revue) et peut s'adapter à des modèles non gaussiens en recourant à des techniques de types MCMC (Hadfield and Nakagawa, 2010).

4.7 Sélection de modèle

De nombreux modèles ont été présentés dans cette section. Chacun capture des caractéristiques différentes des processus évolutifs à l'œuvre et il peut être difficile de choisir en pratique le *meilleur* modèle pour un problème donné. Ce problème n'est pas spécifique à l'étude de traits continus. Il a déjà été rencontré pour les traits discrets : choix du modèle d'évolution pour les nucléotides (JC, HKY, GTR, etc.), de la matrice de transition pour les acides aminés (BLOSUM, LG, etc.), inclusion d'une composante $+\Gamma$ (voir chapitre XXXX).

Comme dans le cas discret, l'approche la plus courante consiste à utiliser des critères de sélection de modèle basés sur une vraisemblance pénalisée, de type AIC (« *Akaike's Information Criterion* », Akaike, 1974) ou BIC (« *Bayesian Information Criterion* »). Ces critères sélectionnent le compromis optimal entre complexité du modèle et ajustement aux données et permettent de choisir le meilleur modèle parmi une liste choisie à l'avance (par exemple : OU, MB, EB ou ACDC). Des versions plus sophistiquées de ces critères permettent aussi de sélectionner le nombre optimal de régimes dans le cadre de la détection automatique de changement de régimes (section 4.5). Le lecteur intéressé pourra se référer à Giraud (2014) pour une introduction à ces techniques de sélection de modèle.

5 Extensions et généralisations

La section 4 se concentre sur les modèles gaussiens, avec un nombre modéré de traits, sans interactions entre traits ou traits et phylogénie, et pour lesquels on sait généralement écrire la vraisemblance et ajuster le modèle de façon efficace. Nous présentons ici quelques

extensions à des grandes classes de processus non-gaussiens, avec interactions ou en grande dimension.

5.1 Modèles non gaussiens

Le processus de Lévy peut-être vu comme un MB (éventuellement avec tendance) perturbé par un processus de Poisson. Le processus de Poisson induit des sauts de position et d'intensité aléatoire censés capturer une évolution de type simpsonienne. Ce modèle a été utilisé dans un cadre bayésien pour analyser des données morphologiques de grands singes (Landis et al., 2013) et identifier la position la plus probable de sauts adaptatifs. La vraisemblance du modèle se factorise sur les branches de l'arbre, à l'instar des modèles gaussiens et peut donc être calculée via l'algorithme récursif décrit en section 2.3. Cependant, et contrairement aux processus gaussiens, la vraisemblance du processus de Lévy le long d'une branche n'est pas analytique et nécessite une étape d'intégration numérique.

Boucher et al. (2018) propose une extension à des modèles plus généraux dans laquelle l'évolution du trait n'est plus régie par des processus simples (comme le MB, le processus OU ou le processus de Lévy) mais par des équations aux dérivées partielles de type Fokker - Planck. Ces dernières sont largement utilisées en génétique des populations et permettent de modéliser des phénomènes complexes (tendances évolutives, adaptations rapides, existence de multiples valeurs optimales, etc). Ces modèles souffrent, comme les processus de Lévy, de limitations techniques (uniquement disponibles pour des traits univariés, estimation longue) mais constituent un champ de recherche actif.

5.2 Interactions entre l'arbre et le trait

Tous les modèles présentés jusqu'ici supposent que le trait évolue sur un arbre *fixe* connu *a priori*. La famille de modèles dits *xxSSE* (pour *State Speciation Extinction*) relâche cette hypothèse : la valeur du trait influence désormais les taux de de spéciation et d'extinction et donc *in fine* la topologie de l'arbre. Maddison et al. (2007) introduisent BiSSE (pour *Binary SSE*), le premier représentant de la famille. Dans ce modèle, un trait latent binaire (0/1) évolue sur un arbre connu et complet selon une chaîne de Markov à temps continu avec taux de transitions q_{01} et q_{10}. Une lignée dans l'état i ($i \in \{0,1\}$) a un taux de spéciation λ_i et d'extinction μ_i. Conditionnellement au processus binaire, le trait est indépendant des évènements de spéciation et d'extinction. Maddison et al. (2007) décrivent un système d'équations différentielles qui peuvent être intégrées numériquement pour calculer la vraisemblance des données en utilisant les même algorithmes récursifs que précédemment.

De nombreux représentants ont enrichi la famille depuis : FitzJohn et al. (2009) a montré comment ajuster le modèle même si l'arbre n'est pas complètement observé, avant de l'étendre à des traits latents catégoriels (FitzJohn, 2010) ou continus (FitzJohn, 2012) et Goldberg et al. (2011) en a proposé une version biogéographique. Enfin, Rabosky and Huang (2016) ont développé un test robuste, applicable au delà des modèles *xxSSE*, pour détecter de la dépendance entre valeur de trait et taux de spéciation.

Dans une autre optique, il est possible de modéliser conjointement l'évolution des séquences moléculaires et des traits continus, dans un cadre bayésien intégratif. L'arbre phylogénétique est alors inféré en utilisant conjointement ces deux sources d'information, et il n'est pas nécessaire de procéder à une étude en deux étapes, comme on l'a fait jusqu'à présent (en inférant d'abord l'arbre à partir des séquences, puis en utilisant cet arbre comme donnée pour étudier les traits quantitatifs). Ces approches, parfois qualifiées en anglais de « total evidence », reposent sur l'hypothèse fondamentale que, conditionnellement à l'arbre phylogénétique, les deux processus d'évolution des séquences et des traits continus sont indépendants (Lemey et al., 2010; Pybus et al., 2012; Cybis et al., 2015).

5.3 Interactions entre espèces

Dans les modèles gaussiens, de Lévy ou *xxSSE*, l'influence des autres espèces sur une espèce d'intérêt est implicite. Elle peut se lire dans des changements de valeur du trait mais n'est pas codée explicitement. En particulier, il est impossible de modéliser explicitement des traits dont la valeur optimale dépend des traits d'autres espèces. Manceau et al. (2016) ont proposé un cadre conceptuel très général pour estimer des interactions entre espèces, qui étend considérablement les modèles gaussiens. Ce cadre considère que les traits de *l'ensemble des espèces* existant à un instant t dans le passé évoluent comme un processus gaussien multivarié. L'idée est séduisante et permet notamment de modéliser de la coévolution entre espèces. Ce modèle souffre cependant comme ceux évoquées dans cette section d'une complexité calculatoire élevée. Le temps depuis la racine doit être découpé en segments – délimités par les évènements de spéciation et d'extinction – et la vraisemblance doit être calculée différemment sur chaque segment. De plus, si n espèces existent sur un segment donné, remplacer n traits indépendants de taille p par un unique trait de taille np fait passer la complexité de $O(np^{2.373})$ à $O((np)^{2.373})$.

L'étude de ce type de processus avec interactions est cependant un domaine actif de recherche, et plusieurs méthodes ont été proposées pour modéliser des types d'interactions spécifiques, mutualistes ou de compétition (voir par exemple Nuismer and Harmon, 2015; Drury et al., 2016; Bartoszek et al., 2017; Drury et al., 2018; Aristide and Morlon, 2019).

5.4 Trait de grande dimension

Les algorithmes récursifs de calcul de vraisemblance sont linéaires en n mais supra quadratique en p et nécessitent donc des adaptations pour les traits de grande dimension. Nous en présentons deux : l'utilisation de facteurs latents et les approches de type pseudo-vraisemblance.

5.4.1 Modèle à facteurs et données mixtes

Lorsque de nombreux traits sont mesurés, il est probable que certains soient très corrélés et redondants. Une méthode pour réduire la dimension du processus sur l'arbre est d'utiliser une « analyse à facteur phylogénétique » (Tolkoff et al., 2018). Dans ce modèle, on suppose que les p traits d'origine se déduisent linéairement d'un nombre restreint $q \ll p$ de facteurs, qui évoluent comme un MB sur l'arbre, à l'aide d'une matrice de poids.

D'un point de vue algorithmique, cela revient à effectuer la récursion sur les q facteurs latents et permet de se ramener à une complexité $\mathcal{O}(nq^{2.373}+qp)$. Le nombre de traits latents q est choisi par une méthode de vraisemblance marginale dans Tolkoff et al. (2018). La formulation à base de facteurs latents permet également de traiter de façon unifiée les données discrètes et continues (Cybis et al., 2015), à l'aide d'un modèle à seuil (en anglais, « *threshold model* », dit aussi « *latent liability model* », Felsenstein, 2005, 2012). Le trait observé est alors soit une combinaison linéaire des facteurs (cas continu) soit une discrétisation de cette combinaison (cas discret). Le trait discret ne change de valeur que si la combinaison latente franchit un seuil. Un tel processus permet de modéliser dans un cadre latent markovien un caractère discret « avec mémoire », c'est-à-dire non markovien (Goldberg and Foo, 2020).

5.4.2 Pseudo-vraisemblances

D'autres approches ont été proposées récemment pour traiter directement des données de grande dimension, sans passer par des traits latents. Goolsby (2016) utilise une pseudo-vraisemblance composite par paires (« *pairwise composite likelihood* ») sur les traits pour n'avoir à traiter dans les calculs que des données de dimension deux. Une procédure de

Progiciel	Modèles et Méthodes	Publication
ape	Manipulation d'arbres (utilisé par tous les autres progiciels cités), contrastes	Paradis et al. (2004)
phytools	ACPp et autres MPC classiques	Revell (2012)
geiger	MPC univariées (dont MB, OU, ACDC et transformations de Pagel)	Pennell et al. (2014)
mvMORPH	MPC multivariées (avec sauts connus)	Clavel et al. (2015)
phylolm	Régression phylogénétique généralisée	Ho and Ané (2014a)
Rphylopars	MB et OU avec variance intra-spécifique	Goolsby et al. (2015)
ouch	OU avec sauts (connus) dans β	Butler and King (2004)
OUwie	OU avec sauts (connus) dans β ou σ^2	Beaulieu et al. (2012)
mvSLOUCH	OU, OUBM et OUOU multivariés	Bartoszek et al. (2012)
bayou	OU avec sauts (inconnus) dans β	Uyeda et al. (2015)
l1ou	OU multivarié indépendant avec sauts (inconnus) dans β	Khabbazian et al. (2016)
PhylogeneticEM	MB, OU multivariés avec avec sauts (inconnus) dans β	Bastide et al. (2017)
PCMbase	Modèles gaussiens, avec sauts (connus) dans les différents paramètres	Mitov et al. (2020)
Diversitree	BiSSE, MultSSE, GeoSSE	FitzJohn (2012)
BAMMtools	ACDC avec sauts (inconnus) dans σ_0 et variante SSE	Rabosky (2014)

Tab. 1: Progiciels R cités dans ce chapitre.

test statistique pour choisir un modèle dans ces conditions est proposé. Clavel et al. (2019) proposent quant à eux une procédure de vraisemblance régularisée basée sur des pénalités de type L_1 ou L_2. Cela leur permet notamment d'inférer des matrices de covariances $\mathbf{R}$ parcimonieuses, qui imposent un faible nombre d'interactions entre les traits.

6 Références utiles

Le tableau 1 référence l'ensemble des progiciels `R` cités dans ce chapitre. Cette liste n'a cependant pas vocation à être exhaustive, et l'on renvoie le lecteur intéressé à la « *Task View* » du CRAN dédié aux méthodes phylogénétiques comparatives pour un tour d'horizon plus complet des outils disponibles : `cran.r-project.org/web/views/Phylogenetics.html`

Les deux ouvrages suivants (en anglais) constituent également de bonnes ressources complémentaires sur le sujet :

- Felsenstein (2004) pour une introduction générale des modèles d'évolution de traits continus, voir les chapitres 23 à 25.
- Harmon (2019) pour une présentation plus à jour des dernières méthodes. L'ouvrage est publié sous une licence libre (CC-BY-4.0), et est disponible gratuitement en ligne à l'adresse : `lukejharmon.github.io/pcm/`.

7 Remerciements

Les auteurs remercient Héloïse Bastide et Amir Yassin, premiers lecteurs de ce texte, et qui ont contribué grandement à son amélioration linguistique et scientifique. Nous remercions également Gilles Didier et Stéphane Guindon, coordinateurs de l'ouvrage, et sans qui

ce document n'aurait jamais vu le jour. PB a bénéficié d'un financement du *Fonds Wetenschappelijk Onderzoek* (FWO, Belgique) lors de son post-doctorat à l'université KU Leuven (Department of Microbiology and Immunology, Rega Institute).

Références

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6) :716–723.

Alizon, S., von Wyl, V., Stadler, T., Kouyos, R. D., Yerly, S., Hirschel, B., Böni, J., Shah, C., Klimkait, T., Furrer, H., Rauch, A., Vernazza, P. L., Bernasconi, E., Battegay, M., Bürgisser, P., Telenti, A., Günthard, H. F., and Bonhoeffer, S. (2010). Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathogens*, 6(9).

Aristide, L., Bastide, P., dos Reis, S. F., Pires dos Santos, T. M., Lopes, R. T., and Perez, S. I. (2018). Multiple factors behind early diversification of skull morphology in the continental radiation of New World monkeys. *Evolution*, 72(12) :2697–2711.

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.

Aristide, L. and Morlon, H. (2019). Understanding the effect of competition during evolutionary radiations : an integrated model of phenotypic and species diversification. *Ecology Letters*, page ele.13385.

Bartoszek, K., Glémin, S., Kaj, I., and Lascoux, M. (2017). Using the ornstein–uhlenbeck process to model the evolution of interacting populations. *Journal of Theoretical Biology*, 429 :35–45.

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S., and Hansen, T. F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *Journal of Theoretical Biology*, 314 :204–215.

Bastide, P., Ané, C., Robin, S., and Mariadassou, M. (2018a). Inference of Adaptive Shifts for Multivariate Correlated Traits. *Systematic Biology*, 67(4) :662–680.

Bastide, P., Mariadassou, M., and Robin, S. (2017). Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 79(4) :1067–1093.

Bastide, P., Solís-Lemus, C., Kriebel, R., Sparks, K. W., and Ané, C. (2018b). Phylogenetic comparative methods on phylogenetic networks with reticulations. *Systematic Biology*, 67(5) :800–820.

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C., and O'Meara, B. C. (2012). Modeling stabilizing selection : Expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, 66(8) :2369–2383.

Blanquart, F., Wymant, C., Cornelissen, M., Gall, A., Bakker, M., Bezemer, D., Hall, M., Hillebregt, M., Ong, S. H., Albert, J., Bannert, N., Fellay, J., Fransen, K., Gourlay, A. J., Grabowski, M. K., Gunsenheimer-Bartmeyer, B., Günthard, H. F., Kivelä, P., Kouyos, R., Laeyendecker, O., Liitsola, K., Meyer, L., Porter, K., Ristola, M., van Sighem, A., Vanham, G., Berkhout, B., Kellam, P., Reiss, P., and Fraser, C. (2017). Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biology*, 15(6) :1–26.

Blomberg, S. P., Garland, T., and Ives, A. R. (2003). Testing for Phylogenetic Signal in Comparative Data : Behavioral Traits Are More Labile. *Evolution*, 57(4) :717–745.

Boucher, F. C., Démery, V., Conti, E., Harmon, L. J., and Uyeda, J. (2018). A General Model for Estimating Macroevolutionary Landscapes. *Systematic Biology*, 67(2) :304–319.

Butler, M. A. and King, A. A. (2004). Phylogenetic Comparative Analysis : A Modeling Approach for Adaptive Evolution. *The American Naturalist*, 164(6) :683–695.

Ceccarelli, F. S., Koch, N. M., Soto, E. M., Barone, M. L., Arnedo, M. A., and Ramírez, M. J. (2018). The Grass was Greener : Repeated Evolution of Specialized Morphologies and Habitat Shifts in Ghost Spiders Following Grassland Expansion in South America. *Systematic Biology*, 68(1) :63–77.

Clavel, J., Aristide, L., and Morlon, H. (2019). A Penalized Likelihood Framework for High-Dimensional Phylogenetic Comparative Methods and an Application to New-World Monkeys Brain Evolution. *Systematic Biology*, 68(1) :93–116.

Clavel, J., Escarguel, G., and Merceron, G. (2015). mvmorph : an r package for fitting multivariate evolutionary models to morphometric data. *Methods in Ecology and Evolution*, 6(11) :1311–1319.

Cooper, N., Thomas, G. H., Venditti, C., Meade, A., and Freckleton, R. P. (2016). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biological Journal of the Linnean Society*, 118(1) :64–77.

Crow, J. F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper & Row, New York.

Cybis, G. B., Sinsheimer, J. S., Bedford, T., Mather, A. E., Lemey, P., and Suchard, M. A. (2015). Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. *The Annals of Applied Statistics*, 9(2) :969–991.

Drury, J., Clavel, J., Manceau, M., and Morlon, H. (2016). Estimating the Effect of Competition on Trait Evolution Using Maximum Likelihood Inference. *Systematic Biology*, 65(4) :700–710.

Drury, J. P., Grether, G. F., Garland, T., and Morlon, H. (2018). An Assessment of Phylogenetic Tools for Analyzing the Interplay Between Interspecific Interactions and Phenotypic Evolution. *Systematic Biology*, 67(3) :413–427.

Duchen, P., Leuenberger, C., Szilágyi, S. M., Harmon, L. J., Eastman, J. M., Schweizer, M., and Wegmann, D. (2017). Inference of Evolutionary Jumps in Large Phylogenies using Lévy Processes. *Systematic Biology*, 66(6) :950–963.

Eastman, J. M., Alfaro, M. E., Joyce, P., Hipp, A. L., and Harmon, L. J. (2011). A Novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution*, 65(12) :3578–3589.

Eastman, J. M., Wegmann, D., Leuenberger, C., and Harmon, L. J. (2013). Simpsonian 'Evolution by Jumps' in an Adaptive Radiation of Anolis Lizards. *arXiv e-print*, (1305.4216).

Felsenstein, J. (1973). Maximum Likelihood and Minimum-Steps Methods for Estimating Evolutionary Trees from Data on Discrete Characters. *Systematic Biology*, 22(3) :240–249.

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

Felsenstein, J. (2004). *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

Felsenstein, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 360(1459) :1427–1434.

Felsenstein, J. (2008). Comparative Methods with Sampling Error and Within-Species Variation : Contrasts Revisited and Revised. *The American Naturalist*, 171(6) :713–725.

Felsenstein, J. (2012). A Comparative Method for Both Discrete and Continuous Characters Using the Threshold Model. *The American Naturalist*, 179(2) :145–156.

FitzJohn, R. G. (2010). Quantitative traits and diversification. *Systematic Biology*, 59(6) :619–633.

FitzJohn, R. G. (2012). Diversitree : Comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, 3(6) :1084–1092.

FitzJohn, R. G., Maddison, W. P., and Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, 58(6) :595–611.

Freckleton, R. P. (2012). Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution*, 3(5) :940–947.

Giraud, C. (2014). *Introduction to High-Dimensional Statistics*. CRC Press, Hoboken, NJ.

Goldberg, E. E. and Foo, J. (2020). Memory in trait macroevolution. *The American Naturalist*, 195(2) :705992.

Goldberg, E. E., Lancaster, L. T., and Ree, R. H. (2011). Phylogenetic Inference of Reciprocal Effects between Geographic Range Evolution and Diversification. *Systematic Biology*, 60(4) :451–465.

Goolsby, E. W. (2016). Likelihood-Based Parameter Estimation for High-Dimensional Phylogenetic Comparative Models : Overcoming the Limitations of "Distance-Based" Methods. *Systematic Biology*, 65(5) :852–870.

Goolsby, E. W., Bruggeman, J., and Ané, C. (2015). R-PhyloPars : an R package for Phylogenetic Comparative Methods for Missing Data and Within-Species Variation.

Goolsby, E. W., Bruggeman, J., and Ané, C. (2017). Rphylopars : fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, 8(1) :22–27.

Grafen, A. (1989). The Phylogenetic Regression. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 326(1233) :119–157.

Grafen, A. (1992). The uniqueness of the phylogenetic regression. *Journal of Theoretical Biology*, 156(4) :405–423.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4) :711–732.

Hadfield, J. D. and Nakagawa, S. (2010). General quantitative genetic methods for comparative biology : phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology*, 23(3) :494–508.

Hansen, T. F. (1997). Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution*, 51(5) :1341.

Hansen, T. F. and Houle, D. (2004). Evolvability, Stabilizing Selection, and the problem of stasis. In Pigliucci, M. and Preston, K., editors, *Phenotypic integration : studying the ecology and evolution of complex phenotypes*, chapter 6, pages 130–154. Oxford University Press, New York.

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.

Hansen, T. F. and Orzack, S. H. (2005). Assessing Current Adaptation and Phylogenetic Inertia as Explanations of Trait Evolution : The Need for Controlled Comparisons. *Evolution*, 59(10) :2063–2072.

Hansen, T. F., Pienaar, J., and Orzack, S. H. (2008). A Comparative Method for Studying Adaptation to a Randomly Evolving Environment. *Evolution*, 62(8) :1965–1977.

Harmon, L. J. (2019). *Phylogenetic Comparative Methods : Learning From Trees*. Center for Open Science, version 1. edition.

Harmon, L. J., Losos, J. B., Davies, T. J., Gillespie, R. G., Gittleman, J. L., Jennings, W. B., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J., Purvis, A., Ricklefs, R. E., Schluter, D., James A Schulte, I. I., Seehausen, O., Sidlauskas, B. L., Torres-Carvajal, O., Weir, J. T., and Mooers, A. O. (2010). EARLY BURSTS OF BODY SIZE AND SHAPE EVOLUTION ARE RARE IN COMPARATIVE DATA. *Evolution*, 64(8) :2385–2396.

Henderson, C. R. (1976). A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. *Biometrics*, 32(1) :69–83.

Hiscott, G., Fox, C., Parry, M., and Bryant, D. (2016). Efficient Recycled Algorithms for Quantitative Trait Models on Phylogenies. *Genome Biology and Evolution*, 8(5) :1338–1350.

Ho, L. S. T. and Ané, C. (2014a). A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. *Systematic Biology*, 63(3) :397–408.

Ho, L. S. T. and Ané, C. (2014b). Intrinsic inference difficulties for trait evolution with Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 5(11) :1133–1146.

Housworth, E. a., Martins, E. P., and Lynch, M. (2004). The phylogenetic mixed model. *The American Naturalist*, 163(1) :84–96.

Hunt, G., Bell, M. A., and Travis, M. P. (2008). Evolution toward a New Adaptive Optimum : Phenotypic Evolution in a Fossil Stickleback Lineage. *Evolution*, 62(3) :700–710.

Hunt, G. and Rabosky, D. L. (2014). Phenotypic Evolution in Fossil Species : Pattern and Process. *Annual Review of Earth and Planetary Sciences*, 42(1) :421–441.

Ives, A. R., Midford, P. E., Garland, T., and Oakley, T. (2007). Within-Species Variation and Measurement Error in Phylogenetic Comparative Methods. *Systematic Biology*, 56(2) :252–270.

Jaffe, A. L., Slater, G. J., and Alfaro, M. E. (2011). The evolution of island gigantism and body size variation in tortoises and turtles. *Biology Letters*, 7(4) :558–561.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424) :444–448.

Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 7(7) :811–824.

Kim, H. and Perl, J. (1983). A computational model for combined causal and diagnostic reasoning in inference systems. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*. Morgan-Kaufmann, San Mateo, CA.

Labra, A., Pienaar, J., and Hansen, T. F. (2009). Evolution of Thermal Physiology in Liolaemus Lizards : Adaptation, Phylogenetic Inertia, and Niche Tracking. *The American Naturalist*, 174(2) :204–220.

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution*, 30(2) :314.

Landis, M. J., Schraiber, J. G., and Liang, M. (2013). Phylogenetic analysis using Lévy processes : Finding jumps in the evolution of continuous traits. *Systematic Biology*, 62(2) :193–204.

Lartillot, N. (2014). A phylogenetic Kalman filter for ancestral trait reconstruction using molecular data. *Bioinformatics*, 30(4) :488–496.

Law, C. J., Slater, G. J., and Mehta, R. S. (2018). Lineage Diversity and Size Disparity in Musteloidea : Testing Patterns of Adaptive Radiation Using Molecular and Fossil-Based Methods. *Systematic Biology*, 67(1) :127–144.

Lemey, P., Rambaut, A., Welch, J. J., and Suchard, M. A. (2010). Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, 27(8) :1877–1885.

Leventhal, G. E. and Bonhoeffer, S. (2016). Potential Pitfalls in Estimating Viral Load Heritability. *Trends in Microbiology*, 24(9) :687–698.

Lynch, M. (1991). Methods for the Analysis of Comparative Data in Evolutionary Biology. *Evolution*, 45(5) :1065–1080.

Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Systematic biology*, 56(5) :701–710.

Mahler, D. L., Ingram, T., Revell, L. J., and Losos, J. B. (2013). Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. *Science*, 341(6143) :292–295.

Manceau, M., Lambert, A., and Morlon, H. (2016). A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Systematic Biology*, 66(4) :syw115.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate analysis*. Probability and mathematical statistics. Academic Press, New York.

Méléard, S. (2016). *Modèles aléatoires en Ecologie et Evolution*, volume 77 of *Mathématiques et Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg.

Meredith, R. W., Janecka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simao, T. L. L., Stadler, T., Rabosky, D. L., Honeycutt, R. L., Flynn, J. J., Ingram, C. M., Steiner, C., Williams, T. L., Robinson, T. J., Burk-Herrick, A., Westerman, M., Ayoub, N. A., Springer, M. S., and Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science*, 334(6055) :521–524.

Meucci, A. (2009). Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck. *SSRN Electronic Journal*, page 20.

Mitov, V., Bartoszek, K., Asimomitis, G., and Stadler, T. (2020). Fast likelihood calculation for multivariate Gaussian phylogenetic models with shifts. *Theoretical Population Biology*, 131 :66–78.

Mitov, V. and Stadler, T. (2018). A Practical Guide to Estimating the Heritability of Pathogen Traits. *Molecular Biology and Evolution*, 35(3) :756–772.

Nuismer, S. L. and Harmon, L. J. (2015). Predicting rates of interspecific interaction from phylogenetic trees. *Ecology Letters*, 18(1) :17–27.

O'Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5) :922–933.

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

Paradis, E., Claude, J., and Strimmer, K. (2004). APE : Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20(2) :289–290.

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., and Harmon, L. J. (2014). geiger v2.0 : an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15) :2216–2218.

Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W., Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P., and Delwart, E. L. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences*, 109(37) :15066–15071.

Rabosky, D. L. (2014). Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE*, 9(2).

Rabosky, D. L. and Huang, H. (2016). A Robust Semi-Parametric Test for Detecting Trait-Dependent Diversification. *Systematic Biology*, 65(2) :181–193.

Raz, R. (2003). On the Complexity of Matrix Product. *SIAM Journal on Computing*, 32(5) :1356–1369.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution*, 63(12) :3258–3268.

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

Revell, L. J. (2012). phytools : An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2) :217–223.

Revell, L. J., Harmon, L. J., and Collar, D. C. (2008). Phylogenetic Signal, Evolutionary Process, and Rate. *Systematic Biology*, 57(4) :591–601.

Rose, J. P., Kriebel, R., and Sytsma, K. J. (2016). Shape analysis of moss (Bryophyta) sporophytes : Insights into land plant evolution. *American Journal of Botany*, 103(4) :652–662.

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B., and Salamin, N. (2015). Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods in Ecology and Evolution*, 6(3) :340–346.

Slater, G. J. and Pennell, M. W. (2014). Robust Regression and Posterior Predictive Simulation Increase Power to Detect Early Bursts of Trait Evolution. *Systematic Biology*, 63(3) :293–308.

Stuart, Y. E., Campbell, T. S., Hohenlohe, P. A., Reynolds, R. G., Revell, L. J., and Losos, J. B. (2014). Rapid evolution of a native species following invasion by a congener. *Science*, 346(6208) :463–466.

Thompson, E. A. (2000). *Statistical Inference from Genetic Data on Pedigrees*, volume 6. Institute of Mathematical Statistics, Beachwood, nsf-cbms r edition.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1) :267–288.

Tolkoff, M. R., Alfaro, M. E., Baele, G., Lemey, P., and Suchard, M. A. (2018). Phylogenetic Factor Analysis. *Systematic Biology*, 67(3) :384–399.

Uyeda, J. C., Caetano, D. S., and Pennell, M. W. (2015). Comparative Analysis of Principal Components Can be Misleading. *Systematic Biology*, 64(4) :677–689.

Uyeda, J. C. and Harmon, L. J. (2014). A Novel Bayesian Method for Inferring and Interpreting the Dynamics of Adaptive Landscapes from Phylogenetic Comparative Data. *Systematic Biology*, 63(6) :902–918.

Vrancken, B., Lemey, P., Rambaut, A., Bedford, T., Longdon, B., Günthard, H. F., and Suchard, M. A. (2015). Simultaneously estimating evolutionary history and repeated traits

phylogenetic signal : Applications to viral and host phenotypic evolution. *Methods in Ecology and Evolution*, 6(1) :67–82.