

Construire un modèle individu centré en macro-évolution

Marc Manceau

Mars 2015

Introduction

Objectif des études de macro-évolution

Approches espèce-centrée ou individu-centrée

NTB utilisée en macro-évolution

Plusieurs propositions de modèles

Support empirique

Relier la généalogie et la phylogénie : SGD

Les hypothèses du modèle

Simuler et inférer sous ce modèle

Conclusion

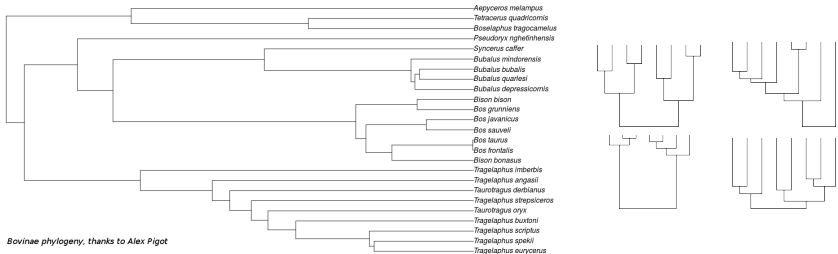
Résumé

Perspectives



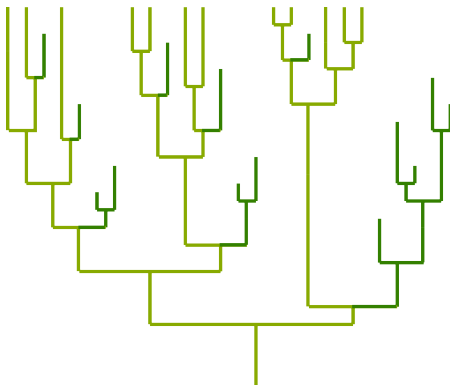
Patterns de macro-évolution

- Objectif : utiliser une phylogénie pour inférer ce qui s'est passé au cours de l'évolution.
- Réalisation : créer un modèle probabiliste produisant des phylogénies.
- Evaluation : comparer la forme des phylogénies produites aux phylogénies empiriques.



- Les phylogénies empiriques sont *stemmy* et *unbalanced*.

Modèles centrés sur les lignées



- ▶ Naissance-mort des particules "espèces"
- ▶ Taux constants $\rightarrow$ arbres trop *tippy* et *balanced*
- ▶ Arbres *stemmy* $\Rightarrow$ diminution du taux de speciation au présent
- ▶ Arbres *unbalanced* $\Rightarrow$ les taux changent dans des sous-arbres

Introduction

Objectif des études de macro-évolution

Approches espèce-centrée ou individu-centrée

NTB utilisée en macro-évolution

Plusieurs propositions de modèles

Support empirique

Relier la généalogie et la phylogénie : SGD

Les hypothèses du modèle

Simuler et inférer sous ce modèle

Conclusion

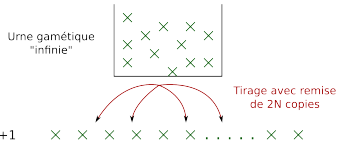
Résumé

Perspectives

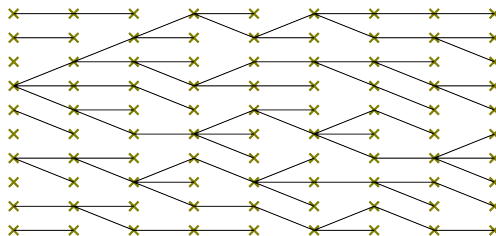
Modèle généalogique de Wright-Fisher

1. Taille de métapopulation N constante.
2. Tirage aléatoire des relations de parenté.

Génération k × × × × × × × × $2N$ copies



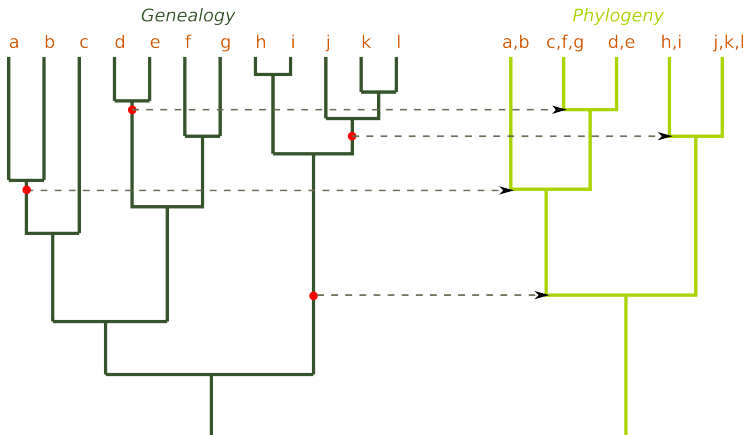
Génération $k+1$ × × × × × × × × $2N$ copies



Speciation par mutation ponctuelle

la première proposition d'Hubbell

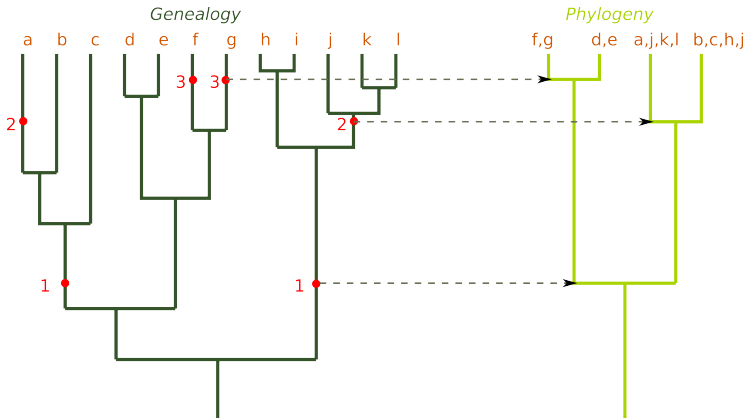
- Processus de Poisson ponctuel sur la généalogie.
- Chaque individu muté donne une nouvelle espèce.



Mode de spéciation par fission

chaque espèce est fondée par un groupe d'individus

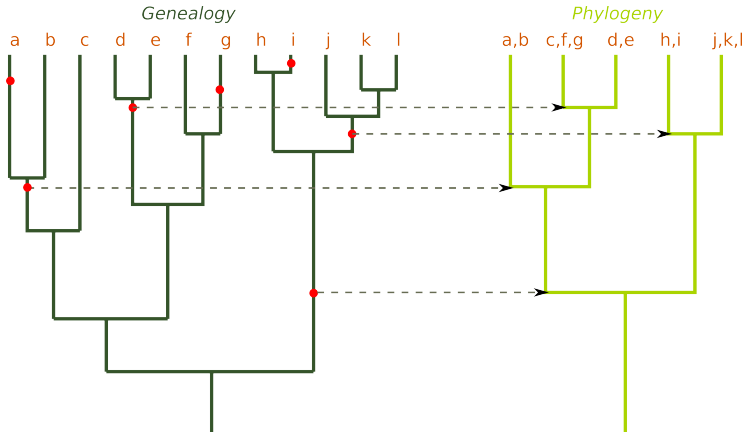
- ▶ Le taux de spéciation est une fonction linéaire de l'abondance d'une espèce.
- ▶ La nouvelle espèce est formée par fission de l'espèce fondatrice.



Mode de spéciation protractée

les espèces trop jeunes ne sont pas visibles

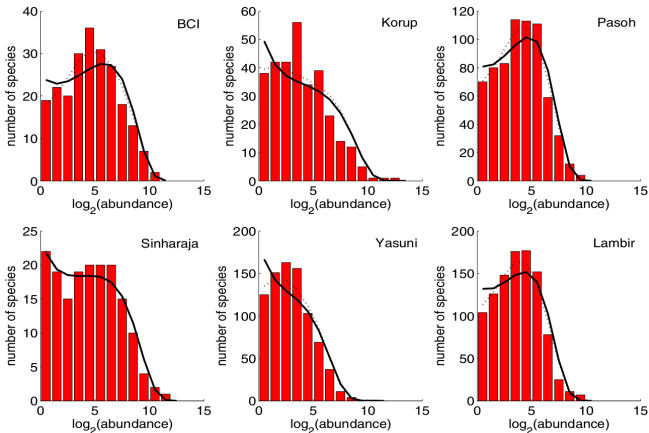
- Processus de Poisson ponctuel sur la généalogie → espèce naissante.
- Une espèce naissante ne devient une vraie espèce qu'après un certain temps.



Distribution d'abondance des espèces sous NTB

Etienne & Haegeman, Theoretical Ecology, 2011

- ▶ Bon fit aux SADs.
- ▶ Spéciation par mutation ponctuelle (pointillés) plus réaliste que par fission (traits pleins).

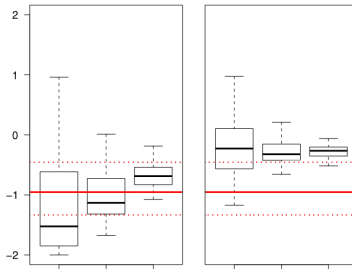


Forme d'arbres non-réaliste

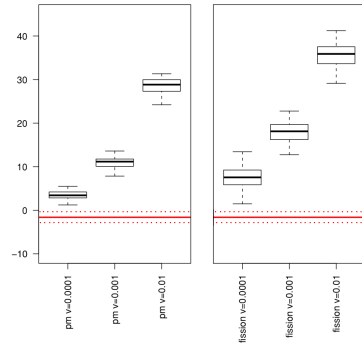
Davies et al., Evolution, 2011

- Bonne *imbalance*.
- ...mais les phylogénies empiriques sont plus *stemmy*.

Index of tree imbalance



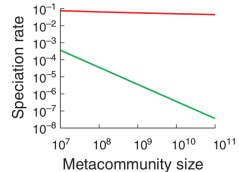
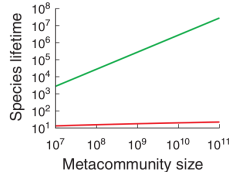
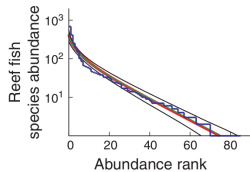
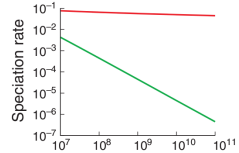
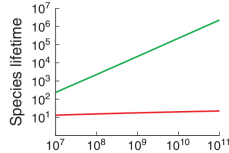
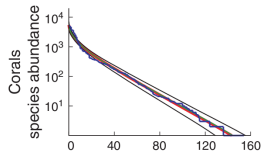
Index of the branch-length distribution



Amélioration apportée par la spéciation protractée

Rosindell et al., Ecology Letters, 2010

- ▶ Bon fit aux SADs.
- ▶ Phylogénies plus réalistes.



Introduction

Objectif des études de macro-évolution

Approches espèce-centrée ou individu-centrée

NTB utilisée en macro-évolution

Plusieurs propositions de modèles

Support empirique

Relier la généalogie et la phylogénie : SGD

Les hypothèses du modèle

Simuler et inférer sous ce modèle

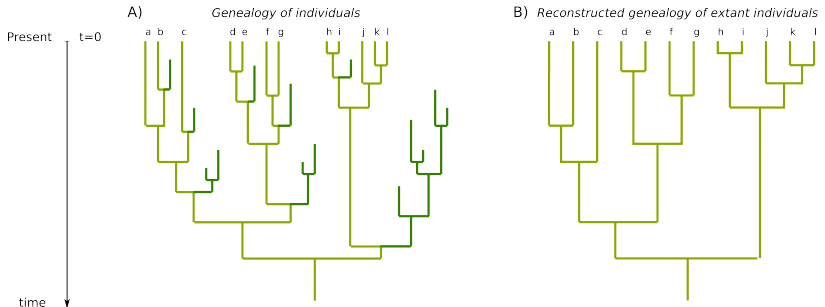
Conclusion

Résumé

Perspectives

Généalogie d'un processus de naissance-mort

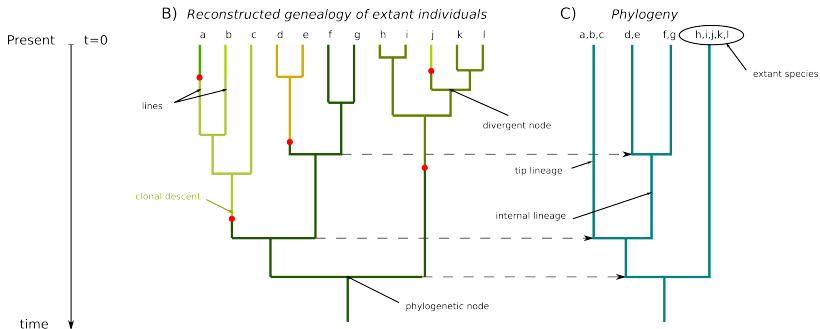
1. Les particules (individus) naissent et meurent à taux $b(t)$ et $d(t)$.
2. On appelle **généalogie reconstruite** la généalogie sans branches mortes.



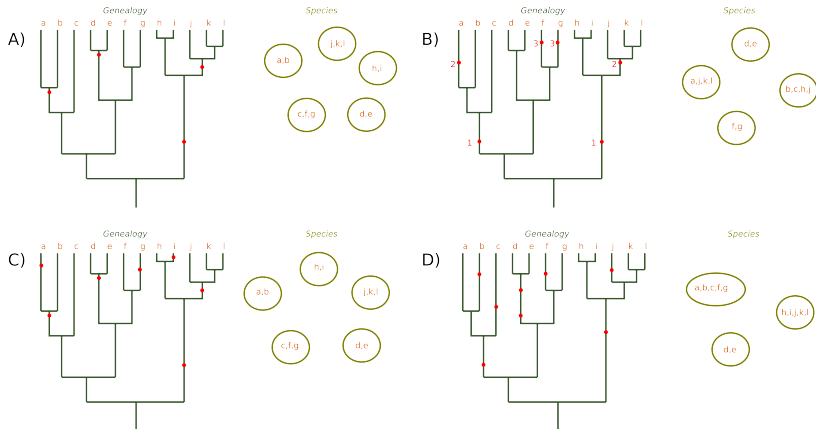
SGD : Speciation by Genetic Differentiation

Insiste sur la monophylie des espèces

- Processus de Poisson ponctuel de mutation → type génétique.
- Les espèces sont les plus petits groupes monophylétiques d'individus tels que deux individus de même type appartiennent au même groupe.



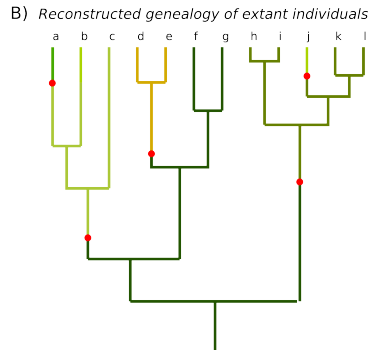
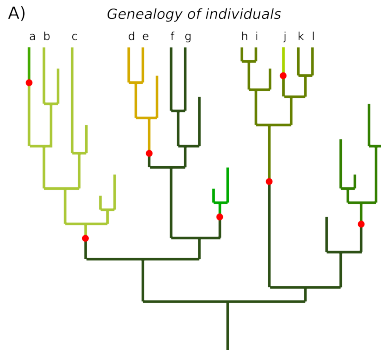
Discordance entre généalogies et phylogénies



Probabilités importantes dans le modèle

La probabilité d'extinction avant temps t d'une population descendant d'un individu et suivant un processus de naissance-mort est :

$$u_{b,d}(t) = \frac{1 + \int_0^t b(s) e^{\int_0^s b(z) - d(z) dz} ds - e^{\int_0^t b(z) - d(z) dz}}{1 + \int_0^t b(s) e^{\int_0^s b(z) - d(z) dz} ds} \quad (1)$$



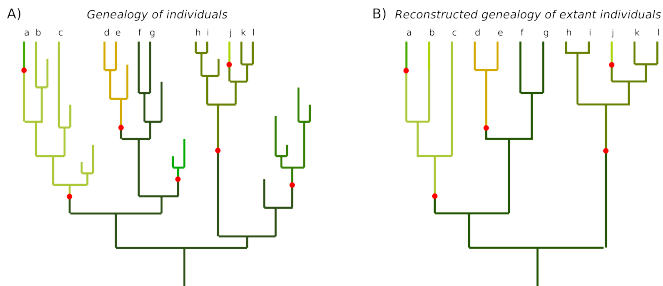
Probabilités importantes dans le modèle

Le taux de branchement sur la généalogie reconstruite est :

$$g(t)dt = b(t)(1 - u_{b,d}(t))dt \quad (2)$$

La probabilité de survie d'une population clonale, conditionnée à la survie de la population totale, est :

$$m(t) = \frac{1 - u_{b,d+\nu}(t)}{1 - u_{b,d}(t)} \quad (3)$$

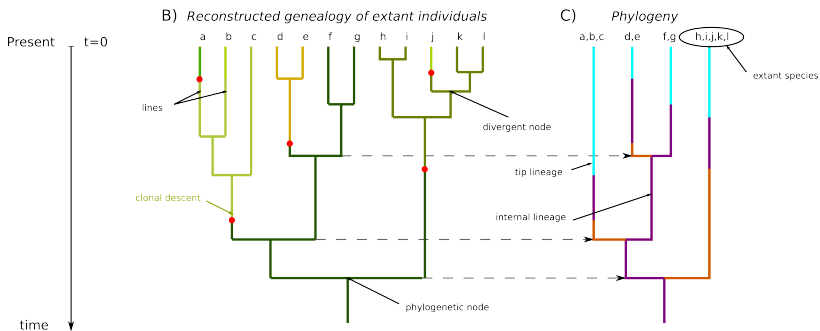


La phylogénie : un processus de branchement inhomogène à 3 types

Type 0 lorsque le type génétique de la branche est vivant au présent.

Type 1 lorsque le type génétique de la branche est éteint au présent.

Frozen lorsque le lignage est figé.



Taux de transitions

Une lignée de type 1 passe en type 0 :

$$\rho_{1 \rightarrow 0}(t)dt = \frac{\nu(t)m(t)}{(1 - m(t))} dt$$

Une lignée de type 1 donne deux lignées de type 1 :

$$\rho_{1 \rightarrow +1}(t)dt = g(t)(1 - m(t))dt$$

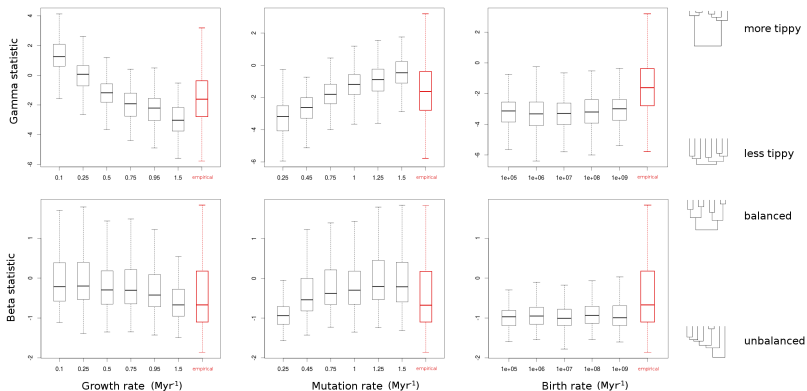
Une lignée de type 0 donne une lignée de type 0 et une lignée de type 1 :

$$\rho_{0 \rightarrow +1}(t)dt = 2g(t)(1 - m(t))dt$$

Une lignée de type 0 "freeze" :

$$\rho_{0 \rightarrow \emptyset}(t)dt = g(t)m(t)dt$$

La forme des phylogénies est réaliste



Méthode d'obtention de la vraisemblance

On suppose que chaque espèce est échantillonnée avec probabilité f .

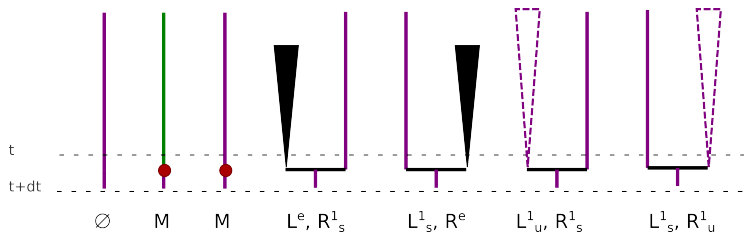
On s'intéresse à :

$$\mathcal{L}_{A,f}^i(t) = P(\text{arbre échantillonné commençant avec un individu au temps } t \\ \text{a la forme de } A \text{ et le type } i \\ | \text{ survie jusqu'à } t, \text{ et les paramètres } (b, d, \nu))$$

Et on a aussi besoin de :

$$w_f^i(t) = P(\text{arbre commençant avec un individu au temps } t \\ \text{a le type } i \text{ et aucune espèce échantillonnée} \\ | \text{ survie jusqu'à } t, \text{ et les paramètres } (b, d, \nu))$$

Méthode d'obtention de la vraisemblance



$$\mathcal{L}_{L,f}^1(t+dt) = \mathcal{L}_{L,f}^1(t) \left[1 - \nu dt - b \frac{1-u^2}{1-u} dt + 2budt + 2b(1-u)w_f^1 dt + \nu dt \right] + \mathcal{L}_{L,f}^0(t)\nu dt$$

$$\frac{d\mathcal{L}_{T,f}^1}{dt} = g\mathcal{L}_{T,f}^1(2w_f^1 - 1) + \nu\mathcal{L}_{T,f}^0$$

Set final d'EDOs

Vraisemblance le long des branches :

$$\frac{dw_f^0}{dt} = w_f^0 [-\nu - g + 2gw_f^1] + gm^2(1 - f)$$

$$\frac{dw_f^1}{dt} = -gw_f^1(1 - w_f^1) + \nu w_f^0$$

$$\frac{d\mathcal{L}_{T,f}^0}{dt} = \mathcal{L}_{T,f}^0 [-\nu - g + 2gw_f^1] + \mathcal{L}_{T,f}^1 2gw_f^0 + gm^2 f$$

$$\frac{d\mathcal{L}_{T,f}^1}{dt} = g\mathcal{L}_{T,f}^1(2w_f^1 - 1) + \nu\mathcal{L}_{T,f}^0$$

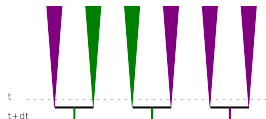
$$\frac{d\mathcal{L}_{l,f}^0}{dt} = \mathcal{L}_{l,f}^0 [-\nu - g + 2gw_f^1] + 2\mathcal{L}_{l,f}^1 gw_f^0$$

$$\frac{d\mathcal{L}_{l,f}^1}{dt} = g\mathcal{L}_{l,f}^1(2w_f^1 - 1) + \nu\mathcal{L}_{l,f}^0$$

Vraisemblance au noeud :

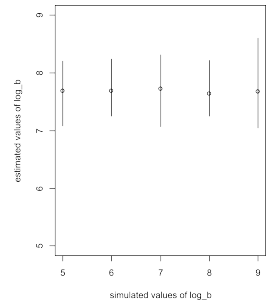
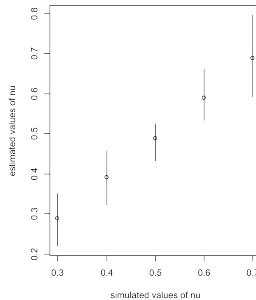
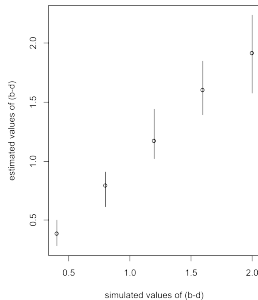
$$\mathcal{L}_{N,f}^0(t) = g(\mathcal{L}_{A,f}^0\mathcal{L}_{B,f}^1 + \mathcal{L}_{A,f}^1\mathcal{L}_{B,f}^0)$$

$$\mathcal{L}_{N,f}^1 = g\mathcal{L}_{A,f}^1\mathcal{L}_{B,f}^1$$



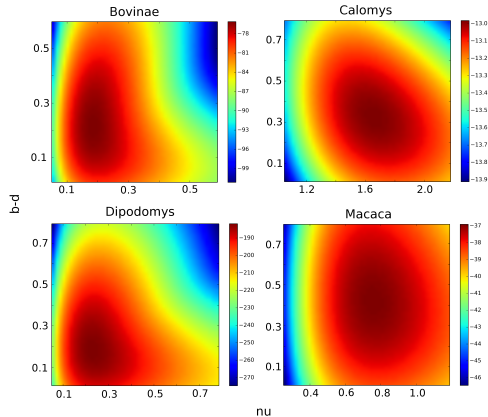
Inférence sur les simulations

- ▶ estimateurs non-biaisés de $b - d$ et ν .
- ▶ ...pas assez de signal pour inférer b .



Inférence sur arbre réel

- On peut prendre en compte un échantillonnage incomplet.
- Assez de signal pour retrouver ν et $b - d$.



Introduction

Objectif des études de macro-évolution

Approches espèce-centrée ou individu-centrée

NTB utilisée en macro-évolution

Plusieurs propositions de modèles

Support empirique

Relier la généalogie et la phylogénie : SGD

Les hypothèses du modèle

Simuler et inférer sous ce modèle

Conclusion

Résumé

Perspectives

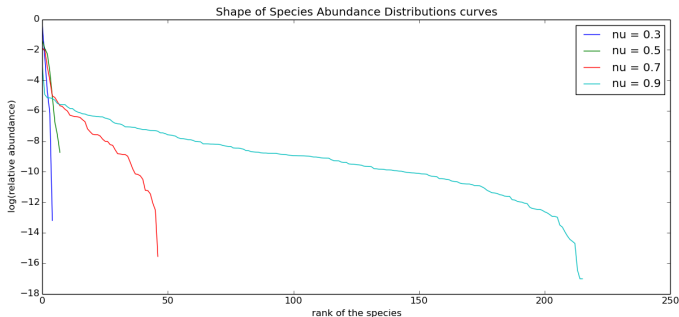
Modèles individu-centrés proposés

Assumptions		Empirical evaluation	
Speciation mode	Metacommunity size	Phylogenies	SAD
Point mutation	Fixed	×	✓
Random fission		×	×
Protracted speciation		✓	✓
SGD	Varying	✓	?

- ▶ Métacommunauté constante ou croissante à large échelle temporelle ?
- ▶ Besoin de considérer des groupes monophylétiques d'individus.
- ▶ ...mais définir les espèces à posteriori peut ne pas être pratique.

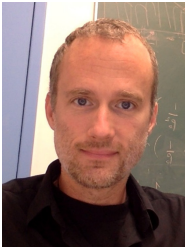
Perspectives

- Quelles sont les prédictions macro-écologiques sous SGD ?
- Utiliser des taux non-constants $b(t)$, $d(t)$, $\nu(t)$: peut-on espérer inférer des larges tendances de dynamique de métacommunauté ?



Remerciements

Travail réalisé sous la direction d'Amaury Lambert (UPMC et Collège de France) et Hélène Morlon (CNRS, ENS).



COLLÈGE
DE FRANCE
—1530—