

Phylogénies d'un marqueur neutre et trait adaptatif sous sélection: un modèle éco-évolutionnaire

Sylvie Méléard, Ecole Polytechnique

Chaire Modélisation mathématique et Biodiversité, 2015



Motivations

Les processus **écologiques** affectent les changements **évolutionnaires** et **l'évolution** influence la structure des systèmes **écologiques**.

Variation d'un **trait adaptif** x dans un formalisme eco-evolutionnaire, avec un point de vue de dynamique adaptative.

Evolution rapide d'une **séquence neutre**: un marqueur u physiquement lié au trait sous sélection dans le génome.

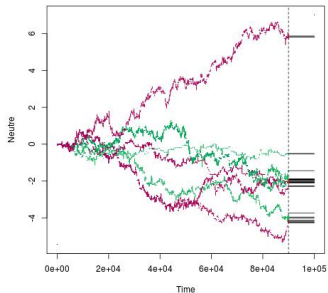
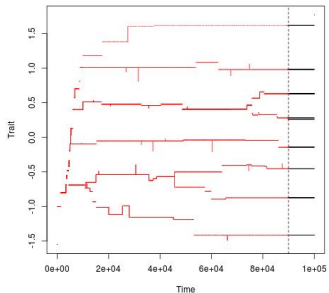
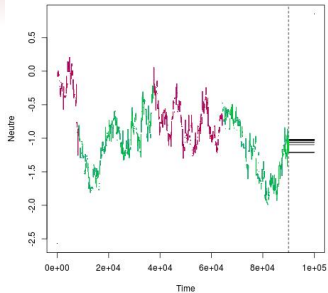
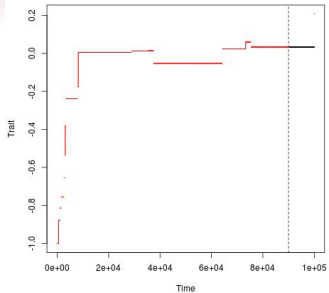
But: *Construire la dynamique stochastique du trait adaptif et du marqueur neutre, dirigée par les feedbacks éco-évolutionnaires, et décrire les phylogénies du marqueur neutre.*

JMB'15 et travail en cours

Résumé des résultats

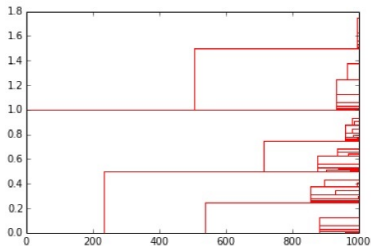
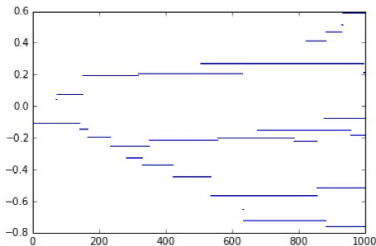
Sens direct du temps:

1. **Polymorphic Evolutionary sequence (PES) pour le trait** (e.g. Metz et al, Champagnat, Champagnat-M.): modélise les invasions successives de traits mutants avantageux.
2. **Phénomène de hitchhiking pour le marqueur** (e.g. Kojima Schaffer, Maynard Smith Haigh, Ohta Kimura, Kaplan, Barton, Etheridge et al., Durrett-Schweinsberg).
3. **Limite diffusive (processus de Wright-Fisher) pour le marqueur** entre les mutations des traits.



Conséquences pour les phylogénies:

4. Les phylogénies sont décrites par les processus de coalescence pour chaque sous-population définie par un trait donné.
5. Coalescence multiple aux temps de mutation des traits.



Travaux antérieurs

Méthodes d'inférence de l'histoire démographique de populations structurées à partir de données de polymorphisme génétique.

- *J.M. Marin, P. Pudlo, C. Robert + co-auteurs*

Comparaison de différents scénarios écologiques donnés a priori. Analyse des séquences (SNP ou microsatellites) et inférence des phylogénies en utilisant des modèles neutres (coalescents).

- *H. Morlon + co-auteurs*

Reconstruction à partir de modèles écologiques (macroscopiques) de diversification-extinction, sans prendre en compte la variabilité génétique.

Modèle individuel

- *Nous considérons un modèle décrivant la dynamique des individus d'une population structurée par les traits x et les marqueurs u , en temps continu, avec compétition, reproduction asexuée et mutations.*

La population: un processus à valeurs mesures ponctuelles qui décrit

N_t^K individus de caractéristiques (x_i, u_i) , $i = 1 \dots N_t^K$.

- Un individu de trait x et marqueur u au temps t :
 - se reproduit au taux $b(x)$
 - meurt au taux

$$d(x) + \frac{\eta(x)}{K} \sum_{i=1}^{N_t^K} C(x, x_i).$$

Mutations

A chaque naissance,

- **Mutation du trait** avec probabilité $p_K \sim 1/K^2$. Le trait mutant est $x + k$ où $k \rightsquigarrow m(x, k)dk$.
- **Mutation du marqueur** avec probabilité q_K et $\lim_{K \rightarrow +\infty} q_K = 0$. Le marqueur mutant est $u + h$ où $h \rightsquigarrow G_K(u, dh)$.

- *Les marqueurs mutent plus vite que les traits:*

$$q_K = \lambda_K p_K = \lambda_K / K^2, \text{ avec}$$

$$\lim_{K \rightarrow +\infty} \lambda_K = +\infty \text{ et } \lim_{K \rightarrow +\infty} \frac{(\log K)^4 \lambda_K}{K^2} = 0.$$

- *Les tailles des mutations du marqueur sont de l'ordre de K/λ_K :* pour tout $u \in \mathcal{U}$,

$$\lim_{K \rightarrow +\infty} \sup_{u \in \mathcal{U}} \left| \frac{\lambda_K}{K} \int_{\mathcal{U}} (f(u+h) - f(u)) G_K(u, dh) - Lf \right| = 0$$

où $(L, \mathcal{D}(L))$ est le générateur d'un semi-groupe de Feller.

Exemples

Exemple 1: $\mathcal{U} = [u_1, u_2]$, $\lambda_K = K^{3/2}$, $q_K = 1/\sqrt{K}$.

$G_K(u, dh)$ est une loi gaussienne conditionnée à rester dans \mathcal{U} , de moyenne μ_K et de variance σ_K , telles que

$$\lim_{K \rightarrow +\infty} \mu_K \frac{\lambda_K}{K} = \mu, \quad \text{et} \quad \lim_{K \rightarrow +\infty} \sigma_K^2 \frac{\lambda_K}{K} = \sigma^2.$$

Alors pour $f \in \mathcal{C}^2$ avec $f'(u_1) = f'(u_2) = 0$,

$$Lf = \frac{\sigma^2}{2} f'' + \mu f'.$$

Exemple 2: $\mathcal{U} = \{A, a\}$, $\lambda_K \sim_{K \rightarrow \infty} \lambda K$, $q_K = \frac{1}{K}$.

Pour le noyau de mutation:

$$G_K(u, dv) = \mathbf{1}_{u=A} q_A \delta_a(dv) + \mathbf{1}_{u=a} q_a \delta_A(dv).$$

Alors :

$$Lf(u) = \lambda \left(\mathbf{1}_{u=A} q_A (f(a) - f(A)) + \mathbf{1}_{u=a} q_a (f(A) - f(a)) \right)$$

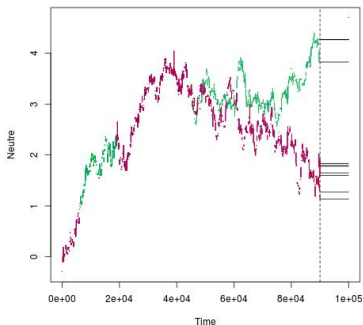
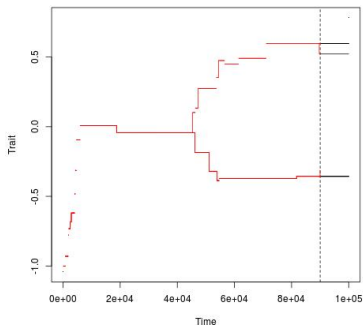
Exemple numérique

Inspiré par le modèle de taille de bec (Dieckmann-Doebeli 1999).

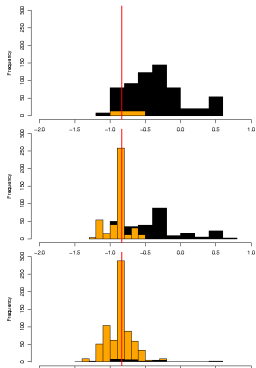
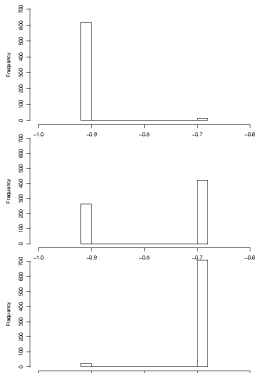
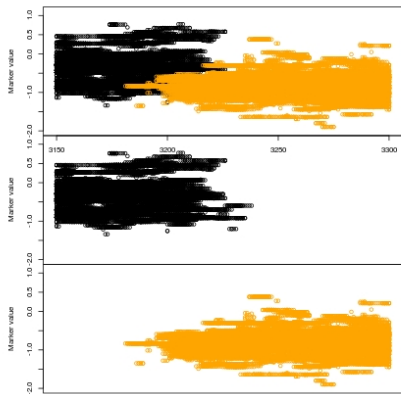
- $\mathcal{X} = [-2, 2]$; $\mathcal{U} = [-2, 2]$; $K = 1000$.
- $p_K = \frac{1}{K^2}$; $\lambda_K = K^{3/2}$; $q_K = \frac{1}{\sqrt{K}}$.
- $m \sim \mathcal{N}(0, 10^{-1})$ et $G_K \sim \mathcal{N}(0, \frac{1}{\sqrt{K}})$ sont deux distributions gaussiennes conditionnées à être dans $[-2, 2]$.
- $b(x) = \exp(-x^2/2\sigma_b^2)$; $\sigma_b = 0.9$, maximum en 0.
- Compétition symétrique pour les ressources.

$$C(x, y) = \exp(-(x - y)^2/2\sigma_C^2), \quad \sigma_C = 0.8.$$

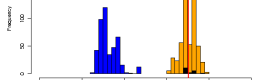
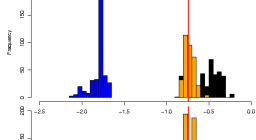
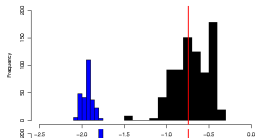
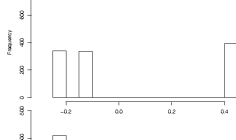
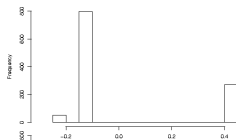
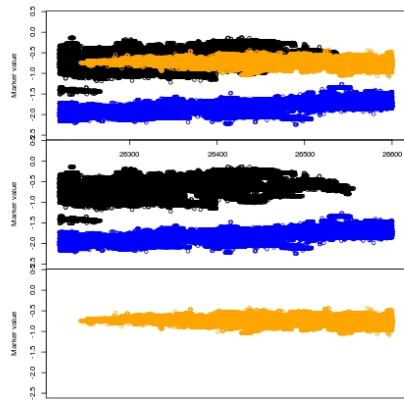
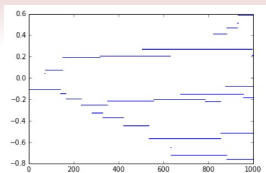
Simulation du processus des traits et marqueurs en temps long



L'effet de hitchhiking



Cas de coexistence.



Echelles de temps

Il y a trois échelles de temps différentes :

- L'échelle de temps de mutation des traits, de l'ordre de

$$K p_K \sim \frac{1}{K}, \text{ est la plus lente.}$$

- L'échelle de temps de mutation des marqueurs, de l'ordre de

$$K q_K \sim \frac{\lambda_K}{K}.$$

- L'échelle de temps écologique (naissances et morts), de l'ordre de K , est la plus rapide.

Approximation en grande population sans changement d'échelle de temps

La probabilité de mutation tend vers zéro.

- *Condition initiale monomorphique en trait: trait x .*

$$\frac{dn_t}{dt} = \left(b(x) - d(x) - \eta(x)C(x, x)n_t \right) n_t$$

qui converge quand $t \rightarrow +\infty$ vers $\hat{n}_x = \frac{b(x) - d(x)}{\eta(x)C(x, x)}$.

- *Condition initiale polymorphique en trait: traits x_1, \dots, x_p .*

$$\frac{dn_t(x_i)}{dt} = \left(b(x_i) - d(x_i) - \eta(x_i) \sum_{j=1}^p C(x_i, x_j) n_t(x_j) \right) n_t(x_i), \quad i \in \{1, \dots, p\}$$

Champagnat-Jabin-Raoul 10': *Si la matrice $(\eta(x_i)C(x_i, x_j))_{i,j}$ est symétrique définie positive, le système de Lotka-Volterra admet un unique point fixe globalement stable sur \mathbb{R}_+^p .*

Vrai pour les noyaux de compétition gaussiens.

Polymorphic Evolution Sequence (échelle de temps Kt des mutations du trait), $K \rightarrow +\infty$

Metz et al. (1996), Champagnat (2006).

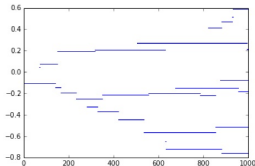
Fitness d'invasion $f(y; x) = b(y) - d(y) - \eta(y)C(y, x)\hat{n}_x$.

Theorem

Sous une hypothèse "*invasion implique fixation*": le processus de population au temps Kt converge vers un processus qui saute de \hat{n}_x individus de trait x à \hat{n}_{x+k} individus de trait $x + k$, avec la mesure de saut

$$b(x)\hat{n}_x \frac{[f(x+k; x)]_+}{b(x+k)} m(x, k) dk.$$

Généralisation au cas de coexistence de traits: PES (Champagnat-M. 2011).



Le marqueur durant l'invasion d'un trait

Considérons un mutant $(x + k, v)$ apparaissant au temps t dans une population monomorphique de trait x et de distribution de marqueurs $\pi_t^K(x, du)$. Supposons que $f(x + k; x) > 0$.

La distribution du marqueur:

$$\pi_t^K(x, du) = \frac{\sum_{i=1}^{N_t^K} \mathbf{1}_{x_i=x} \delta_{u_i}}{\sum_{i=1}^{N_t^K} \mathbf{1}_{x_i=x}}.$$

Il existe $\varepsilon > 0$ tel que pour $(t_K; K \in \mathbb{N}^*)$ avec $\log K \ll t_K \ll K$,

$$\lim_{K \rightarrow +\infty} \mathbb{P}(\langle \nu_{t_K}^K, \mathbf{1}_{x+k} \rangle > \varepsilon) = \frac{f(x + k; x)}{b(x + k)};$$

$$\lim_{K \rightarrow +\infty} \mathbb{P}(\langle \nu_{t_K}^K, \mathbf{1}_{x+k} \rangle = 0) = 1 - \frac{f(x + k; x)}{b(x + k)};$$

$$\lim_{K \rightarrow +\infty} \mathbb{P}(\pi_{t_K}^K(x + k, du) = \delta_v(du)) = \frac{f(x + k; x)}{b(x + k)}.$$

The fixation of the mutant creates a genetical bottleneck.

Processus de substitution de Wright-Fisher:

$$\mathcal{U} = \{a, A\}$$

q_A et q_a sont les probabilités de muter du marqueur A au marqueur a et du marqueur a au marqueur A .

Theorem

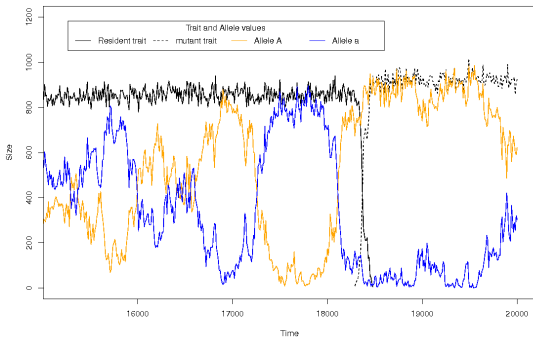
Le processus de population, considéré au temps Kt , converge vers

$$\hat{n}_{Y_t} \left(W_t^a \delta_{(Y_t, a)}(dy, du) + (1 - W_t^a) \delta_{(Y_t, A)}(dy, du) \right).$$

Distribution de a au temps t : Bernouilli de paramètre W_t^a et

$$dW_t^a = \lambda b(Y_t) (q_A(1 - W_t^a) - q_a W_t^a) dt + \sqrt{\frac{2b(Y_t)}{\hat{n}_{Y_t}}} W_t^a (1 - W_t^a) dB_t$$

Le processus limite saute avec le TSS ($Y_t, t \geq 0$) et au temps de saut t , le processus $(W_t^a, 1 - W_t^a)$ va à $(1, 0)$ avec probabilité W_t^a et à $(0, 1)$ avec probabilité $1 - W_t^a$.



Au temps d'apparition du trait mutant, la fréquence de A est de 85%.

Après le temps de fixation du mutant (~ 18490), la population a est éteinte. Elle est régénérée par mutation.

Cas général - la distribution du marqueur

Population initiale monomorphique de trait x et $N_0^K \rightarrow \hat{n}_x$.

Proposition

Avant la première mutation de trait, le processus $(\pi_{Kt}^K(x, dv), t \geq 0)$ converge en loi vers le processus de Fleming-Viot $(F_t^u(x, dv), t \geq 0)$ issue de δ_u et paramétré par x .

Le processus de Fleming-Viot $(F_t^u(x, \cdot), t \geq 0)$ est un processus à valeurs probabilités sur \mathcal{U} issu de δ_u et tel que pour $g \in \mathcal{D}(A)$,

$$\int_{\mathcal{U}} g(v) F_t^u(x, dv) = g(u) + b(x) \int_0^t \langle F_s^u(x, \cdot), Lg \rangle ds + M_t^{(x,u)}(g).$$

$M^{(x,u)}(g)$ est une martingale continue de carré intégrable de variation quadratique

$$\frac{2b(x)}{\hat{n}_x} \int_0^t \left(\langle F_s^u(x, \cdot), g^2 \rangle - \langle F_s^u(x, \cdot), g \rangle^2 \right) ds.$$

Dynamique lente-rapide de traits et marqueurs. Processus de Substitution de Fleming-Viot (SFVP)

Theorem

Le processus de population, considéré au temps Kt converge en loi vers

$$V_t(dy, dv) = \hat{n}_{Y_t} \delta_{Y_t}(dy) F_t^{U_t}(Y_t, dv).$$

Le processus $((Y_t, U_t), t \geq 0)$ sur $\mathcal{X} \times \mathcal{U}$, issu de (x_0, u_0) , saute de (x, u) à $(x + k, v)$ suivant la mesure de saut

$$b(x) \hat{n}_x \frac{[f(x + k; x)]_+}{b(x + k)} F_t^u(x, dv) m(x, k) dk. \quad (1)$$

Extension aux traits coexistants (PES) : des processus de Fleming-Viot indépendants sur chaque branche, **mais les paramètres de ces processus dépendent de toute la population.**

Le problème statistique

Plaçons-nous dans le cas du modèle de Dieckmann-Doebeli.

Les observations: *une population discriminée en sous-populations par un trait soumis à sélection et dotée d'un marqueur neutre observé.*

On connaît

- Le nombre total de sous-populations,
- Les abondances des sous-populations,
- Au sein de chaque sous-population, la distribution en couples trait/marqueurs.

Les paramètres à estimer:

- p probabilité de mutation du trait
- q probabilité de mutation du marqueur
- σ_b
- σ_c
- σ_m écart-type mutation du trait.
- η sensibilité à la compétition.

$$\theta = (p, q, \sigma_b, \sigma_c, \sigma_m, \eta).$$

Approximate Bayesian Computation (ABC)

Beaumont et al. (2002), Marin et al. (2012).

Etude statistique d'un modèle paramétré par une démarche bayésienne: les paramètres sont aléatoires et l'on en cherche une distribution a posteriori, au vu des observations.

Principe de l'ABC:

- Tirer N paramètres suivant une loi *a priori*: $\theta_i, i \in \{1, \dots, N\}$,
- Pour chaque θ_i , simuler le processus.
- Calculer pour chaque i un ensemble de statistiques descriptives S_i obtenues à partir de la i -ème simulation,
- Comparer S_i avec la valeur de la donnée s_{obs} et définir un poids w_i pour θ_i dépendant de la distance $d(S_i, s_{obs})$ (avec corrections possibles).
- L'échantillon pondéré $(\theta_i, w_i)_{i \in \{1, \dots, N\}}$ fournit la distribution *a posteriori*.

La simulation du SFVP

On veut simuler un échantillon d'individus tirés selon la mesure valeur du SFVP au temps $t > 0$.

Simulation pas à pas: trop couteuse.

Soient x_1, \dots, x_p les traits chargés par le SFVP au temps t . On échantillonne k_1, \dots, k_p individus selon les mesures sur l'espace des marqueurs, dans les sous-populations données par le PES, de traits respectifs x_1, \dots, x_p .

1 - Générer le PES et fixer ainsi l'historique des paramètres écologiques et des traits de la population.

2 - Construire la phylogénie avec des coalescents sur chaque branche, forcer les coalescences multiples à chaque saut écologique et mettre à jour les paramètres écologiques contenus dans le PES.

- Une fois la coalescence amenée à $t = 0$, diffuser sur les branches de la coalescence selon les paramètres écologiques pertinents et le générateur L .

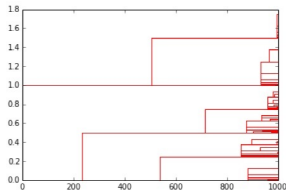
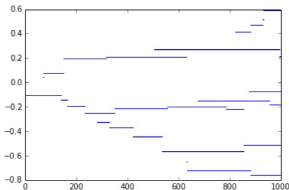
PES et Phylogénie du marqueur

Entre les sauts du PES:

- Les traits et tailles des sous-populations correspondantes calculés à partir des équilibres du système de Lotka-Volterra.
- Dans chaque sous-population, les phylogénies neutres données par des coalescents de Kingman, de paramètres dépendant des traits et distributions d'abondance.

Quand la distribution de trait change, les paramètres changent pour chaque branche.

Quand les phylogénies atteignent un bottleneck, elles coalescent toutes.

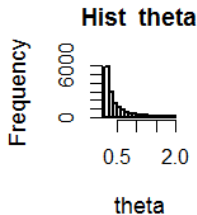
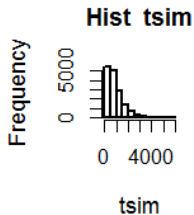
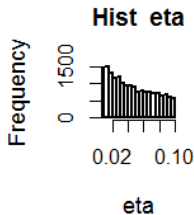
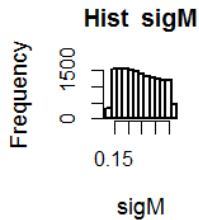
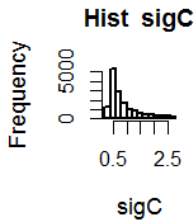
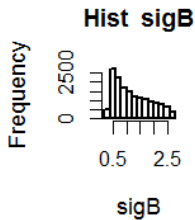
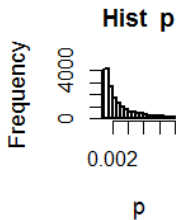


Statistiques descriptives

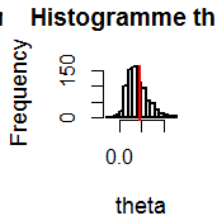
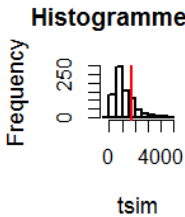
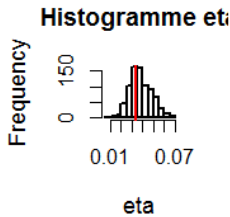
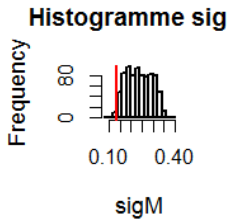
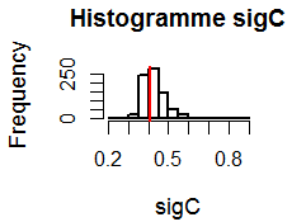
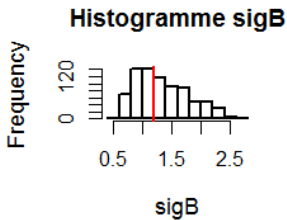
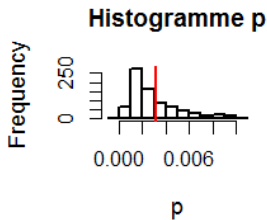
- nombre d'allèles distincts pour le marqueur,
- variance de la distribution des allèles des marqueurs,
- gdiversité génétique (mesure de la largeur du support de la distribution des allèles des marqueurs)
- M-index (rapport du nombre d'allèles pour le marqueur sur la largeur du support de la distribution des allèles des marqueurs)
- nombre de traits dans la population,
- moyenne et variance des traits
- distribution d'abondance des traits.

★ Si on veut comparer plusieurs modèles: indice du modèle choisi et paramètres associés.

Résultats (préliminaires)



Résultats (préliminaires)





**Merci à tous pour la belle aventure de la
chaire MMB!**

