

# Estimer les effets des mutations sur la valeur sélective d'une bactérie

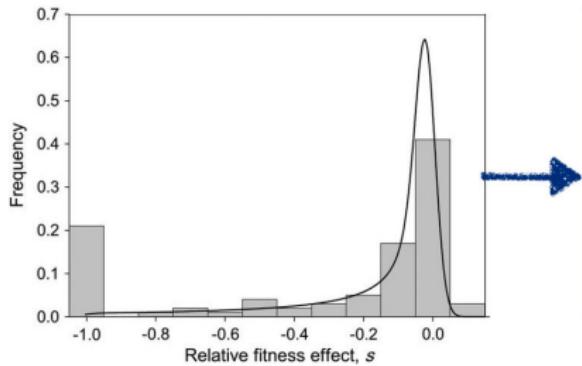
Jérémy Andréoletti et Nathanaël Boutillon

Encadrantes : Marie Doumic et Lydia Robert

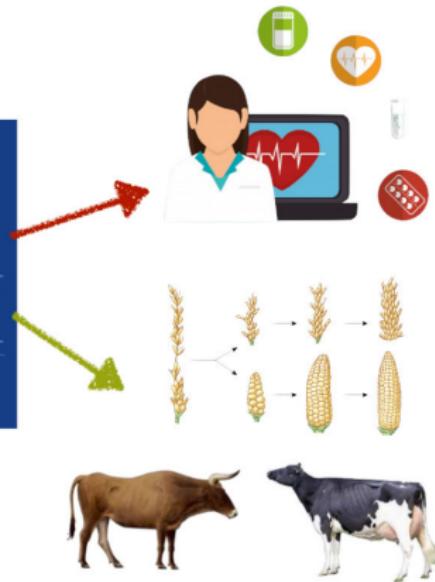
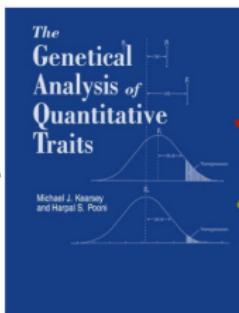
# Objectifs

- Comprendre la dynamique d'apparition des mutations, chez *E. coli*
- Estimer la **DFE = Distribution des Effets des mutations sur la Fitness**

Distribution of fitness effects  
caused by single-nucleotide  
substitutions in bacteriophage f1



Peris et al., Genetics, 2010



# Modèle

On pose

$W_t$	fitness (taux de croissance) au temps $t$
$s_i = \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}}$	effet <i>relatif</i> de la mutation $i$
$N_t$	nombre de mutations avant le temps $t$

# Modèle

On pose

$W_t$	fitness (taux de croissance) au temps $t$
$s_i = \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}}$	effet <i>relatif</i> de la mutation $i$
$N_t$	nombre de mutations avant le temps $t$

**Énoncé du problème :** estimer la loi des  $s_i$  (qui sont iid) sachant que l'on sait estimer la loi des  $W_t$  et  $N_t$  ( $t \geq 0$ ), et sachant que

$$\frac{W_t}{W_0} = \prod_{i=1}^{N_t} (1 - s_i)$$

# Plan

1 Expérience & simulations

2 Problème des moments

- Fonction caractéristique
- Calculs d'erreurs

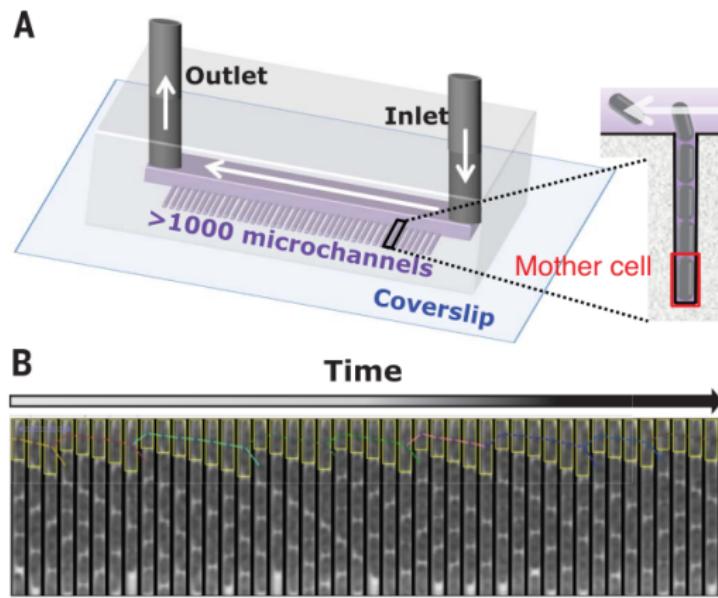
3 EDP

- Nouveau point de vue
- Applications

## Section 1

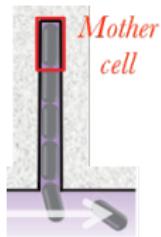
Expérience & simulations

# Expérience : accumulation de mutations



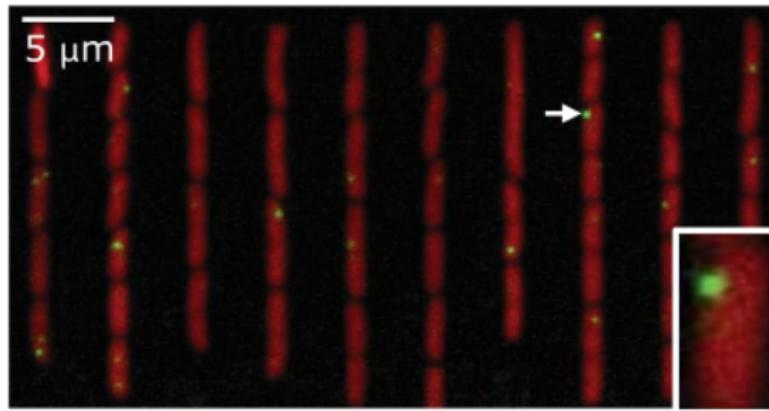
*Robert et al.*

# Expérience : accumulation de mutations



*Robert et al.*

# Expérience : visualisation de mutations

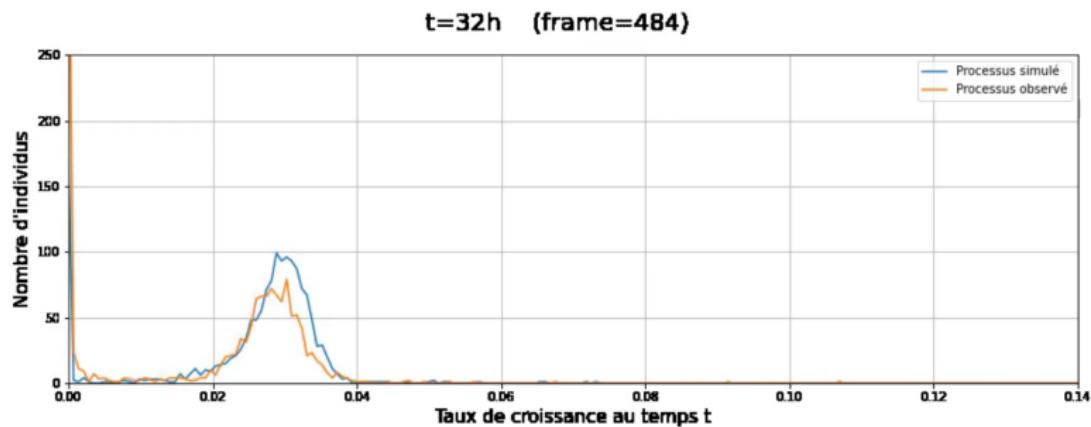


*Robert et al.*

# Expérience : visualisation de mutations

*Robert et al.*

# Simulations : image



# Simulations

## Section 2

Problème des moments

# Estimation des moments

Analyse de [1]

**Moments** de la loi des effets des mutations ;

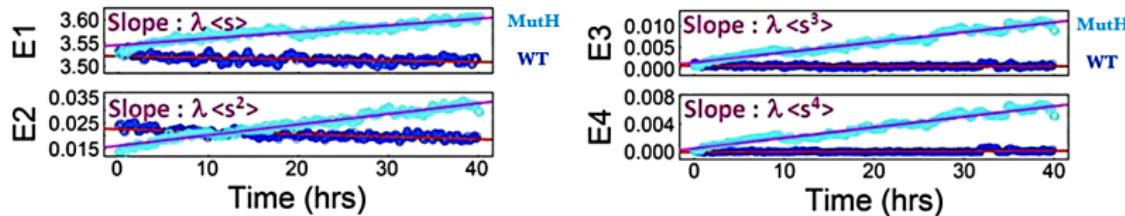
$$E_n(t) := \sum_{k=1}^n \binom{n}{k} (-1)^k \ln \left( \mathbb{E} \left[ W_t^k \right] \right) = (\lambda \mathbb{E}[S^n]) t$$

# Estimation des moments

Analyse de [1]

**Moments de la loi des effets des mutations ;**

$$E_n(t) := \sum_{k=1}^n \binom{n}{k} (-1)^k \ln (\mathbb{E}[W_t^k]) = (\lambda \mathbb{E}[S^n]) t$$



→ estimation facile des  $\mathbb{E}[S^n]$ .

# Méthode par la fonction caractéristique

- Moments  $\rightarrow$  fonction caractéristique  $\varphi_S$  :

$$\varphi_S(\xi) := \mathbb{E}[e^{i\xi S}] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}[S^k]$$

# Méthode par la fonction caractéristique

- Moments  $\rightarrow$  fonction caractéristique  $\varphi_S$  :

$$\varphi_S(\xi) := \mathbb{E}[e^{i\xi S}] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}[S^k]$$

- Fonction caractéristique  $\varphi_S \rightarrow$  densité  $f$  de la loi de  $S$ .

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(\xi) e^{-ix\xi} d\xi$$

# Méthode par la fonction caractéristique

- Moments  $\rightarrow$  fonction caractéristique  $\varphi_S$  :

$$\varphi_S(\xi) := \mathbb{E}[e^{i\xi S}] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}[S^k]$$

- Fonction caractéristique  $\varphi_S \rightarrow$  densité  $f$  de la loi de  $S$ .

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(\xi) e^{-ix\xi} d\xi$$

- Avec les  $N$  premiers moments estimés  $m_k$  :

$$\hat{\varphi}_X(\xi) = \sum_{k=0}^N \frac{(i\xi)^k}{k!} m_k \quad \text{et} \quad \hat{f}(x) = \frac{1}{2\pi} \int_{|\xi| \leq A} \hat{\varphi}_X(\xi) e^{-ix\xi} d\xi$$

# Méthode par la fonction caractéristique

- Moments → fonction caractéristique  $\varphi_S$  :

$$\varphi_S(\xi) := \mathbb{E}[e^{i\xi S}] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}[S^k]$$

- Fonction caractéristique  $\varphi_S$  → densité  $f$  de la loi de  $S$ .

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(\xi) e^{-ix\xi} d\xi$$

- Avec les  $N$  premiers moments estimés  $m_k$  :

$$\hat{\varphi}_X(\xi) = \sum_{k=0}^N \frac{(i\xi)^k}{k!} m_k \quad \text{et} \quad \hat{f}(x) = \frac{1}{2\pi} \int_{|\xi| \leq A} \hat{\varphi}_X(\xi) e^{-ix\xi} d\xi$$

- Espoirs :**  $\varphi_S(\xi) \simeq \hat{\varphi}_S(\xi)$  donc  $f(x) \simeq \hat{f}(x)$ .

# Calculs d'erreurs

## Proposition

Pour tous paramètres  $A > 0$ ,  $k \geq 2$ ,  $N \geq 1$  on a :  $\forall x \in \mathbb{R}$

$$\left| \hat{f}(x) - f(x) \right| \leq \frac{\|f^{(k)}\|_1}{2\pi^2(k-1)A^{k-1}} + \frac{A^{N+1}}{\pi(N+1)!} \mathbb{E} [X^N(e^{AX}-1)] + \frac{\|\varepsilon(N)\|_\infty(e^A-1)}{\pi}$$

# Calculs d'erreurs

## Proposition

Pour tous paramètres  $A > 0$ ,  $k \geq 2$ ,  $N \geq 1$  on a :  $\forall x \in \mathbb{R}$

$$\left| \hat{f}(x) - f(x) \right| \leq \frac{\|f^{(k)}\|_1}{2\pi^2(k-1)A^{k-1}} + \frac{A^{N+1}}{\pi(N+1)!} \mathbb{E} \left[ X^N (e^{AX} - 1) \right] + \frac{\|\varepsilon(N)\|_\infty (e^A - 1)}{\pi}$$

## Proposition

Si :

- erreur sur les  $\mathbb{E}[S^n]$  bornée par  $\varepsilon$  (pour tout  $n \in \mathbb{N}$ ) ;
- $\|f^{(2)}\|_1 < +\infty$ ,

alors :

$$\|f - \hat{f}\|_\infty = O \left( \left| \frac{1}{\ln^2(\varepsilon)} \right| \right) \quad \text{quand } \varepsilon \rightarrow 0$$

## Section 3

EDP

# Un autre point de vue : une EDP

On a

$$\ln W_t = \sum_{i=1}^{N_t} \ln(1 - s_i)$$

Soient :

- $u(t, \cdot) \in C^\infty(\mathbb{R})$  la « densité » de la loi de  $\ln W_t$
- $m(t)$  la probabilité qu'une cellule soit morte au temps  $t$   
     $\implies$  Loi de  $\ln W_t$  :

$$m(t)\delta_{-\infty} + u(t, \cdot)$$

# Un autre point de vue : une EDP

On a

$$\ln W_t = \sum_{i=1}^{N_t} \ln(1 - s_i)$$

Soient :

- $u(t, \cdot) \in C^\infty(\mathbb{R})$  la « densité » de la loi de  $\ln W_t$
- $m(t)$  la probabilité qu'une cellule soit morte au temps  $t$   
 $\implies$  Loi de  $\ln W_t$  :

$$m(t)\delta_{-\infty} + u(t, \cdot)$$

- $f(\cdot) \in C^\infty(\mathbb{R})$  la « densité » de la loi de  $\ln(1 - S)$
- $\mu$  est la proportion de mutations létale  
 $\implies$  Loi de  $\ln(1 - S)$  :

$$\mu\delta_{-\infty} + f(\cdot)$$

# Un autre point de vue : une EDP

On a

$$\ln W_t = \sum_{i=1}^{N_t} \ln(1 - s_i)$$

Soient :

- $u(t, \cdot) \in C^\infty(\mathbb{R})$  la « densité » de la loi de  $\ln W_t$
- $m(t)$  la probabilité qu'une cellule soit morte au temps  $t$   
 $\implies$  Loi de  $\ln W_t$  :

$$m(t)\delta_{-\infty} + u(t, \cdot)$$

- $f(\cdot) \in C^\infty(\mathbb{R})$  la « densité » de la loi de  $\ln(1 - S)$
- $\mu$  est la proportion de mutations létale  
 $\implies$  Loi de  $\ln(1 - S)$  :

$$\mu\delta_{-\infty} + f(\cdot)$$

- $\lambda$  le taux de mutation

# Explication des termes

Donc (\*) :

$$\partial_t u(t, x) = \lambda \left( \int_{\mathbb{R}} f(x - y) u(t, y) dy - \int_{\mathbb{R}} f(y) u(t, x) dy \right) - \lambda \mu u(t, x)$$

Traduction :

- $\partial_t u$  : changement de densité de fitness entre  $t$  et  $t + dt$  ;
- $\int f(x - y) u(t, y) dy$  : arrivées sur la fitness  $x$  ;
- $-\int f(y) u(t, x) dy$  : départs de la fitness  $x$  ;
- $-\lambda \mu u(t, x)$  : morts.

# Explication des termes

Donc (\*) :

$$\partial_t u(t, x) = \lambda \left( \int_{\mathbb{R}} f(x - y) u(t, y) dy - \int_{\mathbb{R}} f(y) u(t, x) dy \right) - \lambda \mu u(t, x)$$

Traduction :

- $\partial_t u$  : changement de densité de fitness entre  $t$  et  $t + dt$  ;
- $\int f(x - y) u(t, y) dy$  : arrivées sur la fitness  $x$  ;
- $-\int f(y) u(t, x) dy$  : départs de la fitness  $x$  ;
- $-\lambda \mu u(t, x)$  : morts.

Conservation de la masse :

$$\int_{\mathbb{R}} u(t, x) dx = N(t) = e^{-\lambda \mu t} N(0)$$

# Solution explicite

La transformée de Fourier sur :

$$\partial_t u(t, x) = \lambda \left( \int_{\mathbb{R}} f(x - y) u(t, y) dy - u(t, x) \int_{\mathbb{R}} f(y) dy \right) - \lambda \mu u(t, x)$$

donne

$$\partial_t \mathcal{F}u_t(\xi) = \lambda \mathcal{F}f(\xi) \mathcal{F}u_t(\xi) - \lambda \mathcal{F}u_t(\xi)$$

# Solution explicite

La transformée de Fourier sur :

$$\partial_t u(t, x) = \lambda \left( \int_{\mathbb{R}} f(x-y) u(t, y) dy - u(t, x) \int_{\mathbb{R}} f(y) dy \right) - \lambda \mu u(t, x)$$

donne

$$\partial_t \mathcal{F}u_t(\xi) = \lambda \mathcal{F}f(\xi) \mathcal{F}u_t(\xi) - \lambda \mathcal{F}u_t(\xi)$$

d'où :

## Proposition

Pour tout  $x \in \mathbb{R}$  :

$$f(x) = \mathcal{F}^{-1} \left( \frac{\partial_t (\mathcal{F}u_t(\xi))}{\lambda \mathcal{F}u_t(\xi)} + 1 \right)$$

# Solution explicite

La transformée de Fourier sur :

$$\partial_t u(t, x) = \lambda \left( \int_{\mathbb{R}} f(x-y) u(t, y) dy - u(t, x) \int_{\mathbb{R}} f(y) dy \right) - \lambda \mu u(t, x)$$

donne

$$\partial_t \mathcal{F}u_t(\xi) = \lambda \mathcal{F}f(\xi) \mathcal{F}u_t(\xi) - \lambda \mathcal{F}u_t(\xi)$$

d'où :

## Proposition

Pour tout  $x \in \mathbb{R}$  :

$$f(x) = \mathcal{F}^{-1} \left( \frac{\partial_t (\mathcal{F}u_t(\xi))}{\lambda \mathcal{F}u_t(\xi)} + 1 \right)$$

Le terme de droite est indépendant du temps :  $\partial_t (\mathcal{F}u_t(\xi)) = a_\xi \mathcal{F}u_t(\xi)$ .

# Vérification

$\partial_t(\mathcal{F}u_t(\xi)) = a_\xi \mathcal{F}u_t(\xi) :$

$$\ln |\mathcal{F}u_t(\xi)| = \ln |\mathcal{F}u_0(\xi)| + t \times \Re(a_\xi)$$

$$\arg(\mathcal{F}u_t(\xi)) = \arg \mathcal{F}u_0(\xi) + t \times \Im(a_\xi)$$

# Vérification

$\partial_t(\mathcal{F}u_t(\xi)) = a_\xi \mathcal{F}u_t(\xi)$  :

$$\ln |\mathcal{F}u_t(\xi)| = \ln |\mathcal{F}u_0(\xi)| + t \times \Re(a_\xi)$$

$$\arg(\mathcal{F}u_t(\xi)) = \arg \mathcal{F}u_0(\xi) + t \times \Im(a_\xi)$$

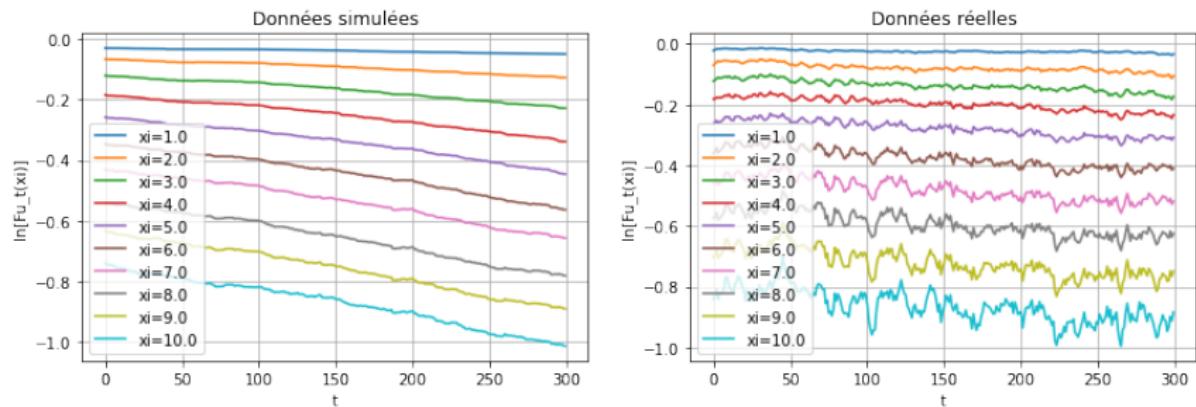
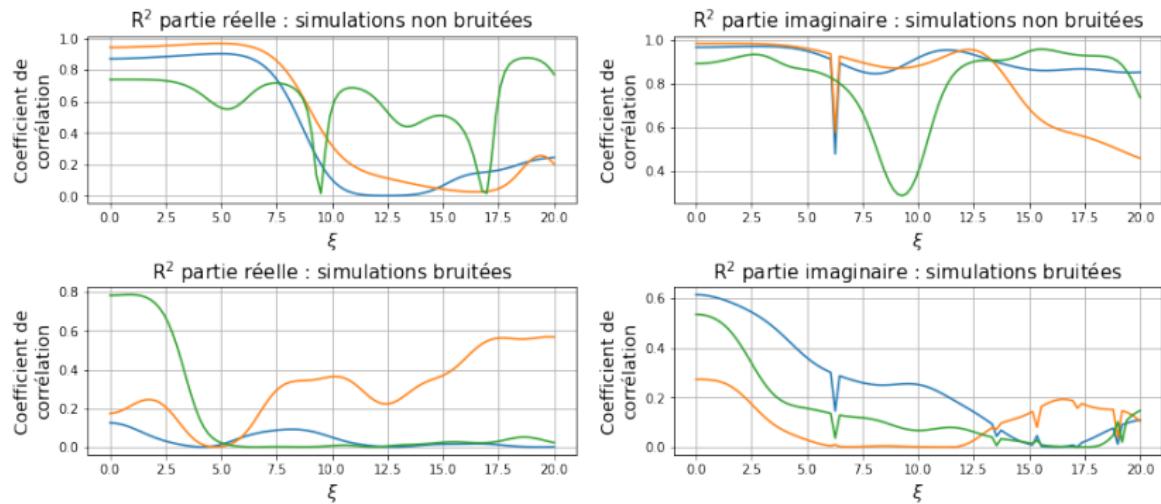


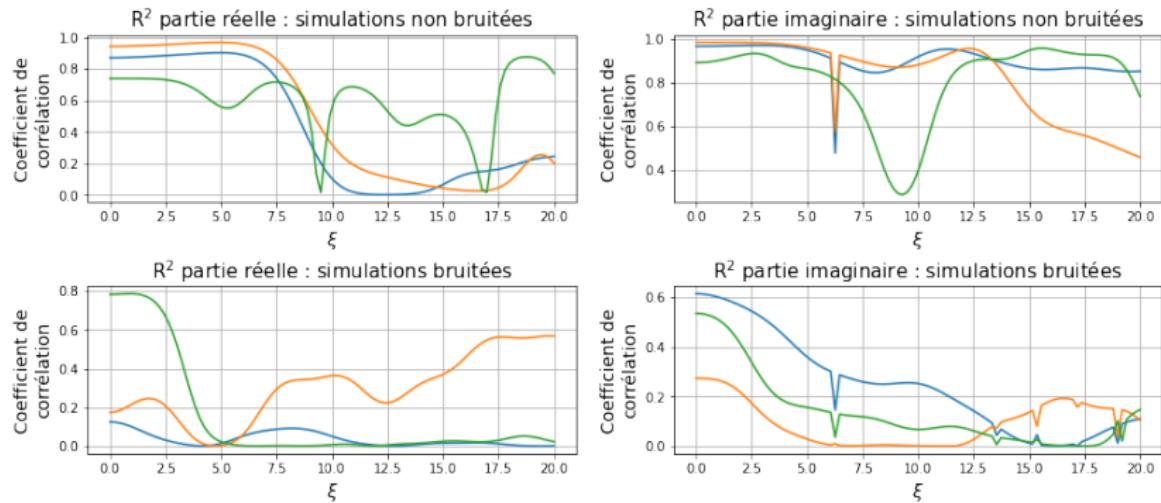
FIGURE – Graphes de  $t \mapsto |\mathcal{F}u_t(\xi)|$  pour données simulées et expérimentales.

# Inférence des $a_\xi$



**FIGURE – Comparaison, selon  $\xi$ , des qualité d'ajustement d'un modèle affine pour  $t \mapsto \partial_t \mathcal{F} u_t(\xi)$  : parties réelles (gauche) ou imaginaires (droite) de  $a_\xi$ , simulations bruitées (bas) ou non (haut).**

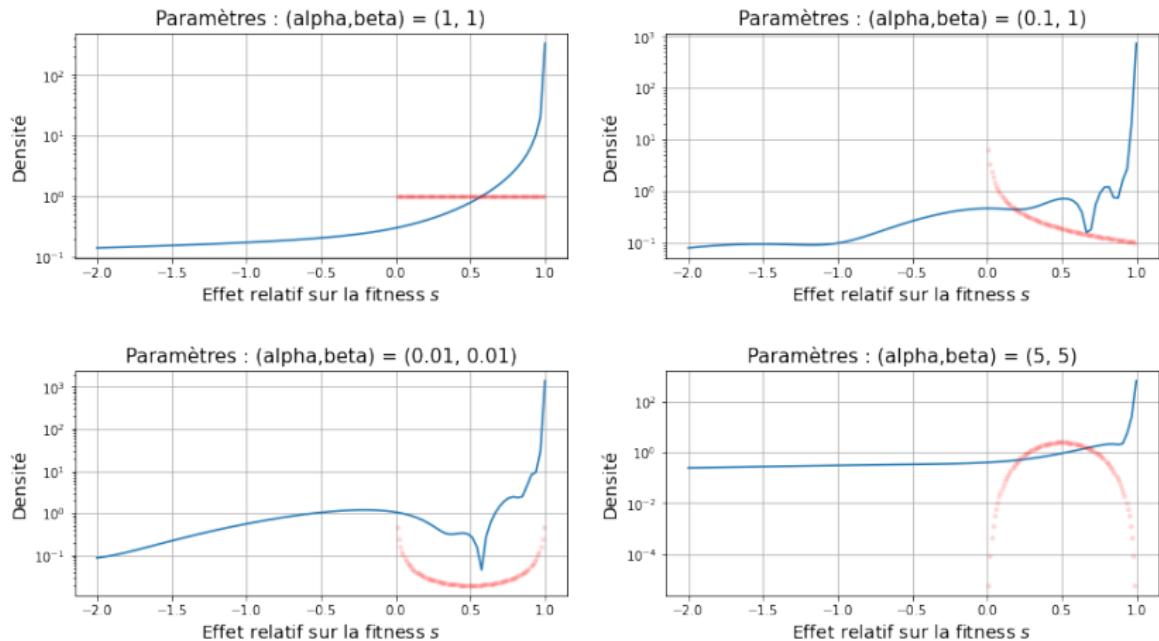
# Inférence des $a_\xi$



**FIGURE – Comparaison, selon  $\xi$ , des qualité d'ajustement d'un modèle affine pour  $t \mapsto \partial_t \mathcal{F} u_t(\xi)$  : parties réelles (gauche) ou imaginaires (droite) de  $a_\xi$ , simulations bruitées (bas) ou non (haut).**

- fonctionne bien pour les petits  $\xi$  ;
- fonctionne très mal pour les grands  $\xi$  ( $\rightarrow$  overfitting ?).

# Mauvais résultats



**FIGURE –** Comparaison entre les DFE inférées (en bleu) et les lois Beta choisies pour effectuer les simulations (en rouge).

# Simulation avec grand nombre de canaux I

On enlève ?

- \* droites magnifiques
- \* stabilité pour  $\xi_{max}$  petit
- \* non-stabilité pour  $\xi_{max}$  grand

# Conclusion et perspectives

- Problème inverse sévèrement mal posé
- 3 approches : Analyses naïves, Problème des moments, EDP
  - Résultats théoriques mais méthodes insuffisantes en pratique
- Notebooks annotés : analyses reproductibles
  - En attente de données plus nombreuses ou moins bruitées
- Améliorations :
  - ★ Combinaison EDP + estimation des moments
  - ★ Similitudes avec les problèmes de fragmentation
  - ★ Propriétés de la DFE
  - Simulations efficaces pour tester ces approches

# Références

Merci pour votre attention !

-  Robert et al., *Mutation dynamics and fitness effects followed in single cells*, Science 359, 1283–1286, 16 March 2018
-  Trudy F. C. Mackay, Eric A. Stone, Julien F. Ayroles, *The genetics of quantitative traits : challenges and prospects*, Nature Reviews Genetics, 565–577, 2009
-  Doumic, Escobedo, *Time asymptotics for a critical case in fragmentation and growth-fragmentation equations*, submitted 2015
-  Beal et al., *The Division of Amyloid Fibrils : Systematic Comparison of Fibril Fragmentation Stability by Linking Theory with Experiments*, iScience, 25 September 2020