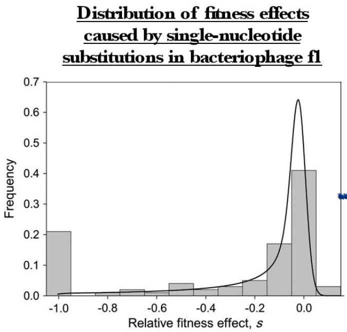
Estimer l'effet des mutations sur la fitness d'une population de bactéries

Jérémy Andréoletti et Nathanaël Boutillon

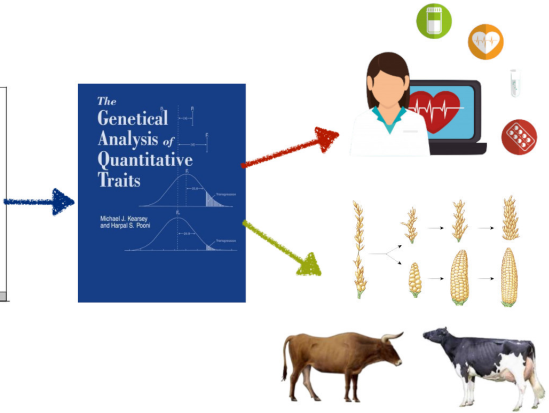
Encadrantes : Marie Doumic et Lydia Robert

Objectifs

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



Peris et al., Genetics, 2010



Modèle

On pose

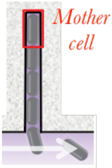
$$
\begin{aligned}
W_t & & \text{fitness (taux de croissance) au temps } t \\
s_i & = \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}} & \text{effet \textit{relatif} de la mutation } i \\
N_t & & \text{nombre de mutations avant le temps } t
\end{aligned}
$$

d'où :

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

$\rightarrow$ But $=$ trouver la loi des s_i.

Experience μ-MA (mutation accumulation)



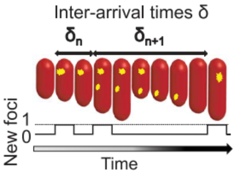
Robert et al.

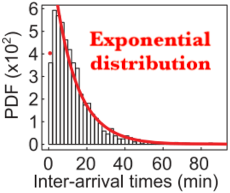
Experience MV (mutation visualization)

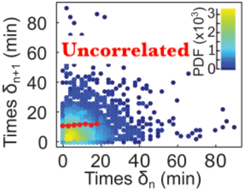
Robert et al.

Résultats de l'article

- les mutations apparaissent selon un **processus de Poisson** ;

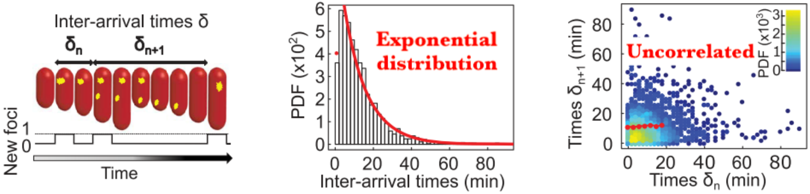




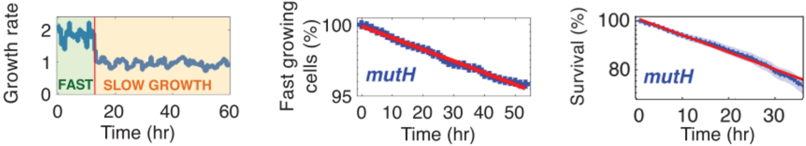


Résultats de l'article

- les mutations apparaissent selon un **processus de Poisson** ;



- 0,3% des mutations sont fortement délétères ; 1% sont létales ; leurs effets sont **indépendants** ;



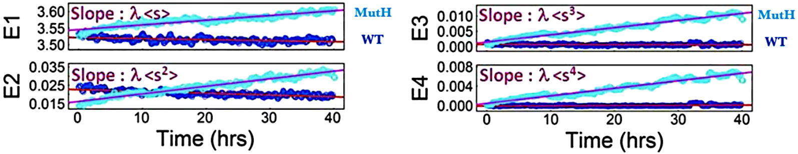
Résultats de l'article

- estimation des **premiers moments** de la loi des effets des mutations ;

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

- estimation des **premiers moments** de la loi des effets des mutations ;

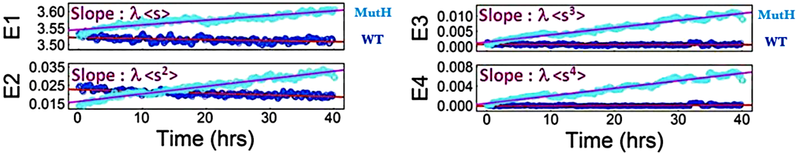
$$E_n(t) := \sum_{k=1}^{n} \binom{k}{n} (-1)^k \ln\left(\mathbb{E}\left[W_t^k\right]\right) = (\lambda \mathbb{E}[s^n])\, t$$



Résultats de l'article

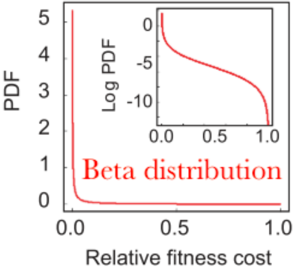
- estimation des **premiers moments** de la loi des effets des mutations ;

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

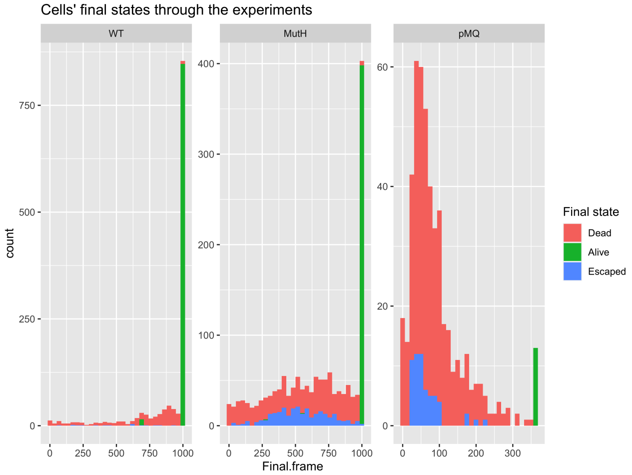


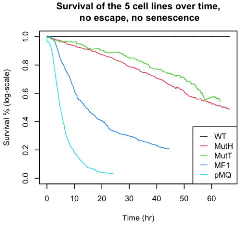
Non-parametric DFE characterization

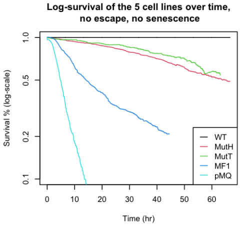
Mean	$3.1\text{x}10^{-3}$ +/- $0.4\text{x}10^{-3}$	**Skewness**	16.6 +/- 0.7
CV	9.5 +/- 1.2	**Kurtosis**	360 +/- 90



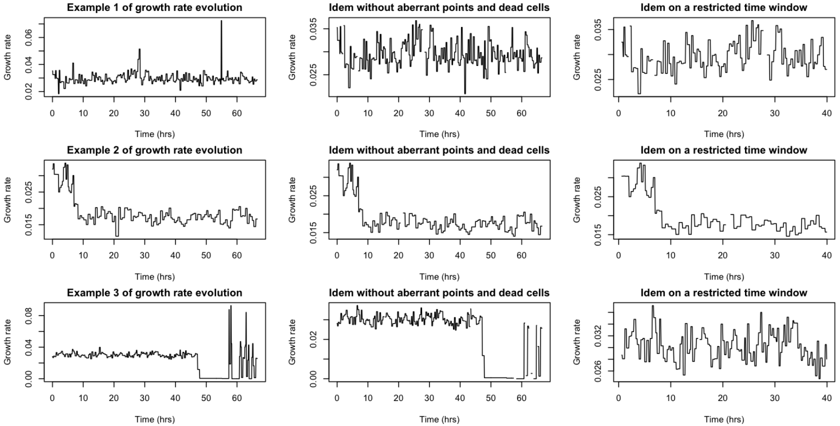
Analyses préliminaires - Survie





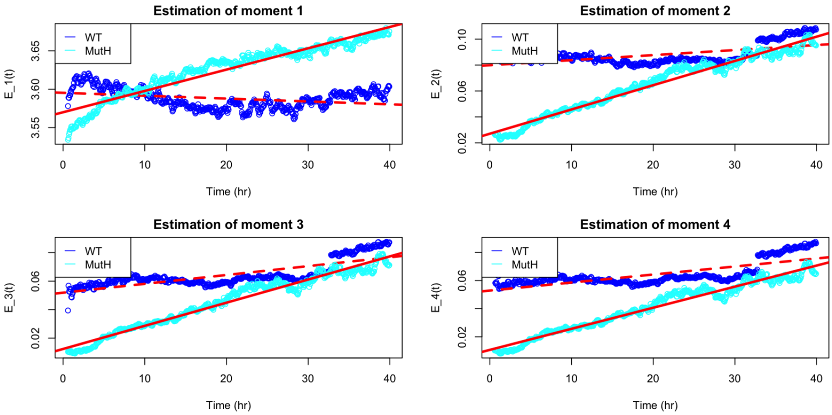


Analyses préliminaires - Taux de croissance



Analyses préliminaires - Taux de croissance

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



Choses à faire

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 ($t \geqslant 0$), et sachant que

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

Choses à faire

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 ($t \geqslant 0$), et sachant que

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

- Trouver la loi de s à partir des $\mathbb{E}[s^n]$: semble délicat dans notre cas : on n'obtient qu'une approximation très grossière.
- Faire un programme en Python pour tester les résultats que l'on obtiendra en partant d'une distribution fixée pour les s_i ;
- Utiliser le fait que $\ln W_t$ est une somme de variables aléatoires iid pour trouver la loi de ces variables.

Références

Robert et al., *Mutation dynamics and fitness effects followed in single cells*, Science 359, 1283–1286, 16 March 2018

Peris et al., *Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1*, Genetics, 2010

Eyre-Walker Keightley, *The distribution of fitness effects of new mutations*, Nat Rev Genet 8, 610–618, 2007