

# Un modèle d'évolution stochastique

Hervé Guiol<sup>1</sup>, Fábio Machado<sup>2</sup>, Rinaldo Schinazi<sup>3</sup>

<sup>1</sup>Université de Grenoble

<sup>2</sup>IME-USP - Brasil

<sup>3</sup>UCCS - USA

Bordeaux, Journées MAS, 3 Septembre 2010

# Résumé

## Modèle d'arbre phylogénétique de Liggett-Schinazi 2009

Motivation

Description et résultat

## Modèle d'évolution de Bak et Sneppen

Modèle de Bak et Sneppen : Motivation

Description

Phénomène de cutoff

## Modèle alternatif

Description

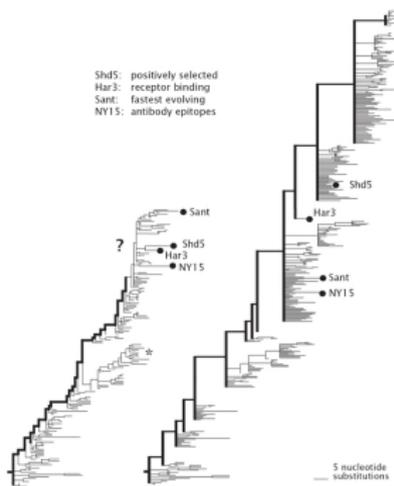
Résultats

Conclusion

## Modèle de Liggett-Schinazi 2009 : Motivation

Arbre phylogénétique = schématise les relations de parenté entre entités ayant un ancêtre commun. Les distances entre nœuds et branches tiennent compte du nombre de mutations de nucléotides entre isolats.

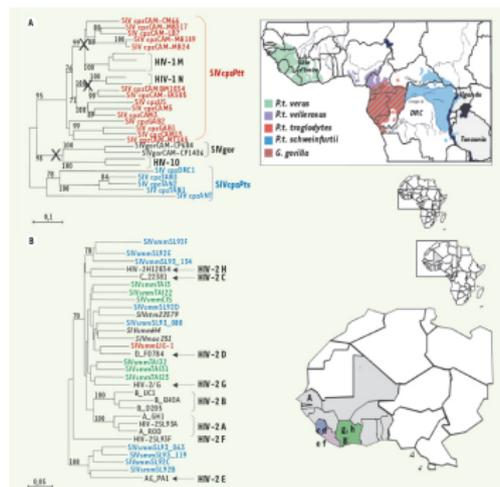
Pour H3 sous type de l'hémaglutine (HA) (protéine impliquée dans l'entrée du virus de la grippe dans les cellules) Bush et al 1999 (Science 286) montrent sur un jeu de données allant de 1983 à 1997 :



On remarquera un type d'évolution contrainte "en cactus". On a un type dominant pour une période assez longue alors que les autres disparaissent rapidement. Puis le type dominant disparaît et est substitué par un autre.

## Modèle de Liggett-Schinazi 2009 : Motivation

L'observation de la phylogénie des SIV et VIH, M. Peeters et al. 2008 *Medecine/Sciences* 24, montre des type d'arbres beaucoup plus fourni et équilibré et n'indique pas ou peu de compétition entre les lignées



Il est bon de préciser que le virus de la grippe est moins varié dans le monde entier que le VIH à Amsterdam (Korber et al. 2001).

# Modèle de Liggett-Schinazi 2009 : Motivation

Le but de Liggett et Schinazi était de trouver un modèle stochastique très simple pouvant expliquer ces phénomènes.

## Modèle de Liggett & Schinazi : description

On considère un processus de naissance et mort dans lequel chaque naissance représente la création d'un nouveau type (mutation) ; on suppose les taux linéaires :

$$\begin{aligned}n &\rightarrow n + 1 \text{ au taux } n\lambda ; \\n &\rightarrow n - 1 \text{ au taux } n \text{ si } n \geq 2.\end{aligned}$$

## Modèle de Liggett & Schinazi : description

On considère un processus de naissance et mort dans lequel chaque naissance représente la création d'un nouveau type (mutation) ; on suppose les taux linéaires :

$$\begin{aligned}n &\rightarrow n + 1 \text{ au taux } n\lambda ; \\n &\rightarrow n - 1 \text{ au taux } n \text{ si } n \geq 2.\end{aligned}$$

- ▶ On utilise une approche gros grain : lorsque la  $i$ -ème mutation apparaît on lui associe une variable aléatoire qui représentera la fitness  $f_i$  (résistance) du nouveau type portant la mutation. On suppose les  $f_i$  *i.i.d.* de loi uniforme sur  $[0, 1]$ .

## Modèle de Liggett & Schinazi : description

On considère un processus de naissance et mort dans lequel chaque naissance représente la création d'un nouveau type (mutation) ; on suppose les taux linéaires :

$$\begin{aligned}n &\rightarrow n + 1 \text{ au taux } n\lambda ; \\n &\rightarrow n - 1 \text{ au taux } n \text{ si } n \geq 2.\end{aligned}$$

- ▶ On utilise une approche gros grain : lorsque la  $i$ -ème mutation apparaît on lui associe une variable aléatoire qui représentera la fitness  $f_i$  (résistance) du nouveau type portant la mutation. On suppose les  $f_i$  *i.i.d.* de loi uniforme sur  $[0, 1]$ .
- ▶ Interaction : chaque mort correspond à la disparition du type ayant la plus petite fitness  $f_i$ .

# Modèle de Liggett & Schinazi : résultat

## Theorem

Soit  $\alpha \in (0, 1)$ . La limite suivante existe

$$\lim_{t \rightarrow \infty} \mathbb{P}(\text{les types } \textit{dominants} \text{ aux temps } \alpha t \text{ et } t \text{ sont les mêmes}) \\ = \alpha \mathbf{1}_{(0,1]}(\lambda).$$

En particulier si  $\lambda < 1$  le type dominant au temps  $t$  est vraisemblablement présent depuis un temps d'ordre  $t$  et il n'y aura pas beaucoup des types : consistant avec l'arbre de la grippe.

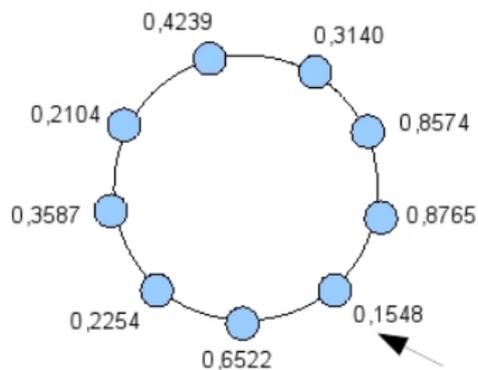
Par contre pour  $\lambda > 1$  le type dominant au temps  $t$  n'était là que depuis un temps d'ordre plus petit que  $t$  et à n'importe quel moment il y aura toujours beaucoup de types : consistant avec l'arbre du VIH tree.

# Motivation

- ▶ Gould et Eldredge (1972) ont proposé une théorie dans laquelle l'évolution comprendrait de longues périodes d'équilibres ponctués et de brèves périodes de changements importants comme la spéciation ou les extinctions (cette idée se retrouvait déjà chez Darwin...)
- ▶ Bak et Sneppen (1993) en proposent un modèle sous forme d'automate cellulaire. Leurs résultats basés entièrement sur des simulations montrent des phénomènes d'équilibres ponctués et de cutoff.

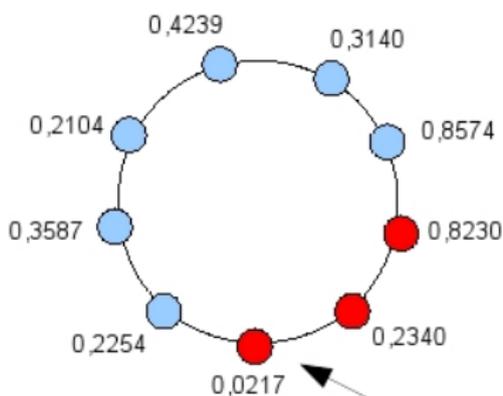
## Modèle de Bak-Sneppen 1993 : Description

- ▶ Nombre d'espèce fixé :  $N$  espèces sont arrangées sur le tore  $\mathbb{T}_N = \{1, \dots, N\}$ .
- ▶ A chaque espèce  $i$  on attribue  $i$  une barrière aléatoire  $f_i$  iid uniforme sur  $[0, 1]$  ;
- ▶ Dynamique comme suit



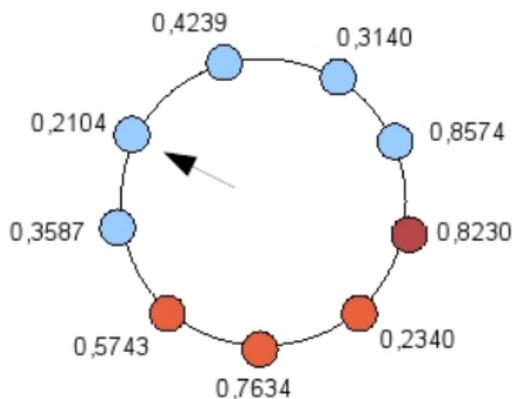
## Modèle de Bak-Sneppen 1993 : Description

- ▶ Nombre d'espèce fixé :  $N$  espèces sont arrangées sur le tore  $\mathbb{T}_N = \{1, \dots, N\}$ .
- ▶ A chaque espèce  $i$  on attribue  $i$  une barrière aléatoire  $f_i$  iid uniforme sur  $[0, 1]$  ;
- ▶ Dynamique comme suit



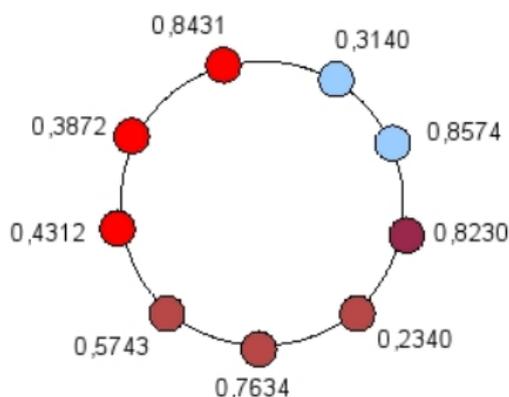
## Modèle de Bak-Sneppen 1993 : Description

- ▶ Nombre d'espèce fixé :  $N$  espèces sont arrangées sur le tore  $\mathbb{T}_N = \{1, \dots, N\}$ .
- ▶ A chaque espèce  $i$  on attribue  $i$  une barrière aléatoire  $f_i$  iid uniforme sur  $[0, 1]$  ;
- ▶ Dynamique comme suit



## Modèle de Bak-Sneppen 1993 : Description

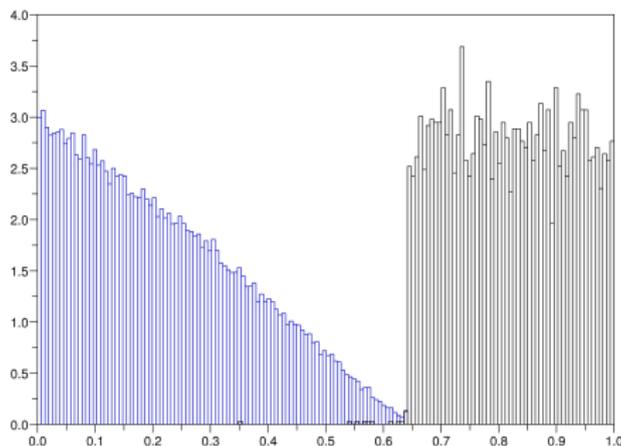
- ▶ Nombre d'espèce fixé :  $N$  espèces sont arrangées sur le tore  $\mathbb{T}_N = \{1, \dots, N\}$ .
- ▶ A chaque espèce  $i$  on attribue  $i$  une barrière aléatoire  $f_i$  iid uniforme sur  $[0, 1]$  ;
- ▶ Dynamique comme suit



## Phénomène de cutoff

Bak et Sneppen ont conjecturé (pour  $N$  grand) que la loi de répartition de la fitness d'une espèce dans le régime à l'équilibre est uniforme sur  $(c, 1)$  avec  $c \approx 2/3$ . Bien que simple à décrire l'analyse mathématique du modèle est très technique. Meester et Znamenski ont montré que la loi limite est non triviale ( $c < 1$ ) puis sous une condition d'égalité d'exposants critiques (question toujours ouverte) que la conjecture de Bak et Sneppen serait exacte.

$N = 5000$  ; bleu histogramme de la fitness minimale ; noir l'histogramme des fitness après 500000 mutations.

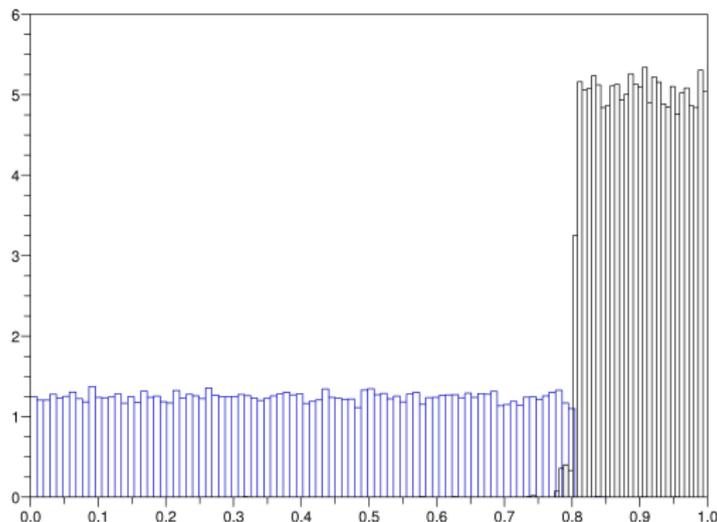


## Un modèle alternatif : description

- ▶ Grain grossier : les espèces sont représentées par leur fitness.
- ▶ Taille de population stochastique : à chaque temps  $n \geq 1$  il y a soit une naissance d'une espèce avec probabilité  $p > 1/2$  soit une mort (si le nombre d'espèce n'est pas 0).
- ▶ Les fitness sont des v.a. *i.i.d.* r.v. de loi uniforme sur  $(0, 1)$ .
- ▶ Une seule interaction globale : lorsqu'une espèce meurt c'est celle de plus petite fitness.

# Simulations

Pour  $p = 5/9$  et 200000 naissances et morts. Bleue : histogramme de la fitness minimale ; en noir : histogramme des fitness des espèces survivantes. Seuil à  $f_c = 0.8 = 1 - p/p$ .



# Résultats

**Théorème** On suppose  $p > 1/2$  et on pose  $f_c = \frac{1-p}{p}$ .

- (a) Le nombre  $|L_n|$  d'espèces dont la fitness est en dessous de  $f_c$  au temps  $n$  est une chaîne de naissance et mort récurrente nulle. En particulier l'ensemble  $L_n$  est vide infiniment souvent avec probabilité 1.
- (b) Pour tous  $f_c < a < b < 1$  l'ensemble  $R_n$  des fitness des espèces survivantes au temps  $n$  vérifie

$$\lim_{n \rightarrow \infty} \frac{1}{n} |R_n \cap (a, b)| = p(b - a) \text{ a.s.}$$

# Conclusion

- ▶ En conséquence
  - ▶ Plus grand est  $p$  plus favorable est l'environnement aux nouvelles espèces. Si  $p$  est proche de 1 alors  $f_c$  est proche de zéro et même les espèces avec une fitness relativement faible survivront.
  - ▶ Si  $p$  est seulement un peu au dessus de  $1/2$  alors  $f_c$  est proche de 1 et seulement les espèces dont la fitness est très grande survivront.
- ▶ En conclusion nous avons un modèle mathématique d'évolution ayant les mêmes propriétés de cutoff que le modèle de Bak et Sneppen. Ce modèle est d'analyse mathématique très simple. Le phénomène de cutoff ne nécessite pas d'interaction locale.

**P. Bak and K. Sneppen** (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, **74**, 4083-4086.

**B. Korber, B. Gaschen, K. Yusim et al.** (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *British Medical Bulletin*, **58**, 19-42.

**T. Liggett and R. B. Schinazi** (2009). A stochastic model for phylogenetic trees. *J. Appl. Probab.* **46**, 601-607.

**R. Meester and D. Znamenski** (2003). Limit behavior of the Bak-Sneppen evolution model. *The Annals of Probability*, **31**, 1986-2002.

**R. Meester and D. Znamenski** (2004). Critical thresholds and the limit distribution in the Bak-Sneppen model. *Communications in Mathematical Physics* **246**, 63-86.