

# Généalogies : dynamique et effet de la sélection

Damien SIMON

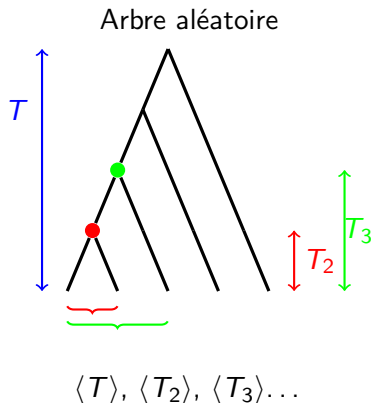
Laboratoire de Physique Statistique  
École Normale Supérieure (Paris)

GT *MEV*, Mars 2008



# Sommaire

- Généalogies en absence de sélection
- Temps de coalescence en présence de sélection et fronts bruités
- un modèle de sélection avec un seuil de *fitness* : fronts déterministes



# Evolution d'une population

- reproduction et mort : *processus aléatoire*.
- effet de saturation et compétition : la taille est *régulée*

## Plus précisément :

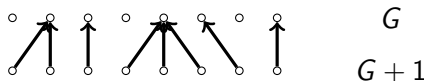
- une seule population (pas d'espèces différentes)
- taille constante



# Sans sélection : modèle de Wright-Fisher

Kingman 1982

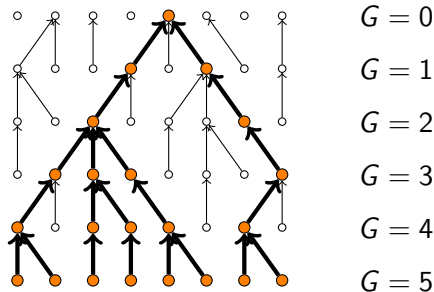
- Taille constante  $N$
- Générations sans recouvrement
- Chaque individu a un parent tiré au hasard uniformément dans la génération précédente
- Processus obtenu en remontant le temps : *généalogies* (coalescence)



# Sans sélection : modèle de Wright-Fisher

Kingman 1982

- Taille constante  $N$
- Générations sans recouvrement
- Chaque individu a un parent tiré au hasard uniformément dans la génération précédente
- Processus obtenu en remontant le temps : *généalogies* (coalescence)



# Généalogies

Grande taille  $N \gg 1$  :

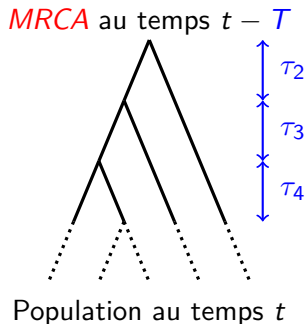
- petit groupe  $p \ll N$  : seulement des coalescences **binaires**
- échelle de temps  $\propto N$



# Généalogies

Grande taille  $N \gg 1$  :

- petit groupe  $p \ll N$  : seulement des coalescences **binaires**
- échelle de temps  $\propto N$
- **MRCA** : Ancêtre commun le plus récent (âge  $T$ ).  
(*Most Recent Common Ancestor*)
- temps de première coalescence  $\tau_p$  de  $p$  individus
- $T = \tau_2 + \tau_3 + \dots$



# Distribution stationnaire ( $N \rightarrow \infty$ ) : revue

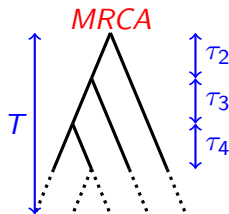
- Temps de coalescence  $\tau_p$  :

$$\rho_p(\tau_p) = \frac{p(p-1)}{2} e^{-\frac{p(p-1)}{2} \tau_p}$$

- Âge  $T$  du *MRCA* :

$$T = \tau_2 + \tau_3 + \tau_4 + \dots$$

$$\rho(T) = \sum_{p=2}^{\infty} \frac{(-1)^p (2p-1) p(p-1)}{2} e^{-\frac{p(p-1)}{2} T}$$



Kingman 1982





# Distribution stationnaire ( $N \rightarrow \infty$ ) : revue

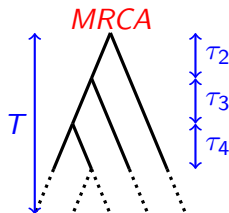
- Temps de coalescence  $\tau_p$  :

$$\rho_p(\tau_p) = \frac{p(p-1)}{2} e^{-\frac{p(p-1)}{2} \tau_p}$$

- Âge  $T$  du **MRCA** :

$$T = \tau_2 + \tau_3 + \tau_4 + \dots$$

$$\rho(T) = \sum_{p=2}^{\infty} \frac{(-1)^p (2p-1) p(p-1)}{2} e^{-\frac{p(p-1)}{2} T}$$



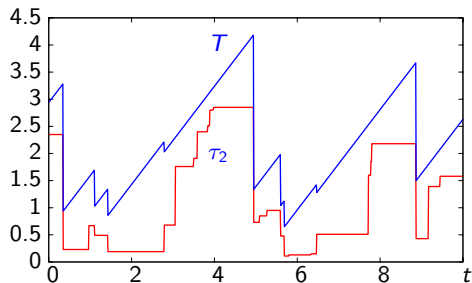
Kingman 1982



# Propriétés dynamiques des généalogies

- **extinctions** de familles  
→ évolution de  $T(t)$  et des  $\tau_p(t)$

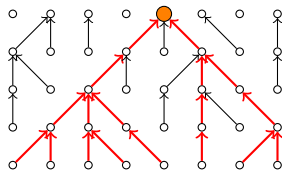
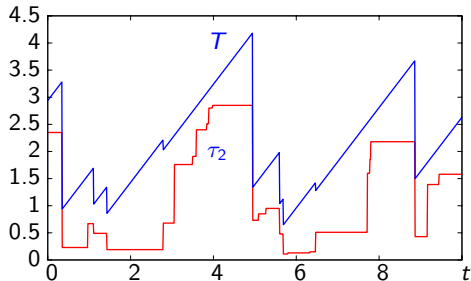
Serva 2005  
S. & Derrida, 2006



# Propriétés dynamiques des généalogies

- **extinctions** de familles  
→ évolution de  $T(t)$  et des  $\tau_p(t)$

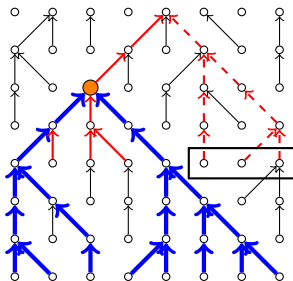
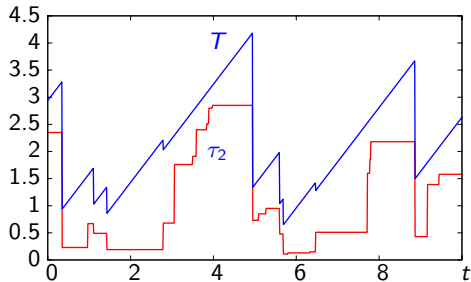
Serva 2005  
S. & Derrida, 2006



# Propriétés dynamiques des généalogies

- **extinctions** de familles  
→ évolution de  $T(t)$  et des  $\tau_p(t)$

Serva 2005  
S. & Derrida, 2006

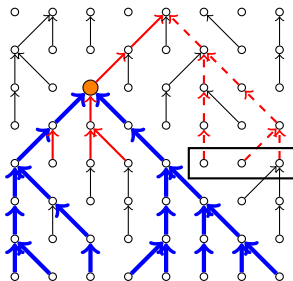
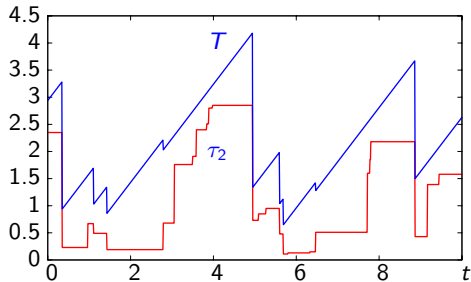


# Propriétés dynamiques des généalogies

- **extinctions** de familles

→ évolution de  $T(t)$  et des  $\tau_p(t)$

Serva 2005  
S. & Derrida, 2006

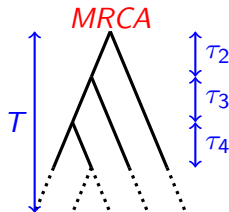


$$\text{Lors d'une discontinuité : } \begin{cases} T \mapsto T - \tau_2 \\ \tau_2 \mapsto \tau_3 \\ \tau_3 \mapsto \tau_4 \dots \end{cases}$$



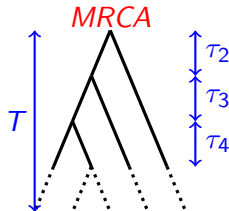
# Description par un processus markovien

- Description des  $\tau_p(t)$  par un processus de Markov
- Taux obtenus à partir de la statistique des arbres



# Description par un processus markovien

- Description des  $\tau_p(t)$  par un processus de Markov
- Taux obtenus à partir de la statistique des arbres



Cascade des  $\tau_p(t)$  avec probabilité  $kdt$  :

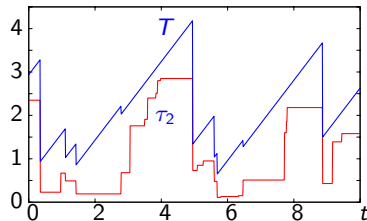
$$\begin{cases} \tau_p(t + dt) = \tau_p(t) & \text{pour } p < k \\ \tau_k(t + dt) = \tau_k(t) + \tau_{k+1}(t) \\ \tau_p(t + dt) = \tau_{p+1}(t) & \text{pour } p > k \end{cases}$$

→ Reproduit les dynamiques observées pour  $T, \tau_2, \dots$



# Fonctions de corrélation de $T(t)$ et des $\tau_p(t)$

- Corrélations  $\langle \tau_p(t) \tau_{p'}(0) \rangle \dots$
- Corrélation de l'âge du MRCA  
 $T = \tau_2 + \tau_3 + \dots :$



S. & Derrida 2006

$$\langle T(t)T(0) \rangle - \langle T(t) \rangle \langle T(0) \rangle = \sum_{p=2}^{\infty} K_p e^{-\frac{p(p-1)}{2}t}$$

avec des coefficients combinatoires :

$$K_p = 4(-1)^p(2p+1) \sum_{q=p}^{\infty} \frac{1}{q} \frac{(q-2)!}{(q-p)!(q+p+1)!}$$

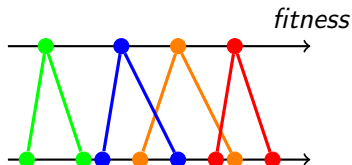




# Modèles avec sélection

- Population de taille  $N$  **constante**
- chaque individu a  $k \geq 1$  enfants dans la génération suivante
- un individu est caractérisé par un seul paramètre, son *fitness*
- le *fitness* est transmis à des fluctuations près : **mutations**
- à chaque génération, seuls les  $N$  **meilleurs** individus sont **conservés**.

Snydner 2003  
Kloster 2005  
Brunet Derrida Mueller  
Munier 2006-07



# Modèles avec sélection

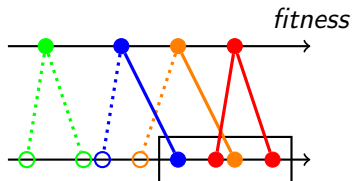
- Population de taille  $N$  **constante**
- chaque individu a  $k \geq 1$  enfants dans la génération suivante
- un individu est caractérisé par un seul paramètre, son *fitness*
- le *fitness* est transmis à des fluctuations près : **mutations**
- à chaque génération, seuls les  $N$  **meilleurs** individus sont **conservés**.

Snydner 2003

Kloster 2005

Brunet Derrida Mueller

Munier 2006-07



# Questions

- vitesse  $\langle v_N \rangle$  de la *fitness* de la population.
- Généalogies :
  - ▶ âge  $T_k$  de l'ancêtre commun le plus récent de  $k$  individus choisis aléatoirement,
  - ▶ forme des arbres.



# Questions

- vitesse  $\langle v_N \rangle$  de la *fitness* de la population.
- Généalogies :
  - ▶ âge  $T_k$  de l'ancêtre commun le plus récent de  $k$  individus choisis aléatoirement,
  - ▶ forme des arbres.

## Outils :

- **fronts bruités** (équation F-KPP)  
et **marches aléatoires avec branchements** ;

Doering, Mueller, Smerecka 2003

- au moins un modèle exactement soluble : le **modèle exponentiel**



# Modèle exponentiel

Brunet Derrida Mueller Munier 2006-2007

- Population de taille  $N$  constante
- à la génération suivante, chaque individu a un nombre infini d'enfants
- *Poisson point process* : un individu à la position  $x$  a un enfant dans l'intervalle  $[y, y + dy]$  avec une probabilité  $e^{-(y-x)} dy$
- les  $N$  meilleurs individus (les plus à droite) sont conservés.

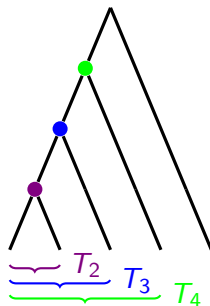
Exactement soluble !



# Temps de coalescence $T_k$

$T_k$  : âge de l'ancêtre commun le plus récent de  $k$  individus choisis au hasard

- Sans sélection :  $\langle T_k \rangle \propto N$
- Avec sélection (BRW) :  $\langle T_k \rangle \propto \ln^3 N$
- Modèle exponentiel :  $\langle T_k \rangle \propto \ln N$

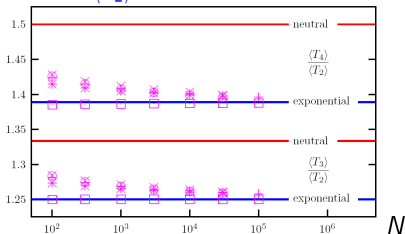


Brunet Derrida Mueller Munier 2006-2007



# Rapports des temps de coalescence $T_k$

Ratios  $\frac{\langle T_p \rangle}{\langle T_2 \rangle}$



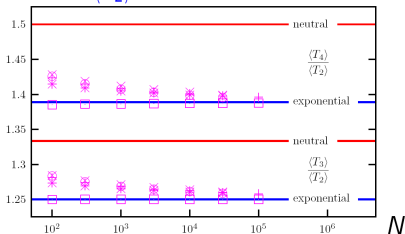
	Neutre	Sélection	Expo.
$\frac{\langle T_3 \rangle}{\langle T_2 \rangle}$	$\frac{4}{3}$	$\frac{5}{4}$	$\frac{5}{4}$
$\frac{\langle T_4 \rangle}{\langle T_2 \rangle}$	$\frac{3}{2}$	$\frac{25}{18}$	$\frac{25}{18}$

Valeurs universelles



# Rapports des temps de coalescence $T_k$

Ratios  $\frac{\langle T_p \rangle}{\langle T_2 \rangle}$



	Neutre	Sélection	Expo.
$\frac{\langle T_3 \rangle}{\langle T_2 \rangle}$	$\frac{4}{3}$	$\frac{5}{4}$	$\frac{5}{4}$
$\frac{\langle T_4 \rangle}{\langle T_2 \rangle}$	$\frac{3}{2}$	$\frac{25}{18}$	$\frac{25}{18}$

Valeurs **universelles**

Remarques :






- **sélection** : même valeur que pour le **coalescent de Bolthausen-Sznitman** et les arbres dans les verres de spin.

Parisi 1979-80  
Bolthausen-Sznitman 1998





# Forme des arbres

Arbre	Neutre	Sélection
	$\frac{2}{3}$	$\frac{1}{3}$
	$\frac{1}{3}$	$\frac{1}{6}$
	0	$\frac{1}{6}$
	0	$\frac{2}{9}$
	0	$\frac{1}{9}$

## Neutre

- coalescences binaires uniquement

## Sélection :

- coalescences multiples autorisées
- $\equiv$  verres de spin en champ moyen



# Quelques relations en présence de sélection

Marches aléatoires avec branchements (image phénoménologique) :

- vitesse en présence de sélection (équation de front KPP avec bruit) :

$$v_N = v_c - \frac{A}{\ln^2 N} + \dots$$

- Temps de coalescence  $\langle T_k \rangle \propto \ln^3 N$



## Quelques relations en présence de sélection

Marches aléatoires avec branchements (image phénoménologique) :

- vitesse en présence de sélection (équation de front KPP avec bruit) :

$$v_N = v_c - \frac{A}{\ln^2 N} + \dots$$

- Temps de coalescence  $\langle T_k \rangle \propto \ln^3 N$

Expressions équivalentes :

$$\langle T_k \rangle \propto (v_c - v)^{-3/2} \quad D_N \times \langle T_2 \rangle \simeq \text{Conste}$$

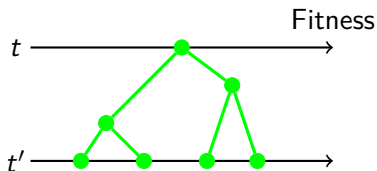
→ Vérification de la première relation dans un modèle exactement soluble ?



# Évolution avec un seuil de *fitness*

S. & Derrida 2007

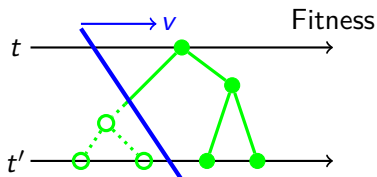
- Chaque individu a un *fitness* qui **diffuse** sur la ligne réelle.
- chaque individu **se divise** en deux individus avec une probabilité  $\beta dt$
- **Seuil de *fitness*** : un individu qui passe sous le seuil **meurt** instantanément
- le seuil se déplace avec une vitesse  $v$



# Évolution avec un seuil de *fitness*

S. & Derrida 2007

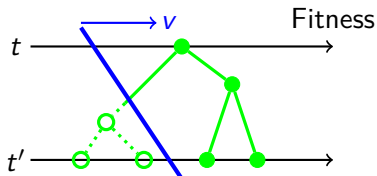
- Chaque individu a un *fitness* qui **diffuse** sur la ligne réelle.
- chaque individu **se divise** en deux individus avec une probabilité  $\beta dt$
- **Seuil** de *fitness* : un individu qui passe sous le seuil **meurt** instantanément
- le seuil se déplace avec une vitesse  $v$



# Évolution avec un seuil de *fitness*

S. & Derrida 2007

- Chaque individu a un *fitness* qui **diffuse** sur la ligne réelle.
- chaque individu **se divise** en deux individus avec une probabilité  $\beta dt$
- **Seuil** de *fitness* : un individu qui passe sous le seuil **meurt** instantanément
- le seuil se déplace avec une vitesse  $v$



- Taille fluctuante  $N$  mais vitesse constante  $v$ .
- $N$  peut s'annuler : **probabilité de survie**  $Q_s(x, t)$



# Probabilité de survie $Q_s(x, t)$

- $x \geq 0$  : position de l'individu initial dans le référentiel du seuil
- diffusion  $D = 1$  & division avec un taux  $\beta$

Équation F-KPP pour  $Q_s(x, t)$  :

Fisher 1937

Kolmogorov Petrovsky Piscounov 1937

McKean 1975

$$\partial_t Q_s = \partial_x^2 Q_s - v \partial_x Q_s + \beta(Q_s - Q_s^2)$$



# Probabilité de survie $Q_s(x, t)$

- $x \geq 0$  : position de l'individu initial dans le référentiel du seuil
- diffusion  $D = 1$  & division avec un taux  $\beta$

Équation F-KPP pour  $Q_s(x, t)$  :

Fisher 1937

Kolmogorov Petrovsky Piscounov 1937

McKean 1975

$$\partial_t Q_s = \partial_x^2 Q_s - v \partial_x Q_s + \beta(Q_s - Q_s^2)$$

avec les conditions initiales :

$$\begin{cases} Q_s(x, 0) = 1 & \text{pour } x > 0 \\ Q_s(0, t) = 0 \end{cases}$$

→ front (déterministe) avec des conditions aux bords modifiées !





# Comportement aux temps longs de $Q_s(x, t)$

- pour  $v > v_c$ ,  $Q_s(x, t) \rightarrow 0$  aux temps longs ;
- si  $v = v_c$  :

Kesten 1978

$$Q_s(x, t) \underset{t \rightarrow \infty}{\propto} \exp(-Ct^{1/3})$$

- pour  $v < v_c$ ,  $Q_s(x, t) \rightarrow Q_s^*(x)$  aux temps longs avec :

$$Q_s^*(x) \propto \exp(-\pi(v_c - v)^{-1/2})$$



# Comportement aux temps longs de $Q_s(x, t)$

- pour  $v > v_c$ ,  $Q_s(x, t) \rightarrow 0$  aux temps longs ;
- si  $v = v_c$  :

Kesten 1978

$$Q_s(x, t) \underset{t \rightarrow \infty}{\propto} \exp(-Ct^{1/3})$$

- pour  $v < v_c$ ,  $Q_s(x, t) \rightarrow Q_s^*(x)$  aux temps longs avec :

$$Q_s^*(x) \propto \exp(-\pi(v_c - v)^{-1/2})$$



# Comportement aux temps longs de $Q_s(x, t)$

- pour  $v > v_c$ ,  $Q_s(x, t) \rightarrow 0$  aux temps longs ;
- si  $v = v_c$  :

$$Q_s(x, t) \underset{t \rightarrow \infty}{\propto} \exp(-Ct^{1/3})$$

- pour  $v < v_c$ ,  $Q_s(x, t) \rightarrow Q_s^*(x)$  aux temps longs avec :

$$Q_s^*(x) \propto \exp(-\pi(v_c - v)^{-1/2})$$

Kesten 1978



# Comportement aux temps longs de $Q_s(x, t)$

- pour  $v > v_c$ ,  $Q_s(x, t) \rightarrow 0$  aux temps longs ;
- si  $v = v_c$  :

Kesten 1978

$$Q_s(x, t) \underset{t \rightarrow \infty}{\propto} \exp(-Ct^{1/3})$$

- pour  $v < v_c$ ,  $Q_s(x, t) \rightarrow Q_s^*(x)$  aux temps longs avec :

$$Q_s^*(x) \propto \exp(-\pi(v_c - v)^{-1/2})$$

et des temps de relaxation :

Derrida & S. 2007

$$\tau_1 \propto (v_c - v)^{-3/2}$$

(cf. le cas de taille constante...)



# Régime quasi-stationnaire

S. & Derrida 2007

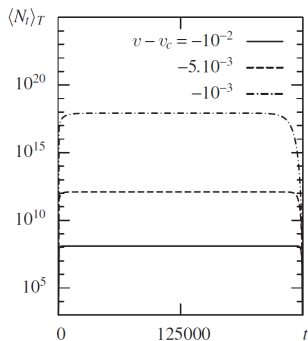
- mur absorbant qui se déplace à la vitesse  $v < v_c$
- **conditionnement** sur une taille  $N = 1$  au temps final  $T$
- propriétés lorsque  $t < T$  ?



# Régime quasi-stationnaire

S. & Derrida 2007

- mur absorbant qui se déplace à la vitesse  $v < v_c$
- **conditionnement** sur une taille  $N = 1$  au temps final  $T$
- propriétés lorsque  $t < T$  ?



- Si  $t \gg 1$  et  $T - t \gg 1$ , mesure indépendante de  $t$  : **régime quasi-stationnaire**
- Taille moyenne quasi-stationnaire  $\langle N \rangle_{qs}$



# Marches aléatoires avec branchements : relations entre les modèles

	Taille constante $N$	Seuil de vitesse $v$
	(stationnaire)	
Taille-vitesse	$v_N - v_c \simeq \frac{A_1}{\ln^2 N} + \frac{A_2 \ln \ln N}{\ln^3 N}$	
Temps	$T \propto (v_c - v)^{-3/2}$	$\tau_1 \propto (v_c - v)^{-3/2}$
Outils	F-KPP bruité	F-KPP déterministe



# Marches aléatoires avec branchements : relations entre les modèles

	Taille constante $N$	Seuil de vitesse $v$
	(stationnaire)	(quasi-stationnaire)
Taille-vitesse	$v_N - v_c \simeq \frac{A_1}{\ln^2 N} + \frac{A_2 \ln \ln N}{\ln^3 N}$	$\langle N \rangle_{qs} \propto e^{-C(v_c - v)^{-1/2}}$
Temps	$T \propto (v_c - v)^{-3/2}$	$\tau_1 \propto (v_c - v)^{-3/2}$
Outils	F-KPP bruité	F-KPP déterministe

→ Équivalence plus générale ?





# Conclusion

- Sans sélection :
  - ▶ temps de coalescence (coalescent de Kingman)
  - ▶ âge de l'ancêtre commun le plus récent d'une population
  - ▶ **corrélations dynamiques des généalogies**
  
- Avec sélection
  - ▶ les temps de coalescence (Bolthausen-Sznitman) sont différents
  - ▶ liens avec **équations de front** (F-KPP avec bruit)
  - ▶ équivalence avec le **régime quasi-stationnaire** et des fronts déterministes



# Conclusion

- Sans sélection :
  - ▶ temps de coalescence (coalescent de Kingman)
  - ▶ âge de l'ancêtre commun le plus récent d'une population
  - ▶ **corrélations dynamiques des généalogies**
  
- Avec sélection
  - ▶ les temps de coalescence (Bolthausen-Sznitman) sont différents
  - ▶ liens avec **équations de front** (F-KPP avec bruit)
  - ▶ équivalence avec le **régime quasi-stationnaire** et des fronts déterministes



# Conclusion

- Sans sélection :
  - ▶ temps de coalescence (coalescent de Kingman)
  - ▶ âge de l'ancêtre commun le plus récent d'une population
  - ▶ **corrélations dynamiques des généalogies**
- Avec sélection
  - ▶ les temps de coalescence (Bolthausen-Sznitman) sont différents
  - ▶ liens avec **équations de front** (F-KPP avec bruit)
  - ▶ équivalence avec le **régime quasi-stationnaire** et des fronts déterministes

## Travail en cours :

- structure spatiale et coalescence :

$$\left\{ \begin{array}{l} \text{neutre + espace : Kingman for } d \geq 3 \\ \text{sélection + espace ? Polymères dirigés...} \end{array} \right.$$



# Articles

Généalogies dans le cas neutre :

- D. S. & B. Derrida, *Evolution of the most recent common ancestor of a population with noselection*, J. Stat. Mech. (2006) P05002.

Temps de coalescence  $T_k$  avec sélection

- É. Brunet, B. Derrida, A.H. Mueller, S. Munier, *Noisy traveling waves : effect of selection on genealogies*, Europhys. Lett., **76** (2006).

Sélection avec un seuil de vitesse  $v$  et régime quasi-stationnaire :

- D. S. & B. Derrida, *Quasi-stationary regime of a branching random walk in presence of an absorbing wall*, J. Stat. Phys. (2008).

Coalescences avec structure spatiale :

- É. Brunet, B. Derrida, D. S., *Universal tree structures in directed polymers and models of selection*, en préparation (2008).

