# Adaptive dynamics in the limit of small populations in a stochastic individual-based model

Nicolas Champagnat, IECL and Inria Nancy



Workshop “Population Dynamics and Statistical Physics in Sinergy”,
Eurandom, Eindhoven, August 28, 2014

# Adaptive dynamics

Darwinian evolution: Three main ingredients

- Heredity: transmissions of individual characteristics from a generation to the next one.
- Mutation: cause of the variability in individual characteristics.
- Selection: consequence of the interactions between individuals and their environment, including the rest of the population (ecology).

Adaptive dynamics (since the 90s): Hofbauer and Sigmund (1990), Metz, Geritz et al. (1992,1996), Dieckmann and Law (1996)...

- Focus on the interplay between ecology and evolution
- Ecological interactions modeled in detail
- Heredity is simplified as much as possible: asexual (clonal) reproduction

# Adaptive dynamics

$\rightsquigarrow$ Density-dependent individual-based models where no fitness is given. The fitness landscape has to be constructed from the parameters of the model.

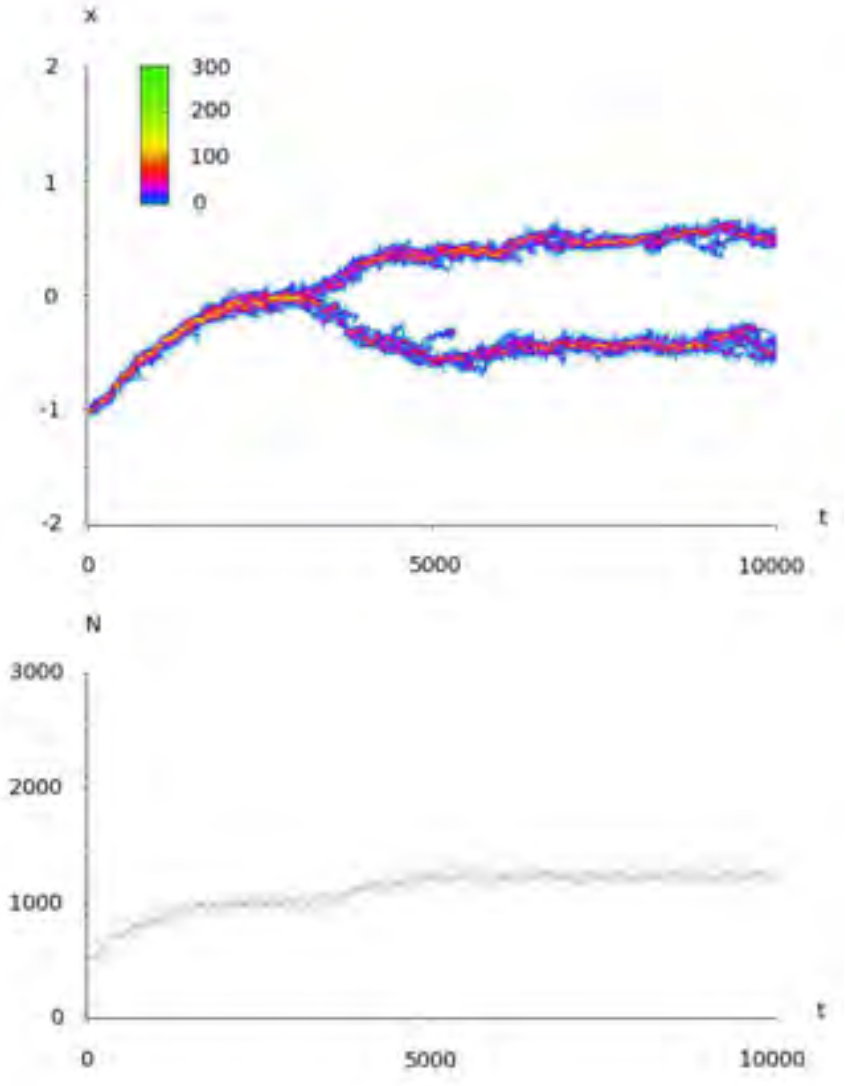
$\rightsquigarrow$ New phenomenon of evolutionary branching (Metz et al., 1996)

- Transition from a population concentrated around a single phenotype to a population concentrated around several distinct phenotypes, still under ecological interaction
- Mechanism of diversification
- Can lead to speciation without geographical separation (Dieckmann and Doebeli, 1999)

$\rightsquigarrow$ Three biological assumptions (Metz et al., 1996):

- large populations
- rare mutations
- small mutation steps

# Evolutionary banching



# Individual-based model

Birth-death-competition-mutation process (Metz et al. 1996, Bolker-Pacala 97, Kisdi 99, Dieckmann-Law 00, Doebeli-Dieckmann 01, Fournier-Méléard 04, C.-Ferrière-Méléard 06...).

- Each individual characterized by a continuous phenotypic trait $x \in \mathcal{X} \subset \mathbb{R}$ (individual size, age at maturity, rate of food intake...).
- $K$ scales the size of the population
- $\mu$ scales the probability of mutation
- $\sigma$ scales the size of mutation steps
- At time $t$, the population is composed of $N_K(t)$ individuals with weights $\frac{1}{K}$ and traits $x_1, \ldots, x_{N_K(t)} \in \mathcal{X}$:

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_K(t)} \delta_{x_i}.$$

# Transition rates for an individual with trait $x$

- Reproduction at rate $b(x)$:
  - With probability $1-\mu$, clonal reproduction (offspring with trait $x$).
  - With probability $\mu$, mutation, and the mutant trait is $x+\sigma H$, where $H \sim m(h)dh$, symmetric w.r.t. 0 (e.g. Gaussian distribution).
- Death without competition at rate $d(x)$.
- Death from competition with any other individual of trait $y$ at rate $\frac{1}{K}c(x,y)$.

$\rightsquigarrow$ an individual with trait $x$ dies at density dependent rate

$$d(x) + \frac{1}{K}\sum_{i=1}^{N_K(t)} c(x,x_i) - c(x,x)$$

$$= d(x) + \int_{\mathcal{X}} c(x,y)\left(\nu_t^K(dy) - \frac{1}{K}\delta_x(dy)\right).$$

# On the limit $K \to +\infty$

Assume $b(x) \equiv b$, $d(x) \equiv d$ and $c(x,y) \equiv c$ (neutral case).
Then the total number of individuals $N_t^K$ is a Markov birth and death process with

- birth rate $bn$
- death rate $dn + cn\frac{n-1}{K}$ when $N_t^K = n$.

It is well known that $N_t^K/K$ converges when $K \to +\infty$ to the solution of the logistic equation

$$\dot{n} = n(b - d - cn).$$

Remark: also true if $\nu_0^K \to n_0\delta_x$ and $\mu = 0$, with $b = b(x)$, $d = d(x)$ and $c = c(x,y)$ (monomorphic case). We will use the notation

$$\bar{n}(x) = \frac{b(x) - d(x)}{c(x,x)}$$

the equilibrium of the logistic equation (monomorphic equilibrium).

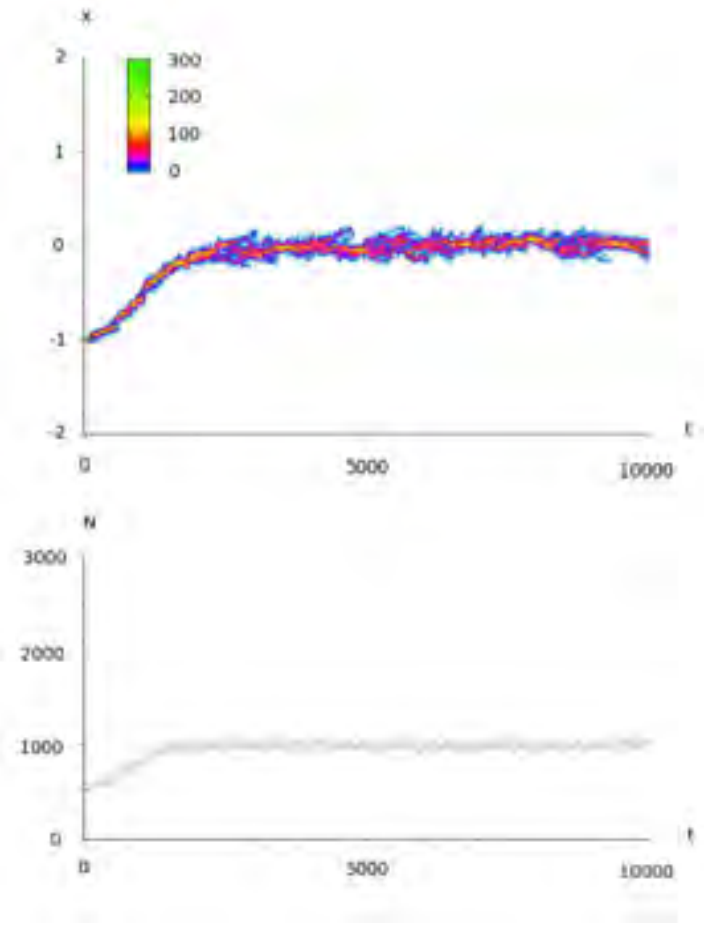
# Example

Roughgarden (1976,1979), Dieckmann-Doebeli (1999)

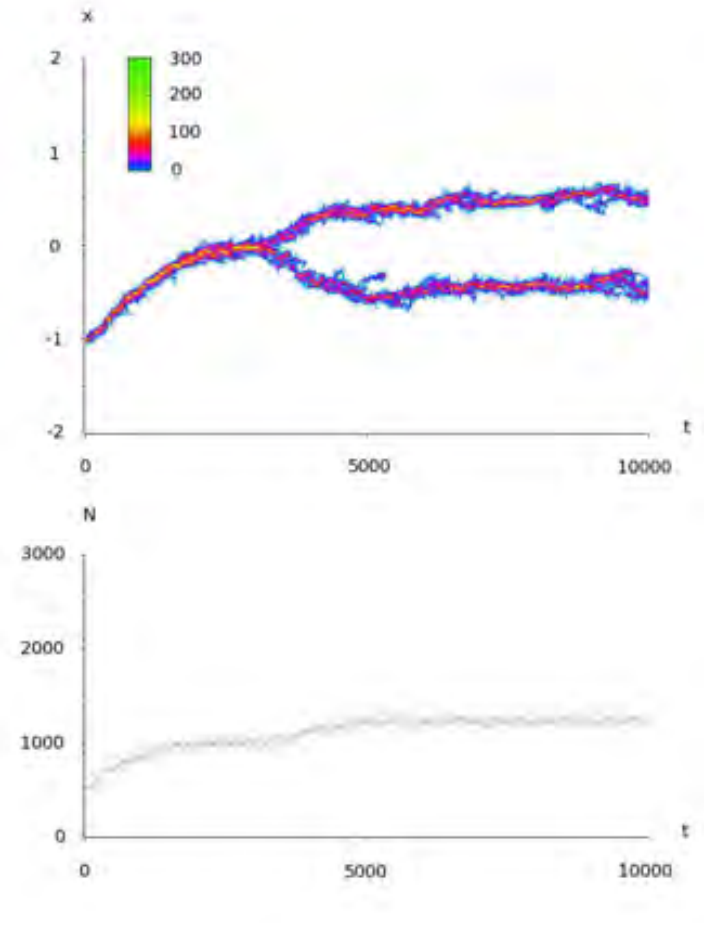
- $\mathcal{X} = [-2, 2] \quad d(x) \equiv 0 \quad u_K = 1 \quad p(x) = p.$
- $m(h)dh = \mathcal{N}(0, 1) \quad$ (conditioned on $x + h \in \mathcal{X}$).
- $b(x) = \exp\left(-\frac{x^2}{2\sigma_b^2}\right)$, maximum at 0.
- Symmetric competition for resources:

$$\alpha(x, y) = \alpha(x - y) = \exp\left(-\frac{(x - y)^2}{2\sigma_\alpha^2}\right).$$

# Simulations

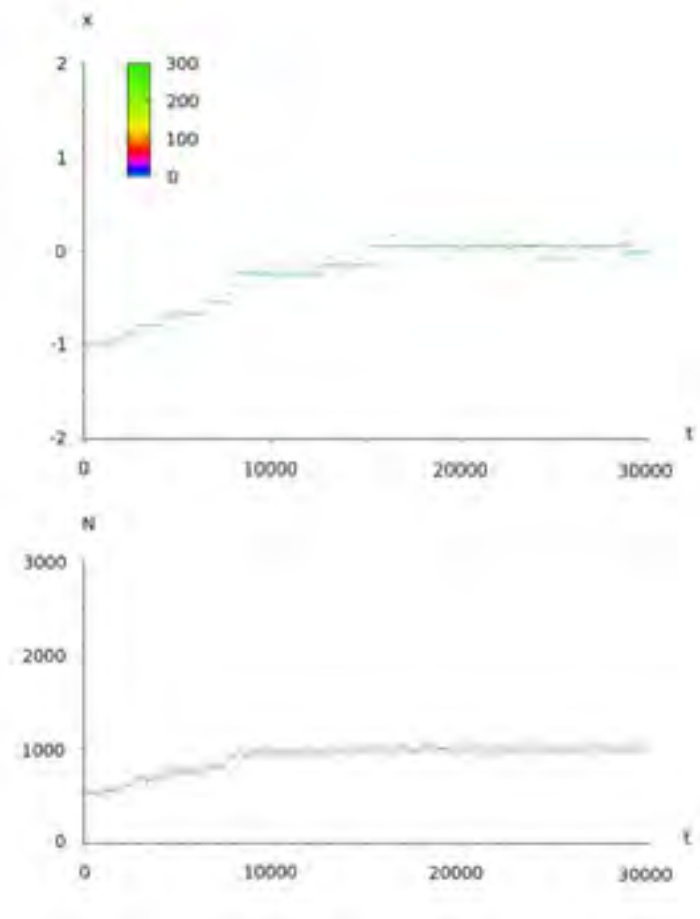



$\mu = 0.1,\ K = 1000,\ \sigma = 0.01,$
$\sigma_b = 0.9,\ \sigma_\alpha = 1.0.$

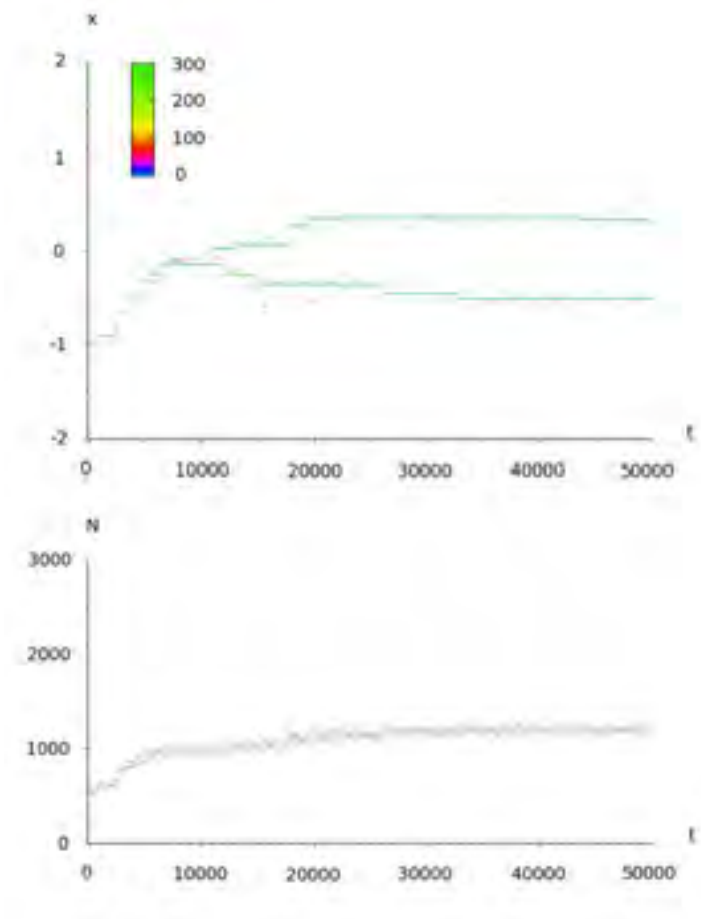



$\mu = 0.1,\ K = 1000,\ \sigma = 0.01,$
$\sigma_b = 0.9,\ \sigma_\alpha = 0.7.$

# Simulations: rare mutations




$\mu = 0.0001,\ K = 1000,$
$\sigma = 0.08,\ \sigma_b = 0.9,\ \sigma_\alpha = 1.0.$




$\mu = 0.0001,\ K = 1000,$
$\sigma = 0.08,\ \sigma_b = 0.9,\ \sigma_\alpha = 0.7.$

# Limit of rare mutations: Metz et al. 1996

- The selection process has sufficient time between two mutations to eliminate disadvantaged traits.
- Large population assumption: (nearly) deterministic population dynamics between mutations, so that one can predict the outcome of competition between the traits.

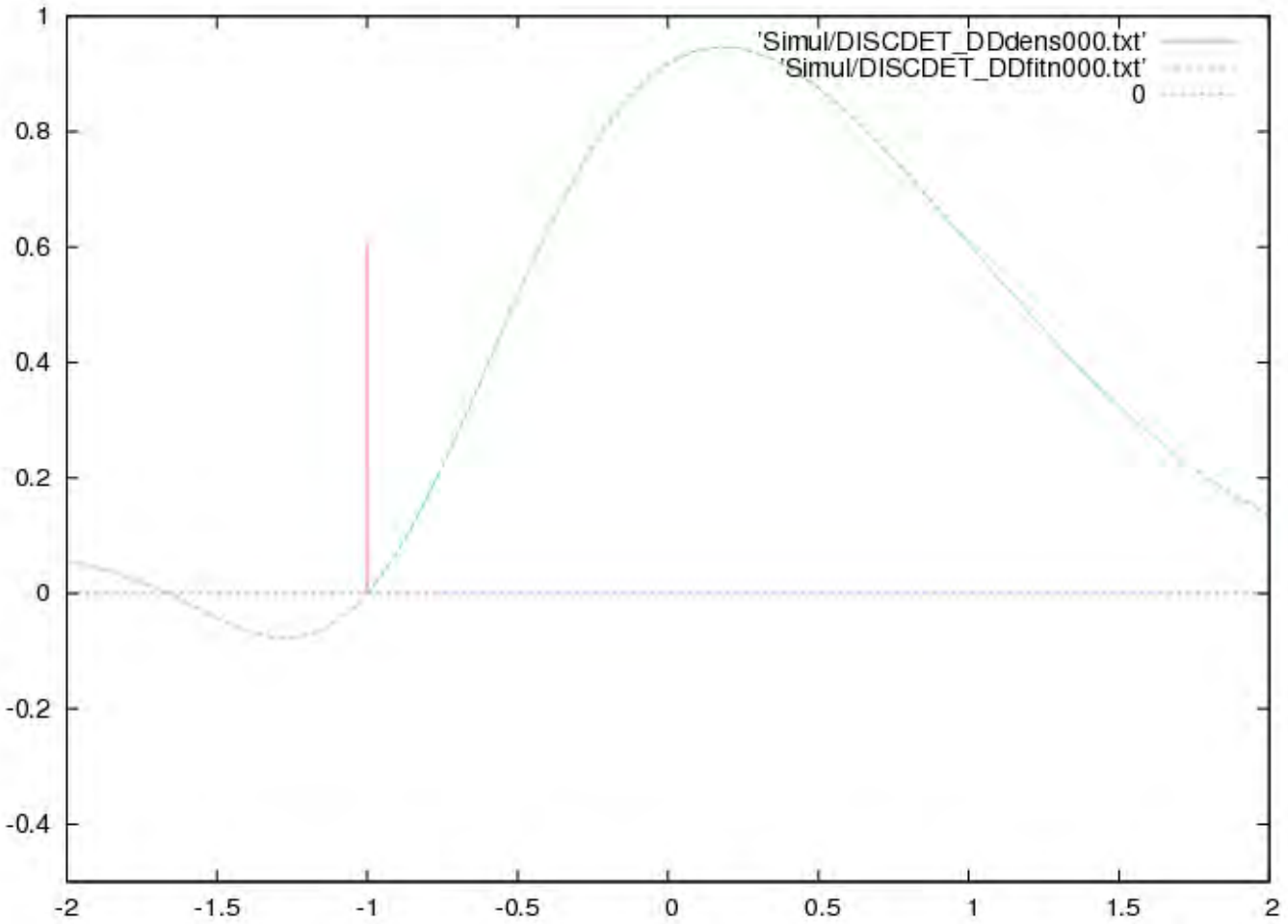
$\rightsquigarrow$ Succession of phases of (random) mutant invasion, and phases of (fast, deterministic) competition between traits.

Adaptive walk in a fitness landscape that depends on the current state of the population: fitness of a mutant trait $y$ in a population $x$ at equilibrium

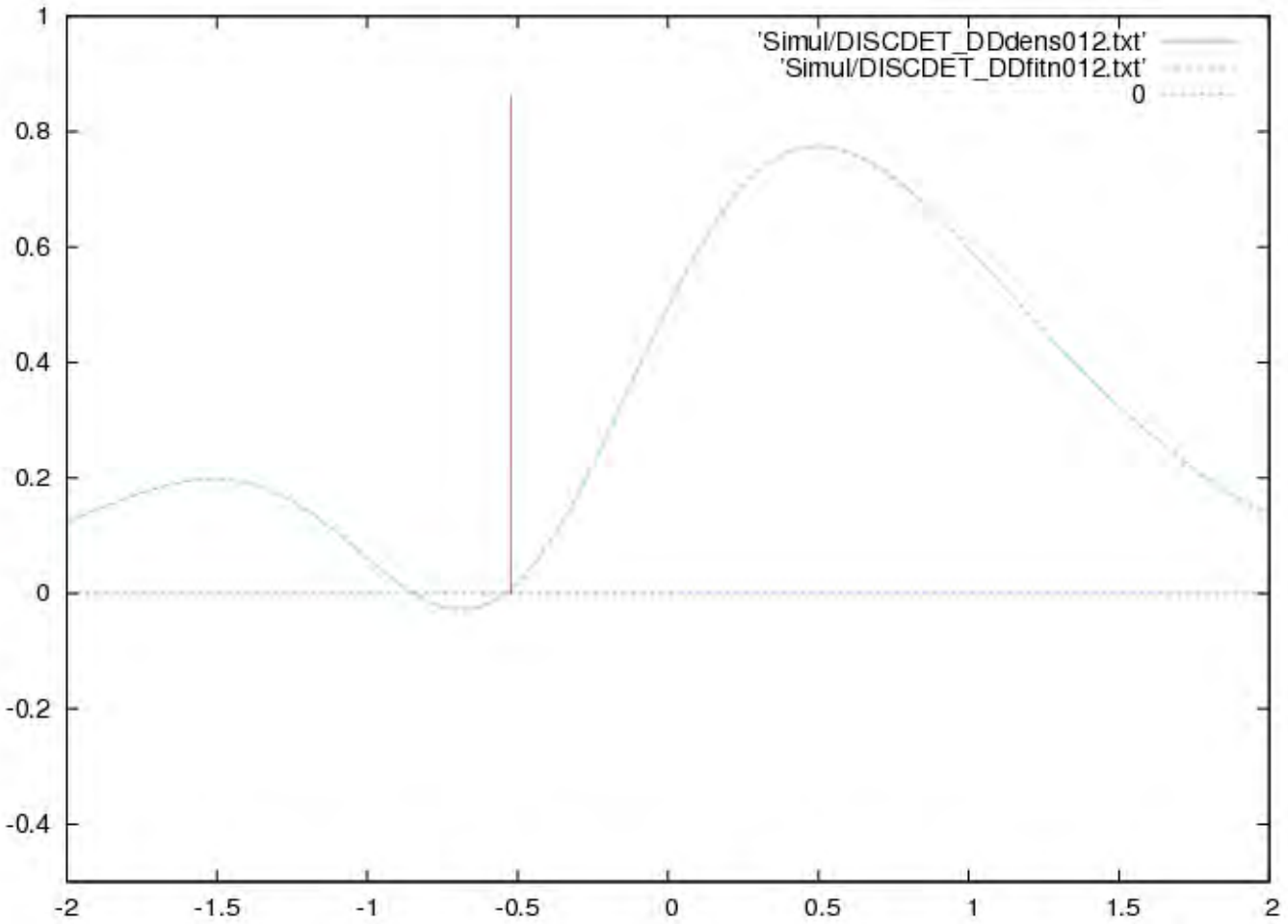
$$f(y,x) = b(y) - d(y) - c(y,x)\bar{n}(x)$$

(C., 2006, C. and Méléard 2011, Baar, Bovier, C. $\rightsquigarrow$ next talk !)

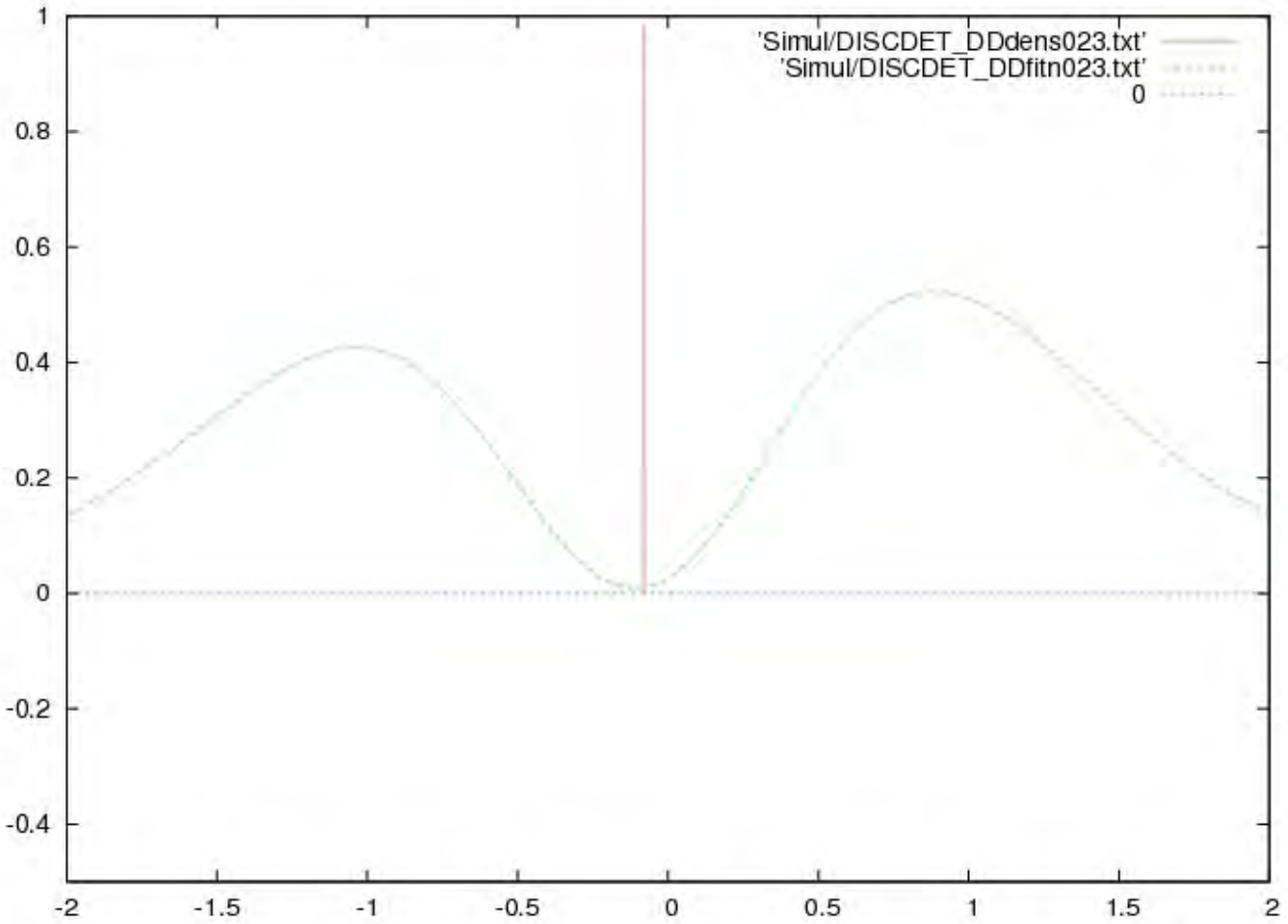
# Coevolution with the fitness landscape



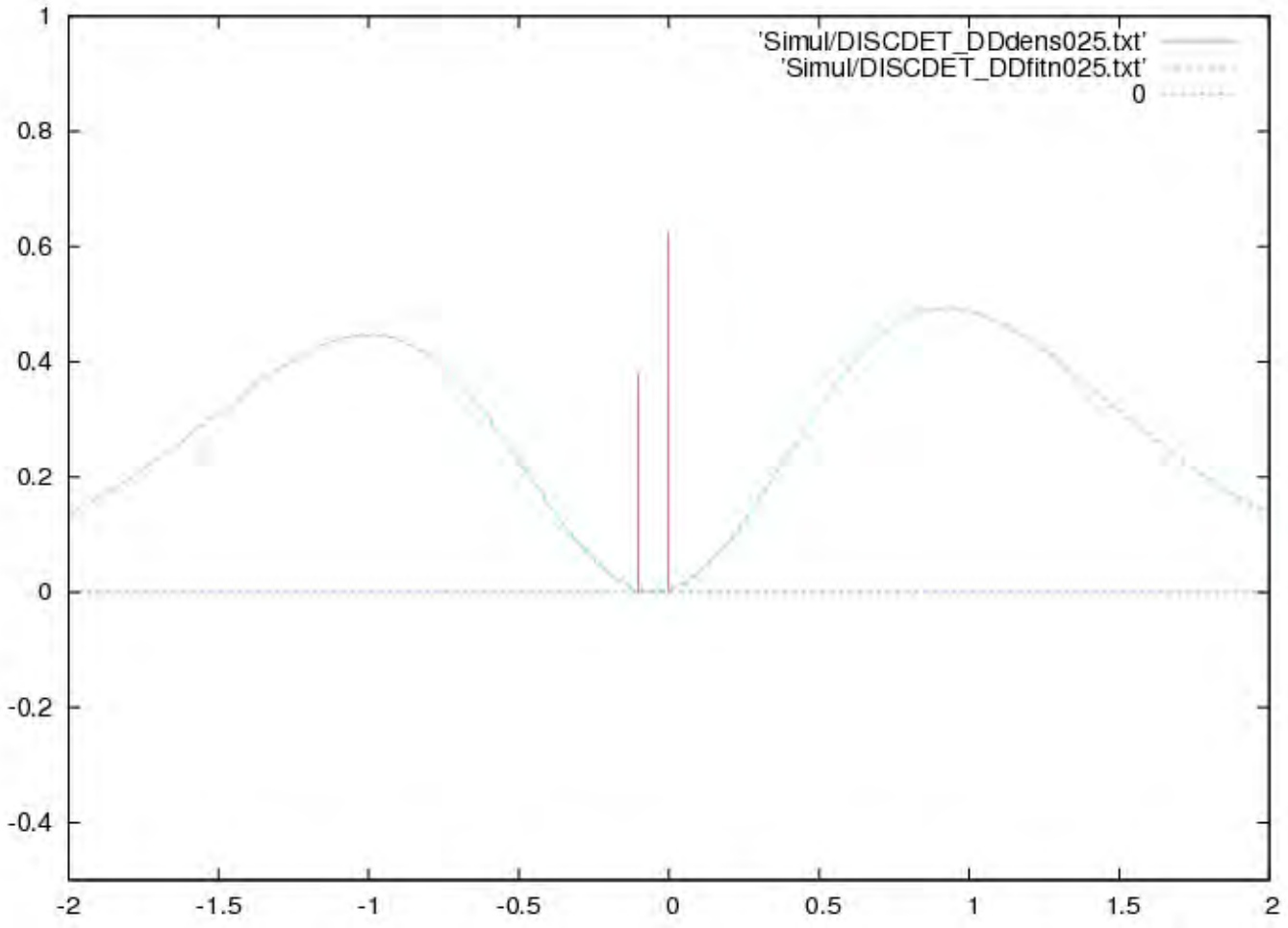
# Coevolution with the fitness landscape



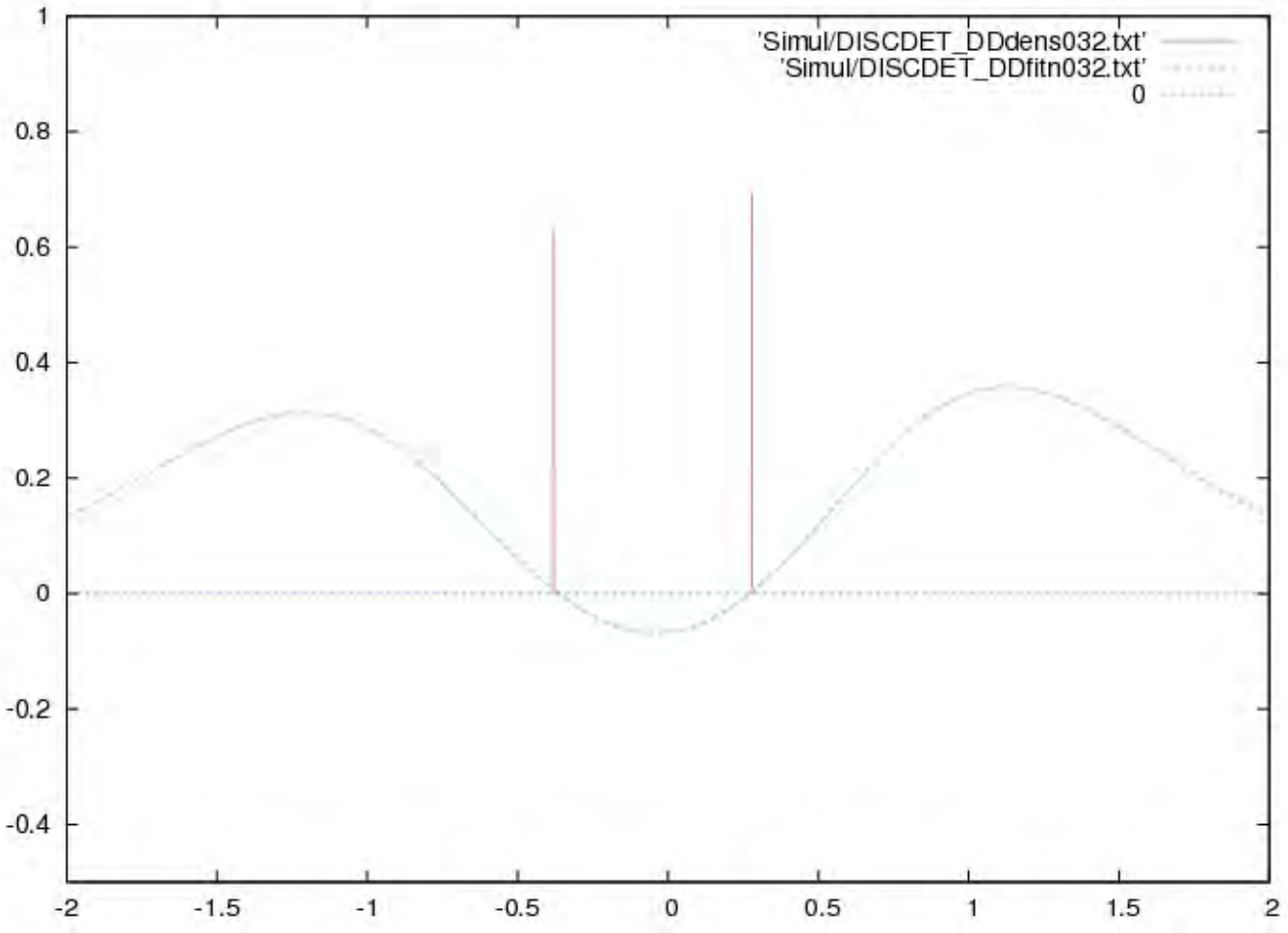
# Coevolution with the fitness landscape



# Coevolution with the fitness landscape



# Coevolution with the fitness landscape



# Canonical Equation of Adaptive Dynamics (first form)

Adaptive walk with small mutations:
When $\mu \to 0$, $K \to +\infty$ and $\sigma \to 0$ conveniently (Baar, Bovier, C.), on the time scale $\frac{t}{K\mu\sigma^2}$, the individual-based model converges to $(\bar{n}(x(t))\delta_{x(t)}, t \geq 0)$, where $x$ is solution of the ODE

$$\frac{dx}{dt} = \int h^2 m(h)dh\, \bar{n}(x)\partial_1 f(x;x).$$

- "hill-climbing" process in the fitness landscape (Dieckmann and Law, 1996).
- evolutionary branching can also be described with this approach (C., Méléard, 2011)
- A criticism (Waxman, Gavrilets, 2005): this approach with rare mutations leads to evolution on a too long time-scale (mutations are too rare).

# Large population limit

Fournier and Méléard, 2004: assuming that $\nu_0^K$ converges in law to the measure $u_0(x)dx$ for the weak topology when $K \to +\infty$, then $(\nu_t^K, t \geq 0)$ converges in law to $(u(t,x)dx, t \geq 0)$, where $u(t,x)$ is solution to the PDE

$$\partial_t u(t,x) = u(t,x)\left((1-\mu)b(x) - d(x) - \int_{\mathcal{X}} c(x,y)u(t,y)\,dy\right) + \int_{\mathcal{X}} b(y)\mu u(t,y) m\left(\frac{x-y}{\sigma}\right)\frac{dy}{\sigma}.$$

Assuming $\mu = 1$ and $\sigma = \varepsilon$, this PDE can be written as

$$\partial_t u(t,x) = u(t,x)\left(r(x) - \int_{\mathbb{R}^\ell} c(x,y)u(t,y)\,dy\right) + \int_{\mathbb{R}^\ell} m(h)(u(t,x+\varepsilon h) - u(t,x))\,dh,$$

# Large population limit

Fournier and Méléard, 2004: assuming that $\nu_0^K$ converges in law to the measure $u_0(x)dx$ for the weak topology when $K \to +\infty$, then $(\nu_t^K, t \geq 0)$ converges in law to $(u(t,x)dx, t \geq 0)$, where $u(t,x)$ is solution to the PDE

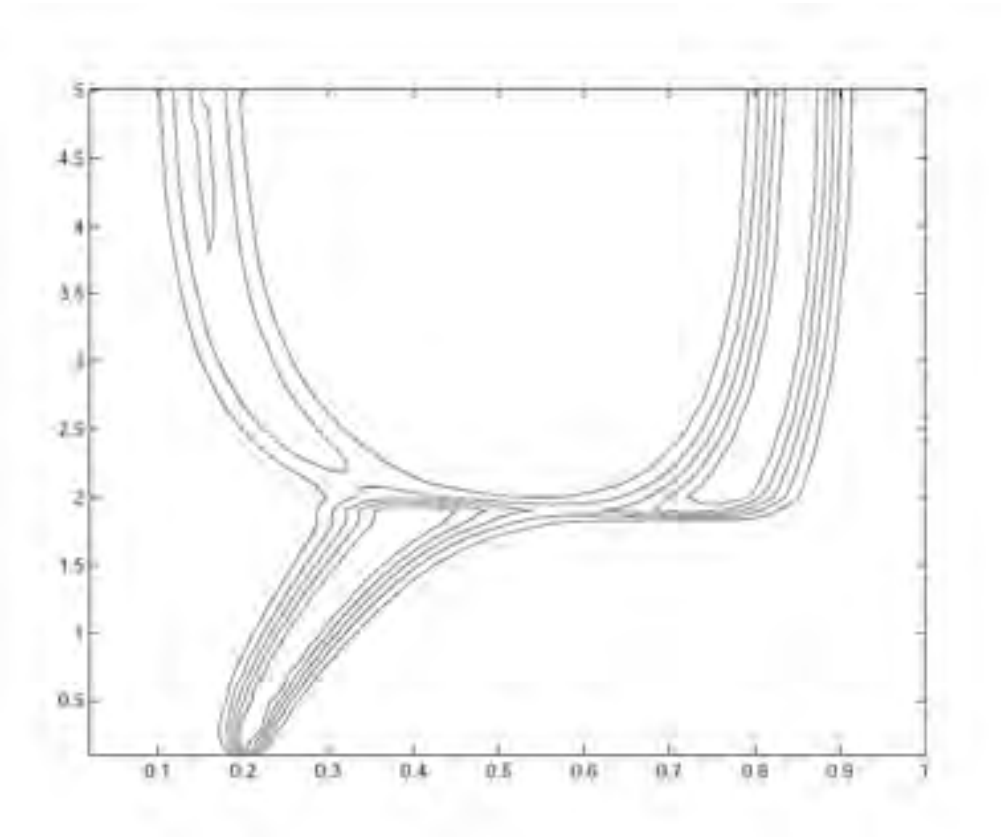
$$\partial_t u(t,x) = u(t,x)\left((1-\mu)b(x) - d(x) - \int_{\mathcal{X}} c(x,y)u(t,y)\,dy\right) + \int_{\mathcal{X}} b(y)\mu u(t,y) m\left(\frac{x-y}{\sigma}\right)\frac{dy}{\sigma}.$$

Assuming $\mu = 1$ and $\sigma = \varepsilon$, this PDE can be written as

$$\partial_t u(t,x) = \frac{1}{\varepsilon}u(t,x)\left(r(x) - \int_{\mathbb{R}^\ell} c(x,y)u(t,y)\,dy\right) + \frac{1}{\varepsilon}\int_{\mathbb{R}^\ell} m(h)(u(t,x+\varepsilon h) - u(t,x))\,dh,$$

scaling time as $t/\varepsilon$ (large time, small mutations)

# Simulation




Competition for two resources

(Diekmann, Jabin, Mischler, Perthame, 2005)

# Limit Hamilton-Jacobi equation (1)

Diekmann et al., 2005: defining (WKB ansatz)

$$u_\varepsilon(t,x) = \exp\left(\frac{\varphi_\varepsilon(t,x)}{\varepsilon}\right),$$

the PDE becomes

$$\partial_t \varphi_\varepsilon(t,x) = r(x) - \int_{\mathbb{R}^\ell} c(x,y) u_\varepsilon(t,y)\, dy + \int_{\mathbb{R}^\ell} m(h)\left[\exp\left(\frac{\varphi_\varepsilon(t,x+\varepsilon h)-\varphi_\varepsilon(t,x)}{\varepsilon}\right) - 1\right] dh.$$

This suggests the convergence of $\varphi_\varepsilon$ to the solution of

$$\partial_t \varphi(t,x) = r(x) - \int_{\mathbb{R}^\ell} c(x,y)\mu_t(dy) + \beta H(\nabla_x \varphi(t,x)),$$

where

$$H(p) = \int_{\mathbb{R}^\ell} \overline{m}(h)(e^{p\cdot h} - 1)\, dh$$

and $\mu_t$ is (in some sense) the limit of $u_\varepsilon(t,\cdot)$.

# Limit Hamilton-Jacobi equation (2)

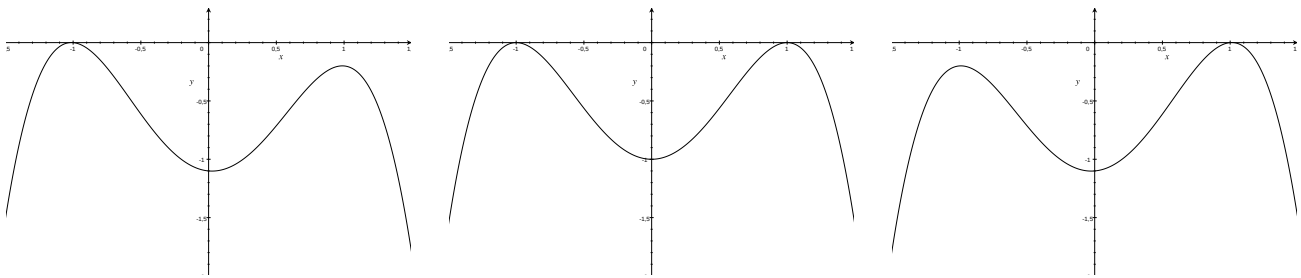
- The total population mass remains bounded
  $\rightsquigarrow$ $\max_x \varphi(t,x) = 0$ for all $t \geq 0$.
- The limit population density at time $t$ is 0 except at the points $x$ where $\varphi(t,x) = 0$ $\rightsquigarrow$ $\mu_t$ has support in $\{\varphi(t,\cdot) = 0\}$.
- The measure $\mu_t$ has to be metastable, i.e.
  - $r(x) - \int c(x,y)\mu_t(dy) \leq 0$ for all $x$ such that $\varphi(t,x) = 0$,
  - $r(x) - \int c(x,y)\mu_t(dy) = 0$ for all $x$ in the support of $\mu_t$.
- Under the assumption that the kernel $c(x,y)$ is positive, these two conditions are satisfied for a unique measure $\mu_t$, and

$$\mu_t = \mu(\{\varphi(t,\cdot) = 0\}),$$

for some well-defined function $\mu$
$\rightsquigarrow$ closed Hamilton-Jacobi equation (C., Jabin, 2011).

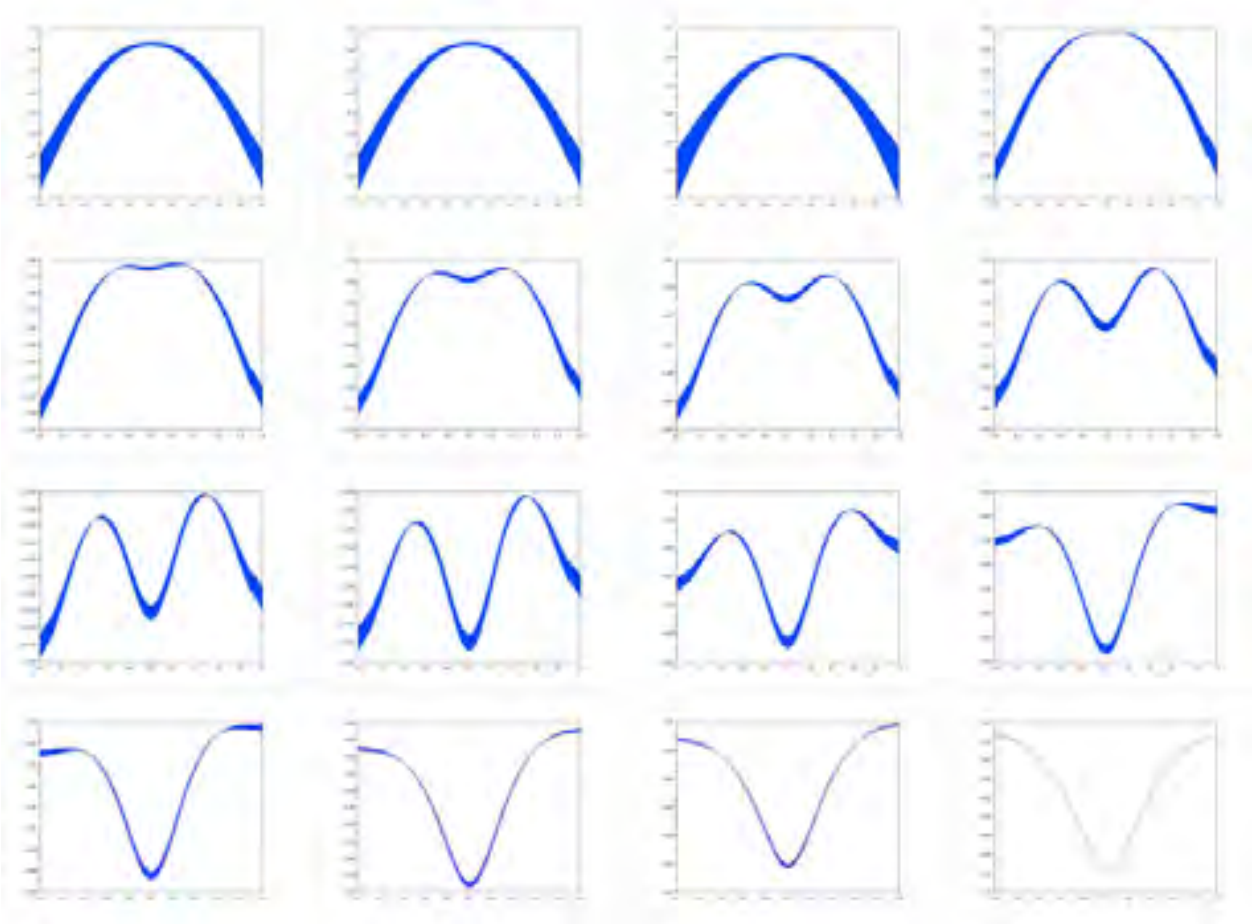
# Some elements of the proof of the theorem

- The existence and uniqueness of a measure $\mu_t$ as above is proved using Lyapunov functions for the PDE without mutations.
- One cannot have existence of a solution to HJ in the strong sense. The support of $\mu_t$ need not be continuous.



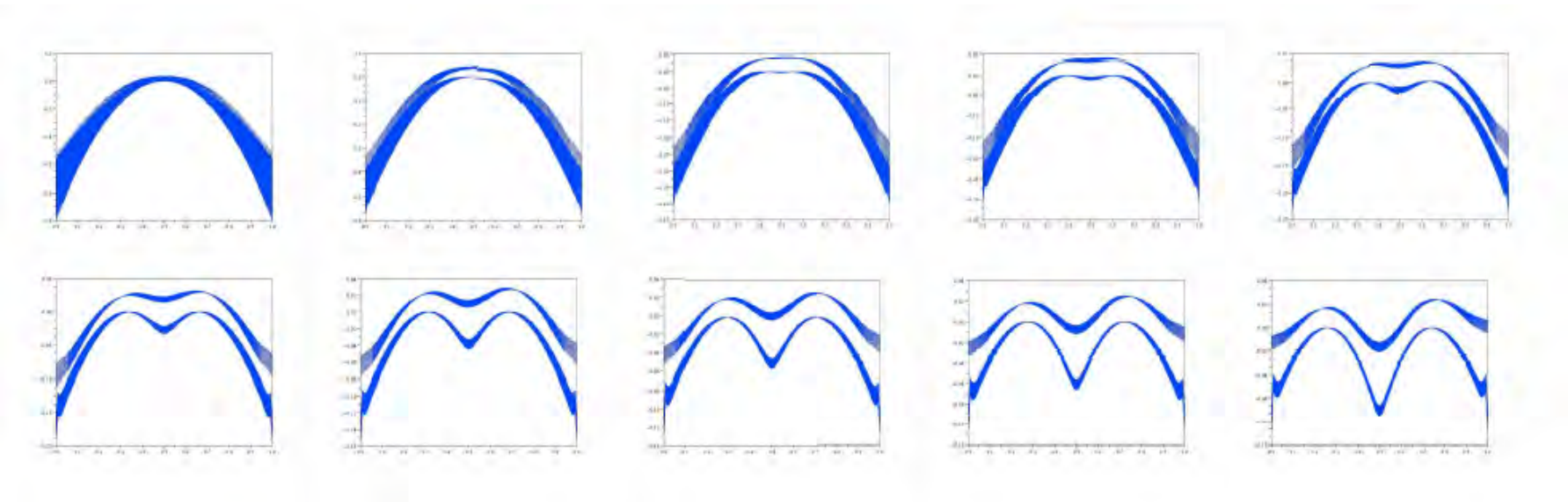
- We use classical a priori estimates for HJ equations, and prove that they hold for $\varphi_\varepsilon$ uniformly in $\varepsilon$. In particular, we prove that $\partial_x \varphi^\varepsilon$ is bounded in $L^\infty([0,T], BV_{\text{loc}}(\mathbb{R}))$ and $\partial_{xx}\varphi_\varepsilon \geq -C$.
- This implies easily the strong convergence of a subsequence of $\varphi_\varepsilon$.
- The difficult part is the convergence of $\int c(x,y)u_\varepsilon(t,y)dy$ to the correct limit. This is done by proving uniform Lebesgue-right-continuity estimates.

# Simulation of the PDE for population densities



(Implicit finite differences)

# Simulation of the HJ equation with constraints



# The canonical equation of adaptive dynamics: second form

- As long as there is only a single point $\bar{x}(t)$ in $\{\varphi(t,\cdot)=0\}$, $\mu_t = \bar{n}(\bar{x}(t))\delta_{\bar{x}(t)}$.
- Since $\partial_x \varphi(t,\bar{x}(t)) = 0$, we have

$$\partial_{tx}\varphi(t,\bar{x}(t)) + \partial_{xx}\varphi(t,\bar{x}(t))\frac{d\bar{x}(t)}{dt} = 0.$$

- Since $\partial_t \varphi(t,x) = f(x,\bar{x}(t)) + H(\partial_x \varphi(t,x))$,

$$\partial_{tx}\varphi(t,\bar{x}(t)) = \partial_x f(x,\bar{x}(t)) + H'(\partial_x\varphi(t,\bar{x}(t)))\partial_{xx}\varphi(t,\bar{x}(t)) = \partial_x f(x,\bar{x}(t))$$

Therefore

$$\frac{d\bar{x}(t)}{dt} = -(\partial_{xx}\varphi(t,\bar{x}(t)))^{-1}\partial_x f(x,\bar{x}(t)).$$

$\rightsquigarrow$ same fitness gradient as for the first form, but different speed.

# Mathematical and biological comments

- Open problem: well-posedness of the Hamilton-Jacobi equation with contraint. Only known in special cases (a single resource or no mutation).
- Smoothness of the solution is only known in special cases
  $\rightsquigarrow$ justification of the canonical equation and of the branching criterion (same criterion as for the first approach) only in special cases.

Biological criticism:

- the dynamics of the population is strongly influenced by very low population densities (so low that there should actually be nobody there)
- evolutionary branching is too fast (also due to very low population densities
- $\rightsquigarrow$ PDE analysts try to modify the model with thresholds or stronger interactions to actually obtain zero densities (see recent works of Jabin, Perthame, Mirrahimi, Lorz).

# Third approach: small mutations and not too large population

This suggests another stochastic approach (already proposed by Meszena, Gyllenberg, Jacobs, Metz, PRL 2005), intermediate between the first two:

- consider the individual-based models with frequent mutations ($\mu = 1$).
- assume that $K$ is large and $\sigma$ is small enough for the diameter of the support of the population to be small (concentration limit, as the first too approaches).
- look at the long term evolution of the population.

# Canonical equation of adaptive dynamics: third form

Assume for simplicity that $d(x) \equiv 0$ (no extinction).

## Theorem

*Assume that $b(\cdot)$ and $c(\cdot,\cdot)$ are $C^1$, positive functions and $m(\cdot)$ has compact support. Assume that $\nu_0^K$ has support $\{x_0\}$ for all $K$. Assume that $\mu = 1$, $\sigma \to 0$ and $K \to +\infty$ so that*

$$\sigma = o(K^{-5/2-a})$$

*for some $a > 0$. Then*

$$(\nu^K_{t/(\sigma^2 K)}, t \geq 0) \Rightarrow (\bar{n}(x(t))\delta_{x(t)}, t \geq 0)$$

*in law for the Skorohod topology on the space of finite positive measures on $\mathcal{X}$ equipped with the weak topology, where $x(0) = x_0$ and*

$$\dot{x}(t) = 2\,Var(m)\bar{n}(x(t))\partial_1 f(x(t), x(t)).$$

# A few comments

- Same form with fitness gradient.
- Very similar to the first form:
  - in the case of symmetric mutations, only a multiplicative contant;
  - in the case of asymmteric mutations, no longer the same.
- If $d \not\equiv 0$, then extinction and possible, and $\sigma$ must not be too small.
- Intermediate time-scale between the first two approaches, but not so fast.

# Idea of the proof

For simplicity, we will assume that $c(x,y) \equiv c > 0$ and that mutation steps are $\pm\sigma$ with probability $1/2$.

Step 1: expected speed of evolution Let $\nu$ be a positive measure on $\mathcal{X}$.

- to simplify the computation, we assume that $b(x) = b_0 + b_1 x$ (OK when the support of $\nu$ is small)
- let $L^K$ be the infinitesimal generator of $\nu^K$, and $\phi(\nu) = \frac{\langle \nu, \mathrm{id} \rangle}{\langle \nu, 1 \rangle}$ the "mean trait value function"; then

$$
\begin{aligned}
L^K \phi(\nu) &= \int \left( \frac{\langle \nu, \mathrm{id} \rangle + \frac{x}{K}}{\langle \nu, 1 \rangle + \frac{1}{K}} - \frac{\langle \nu, \mathrm{id} \rangle}{\langle \nu, 1 \rangle} \right) (b_0 + b_1 x) K \nu(dx) \\
&\quad + \int \left( \frac{\langle \nu, \mathrm{id} \rangle - \frac{x}{K}}{\langle \nu, 1 \rangle - \frac{1}{K}} - \frac{\langle \nu, \mathrm{id} \rangle}{\langle \nu, 1 \rangle} \right) c \left( \langle \nu, 1 \rangle - \frac{1}{K} \right) K \nu(dx) \\
&= b_1 \left( \frac{\langle \nu, \mathrm{id}^2 \rangle}{\langle \nu, 1 \rangle} - \frac{\langle \nu, \mathrm{id} \rangle}{\langle \nu, 1 \rangle (\langle \nu, 1 \rangle + \frac{1}{K})} \right) \\
&= b_1 \mathrm{Var}(\nu) + O(1/K).
\end{aligned}
$$

# Expected speed of evolution

- Hence, when the support of $\nu$ is small, the expected speed of evolution is proportional to the empirical variance of the trait distribution in $\nu$ and the gradient of $b_1$ (= fitness gradient when $c(\cdot,\cdot) \equiv c$)
- To prove that this is true without expectation, one needs to use a version of the ergodic theorem
  $\rightsquigarrow$ needs (approximate, local) stationarity of increments of $\phi(\nu_t^K)$.
- What is (approximately) stationary? the centered trait distribution, i.e.

$$\hat{\nu}_t^K := \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{x_i - \phi(\nu_t^K)}.$$

- This is close to be Markov (locally).
- $\rightsquigarrow$ need to prove that the support of $\nu_t^K$ has small diameter, and that $\hat{\nu}_t^K$ is (nearly) stationary.

# Step 2: the neutral case

Assume that $b(x) \equiv b$.

- Then $N_t^K$ is a birth and death Markov process with birth rate $bn$ and death rate $\frac{c}{K} n(n-1)$ when $N_t^K = n$.
- $(\hat{\nu}_t^K, t \geq 0)$ is also Markov.
- The invariant distribution of $N_t^K$ is a Poisson distribution conditioned to be positive: $\mu_n^K = \frac{\theta^n}{(e^\theta - 1) n!}$, where $\theta = Kb/c$.
- Because of the deterministic limit, large deviations tell us that the exit time of $[\bar{n} - \varepsilon, \bar{n} + \varepsilon]$ is larger that $\exp(C_\varepsilon K)$ with high probability.
- Standard coupling arguments: there exists $\gamma > 0$ s.t., for all $K \geq 1$ and for all $N_0^K$,

$$\| \mathcal{L}(N_t^K) - \mu^K \|_{TV} \leq 2 \exp(-\gamma t / \sqrt{K}).$$

$\rightsquigarrow$ stationary after a time of order $\sqrt{K}$.

# Step 2: the neutral case

A stationary birth and death process is reversible, so after a time $K^{1/2+a}$, the path of $(N_t^K, 0 \leq t \leq T)$ can be constructed backward in time as $(\tilde{N}^K_{T-t}, 0 \leq t \leq T)$ where $\tilde{N}^K$ is a Markov birth and death process with the same transitions and started from distribution $\mu^K$.

It is also possible to construct the genealogical relations between individuals backward in time:

- in case of a birth backward in time, a new edge is added to the genealogical tree
- in case of a death backwards in time, pick at random a pair of branches and make them coalesce.

# Step 2 continued

Hence, the ancestral tree of the people present at some (large) time $T$ can be constructed backwards in time in a Markovian way. In particular, if $\tilde{R}_t^K$ is the number of ancestors at time $T-t$ of the individuals living at time $T$, the process $(\tilde{N}_t^K, \tilde{R}_t^K)_{0\leq t\leq T}$ is Markov with transitions

- from $(n,k)$ to $(n+1,k)$ with rate $bn$
- from $(n,k)$ to $(n-1,k)$ with rate $c\frac{n(n-1)}{K}(1-\frac{k(k-1)}{n(n-1)})$
- from $(n,k)$ to $(n-1,k-1)$ with rate $c\frac{n(n-1)}{K}\frac{k(k-1)}{n(n-1)} = \frac{2c}{K}\frac{k(k-1)}{2}$.

Note that the ancestral process of a stationary logistic birth and death process if exactly a Kingman coalescent ! (for other processes, this is approximately true since the population density belongs to $[\bar{n}-\varepsilon, \bar{n}+\varepsilon]$ with high probability)

# Step 2 continued

- Mutations can then be added also in a Markovian way, since they occur along ancestral lineages either at coalescence times or at death times of the backward birth and death process for which the coalescing pair contains one of the current ancestral lineages.
- When $\tilde{R}_t^K = k$, this occurs on each ancestral line with rate

$$\frac{2c}{K}(\tilde{N}_t^K - k) \approx 2c\bar{n}$$

- Each mutation produces a trait jump of $\pm\sigma$ with probability $1/2$.

# Step 2 continued

Therefore, after rescaling time by $2c/K$, the ancestral process is the Kingman coalescent, whose branches are subject to symmetric random walks with rate $K\bar{n}$. This suggests that

## Proposition

*In the neutral case, and for $t_K \ll K$, $\hat{\nu}^K_{t_K}/\langle \nu^K_{t_k}, 1\rangle$ scaled in space by $\frac{\sqrt{2}}{\sigma\sqrt{K\bar{n}}}$ converges in distribution for the weak topology to the centered distribution of types in the Kingman coalescent with standard Brownian motion along the branches.*
*In addition, the probability that the diameter of $\hat{\nu}^K_{t_K}$ is larger than $\sigma K^{1/2+a}$ converges to 0 for all $a > 0$.*

# Step 2 continued

What we actually need can be checked by hands:

- We want to compute the empirical variance of the traits.
- In the Kingman coalescent with Brownian motion, the computation is easy (Blum, Damerval, Manel, Francois, TPB 2004):
  - in the $n$-coalescent, the empirical variance is given by

$$V_n := \frac{1}{n}\sum_i X_i^2 - \frac{1}{n^2}\left(\sum_j X_j\right)^2 = \frac{1}{n^2}\sum_{i<j}(X_i - X_j)^2,$$

    and $\mathbb{E}[(X_i - X_j)^2] = 2$, so that $V_n \to 1$.
- The computation in the case of the ancestral process of $\nu_t^K$ can be done in a very similar way.

Note: this 1st step works for more general models, allowing extinction and for which the limit equation has a single stable equilibrium. One then needs to use quasi-stationary distribution, time-reversed process for quasi-stationary process, and estimates of convergence in the total variation norm to the quasi-stationary distribution (cf. C. ad Villemonais, 2014).

# Step 3: the nearly neutral case

We now come back to the non-neutral case. Our goal is to prove that one can approximate the trait distribution by that of a neutral process as long as its support is not too big.

Coupling argument:

- as long as the support of the population is in some small interval with maximum value $\bar{b}$ and minimal value $\underline{b}$ for $b(x)$,
- the process can be coupled with a neutral process with birth rate $\underline{b}$ until the first time a difference appears: at most with rate $(\bar{b} - \underline{b})K(\bar{n} + \varepsilon)$;
- after this time, I can obtain a process with more individuals by assuming this new (marked) individual grows a population independently of the first neutral process (no interaction $\rightsquigarrow$ fewer death events),
- this independent population will never reach $K(\bar{n} + \varepsilon)$ individuals,
- do the same for the next (marked) arrivals of a Poisson process with rate $(\bar{b} - \underline{b})K(\bar{n} + \varepsilon)$, until a time of order $K$

# Step 3: the nearly neutral case

Now, $\bar{b} - \underline{b}$ can be assumed to be of order $\sigma K^{1/2+a}$, and so the assumption $\sigma = o(K^{-5/2})$ means that $(\bar{b} - \underline{b})K^2 \to 0$.

Hence, the probability to have one maked individual in the population converges to 0. It is then easy to check that the empirical variance of the nearly neutral population is very close to the one of the neutral population.

# Conclusion and comments

- We obtained a third version of the canonical equation of adaptive dynamics on intermediate time scales assuming mutations are small but not rare and that the population is large but no too much.
- The assumption relating $\sigma$ and $K$ might not be optimal, since we only need the convergence of empirical variances. However, it seems hard to obtain a coupling between the neutral and nearly neutral processes and to contral the difference.
- Evolutionary branching should be studied in this case. However, it seems that the current time scaling leads to a too recent MRCA to hope that evolutionary branching could occur on a short time-scale (it is rather a large deviations event) $\rightsquigarrow$ it is important to find a way to relax the assumption on $\sigma$.