Stochastic models of evolution in a population living in a continuum

Amandine Véber

Easter Probability Meeting – March 2023

LABORATOIRE
MAP5

Genetic diversity and spatial structure

<u>Aim:</u> Model and understand the evolution of the genetic diversity of a population having a continuous spatial structure.

⤳ Dimension 2 is the most relevant for applications to biological pop., but the mathematical models are interesting in any dimension.

Footprint of a spatial structure

- ▶ Interactions/reproduction require that individuals should be sufficiently close to each other.

- ▶ Offspring are born in a more or less extended neighbourhood of their parents.

- ▶ The selective advantage offered by certain alleles depends on the environment, which can vary from one region to another.

⤳ Local allelic distributions are correlated, in a specific way that depends on parameters such as the speed of spatial diffusion of genes across the population.

Questions of interest

► (Compound) parameters characterising the genetic diversity in a population and its evolution?

► Correlation pattern between the local allelic distributions at several locations, under different evolutionary scenarios? Inference methods for the estimation of the corresponding key parameters?

► Detection of evolutionary forces in action based on appropriate types of data?

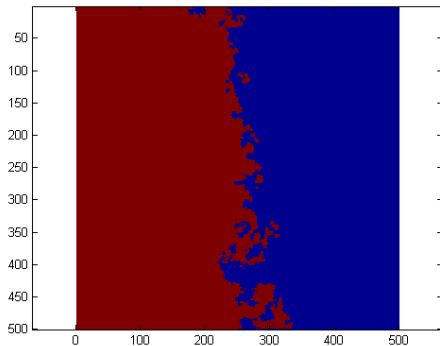# Simulations of the SLFV in two dimensions

Initial configuration:
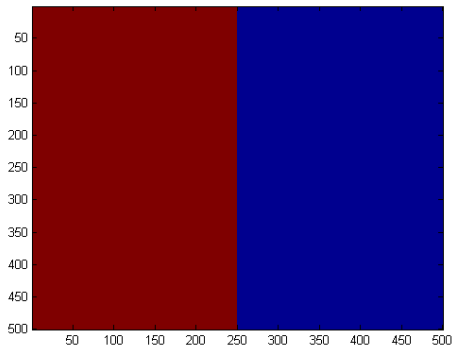
Simulations by H. Saadi. Fixed radius, $u \equiv 1$.
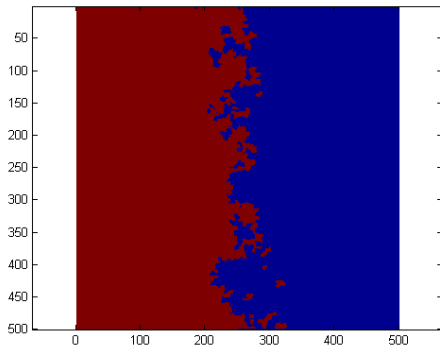
Simulations of the SLFV in two dimensions

After 2.10^6 events:

Simulations by H. Saadi. Fixed radius, $u \equiv 1$.

Simulations of the SLFV in two dimensions

After 3.10^6 events:

Simulations by H. Saadi. Fixed radius, $u \equiv 1$.

Simulations of the SLFV in two dimensions

After 4.10^6 events:

Simulations by H. Saadi. Fixed radius, $u \equiv 1$.
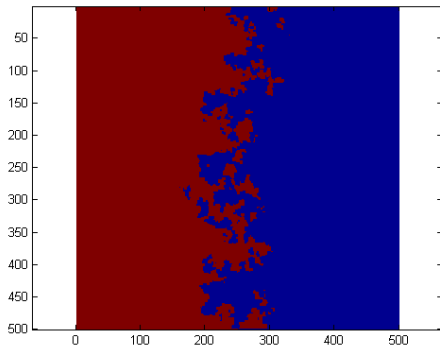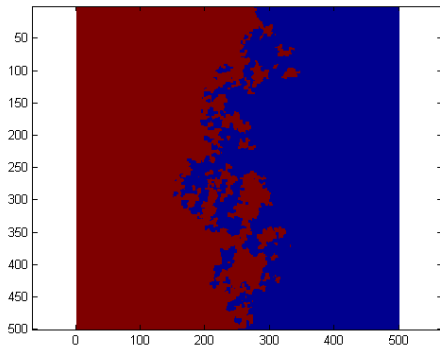
Simulations of the SLFV in two dimensions

After 5.10^6 events:

Simulations by H. Saadi. Fixed radius, $u \equiv 1$.

Long term evolution at an interface
(Berestycki, Etheridge & V., 2013)

Geographical space : $\mathbb{R}^d$ **Allele space:** $\{0, 1\}$

▶ **Case 1: Constant radius**

We fix $R > 0$ and $u \in (0, 1]$. All events have radius R and impact u.

⤳ Reproductions are purely local.

Long term evolution at an interface
(Berestycki, Etheridge & V., 2013)

Geographical space : $\mathbb{R}^d$ **Allele space:** $\{0, 1\}$

- ▶ **Case 1: Constant radius**

  We fix $R > 0$ and $u \in (0, 1]$. All events have radius R and impact u.

  ⤳ Reproductions are purely local.

- ▶ **Case 2: Heavy-tailed radii**

  We fix $\alpha \in (1, 2)$ and $u \in (0, 1]$. Intensity measure on radii given by

  $$\mu(dr) = \frac{\mathbf{1}_{\{r > 1\}}}{r^{d+\alpha+1}} \, dr.$$

  ⤳ Allows the occurrence of rare but very large events.
  ⤳ Ancestral lineages behave like α-stable processes.

Time- and space-scales

- ► Case 1: Constant radius and impact
- ► Case 2: Constant impact, intensity $r^{-(d+\alpha+1)} \, dr$ for radii.

We set $\alpha = 2$ in case 1 and for every $n \geq 1$,

$$w_t^n(x) := w_{nt}(n^{1/\alpha} x).$$

Time- and space-scales

- ▶ Case 1: Constant radius and impact
- ▶ Case 2: Constant impact, intensity $r^{-(d+\alpha+1)}\, dr$ for radii.
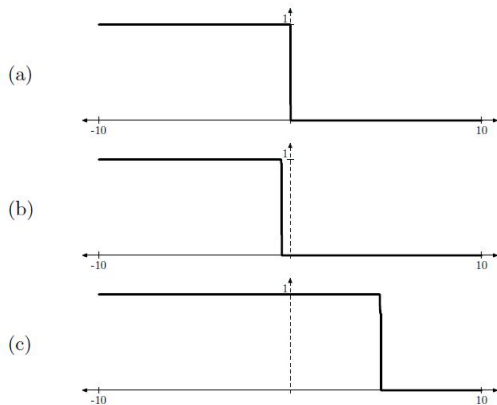
We set $\alpha = 2$ in case 1 and for every $n \geq 1$,

$$w_t^n(x) := w_{nt}(n^{1/\alpha}x).$$

Initial condition: $w_0(x) = \mathbf{1}_H(x)$, where $H = \{x_{(1)} \leq 0\}$.

Questions: How does w_t^n behave when n is large? Width of the interface? Resulting pattern of genetic diversity?

Purely local reproductions, $d = 1$

$u = 0.8$, $r = 0.033$ and $n = 10^3$. Initial condition, after 10^5 events, after 10^7 events.
(Simulations by J. Kelleher, Oxford Univ.)

That is...

Theorem 1 (Berestycki, Etheridge & V., 2013)

▶ There exists a process $(M_t^{(2)},\ t \geq 0)$, with values in $\mathcal{M}_\lambda$, such that

$$M^{n} \overset{(f.d.d.)}{\longrightarrow} M^{(2)} \qquad \text{as } n \to \infty.$$

▶ Moreover, there exists $\tilde{\sigma}^2 > 0$ such that, if X denotes standard Brownian motion and
$$p_t^{(2)}(x) := \mathbf{P}_x\big[X_{u\tilde{\sigma}^2 t} \in H\big], \text{ then}$$

That is...

Theorem 1 (Berestycki, Etheridge & V., 2013)

▶ There exists a process $(M_t^{(2)}, \, t \geq 0)$, with values in $\mathcal{M}_\lambda$, such that

$$M^n \overset{(f.d.d.)}{\longrightarrow} M^{(2)} \qquad \text{as } n \to \infty.$$

▶ Moreover, there exists $\tilde{\sigma}^2 > 0$ such that, if X denotes standard Brownian motion and

$$p_t^{(2)}(x) := \mathbf{P}_x\big[X_{u\tilde{\sigma}^2 t} \in H\big], \text{ then}$$

⤳ If $d = 1$: for all $t > 0$, $w_t^{(2)}$ is a random field of correlated Bernoulli r.v. satisfying

$$\mathbb{E}\big[w_t^{(2)}(x)\big] = p_t^{(2)}(x).$$

That is...

Theorem 1 (Berestycki, Etheridge & V., 2013)

▶ There exists a process $(M_t^{(2)}, t \geq 0)$, with values in $\mathcal{M}_\lambda$, such that

$$M^n \overset{(f.d.d.)}{\longrightarrow} M^{(2)} \qquad \text{as } n \to \infty.$$

▶ Moreover, there exists $\tilde{\sigma}^2 > 0$ such that, if X denotes standard Brownian motion and
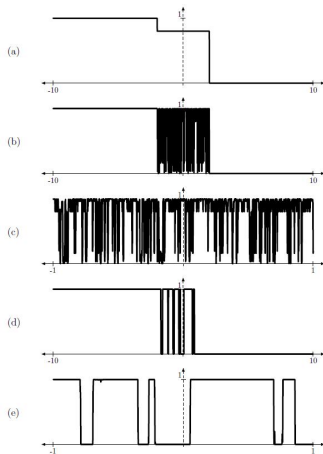
$$p_t^{(2)}(x) := \mathbf{P}_x[X_{u\tilde{\sigma}^2 t} \in H], \text{ then}$$

⤳ If $d = 1$: for all $t > 0$, $w_t^{(2)}$ is a random field of correlated Bernoulli r.v. satisfying
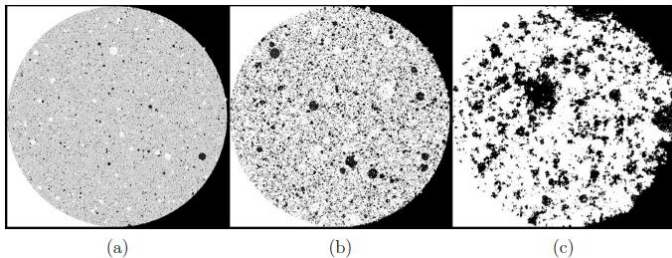
$$\mathbb{E}[w_t^{(2)}(x)] = p_t^{(2)}(x).$$

⤳ If $d \geq 2$: for all $t \geq 0$, $w_t^{(2)}(x) = p_t^{(2)}(x)$ Lebesgue-a.e.

$u = 0.8$, $\alpha = 1.3$ and $n = 10^4$ (Simulations by J. Kelleher).
(a) Initial condition, (b-c) after 100 events, (d-e) after 10^6 events.

In the presence of catastrophes, $d = 2$

(a)   (b)   (c)

$u = 0.8$, $\alpha = 1.3$ and $n = 10^3$. After 10^5, 10^6 and 10^7 events.
(Simulations by J. Kelleher)

Asymptotic behaviour in the presence of large events

Theorem 2 (Berestycki, Etheridge & V., 2013)

▶ There exists a process $(M_t^{(\alpha)}, t \geq 0)$, with values in $\mathcal{M}_\lambda$, such that

$$M^n \overset{(f.d.d.)}{\longrightarrow} M^{(\alpha)} \qquad \text{as } n \to \infty.$$

▶ Moreover, there exists a symmetric α-stable process $X^{(\alpha)}$ such that, if

$$p_t^{(\alpha)}(x) := \mathbf{P}_x\big[X_{ut}^{(\alpha)} \in H\big],$$

then *in any dimension*, for all $t > 0$, $w_t^{(\alpha)}$ is a random field of correlated Bernoulli r.v. satisfying

$$\mathbb{E}\big[w_t^{(\alpha)}(x)\big] = p_t^{(\alpha)}(x).$$

Conclusions

▶ **No local coexistence of alleles** except if $d \geq 2$ and reproductions are purely local.

▶ The correlations between the local allele frequencies are characterised by the genealogical process.
  Correlation length:
  ▶ $\sqrt{n}$ in the case of constant (or bounded) event radius,
  ▶ $n^{1/\alpha}$ in the case of heavy tailed radius distribution.

⤳ Rare but massive extinction/recolonisation events may have a significant impact on the genetic diversity seen in a population.

An example of spatial clustering - *C. elegans* in the wild (M.-A. Félix & H. Teotonio - ENS)
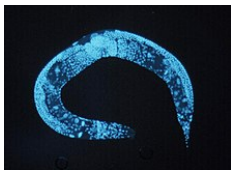

- ▶ A complex life cycle, with an optional *dauer* phase during which it only moves - does not eat/reproduce.

- ▶ Local population dynamics in "boom and bust":
  - ▶ Appearance of a food source ⤳ exponential growth
  - ▶ Exhaustion of the resource ⤳ transition to *dauer* stage

- ▶ Migration possible via individual motion (slow) and/or by hitchhiking in groups (∼10 individuals, potentially moving several meters hung on a snail or an isopod).
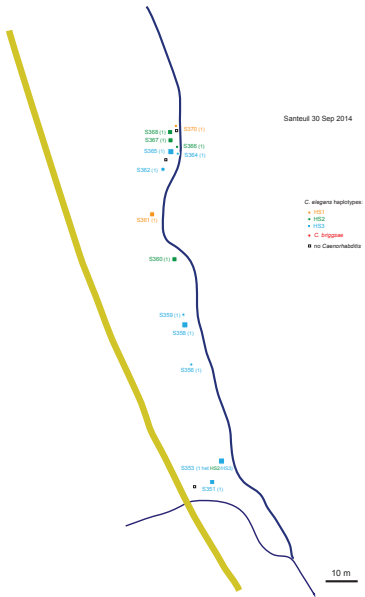
(Richaud et al., 2018)

Sampling location: Santeuil (Paris region)

**Type of data:** Haplotypes of *C. elegans* (orange, green, blue) + *C. briggsae* (red)
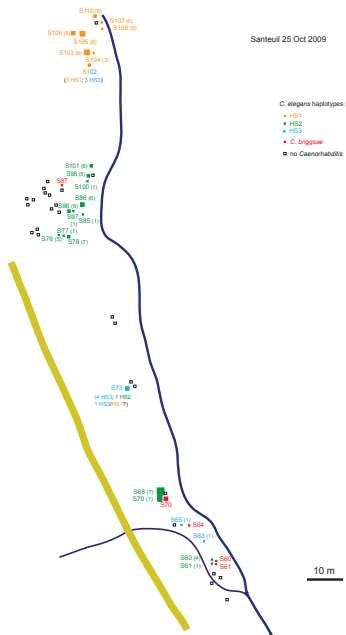
Why you may doubt

- For a 1mm long nematode, a 10m wide strip is not particularly one-dimensional...

- All we can conclude is that a neutral model is compatible with this type of spatial segregation, but maybe (or almost surely...) individuals are in competition for food or other resources.
  ⤳ "Boom and bust" dynamics reinforcing a selective pattern?

- One or two samplings each year and only the most obvious pattern was shown here.

(Richaud et al., 2018) - Other sampling

Genetic diversity of expanding populations
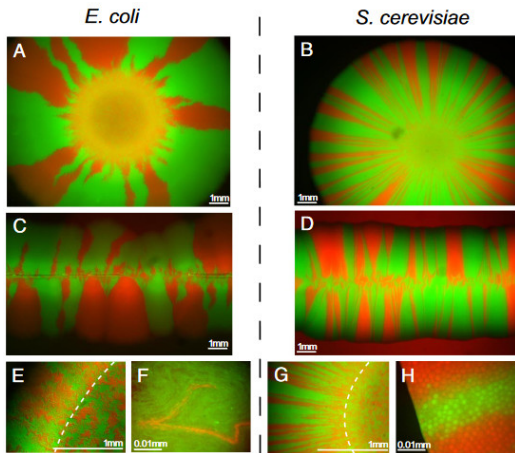
Hallatschek *et al.* (2007), PNAS. Sectors forming after placing a well-mixed droplet (or line) of fluorescent-green/red bacteria (left) or yeast (right).

Genetic diversity in expanding populations

A vast literature:

- PDE models of invasion fronts, notably *Fisher-KPP-type equations*, but also models with *accelerating fronts* ("cane toad model"). Mostly studied in one spatial dimension.

- Stochastic individual-based models (Champagnat & Méléard '07, *e.g.*) focusing on the distribution of a trait that potentially influences individual movement and reproduction chances. Focus put on eco-evo aspects.

Genetic diversity in expanding populations

A vast literature:

- ▶ PDE models of invasion fronts, notably *Fisher-KPP-type equations*, but also models with *accelerating fronts* ("cane toad model"). Mostly studied in one spatial dimension.
- ▶ Stochastic individual-based models (Champagnat & Méléard '07, *e.g.*) focusing on the distribution of a trait that potentially influences individual movement and reproduction chances. Focus put on eco-evo aspects.

Questions:

- ▶ What is the role of stochastic fluctuations at the front when spatial expansion is in 2d? (most relevant dimension for biological applications)
- ▶ Can we detect a recent expansion thanks to the neutral diversity pattern left behind the front?

Genetic diversity in expanding populations

A vast literature:

- ▶ PDE models of invasion fronts, notably *Fisher-KPP-type equations*, but also models with *accelerating fronts* ("cane toad model"). Mostly studied in one spatial dimension.
- ▶ Stochastic individual-based models (Champagnat & Méléard '07, *e.g.*) focusing on the distribution of a trait that potentially influences individual movement and reproduction chances. Focus put on eco-evo aspects.

Questions:

- ▶ What is the role of stochastic fluctuations at the front when spatial expansion is in 2d? (most relevant dimension for biological applications)

Growth speed higher than "expected"
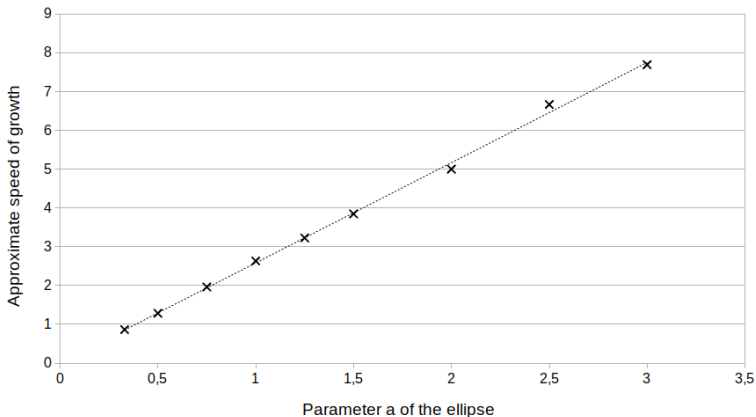
For each value of a, all events have the same parameters $(a, b, 0)$, where b is chosen such that the express chain goes at average speed a. Each cross results from the independent simulation of 30 dual processes. The dotted line is $\widehat{\nu^{-1}}(a) = 2.58a$.

Thank you for your attention!