Evolution de la diversité génétique d'une population ayant une structure spatiale

Amandine Véber

CJC-MA 2023 – 27/09/2023

Genetic diversity and spatial structure

Aim: 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?

The spatial Λ-Fleming-Viot process

(Barton & Etheridge, 2008)

Allele/type space : K compact ($K = \{0, 1\}$, for example).

Population at time t : Measure M_t on $\mathbb{R}^d \times K$ whose first marginal distribution is Lebesgue measure (uniform density of indv.). Possible decomposition:

$$M_t(dx, dk) = dx \, \rho_t(x, dk).$$

The spatial Λ-Fleming-Viot process

(Barton & Etheridge, 2008)

Allele/type space : K compact ($K = \{0, 1\}$, for example).

Population at time t : Measure M_t on $\mathbb{R}^d \times K$ whose first marginal distribution is Lebesgue measure (uniform density of indv.). Possible decomposition:

$$M_t(dx, dk) = dx \, \rho_t(x, dk).$$

Evolution : μ a σ-finite measure on $\mathbb{R}_+^*$, $\{\nu_r, r > 0\}$ a collection of probability measures on $[0, 1]$. Let Π be a Poisson point process on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0, 1]$ with intensity measure $dt \otimes dx \otimes \mu(dr)\nu_r(du)$.

The spatial Λ-Fleming-Viot process

(Barton & Etheridge, 2008)

Allele/type space : K compact ($K = \{0, 1\}$, for example).

Population at time t : Measure M_t on $\mathbb{R}^d \times K$ whose first marginal distribution is Lebesgue measure (uniform density of indv.). Possible decomposition:

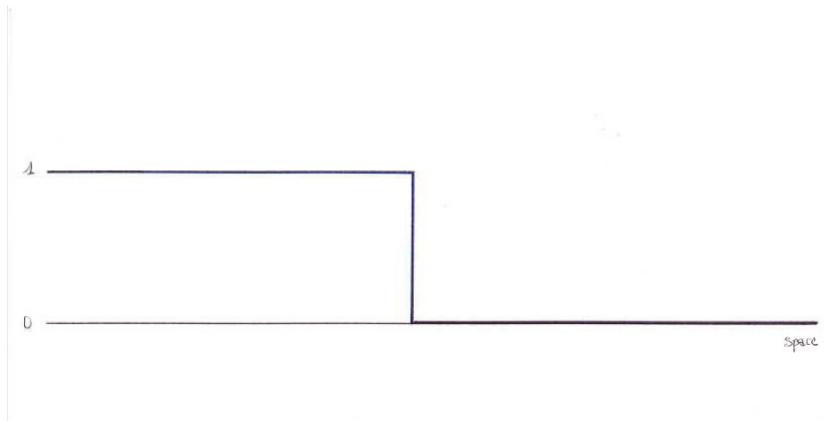$$M_t(dx, dk) = dx\, \rho_t(x, dk).$$

Evolution : μ a σ-finite measure on $\mathbb{R}_+^*$, $\{\nu_r, \, r > 0\}$ a collection of probability measures on $[0, 1]$. Let Π be a Poisson point process on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0, 1]$ with intensity measure $dt \otimes dx \otimes \mu(dr)\nu_r(du)$.

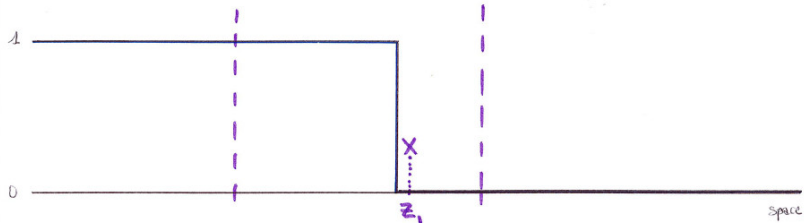If $(t, x, r, u) \in \Pi$, at time t and in $B(x, r)$:

► We choose a parental allele κ according to the allele distribution at time $t-$ within $B(x, r)$;

► For every $y \in B(x, r)$, $\rho_t(y, dk) = (1 - u)\rho_{t-}(y, dk) + u\delta_\kappa(dk)$.
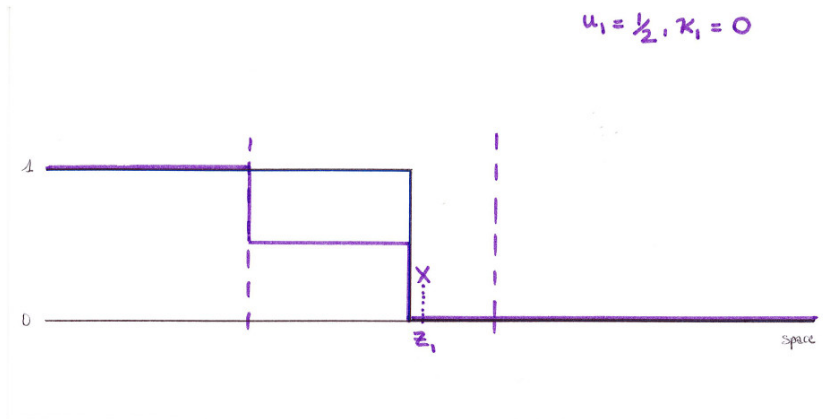
An example
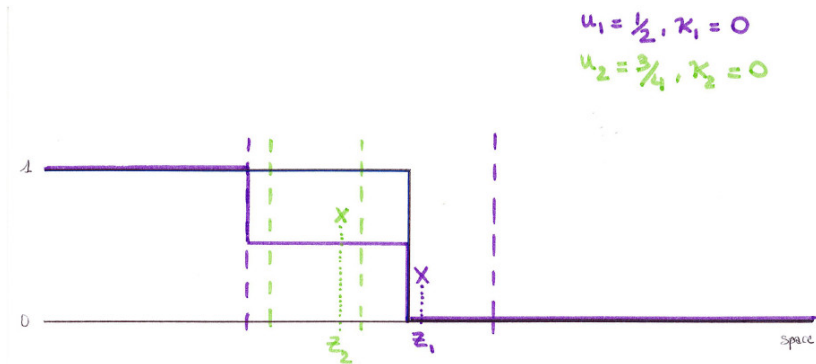
An example

$u_1 = \frac{1}{2}, \; \kappa_1 = 0$

1

0

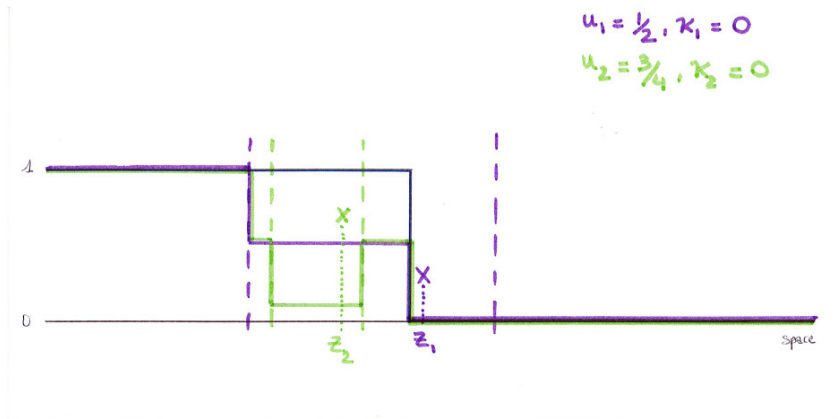$\times$

z_1

Space

An example

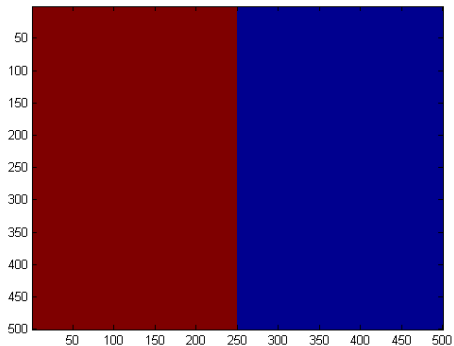$u_1 = \frac{1}{2}, \, \kappa_1 = 0$

An example

An example

$u_1 = \frac{1}{2}, \kappa_1 = 0$

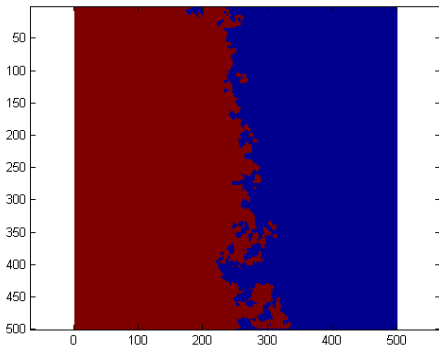$u_2 = \frac{3}{4}, \kappa_2 = 0$

Over larger scales

Initial configuration:

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

Over larger scales

After 2.10^6 events:

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

Over larger scales

After 3.10^6 events:

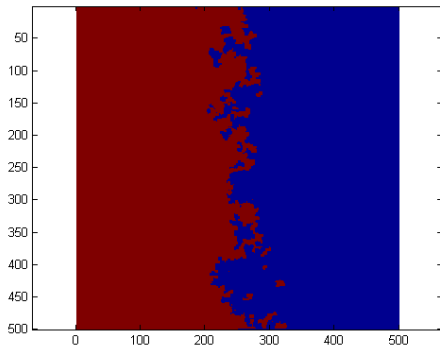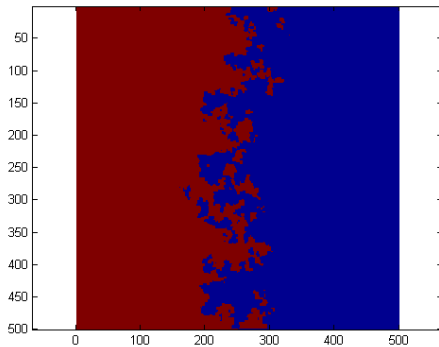Simulations by H. Saadi. Fixed radius, $u \equiv 1$.

Over larger scales

After 4.10^6 events:

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

Over larger scales

After 5.10^6 events:
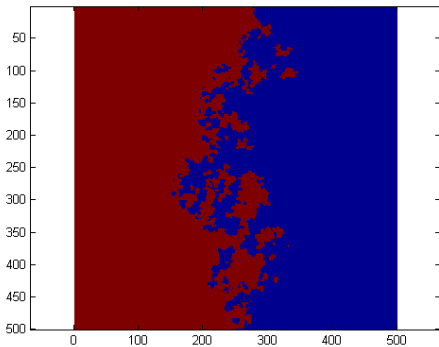
Simulations by H. Saadi. Fixed radius, $u \equiv 1$.

A few remarks

Existence/uniqueness:

▶ Initially obtained in Barton et al. (2010) via an analytical method due to Evans (1997),

▶ then via a particle construction (V. & Wakolbinger, 2015; Etheridge & Kurtz, "2019").

Sufficient condition:

$$\int_0^\infty \int_0^1 u r^d \, \nu_r(du) \mu(dr) < \infty.$$

State space: $\mathcal{M}_\lambda$, endowed with the topology of vague convergence (which makes it a compact space).

A few remarks

Existence/uniqueness:

- ▶ Initially obtained in Barton et al. (2010) via an analytical method due to Evans (1997),
- ▶ then via a particle construction (V. & Wakolbinger, 2015; Etheridge & Kurtz, "2019").

Sufficient condition:

$$\int_0^\infty \int_0^1 ur^d \, \nu_r(du)\mu(dr) < \infty.$$

State space: $\mathcal{M}_\lambda$, endowed with the topology of vague convergence (which makes it a compact space).

Very flexible framework: Gaussian replacements, possibility to add natural selection, recombination, inhomogeneous environment, ...

Back to the high-tech example
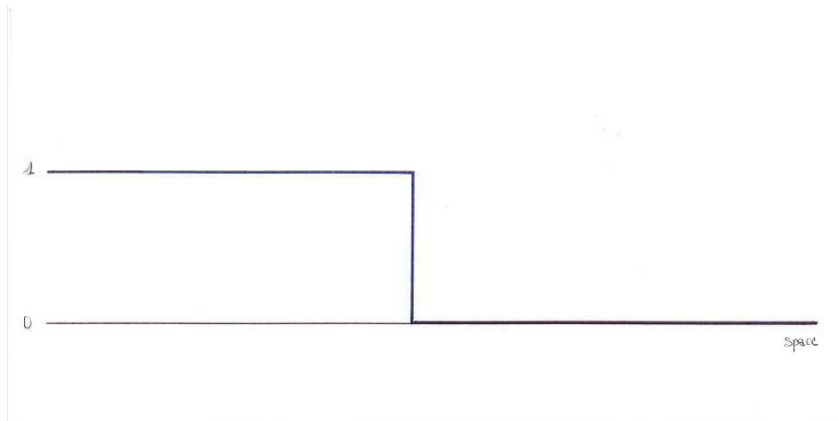
Back to the high-tech example

$u_1 = \frac{1}{2}, \ \kappa_1 = 0$
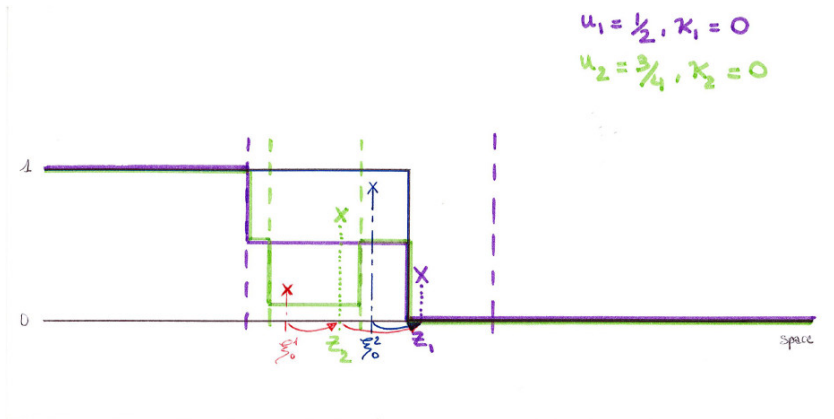
Back to the high-tech example

Back to the high-tech example

Back to the high-tech example

Back to the high-tech example

Back to the high-tech example

Genealogies and duality

▶ We trace back the origin of the alleles observed in a sample of *j* individuals ⇝ "genealogical" process

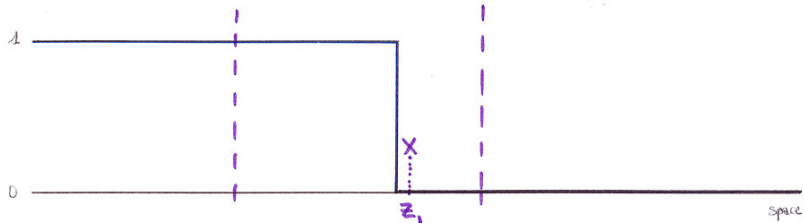▶ $(\{\xi_s^1, \ldots, \xi_s^{N_s}\})_{s \geq 0}$ system of correlated symmetric jump processes, which coalesce/merge when they are *affected* by the same event.

▶ Jump rate of an ancestral lineage (presently at 0):

$$\int_{\mathbb{R}^d} \int_0^\infty \int_0^1 \mathbf{1}_{B(x,r)}(0) u\, \nu_r(du)\mu(dr)dx = \int_0^\infty \int_0^1 u V_r\, \nu_r(du)\mu(dr) < \infty.$$

Genealogies and duality

- We trace back the origin of the alleles observed in a sample of j individuals ⤳ "genealogical" process

- $(\{\xi_s^1, \ldots, \xi_s^{N_s}\})_{s \geq 0}$ system of correlated symmetric jump processes, which coalesce/merge when they are *affected* by the same event.

- Jump rate of an ancestral lineage (presently at 0):

$$\int_{\mathbb{R}^d} \int_0^\infty \int_0^1 \mathbf{1}_{B(x,r)}(0) u\, \nu_r(du) \mu(dr) dx = \int_0^\infty \int_0^1 u V_r\, \nu_r(du) \mu(dr) < \infty.$$

- If $K = \{0, 1\}$ and $w_t(x) := \rho_t(x, \{1\})$: for every $\psi \in C_c((\mathbb{R}^d)^j)$,

$$\mathbb{E}_{w_0} \left[\int_{(\mathbb{R}^d)^j} \psi(x_1, \ldots, x_j) \left\{ \prod_{i=1}^j w_t(x_i) \right\} dx_1 \cdots dx_j \right]$$

$$= \int_{(\mathbb{R}^d)^j} \psi(x_1, \ldots, x_j) \mathbf{E}_{\{x_1, \ldots, x_j\}} \left[\prod_{i=1}^{N_t} w_0(\xi_t^i) \right] dx_1 \cdots dx_j.$$

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, $d = 1$
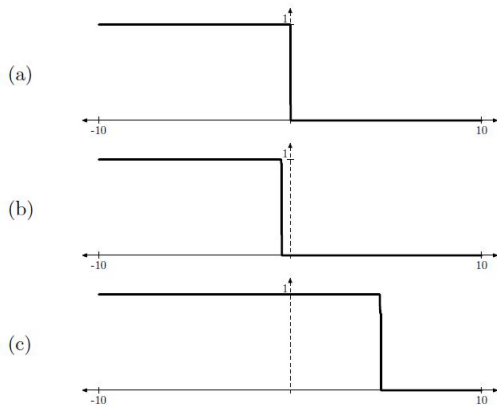
$u = 0.8$, $r = 0.033$ and $n = 10^3$. Initial condition, after 10^5 events, after 10^7 events.
(Simulations by J. Kelleher, Oxford Univ.)

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

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

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

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

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

In the presence of catastrophes, $d = 1$
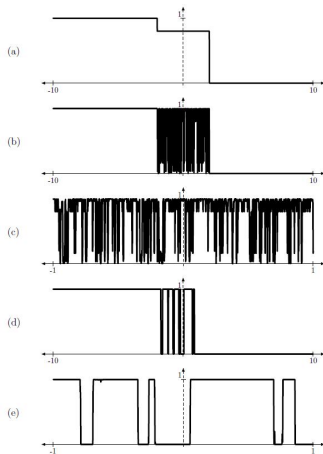
$u = 0.8$, $\alpha = 1.3$ and $n = 10^4$ (Simulations by J. Kelleher).
(a) Initial condition, (b-c) after 100 events, (d-e) after 10^6 events.

In the presence of catastrophes, $d = 2$

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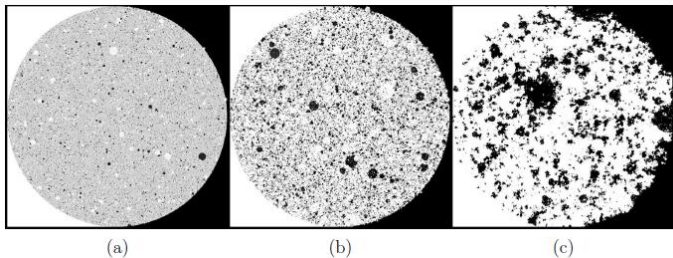
$u = 0.8$, $\alpha = 1.3$ and $n = 10^3$. After 10^5, 10^6 and 10^7 events.
(Simulations by J. Kelleher)

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.

  $\rightsquigarrow$ 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 resources, and subject to selection pressures due to pathogens.

    ⤳ "Boom and bust" dynamics reinforcing a selective pattern?

▶ Still, in "2d" (apple orchard in Orsay), minor haplotypes are maintained and spatial segregation is not observed.

Reference: A. Richaud, G. Zhang, D. Lee, J. Lee, M.-A. Félix (2018). The local coexistence pattern of selfing genotypes in *Caenorhabditis elegans* natural metapopulations. *Genetics*, 208(2), 807–821.

We bias the choice of the parent by giving a slight reproductive advantage to individuals with allele 0 over individuals with allele 1.

The effects of a weak selection pressure

We bias the choice of the parent by giving a slight reproductive advantage to individuals with allele 0 over individuals with allele 1.

▶ **One way to proceed:**

$\leadsto \Pi^N$ PPP on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0,1]$ with intensity $dt \otimes dx \otimes \mu(dr)\nu_r(du)$. Local update of the allele frequencies: we draw κ according to the allelic distr. in the area of the event, then $\forall y \in B(x_i, r_i)$,

$$w_{t_i}(y) = (1 - u_i)w_{t_i-}(y) + u_i \mathbf{1}_{\{\kappa=1\}}.$$

$\leadsto \Pi^S$ PPP on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0,1]$, indep. of Π^N and with intensity $s'dt \otimes dx \otimes \mu(dr)\nu_r(du)$. This time, we choose two alleles κ, κ' indep. and according to the allelic distr. in the area of the event, and then $\forall y \in B(x_i, r_i)$,

$$w_{t_i}(y) = (1 - u_i)w_{t_i-}(y) + u_i \mathbf{1}_{\{\kappa=\kappa'=1\}}.$$

The effects of a weak selection pressure

We bias the choice of the parent by giving a slight reproductive advantage to individuals with allele 0 over individuals with allele 1.

- **One way to proceed:**

  $\rightsquigarrow \Pi^N$ PPP on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0,1]$ with intensity $dt \otimes dx \otimes \mu(dr)\nu_r(du)$. Local update of the allele frequencies: we draw κ according to the allelic distr. in the area of the event, then $\forall y \in B(x_i, r_i)$,

  $$w_{t_i}(y) = (1 - u_i)w_{t_i-}(y) + u_i \mathbf{1}_{\{\kappa=1\}}.$$

  $\rightsquigarrow \Pi^S$ PPP on $\mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_+^* \times [0,1]$, indep. of Π^N and with intensity $s'dt \otimes dx \otimes \mu(dr)\nu_r(du)$. This time, we choose two alleles κ, κ' indep. and according to the allelic distr. in the area of the event, and then $\forall y \in B(x_i, r_i)$,

  $$w_{t_i}(y) = (1 - u_i)w_{t_i-}(y) + u_i \mathbf{1}_{\{\kappa=\kappa'=1\}}.$$

- **Dual process of *potential* ancestors**, where ancestral lineages can also branch during the events of Π^S.

Weak selection and high population density
(Etheridge, V. & Yu, 2020)

$$u_n = \frac{u}{n^\gamma}, \qquad s_n = \frac{s}{n^\delta}, \qquad w_t^n(x) = w_{nt}(n^\beta x),$$

or rather:

$$\overline{w}_t^n(x) := \frac{n^{d\beta}}{V_R} \int_{B(x,n^{-\beta}R)} w_t^n(y)dy \qquad \text{and} \qquad \overline{M}_t^n \text{ with 'density' } \overline{w}_t^n.$$

Weak selection and high population density

(Etheridge, V. & Yu, 2020)

$$u_n = \frac{u}{n^\gamma}, \qquad s_n = \frac{s}{n^\delta}, \qquad w_t^n(x) = w_{nt}(n^\beta x),$$

or rather:

$$\overline{w}_t^n(x) := \frac{n^{d\beta}}{V_R} \int_{B(x, n^{-\beta} R)} w_t^n(y) dy \qquad \text{and} \qquad \overline{M}_t^n \text{ with 'density' } \overline{w}_t^n.$$

Again, for every $n \in \mathbb{N}$

⤳ **Case 1:** all events are of fixed radius $R > 0$ and impact u_n.

⤳ **Case 2:** the intensity measure on radii is

$$\mu(dr) = \frac{\mathbf{1}_{\{r > 1\}}}{r^{d+\alpha+1}} dr, \qquad \alpha \in (1, 2),$$

and the impact u_n is fixed. (And we take $R = 1$ in the def. of $\overline{w}_t^n(x)$.)

Case of purely local reproductions

Theorem 3 (Etheridge, V. & Yu, 2020)

$$u_n = un^{-1/3}, \qquad s_n = sn^{-2/3}, \qquad w_t^n(x) = w_{nt}(n^{1/3}x)$$

If $\overline{M}_0^n$ converges to $M_0 \in \mathcal{M}_\lambda$ as $n \to \infty$, then $(\overline{M}_t^n)_{t \geq 0}$ converges in law in $D_{\mathcal{M}_\lambda}[0, \infty)$ to a Markov process M^∞, with initial value M_0 and characterised by: for all families $\{w_t^\infty, t \geq 0\}$ of representatives of the density of each M_t^∞ and every $f \in C_c^\infty(\mathbb{R}^d)$,

$$\left(\langle w_T^\infty, f \rangle - \langle w_0^\infty, f \rangle - \int_0^T \left\{ \frac{u\Gamma_R}{2} \langle w_t^\infty, \Delta f \rangle - V_R us \langle w_t^\infty(1 - w_t^\infty), f \rangle \right\} dt \right)_{T \geq 0}$$

is a martingale with quadratic variation zero if $d \geq 2$ and

$$4R^2 u^2 \int_0^T \langle w_t^\infty(1 - w_t^\infty), f^2 \rangle dt$$

if $d = 1$, where $\Gamma_R > 0$ depends only on d and R.

In the presence of large events

Theorem 4 (Etheridge, V. & Yu, 2020)

$$u_n = u n^{-\frac{\alpha-1}{2\alpha-1}}, \qquad s_n = s n^{-\frac{\alpha}{2\alpha-1}}, \qquad w_t^n(x) = w_{nt}\big(n^{\frac{1}{2\alpha-1}} x\big)$$

If $\overline{M}_0^n$ converges to $M_0 \in \mathcal{M}_\lambda$ as $n \to \infty$, then $(\overline{M}_t^n)_{t \geq 0}$ converges in law in $D_{\mathcal{M}_\lambda}[0, \infty)$ to a Markov process M^∞, with initial value M_0 and characterised by: for every family $\{w_t^\infty, \, t \geq 0\}$ of representatives of the density of each M_t^∞ and for every $f \in C_c^\infty(\mathbb{R}^d)$,

$$\left(\langle w_T^\infty, f \rangle - \langle w_0^\infty, f \rangle - \int_0^T \left\{ Cu \langle w_t^\infty, (-\Delta)^{\alpha/2} f \rangle - \frac{V_1 u s}{\alpha} \langle w_t^\infty (1 - w_t^\infty), f \rangle \right\} dt \right)_{T \geq 0}$$

is a martingale with quadratic variation zero if $d \geq 2$ and

$$\frac{4u^2}{\alpha - 1} \int_0^T \langle w_t^\infty (1 - w_t^\infty), f^2 \rangle dt$$

if $d = 1$, where C depends only on d.

Next questions

- Methods for parameter estimation (**Barton, Etheridge, Kelleher & V. 2013**, Guindon et al. 2016, Forien et al. 2023)

Next questions

▶ Methods for parameter estimation (**Barton, Etheridge, Kelleher & V. 2013**, Guindon et al. 2016, Forien et al. 2023)

▶ Many other parameter regimes or forms of selection to explore (A. Etheridge, S. Penington, ...)

Next questions

▶ Methods for parameter estimation (**Barton, Etheridge, Kelleher & V. 2013**, Guindon et al. 2016, Forien et al. 2023)

▶ Many other parameter regimes or forms of selection to explore (A. Etheridge, S. Penington, ...)

▶ Effects on linked neutral genes (**Barton, Etheridge, Kelleher & V. 2013b**) and detection of a recent selective sweep?

Next questions

► Methods for parameter estimation (**Barton, Etheridge, Kelleher & V. 2013**, Guindon et al. 2016, Forien et al. 2023)

► Many other parameter regimes or forms of selection to explore (A. Etheridge, S. Penington, ...)

► Effects on linked neutral genes (**Barton, Etheridge, Kelleher & V. 2013b**) and detection of a recent selective sweep?

► Impact of stochasticity on range expansions (Louvet 2023, **Louvet & V. 2023**).

    ↝ A very interesting question is that of *gene surfing* (*cf.* work by O. Hallatschek, D.R. Nelson, ...).

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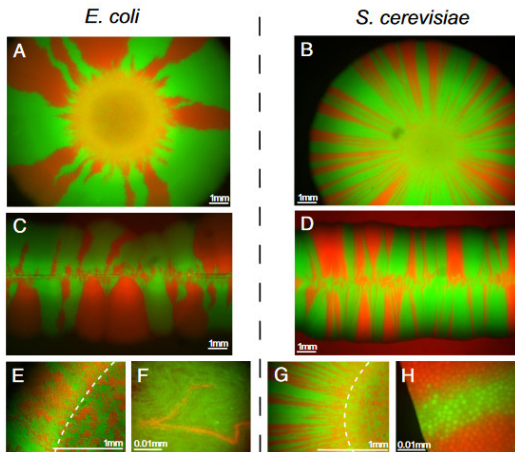
Hallatschek *et al.* (2007), PNAS. Sectors forming after placing a well-mixed droplet (or line) of fluorescent-green/red bacteria (left) or yeast (right).

Merci pour votre attention !