The mathematical challenge

What is the relative importance of mutation, selection, random drift and population subdivision for standing genetic variation?

Evolution in a spatial continuum

Alison Etheridge

University of Oxford
Joint work with Nick Barton (IST Vienna), Jerome Kelleher (Edinburgh) and
Amandine Véber (ENS)

The mathematical challenge

What is the relative importance of mutation, selection, random drift and population subdivision for standing genetic variation?

Use pattern of variation in a sample to infer the genealogical relationships between individuals ⤳ *coalescent models*

Other recruits...

Nathanael Berestycki (Cambridge)
Martin Hutzenthaler (Frankfurt)
Tom Kurtz (Madison)
Habib Saadi (Oxford)
Feng Yu (Bristol)

The mathematical challenge

What is the relative importance of mutation, selection, random drift and population subdivision for standing genetic variation?

Use pattern of variation in a sample to infer the genealogical relationships between individuals ⤳ *coalescent models*

We require consistent

- forwards in time models for evolution of population,
- backwards in time models for genealogical trees relating individuals in a sample from the population.
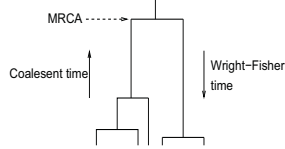

The mathematical challenge

Drift (large population limit)

Neutral (haploid) panmictic population of constant size

Forwards in time,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^3] = O(\delta t)^2$

Backwards in time

$$dp_t = \sqrt{p_t(1-p_t)}dW_t$$

Coalescence rate $\binom{k}{2}$.

$$dp_\tau = \sqrt{\frac{1}{N_e}p_\tau(1-p_\tau)}dW_\tau, \quad \text{Coalescence rate } \frac{1}{N_e}\binom{k}{2}$$
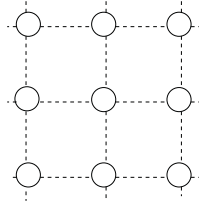

Drift (large population limit)

Neutral (haploid) panmictic population of constant size

Spatial structure

Kimura's stepping stone model

$$dp_i = \sum_j m_{ji}(p_j-p_i)dt + \sqrt{\frac{1}{N_e}p_i(1-p_i)}dW_i$$

System of interacting W-F diffusions

Drift (large population limit)

Neutral (haploid) panmictic population of constant size

Forwards in time,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^3] = O(\delta t)^2$

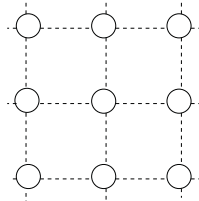$$dp_t = \sqrt{p_t(1-p_t)}dW_t$$

Spatial structure

Kimura's stepping stone model

$$dp_i = \sum_j m_{ji}(p_j-p_i)dt + \sqrt{\frac{1}{N_e}p_i(1-p_i)}dW_i$$

System of interacting W-F diffusions

The coalescent dual process $\underline{n}$ evolves as follows:

- $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ at rate $n_i m_{ji}$

- $n_i \mapsto n_i - 1$ at rate $\frac{1}{2N_e}n_i(n_i-1)$

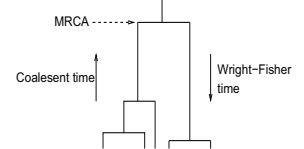## Drift (large population limit)

Neutral (haploid) panmictic population of constant size

Forwards in time,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^3] = O(\delta t)^2$

Backwards in time

$$dp_t = \sqrt{p_t(1-p_t)}dW_t$$

Coalescence rate $\binom{k}{2}$.

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot and Wright (almost) solved this problem in the 1940s: Assume uniform density, independent reproduction

Identity in state between two genes x apart:

$$0 = -2\mu F(x) + \frac{1}{2\rho}(1 - F(x))G_{2\sigma^2}(x) + \int [F(y) - F(x)]\, G_{2\sigma^2}(y - x)dy$$

where individuals leave offspring following a Gaussian G_{σ^2}.

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum.

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot and Wright (almost) solved this problem in the 1940s: Assume uniform density, independent reproduction

Identity in state between two genes x apart:

$$0 = -2\mu F(x) + \frac{1}{2\rho}(1 - F(x))G_{2\sigma^2}(x) + \int [F(y) - F(x)]\, G_{2\sigma^2}(y - x)dy$$

where individuals leave offspring following a Gaussian G_{σ^2}.
$F(0) = \frac{1}{1 + \mathcal{N}/\log(\sqrt{2\mu})}$ where $\mathcal{N} = 4\pi\rho\sigma^2$ is the *neighbourhood size*.
$F(x) \sim \frac{1}{\mathcal{N}}K_0(|x|/l)$ for $|x| \gg \sigma$, $l = \sigma/\sqrt{2\mu}$.

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot and Wright (almost) solved this problem in the 1940s:

Mathematical problems

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot and Wright (almost) solved this problem in the 1940s: Assume uniform density, independent reproduction

Mathematical problems

Felsenstein (1975). The pain in the torus: Independent reproduction
$\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

In 2D the diffusion limit fails over small scales

The obvious backwards model *fails* in 2D

Banff, September 2009 – p.7
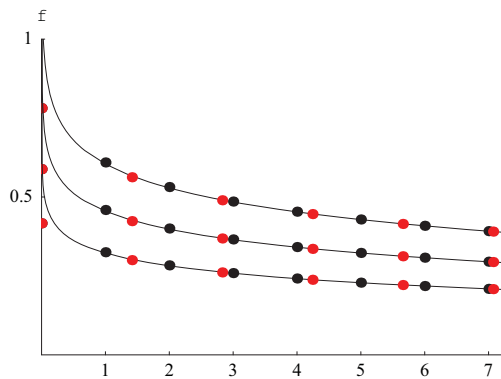

Mathematical problems

Felsenstein (1975). The pain in the torus: Independent reproduction
$\implies$ clumping;

Banff, September 2009 – p.7

. . . but

Banff, September 2009 – p.8

Mathematical problems

Felsenstein (1975). The pain in the torus: Independent reproduction
$\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

Banff, September 2009 – p.7

Biological problems

Banff, September 2009 – p.9

Mathematical problems

Felsenstein (1975). The pain in the torus: Independent reproduction
$\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

In 2D the diffusion limit fails over small scales

Banff, September 2009 – p.7

Biological problems

Genetic diversity much lower than expected from census numbers

Allele frequencies correlated over long distances

Correlations across loci reflect a shared history

Demographic history of many species dominated by large scale extinction-recolonisation events

---

Biological problems

Genetic diversity much lower than expected from census numbers

---

Biological problems

Genetic diversity much lower than expected from census numbers

Allele frequencies correlated over long distances

Correlations across loci reflect a shared history

Demographic history of many species dominated by large scale extinction-recolonisation events

. . . in a spatial continuum, neighbourhood size could be small and then pairwise coalescences may not dominate.

Biological problems

Genetic diversity much lower than expected from census numbers

Allele frequencies correlated over long distances

Λ-coalescents

Pitman (1999), Sagitov (1999)

If there are currently n ancestral lineages, each transition involving j of them merging happens at rate

$$\beta_{n,j} = \int_0^1 u^{j-2}(1-u)^{n-j}\Lambda(du)$$

- Λ a finite measure on $[0, 1]$
- Kingman's coalescent, $\Lambda = \delta_0$

Biological problems

Genetic diversity much lower than expected from census numbers

Allele frequencies correlated over long distances

Correlations across loci reflect a shared history

An individual based model

- Start with Poisson point process intensity λdx

- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x, radius r.

- If region empty, do nothing, otherwise:

Forwards in time

Bertoin & Le Gall (2003)

Suppose there is no Kingman component.
The Λ-coalescent describes the genealogy of a sample from a population evolving according to a Λ-Fleming-Viot process.

- Poisson point process intensity $dt \otimes u^{-2}\Lambda(du)$
- individual sampled at random from population
- proportion u of population replaced by offspring of chosen individual
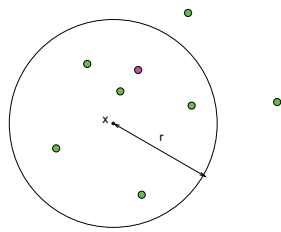

An individual based model

- Start with Poisson point process intensity λdx

- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x, radius r.

- If region empty, do nothing, otherwise:

- Choose parent at random from region

An individual based model

- Start with Poisson point process intensity λdx

An individual based model

- Start with Poisson point process intensity λdx

- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x, radius r.

- If region empty, do nothing, otherwise:

- Choose parent at random from region

- Pick $u \sim \nu_r(du)$. Each individual in region dies with probability u

---

An individual based model

- Start with Poisson point process intensity λdx

- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x, radius r.
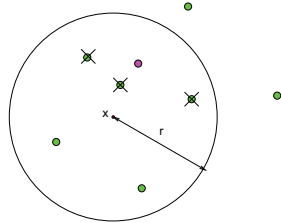
An individual based model
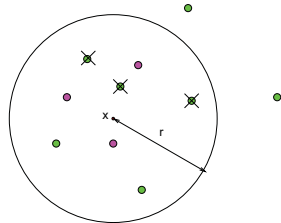
- Start with Poisson point process intensity λdx

- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x, radius r.

- If region empty, do nothing, otherwise:

- Choose parent at random from region

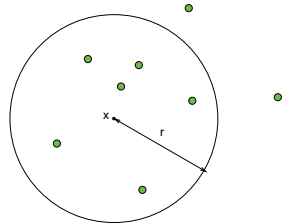- Pick $u \sim \nu_r(du)$. Each individual in region dies with probability u

- New individuals born according to a Poisson $\lambda u \mathbf{1}_{B(x,r)} dx$

A continuum limit

- If λ is sufficiently large, the population survives with positive probability (N. Berestycki, E & Hutzenthaler)

Possibility of empty regions makes formulae cumbersome.

- Replace according to Gaussian density instead of just in disc.

- Let $\lambda \to \infty$. Model retains signature of finite local population density. $\rightsquigarrow$ a *spatial Λ-Fleming-Viot process*

Genealogy of a sample from the population described by a *spatial Λ-coalescent*

Lineages follow coalescing *Lévy* (actually compound Poisson) processes with *multiple* mergers

A continuum limit

- If λ is sufficiently large, the population survives with positive probability (N. Berestycki, E & Hutzenthaler)

Possibility of empty regions makes formulae cumbersome.

The spatial Λ-Fleming-Viot process

State $\{\rho(t, x, \cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$.

A continuum limit

- If λ is sufficiently large, the population survives with positive probability (N. Berestycki, E & Hutzenthaler)

Possibility of empty regions makes formulae cumbersome.

- Replace according to Gaussian density instead of just in disc.

The spatial Λ-Fleming-Viot process

State $\{\rho(t, x, \cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$.

A continuum limit

- If λ is sufficiently large, the population survives with positive probability (N. Berestycki, E & Hutzenthaler)

Possibility of empty regions makes formulae cumbersome.

- Replace according to Gaussian density instead of just in disc.

- Let $\lambda \to \infty$. Model retains signature of finite local population density. $\rightsquigarrow$ a *spatial Λ-Fleming-Viot process*

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

Dynamics: for each $(t,x,r) \in \pi$,

- $u \sim \nu_r(du)$

- $z \sim U(B_r(x))$

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

Dynamics: for each $(t,x,r) \in \pi$,

- $u \sim \nu_r(du)$

- $z \sim U(B_r(x))$

- $k \sim \rho(t-,z,\cdot)$.

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

Dynamics: for each $(t,x,r) \in \pi$,

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

Dynamics: for each $(t,x,r) \in \pi$,

- $u \sim \nu_r(du)$

- $z \sim U(B_r(x))$

- $k \sim \rho(t-,z,\cdot)$.

For all $y \in B_r(x)$,

$$\rho(t,y,\cdot) = (1-u)\rho(t-,y,\cdot) + u\delta_k.$$

The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. π Poisson point process rate $\mu(dr) \otimes dx \otimes dt$. For each $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0,1])$.

Dynamics: for each $(t,x,r) \in \pi$,

- $u \sim \nu_r(du)$

Mixing

(Amandine Véber)

Work on a torus $\mathbb{T}(L)$ of side L in $\mathbb{R}^2$. Two types of event:
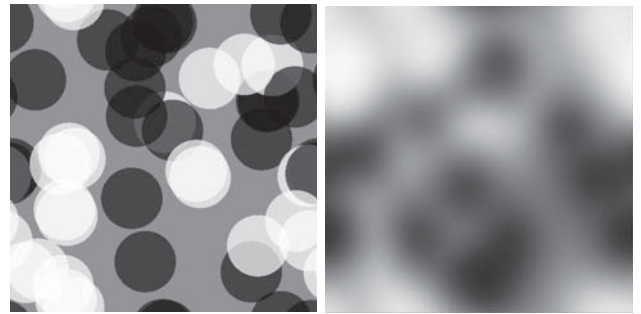
- Small events affecting bounded regions.
- Large events affecting regions of diameter $\mathcal{O}(L^\alpha)$

Each ancestral lineage is hit by a *small* event at rate $\mathcal{O}(1)$, but by a *large* event at rate $\mathcal{O}(1/\rho(L))$.
Sample at random from the whole of $\mathbb{T}(L)$.

What happens to the genealogy as $L \to \infty$?

Patterns of allele frequencies

Case (i): $\alpha < 1$

On a suitable timescale the genealogy converges to a Kingman coalescent (with an effective parameter),

The effective population size can depend on both large and small scale events.

c.f. Zähle, Cox, Durrett for classical stepping stone model.

Mixing

(Amandine Véber)

Work on a torus $\mathbb{T}(L)$ of side L in $\mathbb{R}^2$.

Case (ii): $\alpha = 1$

Three cases:

Mixing

(Amandine Véber)

Work on a torus $\mathbb{T}(L)$ of side L in $\mathbb{R}^2$. Two types of event:

- Small events affecting bounded regions.
- Large events affecting regions of diameter $\mathcal{O}(L^\alpha)$

Each ancestral lineage is hit by a *small* event at rate $\mathcal{O}(1)$, but by a *large* event at rate $\mathcal{O}(1/\rho(L))$.

Case (ii): $\alpha = 1$

Three cases:

- $\rho(L) \approx L^2$, timescale $\rho(L)$, $\rightsquigarrow$ spatial Λ-coalescent in which lineages follow independent Brownian motions in between coalescence events.

- $\rho(L) \approx L^2 \log L$, timescale $\rho(L)$, $\rightsquigarrow$ non-spatial Λ-coalescent.

- $\rho(L) \gg L^2 \log L$, timescale $L^2 \log L$, $\rightsquigarrow$ Kingman coalescent.

c.f. Nordborg & Krone (2002)

Case (ii): $\alpha = 1$

Three cases:

- $\rho(L) \approx L^2$, timescale $\rho(L)$, $\rightsquigarrow$ spatial Λ-coalescent in which lineages follow independent Brownian motions in between coalescence events.

Detecting large scale events

Two ideas:

Case (ii): $\alpha = 1$

Three cases:

- $\rho(L) \approx L^2$, timescale $\rho(L)$, $\rightsquigarrow$ spatial Λ-coalescent in which lineages follow independent Brownian motions in between coalescence events.

- $\rho(L) \approx L^2 \log L$, timescale $\rho(L)$, $\rightsquigarrow$ non-spatial Λ-coalescent.

Detecting large scale events

Two ideas:

- Slow decay in probability of identity

Case (ii): $\alpha = 1$

Three cases:

- $\rho(L) \approx L^2$, timescale $\rho(L)$, $\rightsquigarrow$ spatial Λ-coalescent in which lineages follow independent Brownian motions in between coalescence events.

- $\rho(L) \approx L^2 \log L$, timescale $\rho(L)$, $\rightsquigarrow$ non-spatial Λ-coalescent.

- $\rho(L) \gg L^2 \log L$, timescale $L^2 \log L$, $\rightsquigarrow$ Kingman coalescent.

Adding recombination

(Amandine again)

- Small events: Pick two parents, types ab and AB, say. Write r_L for fraction of recombinants.

Detecting large scale events

Two ideas:

- Slow decay in probability of identity

- Correlations between loci

Adding recombination

(Amandine again)

- Small events: Pick two parents, types ab and AB, say. Write r_L for fraction of recombinants.

$$\rho(t) = (1-u)\rho(t-) + \frac{1}{2}u(1-r_L)(\delta_{AB}+\delta_{ab}) + \frac{1}{2}ur_L(\delta_{aB}+\delta_{Ab})$$

Malécot again
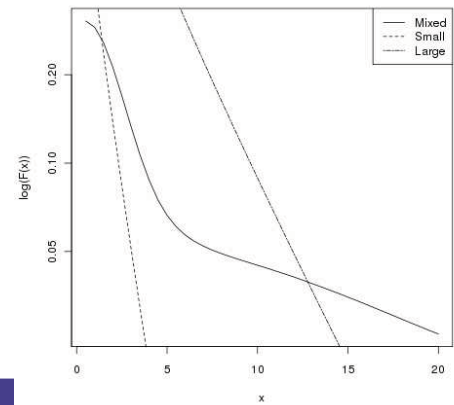
Adding recombination

(Amandine again)

- Small events: Pick two parents, types ab and AB, say. Write r_L for fraction of recombinants.

$$\rho(t) = (1-u)\rho(t-) + \frac{1}{2}u(1-r_L)(\delta_{AB}+\delta_{ab}) + \frac{1}{2}ur_L(\delta_{aB}+\delta_{Ab})$$

- Large events: ignore recombination.

Adding recombination

(Amandine again)

Correlations

Cases... but e.g. $\rho(L) \le L^{2\alpha} \rightsquigarrow$
Start L^β apart, $\beta > \alpha$.

- If
$$\lim_{L\to\infty} \frac{\log(1 + \frac{\log \rho(L)}{r_L \rho(L)})}{2\log(L^{\beta-\alpha})} \le 1$$

  then genealogies asymptotically independent.

- Otherwise genealogies completely correlated up to some time L^η.

Correlations

Cases... but e.g. $\rho(L) \le L^{2\alpha} \rightsquigarrow$

Some work in progress

Correlations

Cases... but e.g. $\rho(L) \le L^{2\alpha} \rightsquigarrow$
Start L^β apart, $\beta > \alpha$.

Some work in progress

- For $d \ge 2$, Cox, Durrett, Perkins rescaled voter model to obtain superBrownian motion. Restrict patch sizes to recover same result. What about heavy tails? (With Nathanael Berestycki & Amandine Véber)

Correlations

Cases... but e.g. $\rho(L) \le L^{2\alpha} \rightsquigarrow$
Start L^β apart, $\beta > \alpha$.

- If
$$\lim_{L\to\infty} \frac{\log(1 + \frac{\log \rho(L)}{r_L \rho(L)})}{2\log(L^{\beta-\alpha})} \le 1$$

  then genealogies asymptotically independent.

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space

Some work in progress

- For $d \geq 2$, Cox, Durrett, Perkins rescaled voter model to obtain superBrownian motion. Restrict patch sizes to recover same result. What about heavy tails? (With Nathanael Berestycki & Amandine Véber)

- Adding selection. After rescaling one can recover the classical Fisher wave. Small noise perturbations? Spatial hitchhiking. (With Amandine Véber and Feng Yu)

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event

Some work in progress

- For $d \geq 2$, Cox, Durrett, Perkins rescaled voter model to obtain superBrownian motion. Restrict patch sizes to recover same result. What about heavy tails? (With Nathanael Berestycki & Amandine Véber)

- Adding selection. After rescaling one can recover the classical Fisher wave. Small noise perturbations? Spatial hitchhiking. (With Amandine Véber and Feng Yu)

- Instead of replacing fraction u of population in a disc, replace according to a distribution (eg Gaussian). (With Nick Barton & Jerome Kelleher)

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event
- Choose spatial position of parents non-uniformly.

Some work in progress

- For $d \geq 2$, Cox, Durrett, Perkins rescaled voter model to obtain superBrownian motion. Restrict patch sizes to recover same result. What about heavy tails? (With Nathanael Berestycki & Amandine Véber)

- Adding selection. After rescaling one can recover the classical Fisher wave. Small noise perturbations? Spatial hitchhiking. (With Amandine Véber and Feng Yu)

- Instead of replacing fraction u of population in a disc, replace according to a distribution (eg Gaussian). (With Nick Barton & Jerome Kelleher)

- Convergence of genealogies. (With Tom Kurtz)

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event
- Choose spatial position of parents non-uniformly.
- Impose spatial motion of individuals not linked directly to the reproduction events.
- Spatially varying population density.
- . . . and many more.

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event
- Choose spatial position of parents non-uniformly.
- Impose spatial motion of individuals not linked directly to the reproduction events.

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event
- Choose spatial position of parents non-uniformly.
- Impose spatial motion of individuals not linked directly to the reproduction events.
- Spatially varying population density.

A framework for modelling

- Replace $\mathbb{R}^2$ by an arbitrary Polish space
- Choose a Poisson number of parents at each reproduction event
- Choose spatial position of parents non-uniformly.
- Impose spatial motion of individuals not linked directly to the reproduction events.
- Spatially varying population density.
- . . . and many more.