



Evolution in a spatial continuum

Alison Etheridge

University of Oxford

Joint work with Nick Barton (IST Vienna) and Amandine Véber (ENS)

Recruits...

Nathanael Berestycki (Cambridge)

Martin Hutzenthaler (Frankfurt)

Jerome Kelleher (Edinburgh)

Tom Kurtz (Madison)

Habib Saadi (Oxford)

Theoretical population genetics

Aim: to understand observed complex patterns of genetic variation.

Theoretical population genetics

Aim: to understand observed complex patterns of genetic variation.

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

Theoretical population genetics

Aim: to understand observed complex patterns of genetic variation.

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

Fisher, Haldane, Wright. . . . DNA sequencing presents a new type of data and requires new models.

Theoretical population genetics

Aim: to understand observed complex patterns of genetic variation.

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

Fisher, Haldane, Wright. ... DNA sequencing presents a new type of data and requires new models.

Use pattern of variation in a sample to infer the genealogical relationships between individuals

Theoretical population genetics

Aim: to understand observed complex patterns of genetic variation.

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

Fisher, Haldane, Wright. . . . DNA sequencing presents a new type of data and requires new models.

Use pattern of variation in a sample to infer the genealogical relationships between individuals

~> **coalescent models.**

The mathematical challenge

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.

Drift

Neutral (haploid) population of constant size N

Wright-Fisher model: new generation determined by multinomial sampling with equal weights.

Drift

Neutral (haploid) population of constant size N

Wright-Fisher model: new generation determined by multinomial sampling with equal weights.

- Two types a and A .
- $p(t)$ = proportion of type a .

Time in units of population size and let $N \rightarrow \infty$

Large population limit

Forwards in time,

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

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

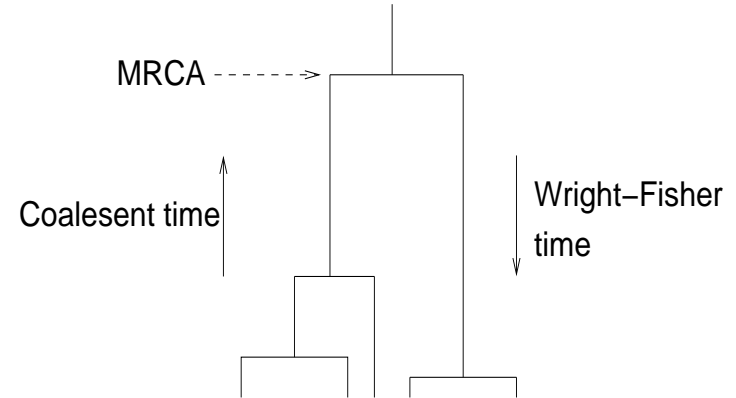
Large population limit

Forwards in time,

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

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

Backwards in time



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

Large population limit

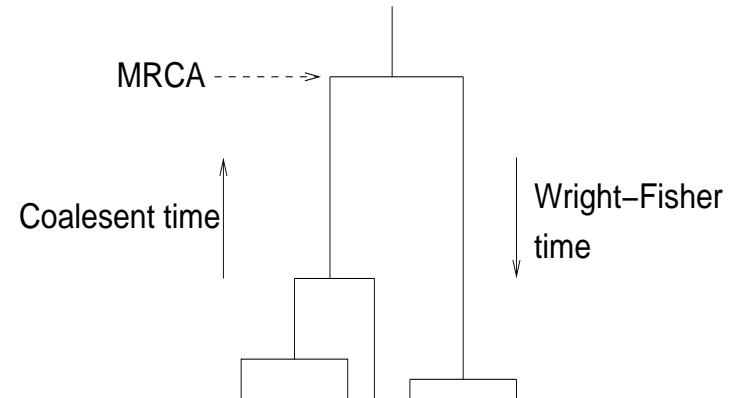
Forwards in time,

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

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

$$dp_\tau = \sqrt{\frac{1}{N_e} p_\tau(1 - p_\tau)} dW_\tau,$$

Backwards in time



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

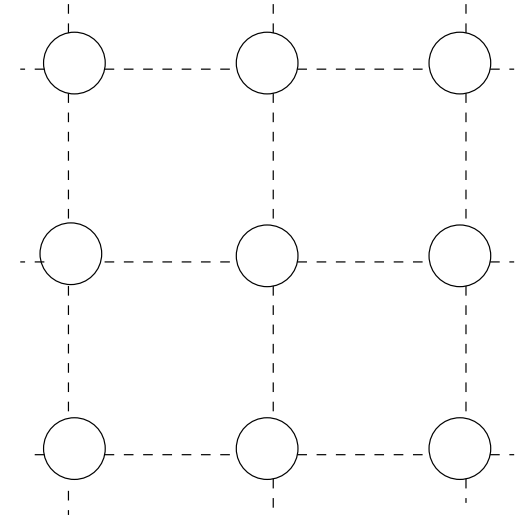
Coalescence rate $\frac{1}{N_e} \binom{k}{2}$

Spatial structure

Kimura's stepping stone model

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

System of interacting W-F diffusions



Spatial structure

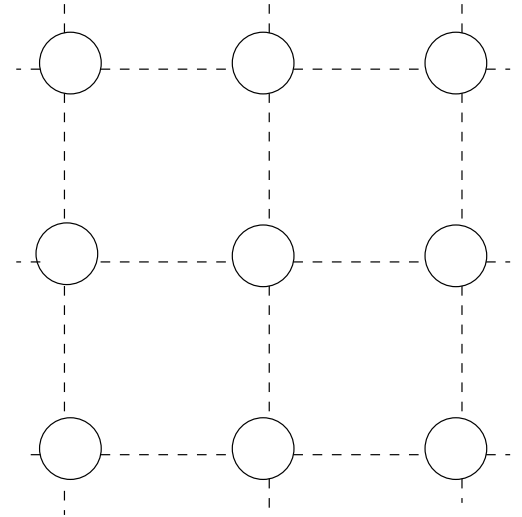
Kimura's stepping stone model

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

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



Evolution in a spatial continuum?

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

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot, Wright. . . . **inconsistent** forwards and backwards in time models.

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot, Wright. ... **inconsistent** forwards and backwards in time models.

- Can we replace the stepping stone model by a stochastic pde?

$$dp = \frac{1}{2} \Delta p dt + \sqrt{\frac{1}{N_e} p(1-p)} dW$$

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot, Wright. ... **inconsistent** forwards and backwards in time models.

- Can we replace the stepping stone model by a stochastic pde?

$$dp = \frac{1}{2} \Delta p dt + \sqrt{\frac{1}{N_e} p(1-p)} dW$$

- Can we model genealogies as coalescing Brownian motions with coalescence rate depending on current separation?

Evolution in a spatial continuum?

For many biological populations it is more natural to consider a spatial continuum. Malécot, Wright. ... **inconsistent** forwards and backwards in time models.

- Can we replace the stepping stone model by a stochastic pde?

$$dp = \frac{1}{2} \Delta p dt + \sqrt{\frac{1}{N_e} p(1-p)} dW$$

- Can we model genealogies as coalescing Brownian motions with coalescence rate depending on current separation?

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

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, $\Lambda = \delta_0$

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

Basic observation

Genetic diversity is orders of magnitude lower than expected from census numbers and genetic drift.

Something else is going on...

Basic observation

Genetic diversity is orders of magnitude lower than expected from census numbers and genetic drift.

Something else is going on...

Real populations experience large scale fluctuations in which the movement and reproductive success of many individuals are correlated.

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 .

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:

An individual based model

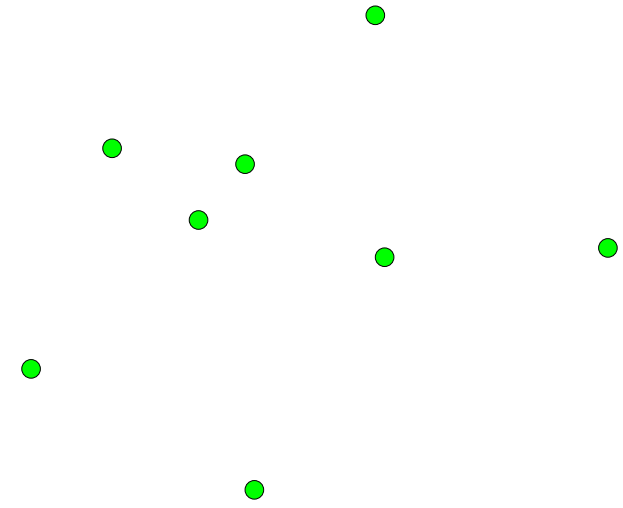
- Start with Poisson point process intensity λdx
- At rate $\mu(dr) \otimes dx \otimes dt$ throw down ball centre x , radius r .
- If region empty, do nothing, otherwise:
- Each individual in region dies with probability $u \sim \nu_r(du)$

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:
- Each individual in region dies with probability $u \sim \nu_r(du)$
- New individuals born according to a Poisson $\lambda u dx$

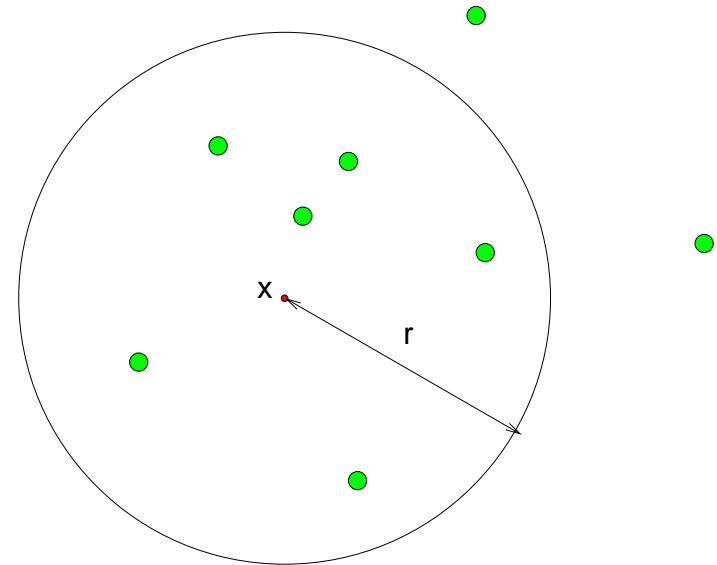
An individual based model

- 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:
- Each individual in region dies with probability $u \sim \nu_r(du)$
- New individuals born according to a Poisson $\lambda u dx$



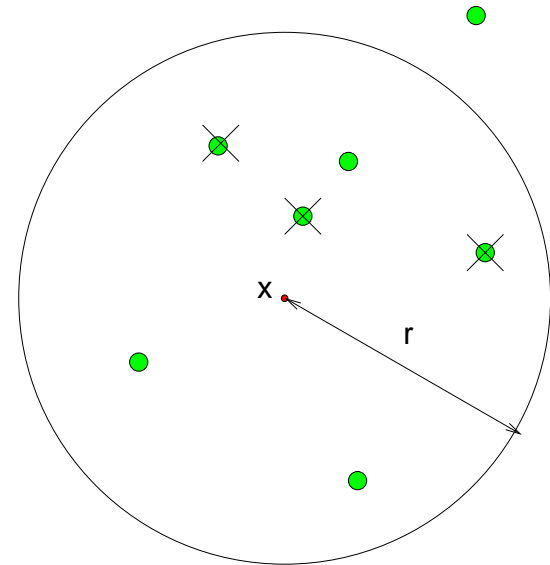
An individual based model

- 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:
- Each individual in region dies with probability $u \sim \nu_r(du)$
- New individuals born according to a Poisson $\lambda u dx$



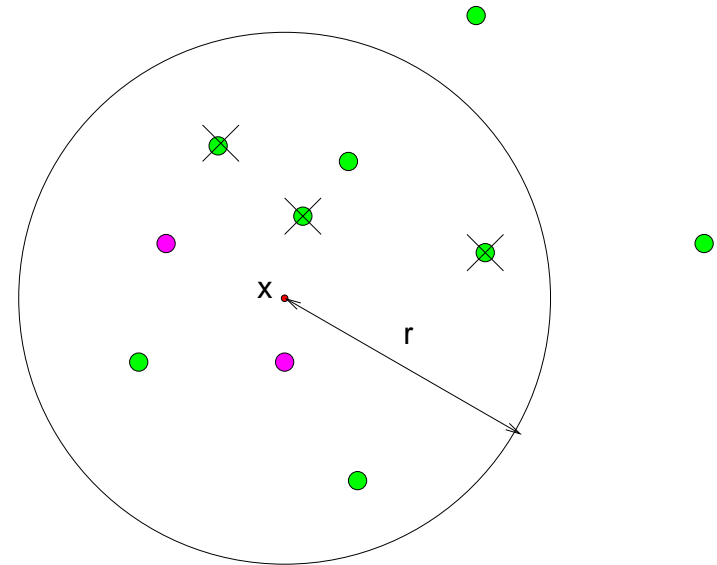
An individual based model

- 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:
- Each individual in region dies with probability $u \sim \nu_r(du)$
- New individuals born according to a Poisson $\lambda u dx$



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:
- Each individual in region dies with probability $u \sim \nu_r(du)$
- New individuals born according to a Poisson $\lambda u dx$



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.

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.

- Let $\lambda \rightarrow \infty$. **Model retains signature of finite local population density**

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.

- Let $\lambda \rightarrow \infty$. **Model retains signature of finite local population density**

\rightsquigarrow a *spatial Λ -Fleming-Viot process*

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.

- Let $\lambda \rightarrow \infty$. **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

The spatial Λ -Fleming-Viot process

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

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 $\mu(dr) \otimes dx \otimes dt$.

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 $(t, x, r) \in \pi$,

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 $(t, x, r) \in \pi$,

- $u \sim \nu_r(du)$

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 $r > 0$, $\nu_r(du) \in \mathcal{M}_1([0, 1])$.

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 $(t, x, r) \in \pi$,

- $u \sim \nu_r(du)$
- $z \sim U(B_r(x))$
- $k \sim \rho(t-, z, \cdot)$.

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

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

Conditions

$$\tilde{\Lambda}(du) = \int_{(0, \infty)} u^2 r^2 \nu_r(du) \mu(dr) \in \mathcal{M}_F([0, 1])$$

Conditions

$$\tilde{\Lambda}(du) = \int_{(0, \infty)} ur^2 \nu_r(du) \mu(dr) \in \mathcal{M}_F([0, 1])$$

Conditions

$$\tilde{\Lambda}(du) = \int_{(0,\infty)} ur^2 \nu_r(du) \mu(dr) \in \mathcal{M}_F([0,1])$$

Writing $L_r(x) = |B_r(0) \cap B_r(x)|$.

$$\int_{\mathbb{R}^2} (1 \wedge |x|^2) \left(\int_{(|x|/2, \infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2} u \nu_r(du) \mu(dr) \right) dx < \infty.$$

Mixing

(Amandine Véber)

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

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 $\mathcal{O}(1/\rho(L))$.

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 $L \rightarrow \infty$?

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.

Case (ii): $\alpha = 1$

Three cases:

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.

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.

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.

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)

-
-
-

Detecting large scale events

(Still Amandine)

Two ideas:

Detecting large scale events

(Still Amandine)

Two ideas:

- Slow decay in probability of identity

Detecting large scale events

(Still Amandine)

Two ideas:

- Slow decay in probability of identity
- Correlations between loci

Adding recombination

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

Adding recombination

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

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

Adding recombination

- 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.

Correlations

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

Correlations

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

Correlations

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

Start L^β apart, $\beta > \alpha$.

- If

$$\lim_{L \rightarrow \infty} \frac{\log\left(1 + \frac{\log \rho(L)}{r_L \rho(L)}\right)}{2 \log(L^{\beta-\alpha})} \leq 1$$

then genealogies asymptotically independent.

Correlations

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

Start L^β apart, $\beta > \alpha$.

- If

$$\lim_{L \rightarrow \infty} \frac{\log\left(1 + \frac{\log \rho(L)}{r_L \rho(L)}\right)}{2 \log(L^{\beta-\alpha})} \leq 1$$

then genealogies asymptotically independent.

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

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)

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)

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)

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

A framework for modelling

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

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.

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.

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.