Evolution in a spatial continuum

Ajison 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.
**Drift (large population limit)**

Neutral (haploid) panmictic population of constant size

Forwards in time,
- $E[Δp] = 0$ (neutrality)
- $E[(Δp)^2] = δtp(1 - p)$
- $E[(Δp)^3] = O(δt)^2$

Forwards in time,

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

Backwards in time,

$dp_r = \sqrt{\frac{1}{N_e}p_r(1 - p_r)}dW_r$, Coalescence rate $\left( \frac{k}{2} \right)$.

**Spatial structure**

Kimura's stepping stone model

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

System of interacting W-F diffusions

**Drift (large population limit)**

Neutral (haploid) panmictic population of constant size

Forwards in time,
- $E[Δp] = 0$ (neutrality)
- $E[(Δp)^2] = δtp(1 - p)$
- $E[(Δp)^3] = O(δt)^2$

Forwards in time,

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

Backwards in time,

$dp_r = \sqrt{\frac{1}{N_e}p_r(1 - p_r)}dW_r$, Coalescence rate $\left( \frac{k}{2} \right)$.

**Spatial structure**

Kimura's stepping stone model

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

System of interacting W-F diffusions

The coalescent dual process evolves as follows:

- $n_i \to n_i - 1$ at rate $n_i m_{ji}$
- $n_j \to n_j + 1$
- $n_i \to n_i - 1$ at rate $\frac{1}{2N_e}n_i(n_i - 1)$
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:

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

where individuals leave offspring following a Gaussian $G_{\sigma^2}$.

$$F(0) = \frac{1}{1 + N / \log(\sqrt{2} \mu)}$$

where $N = 4\pi \rho \sigma^2$ is the neighbourhood size.

$$F(x) \sim \frac{1}{l} K_0(|x|/l) \text{ for } |x| \gg \sigma, l = \sigma / \sqrt{2\mu}.$$
Mathematical problems

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

Local regulation \(\Rightarrow\) correlated reproduction.

In 2D the diffusion limit fails over small scales

The obvious backwards model fails in 2D

... but

Biological problems

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

Local regulation \(\Rightarrow\) correlated reproduction.

In 2D the diffusion limit fails over small scales
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.

Λ-coalescents

Pitman (1999), Sagitov (1999)

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

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

- $\Lambda$ 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 $\Lambda$-coalescent describes the genealogy of a sample from a population evolving according to a $\Lambda$-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 $\lambda dx$

• 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 $\lambda dx$

• 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 $\lambda u 1_{B(x,r)} dx$
A continuum limit

• If $\lambda$ is sufficiently large, the population survives with positive probability (N. Berestycki, E & Hutzenthaler)

Possibility of empty regions makes formulae cumbersome.

• Let $\lambda \to \infty$. Model retains signature of finite local population density. \( \rightsquigarrow \) a spatial $\Lambda$-Fleming-Viot process

Genealogy of a sample from the population described by a spatial $\Lambda$-coalescent

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

The spatial $\Lambda$-Fleming-Viot process

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

Poisson point process rate $\mu(dr) \otimes dx \otimes dt$.

A continuum limit

• If $\lambda$ 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 $\Lambda$-Fleming-Viot process

Genealogy of a sample from the population described by a spatial $\Lambda$-coalescent

Lineages follow coalescing Lévy (actually compound Poisson) processes with multiple mergers.
The spatial \( \Lambda \)-Fleming-Viot process

State \( \{ \rho(t, x, \cdot) \in M_1(K), x \in \mathbb{R}^2, t \geq 0 \} \). \( \pi \) Poisson point process rate \( \mu(dr) \otimes dx \otimes dt \). 

Dynamics: for each \( (t, x, r) \in \pi \),
- \( u \sim \nu_r(du) \)
- \( z \sim U(B_r(x)) \)

For all \( y \in B_r(x) \),
\[
\rho(t, y, \cdot) = (1 - u)\rho(t-, y, \cdot) + u\delta_k.
\]
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 $O(L^\alpha)$

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

Sample at random from the whole of $\mathbb{T}(L)$.

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

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:

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 $O(L^\alpha)$

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

Patterns of allele frequencies
Case (ii): $\alpha = 1$

Three cases:

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

\textit{c.f.} Nordborg & Krone (2002)

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 $\Lambda$-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 $\Lambda$-coalescent.
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.

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

- Large events: ignore recombination.
Correlations

Cases... but e.g. \( p(L) \leq L^{2\alpha} \Rightarrow \)
Start \( L^3 \) apart, \( \beta > \alpha \).

- If
  \[
  \lim_{L \to \infty} \frac{\log(1 + \frac{\log p(L)}{\frac{\beta}{\alpha} L^\beta})}{2 \log(L^{\beta - \alpha})} \leq 1
  \]
  then genealogies asymptotically independent.

- Otherwise genealogies completely correlated up to some time \( L^\gamma \).

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)
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 $\alpha$ of population in a disc, replace according to a distribution (e.g., 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.