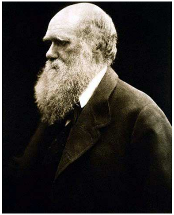
Ashok Maitra Memorial Lectures, March 2022

Some mathematical models of evolution: spatial population models

Alison Etheridge
University of Oxford

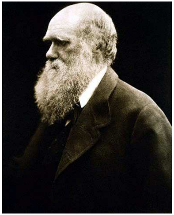
Special thanks to Nick Barton, IST Austria

The Modern Evolutionary Synthesis

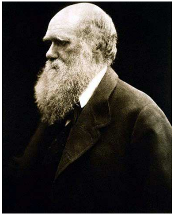


The Modern Evolutionary Synthesis

Darwin: Heritable traits that increase reproductive success will become more common in a population.



The Modern Evolutionary Synthesis

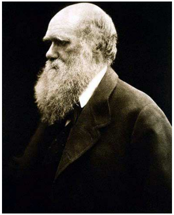


Darwin: Heritable traits that increase reproductive success will become more common in a population.

Requires:

- Variation in population
- Offspring must be similar to parents

The Modern Evolutionary Synthesis



Darwin: Heritable traits that increase reproductive success will become more common in a population.

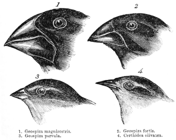
Requires:

- Variation in population
- Offspring must be similar to parents

Mendel: Traits ‘determined’ by genes.

- Genes occur in different types (alleles)
- Offspring inherit genes from parents

The Modern Evolutionary Synthesis



Darwin: Heritable traits that increase reproductive success will become more common in a population.

Requires:

- Variation in population
- Offspring must be similar to parents

Mendel: Traits 'determined' by genes.

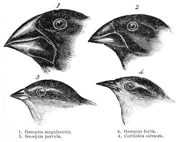
- Genes occur in different types (alleles)
- Offspring inherit genes from parents

The Modern Evolutionary Synthesis

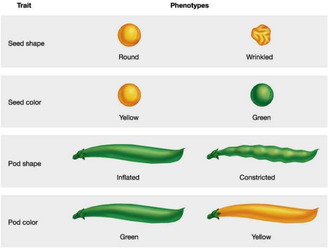
Darwin: Heritable traits that increase reproductive success will become more common in a population.

Requires:

- Variation in population
- Offspring must be similar to parents



1. Geospiza magnirostris.
3. Geospiza parvula.
2. Geospiza fortis.
4. Certhidea olivacea.



Trait	Phenotypes	
Seed shape	Round	Wrinkled
Seed color	Yellow	Green
Pod shape	Inflated	Constricted
Pod color	Green	Yellow

Mendel: Traits ‘determined’ by genes.

- Genes occur in different types (alleles)
- Offspring inherit genes from parents

A fundamental question

At the beginning of the 20th Century, the modern evolutionary synthesis saw the theories of Darwin and Mendel united, but fundamental questions remained.

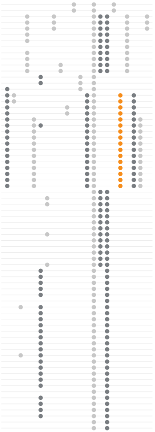
What is the relative importance of:

- natural selection;
- population structure (spatial and genetic);
- genetic drift (randomness due to reproduction in a finite population);
- . . .

A mathematical challenge

The pioneers could only observe genetic variation indirectly through phenotype.

www.alamy.com

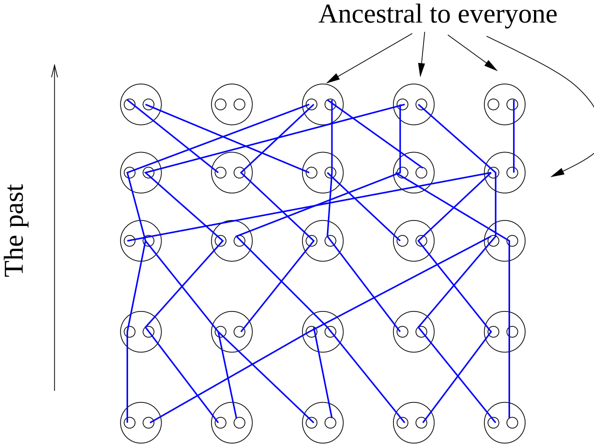


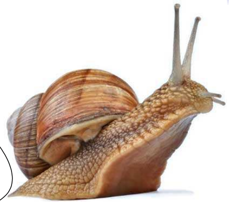
Jonathan Marchini

Modern geneticists use differences in DNA sequences to infer ‘relatedness’ between sampled individuals. We require consistent

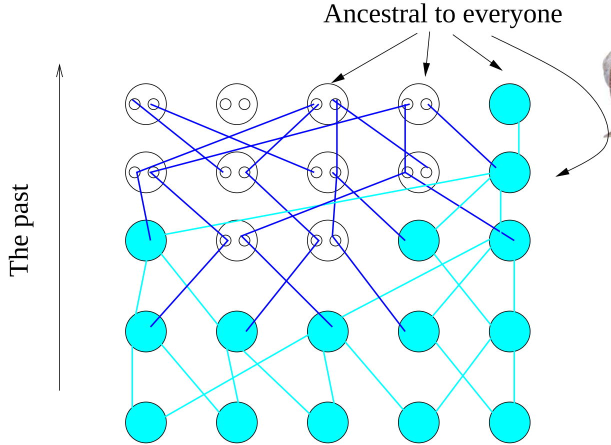
- forwards in time models for the evolution of the population,
- and backwards in time models for the relatedness between individuals in a sample.

A simple model of inheritance




www.snailworld.com

A simple model of inheritance

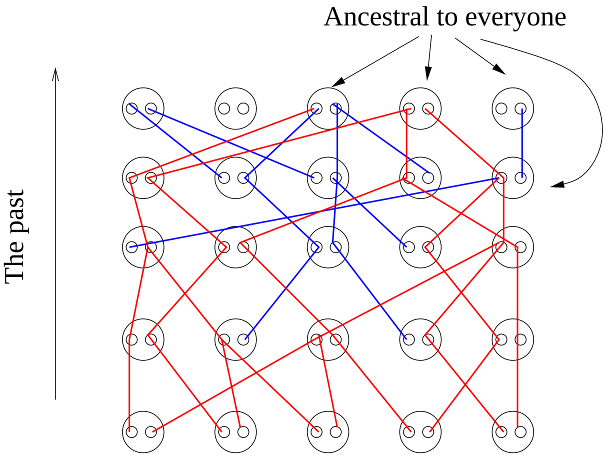


Ancestral to everyone

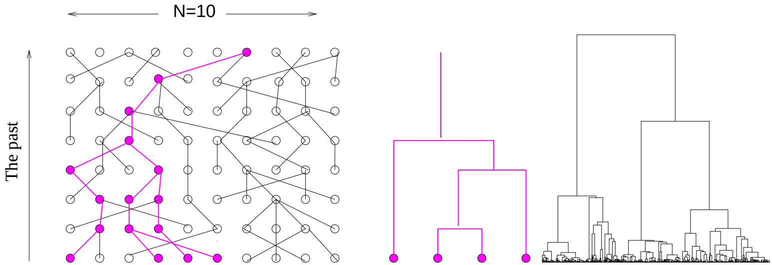
The past

www.snailworld.com

A simple model of inheritance



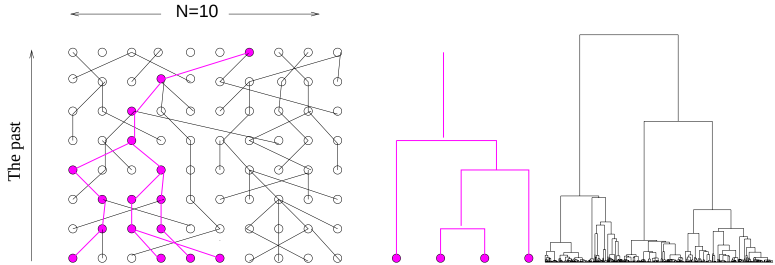
The simplest imaginable model of inheritance



Kingman 1982

$$\mathbb{P}[\text{2 lineages coalesce in previous generation}] \approx \frac{1}{N}$$

The simplest imaginable model of inheritance

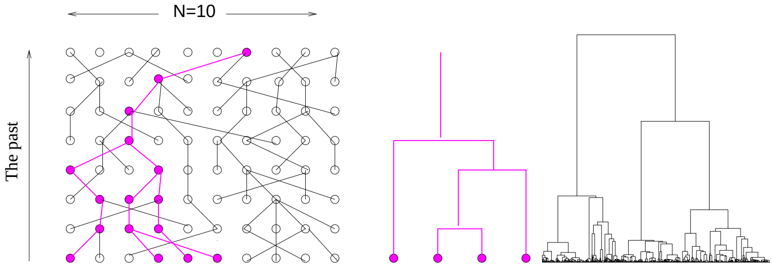


Kingman 1982

$\mathbb{P}[2 \text{ lineages coalesce in previous generation}] \approx \frac{1}{N}$

Time in units of N generations, $N \to \infty$, $\rightsquigarrow$ time to coalescence pair of lineages $\sim \texttt{Exp}(1)$

The simplest imaginable model of inheritance



Kingman 1982

$\mathbb{P}$[2 lineages coalesce in previous generation] $\approx \frac{1}{N}$

The most recent common ancestor in the pedigree was $\approx \log_2 N$ generations in the past. The most recent common *genetic* ancestor was $\approx 2N$ generations ago.

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability one of these events before single pairwise merger

$$\approx \frac{N}{\binom{j}{2}}\binom{j}{2}\binom{j-2}{2}\frac{1}{N^2} \approx \frac{j^2}{2N}$$

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability such an event somewhere in the genealogical tree

$$\approx \sum_{j=1}^{k} \frac{j^2}{N} \approx \frac{k^3}{3N}$$

Sample size k

If currently j ancestral lineages:

- Probability two pairs of lineages merge into separate parents

$$\binom{j}{2}\binom{j-2}{2}\frac{1}{N}\frac{1}{N-1} \approx 6\binom{j}{4}\frac{1}{N^2}$$

- Probability three-merger

$$\approx \binom{j}{3}\frac{1}{N^2}$$

Probability such an event somewhere in the genealogical tree

$$\approx \sum_{j=1}^{k} \frac{j^2}{N} \approx \frac{k^3}{3N}$$

Sample size $k \ll N^{1/3}$, pair of lineages coalesces rate $\approx \frac{1}{N}\binom{k}{2}$

Melfi & Viswanath (2018)

The Kingman coalescent

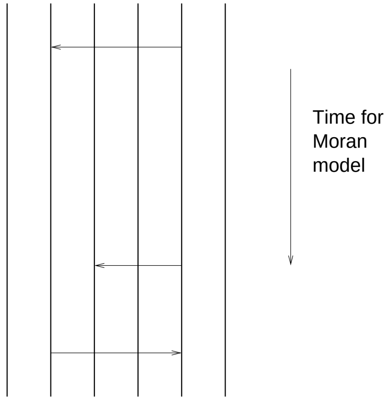
Tracing back in time, if there are currently k ancestral lineages, the next event will occur after an exponentially distributed time with parameter $\binom{k}{2}$, when a pair of lineages (chosen at random) will coalesce.

For a vast array of models in which

- population size large and constant;
- all individuals are equally fit;
- there is no spatial structure;

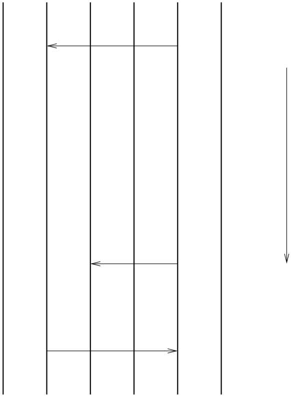
measuring time in units of N_e generations, the genealogy of a sample is well approximated by the Kingman coalescent.

The Moran Model

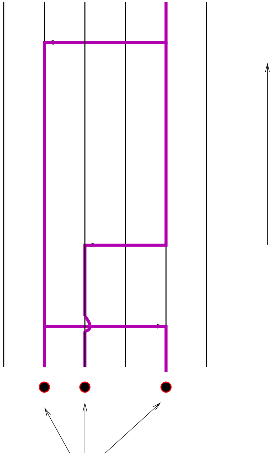


- Events determined by Poisson Process intensity $\binom{N}{2}dt$;
- Pair chosen at random;
- One reproduces, the other dies.

Genealogies under the Moran model



Time for
Moran
model



Time for
Coalescent

Sample

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

- Each individual chooses parent uniformly at random from the previous generation;
- Offspring inherit the type of their parent.

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

- Each individual chooses parent uniformly at random from the previous generation;
- Offspring inherit the type of their parent.

'Alleles' a, A. Proportion p of a alleles among parents.

- Number of a-offspring $\texttt{Bin}(N, p)$.
- $\mathbb{E}[\Delta p] = 0$ (neutral); $\mathbb{E}[(\Delta p)^2] = \frac{1}{N} p(1 - p)$.

Forwards in time: scaling the Wright-Fisher model

Population of fixed size N evolves in discrete generations.

- Each individual chooses parent uniformly at random from the previous generation;
- Offspring inherit the type of their parent.

'Alleles' a, A. Proportion p of a alleles among parents.

- Number of a-offspring $\mathtt{Bin}(N, p)$.
- $\mathbb{E}[\Delta p] = 0$ (neutral); $\mathbb{E}[(\Delta p)^2] = \frac{1}{N}p(1-p)$.

$\rightsquigarrow$ changes in p over timescales $\mathcal{O}(N)$ generations.

Drift (large population limit)

Time in units of N generations, $\delta t = \frac{1}{N}$, $N \to \infty$

Drift (large population limit)

Time in units of N generations, $\delta t = \frac{1}{N}$, $N \to \infty$

Forwards in time, $\Delta p = p_{t+\delta t} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(\delta t)^2$

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

Drift (large population limit)

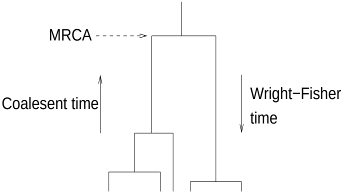
Time in units of N generations, $\delta t = \frac{1}{N}$, $N \to \infty$

Forwards in time, $\Delta p = p_{t+\delta t} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(\delta t)^2$

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

Backwards in time



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

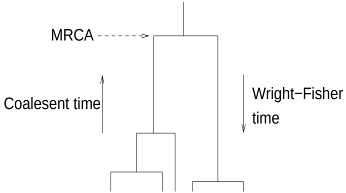
Drift (large population limit)

Time in units of N generations, $\delta t = \frac{1}{N}$, $N \to \infty$

Forwards in time, $\Delta p = p_{t+\delta t} - p_t$,

- $\mathbb{E}[\Delta p] = 0$ (neutrality)
- $\mathbb{E}[(\Delta p)^2] = \delta t p(1-p)$
- $\mathbb{E}[(\Delta p)^4] = O(\delta t)^2$

Backwards in time



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

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

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

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

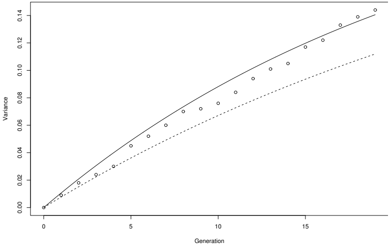
Buri's experiment

Gene that affects eye colour (but not fitness). Two alleles, a, A.

~ 100 populations, 8 males, 8 females. Each started with proportion $p = 0.5$ type a.

Eventually, each population will be entirely one type (with equal probabilities).

Under Wright-Fisher model, variance in p across populations increases from 0 to $1/4$ over time.



Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \le u \le t. \qquad (*)$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

$$\begin{aligned} dp_u^{n(t-u)} = n(t-u)p_u^{n(t-u)-1}&\sqrt{p_u(1-p_u)}dW_u \\ &+ \binom{n(t-u)}{2}p_u^{n(t-u)-2}p_u(1-p_u)du \\ &\qquad - \binom{n(t-u)}{2}\left(p_u^{n(t-u)-1} - p_u^{n(t-u)}\right) \end{aligned}$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

Another relationship between Kingman and Wright-Fisher

Method of duality: Express distribution of one random variable $\underline{p}$ in terms of another (simpler) random variable $\underline{n}$.

$$\frac{d}{du}\mathbb{E}\left[f\left(\underline{p}(u), \underline{n}(t-u)\right)\right] = 0, \quad 0 \leq u \leq t. \qquad (*)$$

$$dp_t = \sqrt{p_t(1-p_t)}dW_t, \quad n_t \mapsto n_t - 1 \text{ rate } \binom{n_t}{2}, \quad f(p,n) = p^n$$

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

Weaker than saying genealogy given by Kingman coalescent

Adding selection (Wright-Fisher setting)

Relative fitnesses:

a	A
$1 - s_N$	1

- Each individual independently chooses parent; probability parent of type a:

$$\frac{(1 - s_N)p}{1 - s_N p} = (1 - s_N)p + s_N p^2 + \mathcal{O}(s_N^2)$$

Adding selection (Wright-Fisher setting)

Relative fitnesses:

a	A
$1 - s_N$	1

- Each individual independently chooses parent; probability parent of type a:

$$\frac{(1-s_N)p}{1-s_Np} = (1-s_N)p + s_Np^2 + \mathcal{O}(s_N^2)$$

Ignoring $\mathcal{O}(s_N^2)$ terms

- Number of a-offspring $\texttt{Bin}\big(N, p - s_Np(1-p)\big)$.
- $\mathbb{E}[\Delta p] = -s_Np(1-p)$; $\mathbb{E}[(\Delta p)^2] = \frac{1}{N}p(1-p)$.

Adding selection (Wright-Fisher setting)

Relative fitnesses:

a	A
$1 - s_N$	1

- Each individual independently chooses parent; probability parent of type a:

$$\frac{(1-s_N)p}{1-s_N p} = (1-s_N)p + s_N p^2 + \mathcal{O}(s_N^2)$$

Ignoring $\mathcal{O}(s_N^2)$ terms

- Number of a-offspring $\texttt{Bin}\big(N, p - s_N p(1-p)\big)$.
- $\mathbb{E}[\Delta p] = -s_N p(1-p)$; $\mathbb{E}[(\Delta p)^2] = \frac{1}{N} p(1-p)$.

$N s_N \to s$, $dp_t = -s p_t(1-p_t)dt + \sqrt{p_t(1-p_t)}dW_t$

Adding selection (Moran model)

Relative fitnesses:

a	A
$1 - s_N$	1

- Events determined by Poisson Process intensity $\binom{N}{2} dt$;
- Pair chosen at random;
- If types a, A chosen, probability a reproduces $\frac{1}{2}(1 - s_N)$.

Adding selection (Moran model)

Relative fitnesses:

a	A
$1 - s_N$	1

- Events determined by Poisson Process intensity $\binom{N}{2}dt$;
- Pair chosen at random;
- If types a, A chosen, probability a reproduces $\frac{1}{2}(1 - s_N)$.

$\mathbb{P}[\text{type } a \text{ reproduces}]$

$$= p^2 + 2p(1-p)\frac{1}{2}(1 - s_N) = p - s_N p(1-p).$$

Adding selection (Moran model)

Relative fitnesses:

a	A
$1-s_N$	1

- Events determined by Poisson Process intensity $\binom{N}{2}dt$;
- Pair chosen at random;
- If types a, A chosen, probability a reproduces $\frac{1}{2}(1-s_N)$.

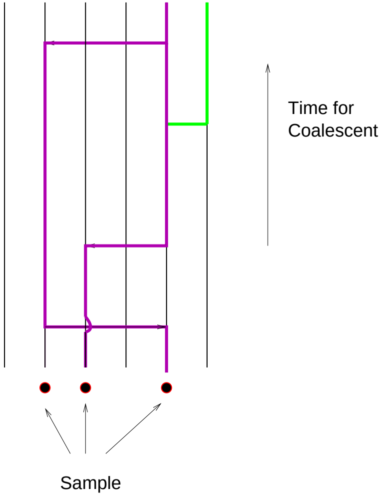
$\mathbb{P}$[type a reproduces]

$$= p^2 + 2p(1-p)\frac{1}{2}(1-s_N) = p - s_N p(1-p).$$

c.f. in Wright-Fisher, probability parent of type a:

$$\frac{(1-s_N)p}{1-s_N p} = (1-s_N)p + s_N p^2 + \mathcal{O}(s_N^2)$$

Adding selection (alternative view)



Neutral events at rate $(1 - s_N)\binom{N}{2}$;

Selective events at rate $s_N\binom{N}{2}$:
if $\{a, A\}$ chosen, A reproduces.

$$\mathbb{P}[\text{type } a \text{ parent}]$$
$$= (1 - s_N)p + s_N p^2 = p - s_N p(1-p)$$

$$N s_N \to s,$$

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

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}].$$

The ancestral selection graph

- $n_t \mapsto n_t - 1$ rate $\binom{n_t}{2}$
- $n_t \mapsto n_t + 1$ rate sn_t

The ancestral selection graph

- $n_t \mapsto n_t - 1$ rate $\binom{n_t}{2}$
- $n_t \mapsto n_t + 1$ rate sn_t

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

The ancestral selection graph

- $n_t \mapsto n_t - 1$ rate $\binom{n_t}{2}$
- $n_t \mapsto n_t + 1$ rate sn_t

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

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

The ancestral selection graph

- $n_t \mapsto n_t - 1$ rate $\binom{n_t}{2}$
- $n_t \mapsto n_t + 1$ rate sn_t

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

Sampling probabilities:

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

All individuals in sample are type a *iff* all their ancestors in the ASG are type a.

What we have so far

In time units of N_e generations,

- (Forwards time) The Wright-Fisher diffusion (with and without selection)

$$dp_t = -sp_t(1-p_t)dt + \sqrt{p_t(1-p_t)}dW_t;$$

- (Backwards time) The Kingman coalescent/ ASG

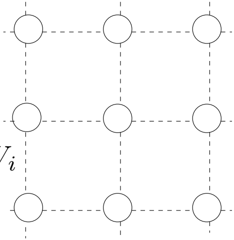
$$n_t \mapsto n_t - 1 \text{ at rate } \binom{n_t}{2}, \quad n_t \mapsto n_t + 1 \text{ at rate } sn_t;$$

- Sampling probabilities

$$\mathbb{E}[p(t)^{n(0)}] = \mathbb{E}[p(0)^{n(t)}]$$

Stronger result holds. Kingman coalescent really describes genealogy of random sample from (neutral) population.

Spatial structure



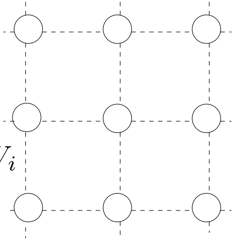
Kimura's stepping stone model

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

System of interacting W-F diffusions

$\sum_j N_e(i)m_{ij} = \sum_j N_e(j)m_{ji}$

Spatial structure



Kimura's stepping stone model

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

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 $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

$\sum_j N_e(i) m_{ij} = \sum_j N_e(j) m_{ji}$

The structured coalescent

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 $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

The structured coalescent

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 $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

If $N_e(i)$ independent of i, backwards in time motion determined by (reversal of) forwards in time dispersal densities m_{ij}

The structured coalescent

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 $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

If $N_e(i)$ independent of i, backwards in time motion determined by (reversal of) forwards in time dispersal densities m_{ij}

In general, lineages drawn to more densely population demes

The structured coalescent

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 $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

If $N_e(i)$ independent of i, backwards in time motion determined by (reversal of) forwards in time dispersal densities m_{ij}

In general, lineages drawn to more densely population demes

More natural to have a model in which population size evolves

The structured coalescent

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

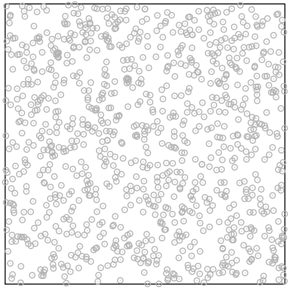
- $\begin{cases} n_i \mapsto n_i - 1 \\ n_j \mapsto n_j + 1 \end{cases}$ at rate $n_i \frac{N_e(j)}{N_e(i)} m_{ji}$
- $n_i \mapsto n_i - 1$ at rate $\frac{1}{N_e(i)} \binom{n_i}{2}$

If $N_e(i)$ independent of i, backwards in time motion determined by (reversal of) forwards in time dispersal densities m_{ij}

In general, lineages drawn to more densely population demes

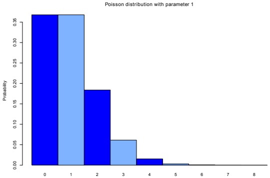
More natural to have a model in which population size evolves
Many populations are distributed across spatial continua

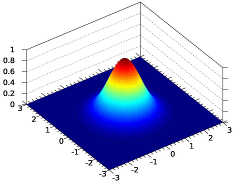
The Wright-Malécot model



$N = 1000$

- Individuals are scattered across a two-dimensional space.
- In each generation, each individual produces a Poisson number of offspring (average one).
- Offspring are scattered in a Gaussian distribution around their parent.



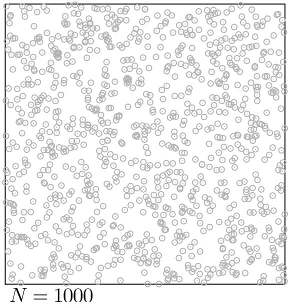


Mitch Gooding
Jerome Kelleher

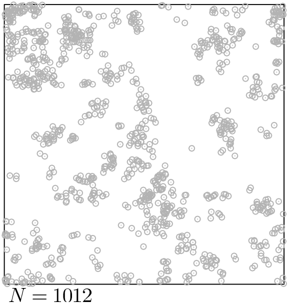
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 0$



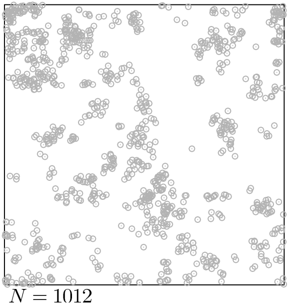
$t = 10$



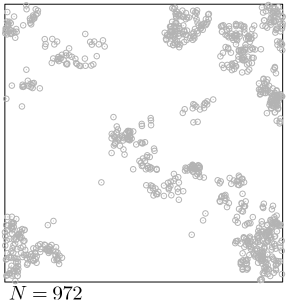
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 10$



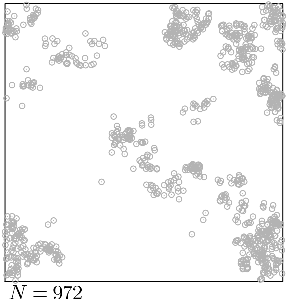
$t = 100$



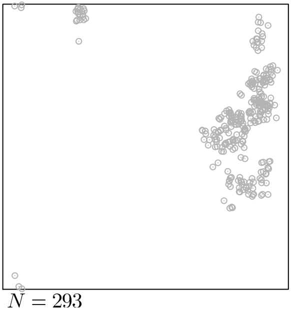
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 100$



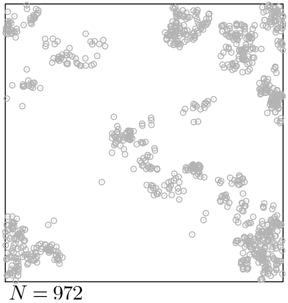
$t = 1000$



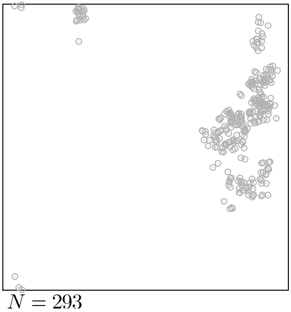
The pain in the torus (Felsenstein, 1975)

With thanks to Jerome Kelleher

$t = 100$



$t = 1000$



In $d = 1, 2$ population exhibits clumping/extinction

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\implies$ clumping;

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1$, 2, independent reproduction $\Longrightarrow$ clumping;

Local regulation $\Longrightarrow$ correlated reproduction.

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1, 2$, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e}p_t(x)(1 - p_t(x))}dW(t, x)$$

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1, 2$, independent reproduction $\Longrightarrow$ clumping;

Local regulation $\Longrightarrow$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e} p_t(x)(1 - p_t(x))} dW(t,x)$$

In 2D the diffusion limit fails over small scales

Mathematical problems

Felsenstein (1975). The pain in the torus: In $d = 1, 2$, independent reproduction $\implies$ clumping;

Local regulation $\implies$ correlated reproduction.

What about modifying the stepping stone model?

$$dp_t(x) = \frac{1}{2}\Delta p_t(x) + \sqrt{\frac{1}{2N_e}p_t(x)(1 - p_t(x))}dW(t, x)$$

In 2D the diffusion limit fails over small scales ... and so does the obvious backwards model.

Probability of identity

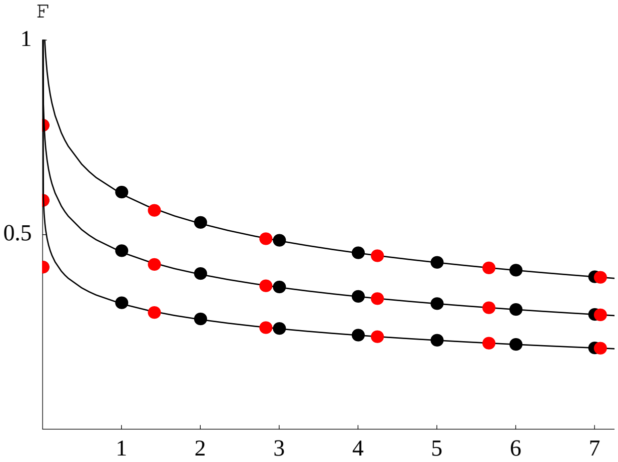
Information about the genealogy in patterns of neutral variation.

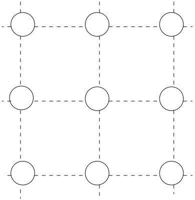
Simplest example:

- Infinitely many alleles model of mutation: each individual in each generation, independently, with small probability μ mutates to a type never before seen in the population
- *Probability of identity by descent* of two individuals, F, is probability no mutation since MRCA
- Equivalently $F = (1 - 2\mu)^T \approx \exp(-2\mu T)$ is the Laplace transform of the distribution of the time to the MRCA.

The neutral mutation rate dictates the timescales over which we can reconstruct information about genealogies.

Malécot-Wright versus Kimura?





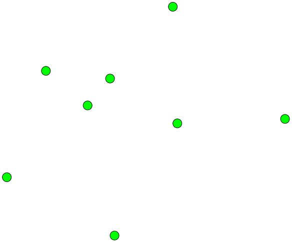
$F = \mathbb{P}[\text{identity}]$

One more observation

In a spatial continuum, a single individual can be parent to a significant proportion of the local population.

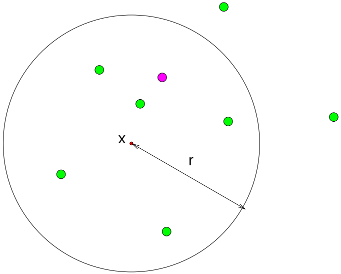
An individual based model

- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.



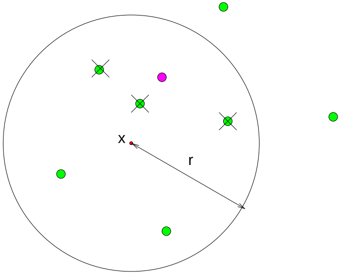
An individual based model

- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.
- If region empty, do nothing, otherwise:
- Choose parent from $B(x, r)$,



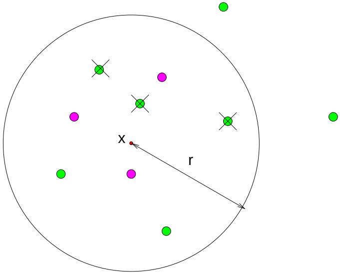
An individual based model

- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.
- If region empty, do nothing, otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,



An individual based model

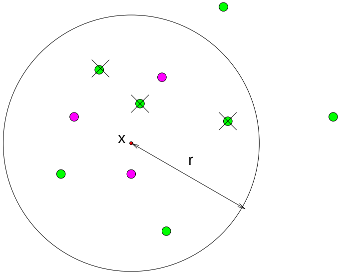
- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.
- If region empty, do nothing, otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,
- New individuals born according to Poisson intensity $\lambda u \mathbf{1}_{B_r(x)}$.



An individual based model

- Start with Poisson intensity λdx. Events rate $dt \otimes dx \otimes \xi(dr, du)$. Throw down ball $B(x, r)$.
- If region empty, do nothing, otherwise:
- Choose parent from $B(x, r)$,
- Each individual in region dies with probability u,
- New individuals born according to Poisson intensity $\lambda u \mathbf{1}_{B_r(x)}$.

Offspring inherit type of parent



$\lambda \to \infty$ limit (no space)

Start from $\mathtt{Poiss}(\lambda)$

If first reproduction event has 'impact' u

- $\mathtt{Poiss}((1-u)\lambda)$ 'survivors';
- $\mathtt{Poiss}(u\lambda)$ offspring.

As $\lambda \to \infty$ proportion u of individuals die and are replaced by offspring of the type of the parent.

The Λ-Fleming-Viot process

State $\{\rho(t,\cdot) \in \mathcal{M}_1(K), t \geq 0\}$. K space of genetic types.

- Poisson Point Process Π intensity $dt \otimes F(du)$
- if $(t,u) \in \Pi$, individual sampled at random from population at time $t-$ (i.e. choose $k \sim \rho(t-)$)
- proportion u of population replaced by offspring of chosen individual

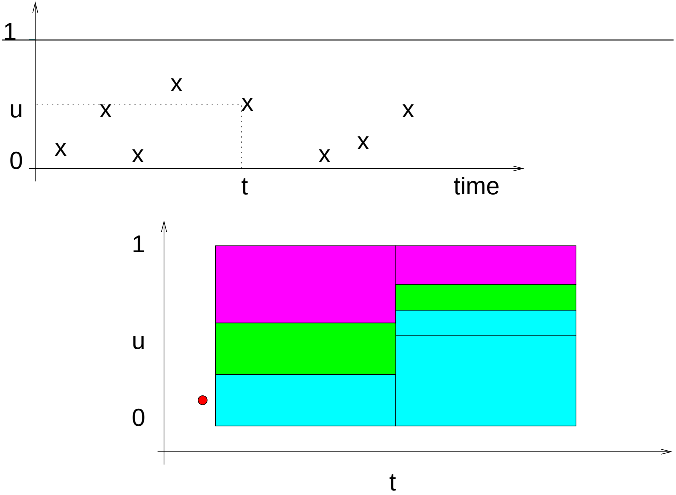
$$\rho(t,\cdot) = (1-u)\rho(t-,\cdot) + u\delta_k.$$

$F(du) = \frac{\Lambda(du)}{u^2}$, Λ finite measure on $[0,1]$.

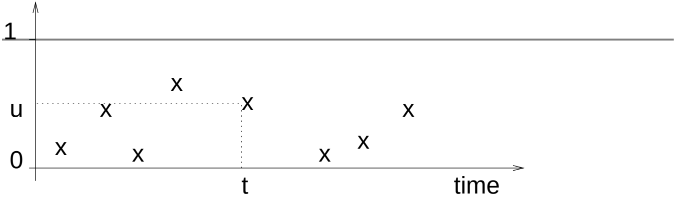
Donnelly & Kurtz (1999)

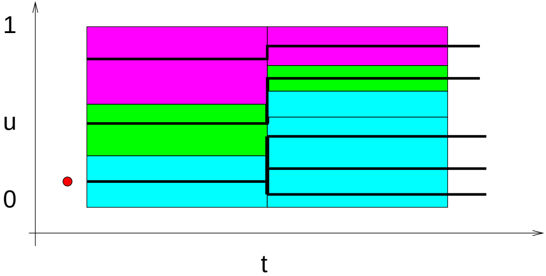
('Generalised Fleming-Viot process', Bertoin & Le Gall 2003)

The Λ-Fleming-Viot process



The Λ-Fleming-Viot process





Λ-coalescents

Donnelly & Kurtz (1999), Pitman (1999), Sagitov (1999)

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

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

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

The spatial Λ-Fleming-Viot process

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

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 $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

The spatial Λ-Fleming-Viot process

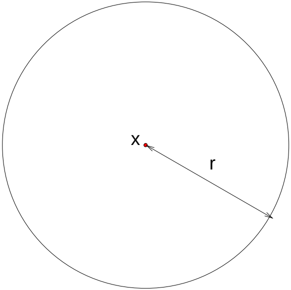
State $\{\rho(t, x, \cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr, du)$ on $[0, \infty) \times \mathbb{R}^2 \times [0, \infty) \times [0, 1]$.

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

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 $dt \otimes dx \otimes \xi(dr, du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

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

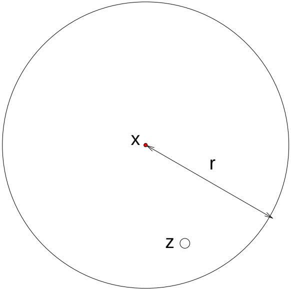


The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

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

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

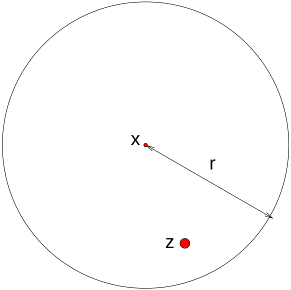


The spatial Λ-Fleming-Viot process

State $\{\rho(t,x,\cdot) \in \mathcal{M}_1(K), x \in \mathbb{R}^2, t \geq 0\}$. Π Poisson point process rate $dt \otimes dx \otimes \xi(dr, du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.

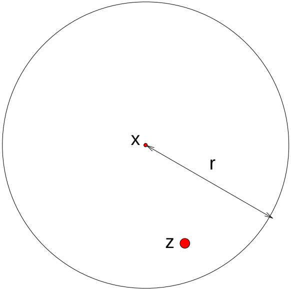
Dynamics: for each $(t,x,r,u) \in \Pi$,

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



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 $dt \otimes dx \otimes \xi(dr,du)$ on $[0,\infty) \times \mathbb{R}^2 \times [0,\infty) \times [0,1]$.



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

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

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

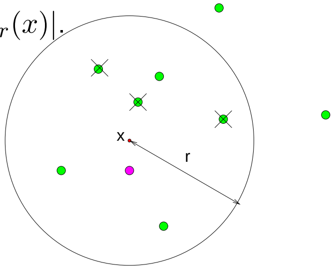
$$\rho(t,y,\cdot) = (1-u)\rho(t-,y,\cdot) + u\delta_k.$$

Backwards in time

- A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2}\, u\, \xi(dr, du) dx$$

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.



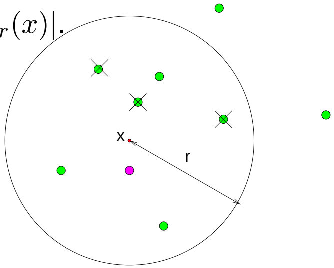
Backwards in time

- A *single* ancestral lineage evolves in series of jumps with intensity

$$dt \otimes \int_{(|x|/2,\infty)} \int_{[0,1]} \frac{L_r(x)}{\pi r^2} \, u \, \xi(dr, du) dx$$

on $\mathbb{R}_+ \times \mathbb{R}^2$ where $L_r(x) = |B_r(0) \cap B_r(x)|$.

- Lineages can coalesce when hit by same ‘event’.



Introducing selection to the SLFV

$K = \{a, A\}$, $w(t, x) = \rho(t, x, a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1 - s)$. If a reproduction event affects a region $B(x, r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is
$$\frac{(1 - s)\overline{w}}{1 - s\overline{w}}$$

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t, x) = \rho(t, x, a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1 - s)$. If a reproduction event affects a region $B(x, r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is
$$\frac{(1-s)\overline{w}}{1-s\overline{w}} = \overline{w}(1-s) + s\overline{w}^2 + \mathcal{O}(s^2).$$

Introducing selection to the SLFV

$K = \{a, A\}$, $w(t,x) = \rho(t,x,a)$ proportion of type a

- (i) Two types, a, A. Weight type a by $(1-s)$. If a reproduction event affects a region $B(x,r)$ in which current proportion of a-alleles is $\overline{w}$, then probability offspring are type a is

$$\frac{(1-s)\overline{w}}{1-s\overline{w}} = \overline{w}(1-s) + s\overline{w}^2 + \mathcal{O}(s^2).$$

- (ii) Neutral events rate $\propto (1-s)$, selective events rate $\propto s$. At selective reproduction events, sample two potential parents. If types aa, then an a reproduces, otherwise an A does.

(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

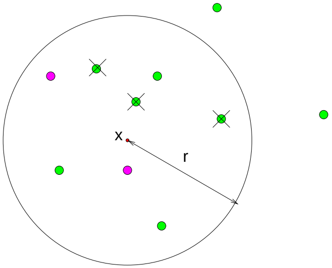
(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

At *selective* events

- *Two 'potential' parents must be traced;*
- Lineages can coalesce when hit by same 'event'.



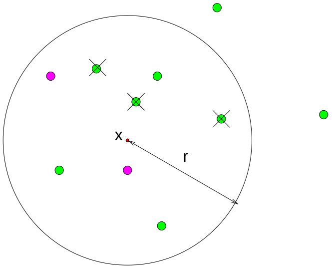
(Spatial) Ancestral selection graph

Evolution of ancestry due to neutral events as before:

- lineages evolve in a series of jumps;
- they can coalesce when covered by same event.

At *selective* events

- *Two 'potential' parents must be traced;*
- Lineages can coalesce when hit by same 'event'.



A sampled individual is type a *iff* all lineages in the corresponding ASG are type a at any previous time.

When can we detect selection?

Neutral mutation rate, μ, sets timescale

- Mutation rates are low;
- Scaling limits are ‘robust’.

Natural question:

When, and over what spatial scales can we expect to observe a signature of natural selection?

Scaling limits: Small ‘neighbourhood size’:

Fix $u \in (0,1)$.

Scaling limits: Small 'neighbourhood size':

Fix $u \in (0, 1)$.

Set $n = 1/\mu$ and rescale: $w(nt, \sqrt{n}x)$.

Scaling limits: Small ‘neighbourhood size’:

Fix $u \in (0,1)$.

Set $n = 1/\mu$ and rescale: $w(nt, \sqrt{n}x)$.

Heuristics:

- At a ‘branching’ event in ASG, two lineages born at separation $\mathcal{O}(1/\sqrt{n})$.
- Probability they separate to $\mathcal{O}(1)$ before coalescing is
  - $d = 1$: $\mathcal{O}(1/\sqrt{n})$,
  - $d = 2$: $\mathcal{O}(1/\log n)$,
  - $d \geq 3$: $\mathcal{O}(1)$.
- Selection will only be visible if expect to see at least one pair ‘separate’ by time 1.

Scaling limits: Small ‘neighbourhood size’:

Fix $u \in (0,1)$.

Set $n = 1/\mu$ and rescale: $w(nt, \sqrt{n}x)$.

Ability to detect selection depends on dimension:

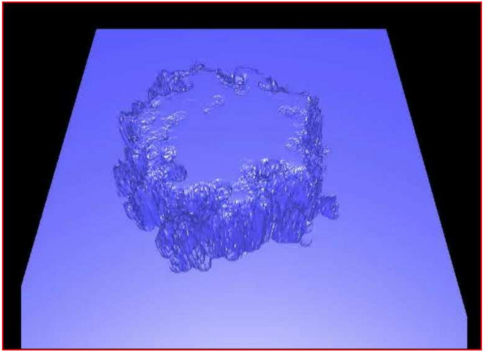
- $d = 1$, selection only visible if $s = \mathcal{O}(1/\sqrt{n}) = \mathcal{O}(\sqrt{\mu})$, *limiting ASG embedded in Brownian net*;
- $d = 2$, selection only visible if $s = \mathcal{O}(\log n/n) = \mathcal{O}(\mu|\log(\mu)|)$, *limiting ASG ‘Branching BM’*;
- $d \geq 3$, selection only visible if $s = \mathcal{O}(1/n) = \mathcal{O}(\mu)$, *limiting ASG Branching BM*.

Technical challenges because $ns_n \to \infty$.

E., Freeman, Penington, Straulino (2017)

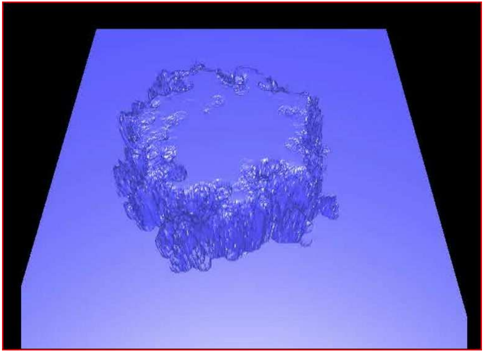
Spread of a favoured allele

Two types, a, A, relative fitnesses $1 : 1 + s$. If a reproduction event affects a region $B(x, r)$ in which current proportion of a-alleles is w, then probability offspring are type a is $\frac{w}{1+s(1-w))}$.



Spread of a favoured allele

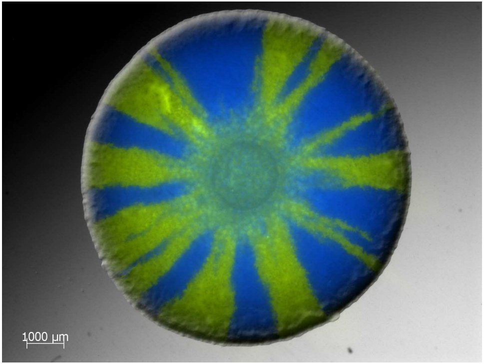
Two types, a, A, relative fitnesses $1 : 1 + s$. If a reproduction event affects a region $B(x, r)$ in which current proportion of a-alleles is w, then probability offspring are type a is $\frac{w}{1+s(1-w))}$.



Alternative interpretation: strong selection $\sim$ range expansion

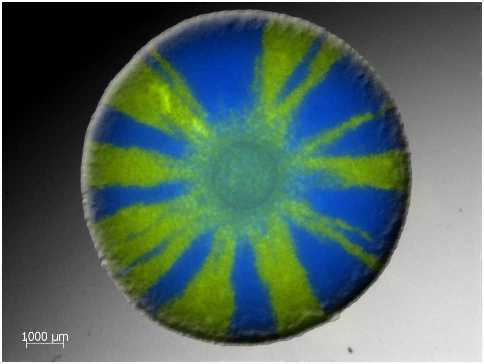
Range expansion

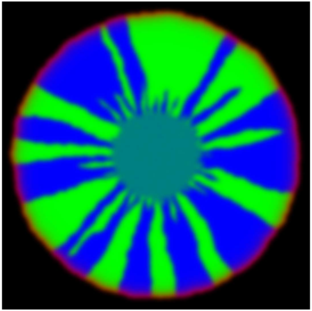
Pseudomanas aeruginosa (Kevin Foster)



Range expansion

Pseudomanas aeruginosa (Kevin Foster)





What's really happening?