Seed bank models in population genetics

Jochen Blath

Joint work with

Adrián González Casanova, Noemi Kurt,
Maite Wilke Berenguer, Eugenio Buzzoni, Jere Koskela...

CIRM Luminy, June 2018

Berlin Mathematical School

PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590

Dormancy and seed banks in population genetics

- *Dormancy* is an adaptive trait that has independently evolved multiple times across the tree of life.
- For example, many microorganisms can enter a reversible state of *vanishing metabolic activity*. These dormant forms may stay inactive for extended periods of time.
- Dormancy allows a population to maintain a reservoir of genotypic and phenotypic diversity (i.e., a *seed bank*) that can contribute to the long-term survival of a population. This "bet hedging" strategy has long been of interest to ecologists.
- However, in population genetics, comparatively little is known about how dormancy influences the *evolutionary forces* of genetic drift, mutation, selection, recombination, and gene flow. See [LENNON & SHOEMAKER 2018] for a recent systematic overview.

Dormancy and seed banks in population genetics

- *Dormancy* is an adaptive trait that has independently evolved multiple times across the tree of life.
- For example, many microorganisms can enter a reversible state of *vanishing metabolic activity*. These dormant forms may stay inactive for extended periods of time.
- Dormancy allows a population to maintain a reservoir of genotypic and phenotypic diversity (i.e., a *seed bank*) that can contribute to the long-term survival of a population. This “bet hedging” strategy has long been of interest to ecologists.
- However, in population genetics, comparatively little is known about how dormancy influences the *evolutionary forces* of genetic drift, mutation, selection, recombination, and gene flow. See [LENNON & SHOEMAKER 2018] for a recent systematic overview.

Dormancy and seed banks in population genetics

- *Dormancy* is an adaptive trait that has independently evolved multiple times across the tree of life.
- For example, many microorganisms can enter a reversible state of *vanishing metabolic activity*. These dormant forms may stay inactive for extended periods of time.
- Dormancy allows a population to maintain a reservoir of genotypic and phenotypic diversity (i.e., a *seed bank*) that can contribute to the long-term survival of a population. This “bet hedging” strategy has long been of interest to ecologists.
- However, in population genetics, comparatively little is known about how dormancy influences the *evolutionary forces* of genetic drift, mutation, selection, recombination, and gene flow. See [LENNON & SHOEMAKER 2018] for a recent systematic overview.

Dormancy and seed banks in population genetics

- *Dormancy* is an adaptive trait that has independently evolved multiple times across the tree of life.
- For example, many microorganisms can enter a reversible state of *vanishing metabolic activity*. These dormant forms may stay inactive for extended periods of time.
- Dormancy allows a population to maintain a reservoir of genotypic and phenotypic diversity (i.e., a *seed bank*) that can contribute to the long-term survival of a population. This “bet hedging” strategy has long been of interest to ecologists.
- However, in population genetics, comparatively little is known about how dormancy influences the *evolutionary forces* of genetic drift, mutation, selection, recombination, and gene flow. See [LENNON & SHOEMAKER 2018] for a recent systematic overview.

Dormancy and seed banks in population genetics

[Lennon & Jones, 2011] investigate dormancy in microbial communities:

- A large fraction of microorganisms in nature seems to be metabolically inactive.
- There is a variety of mechanisms for the initiation and termination of dormancy: Can be triggered by environmental cues (temperature, pH, resources, antibiotics treatment)...
- ... but also seems to happen spontaneously:

"*Spontaneous (individual) switching* vs. *responsive (simultaneous) switching*"

Dormancy and seed banks in population genetics

[LENNON & JONES, 2011] investigate dormancy in microbial communities:

- A large fraction of microorganisms in nature seems to be metabolically inactive.
- There is a variety of mechanisms for the initiation and termination of dormancy: Can be triggered by environmental cues (temperature, pH, resources, antibiotics treatment)...
- ... but also seems to happen spontaneously:

"*Spontaneous (individual) switching* vs. *responsive (simultaneous) switching*"

Dormancy and seed banks in population genetics

[Lennon & Jones, 2011] investigate dormancy in microbial communities:

- A large fraction of microorganisms in nature seems to be metabolically inactive.
- There is a variety of mechanisms for the initiation and termination of dormancy: Can be triggered by environmental cues (temperature, pH, resources, antibiotics treatment)...
- ... but also seems to happen spontaneously:

"*Spontaneous (individual) switching* vs. *responsive (simultaneous) switching*"

Dormancy and seed banks in population genetics

[Lennon & Jones, 2011] investigate dormancy in microbial communities:

- A large fraction of microorganisms in nature seems to be metabolically inactive.
- There is a variety of mechanisms for the initiation and termination of dormancy: Can be triggered by environmental cues (temperature, pH, resources, antibiotics treatment)...
- ... but also seems to happen spontaneously:

"*Spontaneous (individual) switching* vs. *responsive (simultaneous) switching*"

Dormancy in microbial communities

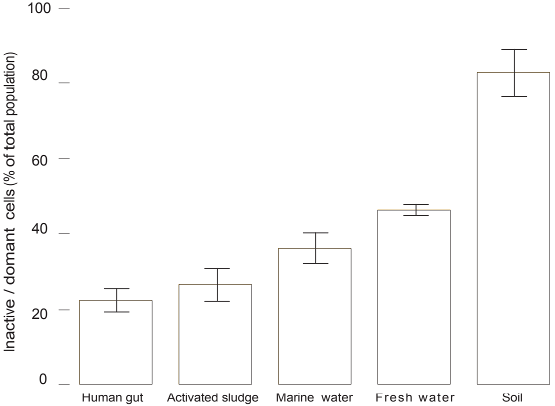


Figure: Percentage of inactive cells in microbial communities, data from [LENNON & JONES, 2011])

Mathematical modeling of seed-banks

Many (empirical) studies. Only relatively recently, systematic derivation of mathematical models. See e.g. (incomplete list)

- [KAJ, KRONE & LASCOUX 2001] (Wright-Fisher model with "weak seed bank"),
- [TELLIER, LAURENT, LAINER, PAVLIDIS, STEPHAN 2011] (inference of "seed bank parameters" in wild tomatoe species),
- [LENNON & JONES 2011] (overview and concepts for seed bank models in microbial communities),
- [BEGCKW 2015, BGCKW 2016,...] ("seed bank coalescent", "seed bank diffusion", some properties and evolutionary parameters),
- [KOOPMANN, MÜLLER, TELLIER, ŽIVKOVIĆ 2017] ("weak" seed banks and selection),
- [DEN HOLLANDER & PEDERZANI 2017] (strong seed bank models on torus, IBD),
- [SHOEMAKER & LENNON 2018] (systematic discussion of existing models and interplay with evolutionary forces).

Mathematical modeling of seed-banks

Many (empirical) studies. Only relatively recently, systematic derivation of mathematical models. See e.g. (incomplete list)

- [KAJ, KRONE & LASCOUX 2001] (Wright-Fisher model with “weak seed bank”),
- [TELLIER, LAURENT, LAINER, PAVLIDIS, STEPHAN 2011] (inference of “seed bank parameters” in wild tomatoe species),
- [LENNON & JONES 2011] (overview and concepts for seed bank models in microbial communities),
- [BEGCKW 2015, BGCKW 2016,...] (“seed bank coalescent”, “seed bank diffusion”, some properties and evolutionary parameters),
- [KOOPMANN, MÜLLER, TELLIER, ŽIVKOVIĆ 2017] (“weak” seed banks and selection),
- [DEN HOLLANDER & PEDERZANI 2017] (strong seed bank models on torus, IBD),
- [SHOEMAKER & LENNON 2018] (systematic discussion of existing models and interplay with evolutionary forces).

Aims of this talk

- Review of known seed bank models in population genetics including "*seed bank coalescent / seed bank diffusion*" (introduced in previous Luminy talk in 2015), based on "spontaneous switching".
- Report progress in understanding of *boundary behaviour* of seed bank diffusion and the relation between "weak" and "strong" seed bank models via a *stochastic delay equation*.
- Separation of timescales: Provide an application of the *method of duality* to identify *degenerate diffusion limits*.
- *Statistical method* to select seed bank model based on IMS data.
- New seed bank model incorporating *responsive switching*, further extending the class of "*on/off coalescents*".

Aims of this talk

- Review of known seed bank models in population genetics including "*seed bank coalescent / seed bank diffusion*" (introduced in previous Luminy talk in 2015), based on "spontaneous switching".
- Report progress in understanding of *boundary behaviour* of seed bank diffusion and the relation between "weak" and "strong" seed bank models via a *stochastic delay equation*.
- Separation of timescales: Provide an application of the *method of duality* to identify *degenerate diffusion limits*.
- *Statistical method* to select seed bank model based on IMS data.
- New seed bank model incorporating *responsive switching*, further extending the class of "*on/off coalescents*".

Aims of this talk

- Review of known seed bank models in population genetics including "*seed bank coalescent / seed bank diffusion*" (introduced in previous Luminy talk in 2015), based on "spontaneous switching".
- Report progress in understanding of *boundary behaviour* of seed bank diffusion and the relation between "weak" and "strong" seed bank models via a *stochastic delay equation*.
- Separation of timescales: Provide an application of the *method of duality* to identify *degenerate diffusion limits*.
- *Statistical method* to select seed bank model based on IMS data.
- New seed bank model incorporating *responsive switching*, further extending the class of "*on/off coalescents*".

Aims of this talk

- Review of known seed bank models in population genetics including "*seed bank coalescent / seed bank diffusion*" (introduced in previous Luminy talk in 2015), based on "spontaneous switching".
- Report progress in understanding of *boundary behaviour* of seed bank diffusion and the relation between "weak" and "strong" seed bank models via a *stochastic delay equation*.
- Separation of timescales: Provide an application of the *method of duality* to identify *degenerate diffusion limits*.
- *Statistical method* to select seed bank model based on IMS data.
- New seed bank model incorporating *responsive switching*, further extending the class of "*on/off coalescents*".

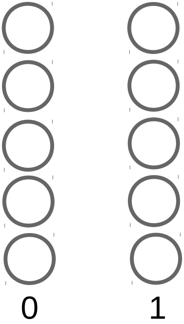
Aims of this talk

- Review of known seed bank models in population genetics including "*seed bank coalescent / seed bank diffusion*" (introduced in previous Luminy talk in 2015), based on "spontaneous switching".
- Report progress in understanding of *boundary behaviour* of seed bank diffusion and the relation between "weak" and "strong" seed bank models via a *stochastic delay equation*.
- Separation of timescales: Provide an application of the *method of duality* to identify *degenerate diffusion limits*.
- *Statistical method* to select seed bank model based on IMS data.
- New seed bank model incorporating *responsive switching*, further extending the class of "*on/off coalescents*".

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

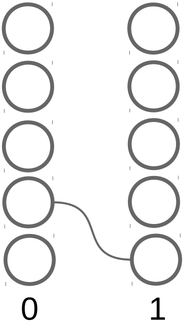


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

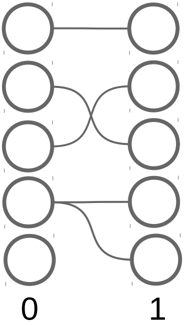


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random “multinomial sampling”

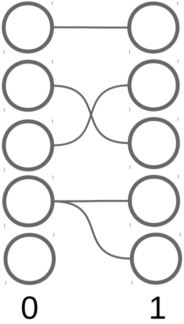


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

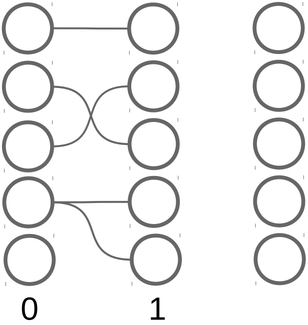


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N (= 5)$

Choose 1 parent uniformly at random "multinomial sampling"

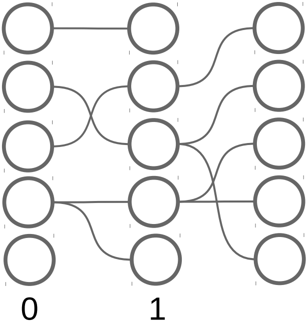


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N (= 5)$

Choose 1 parent uniformly at random "multinomial sampling"

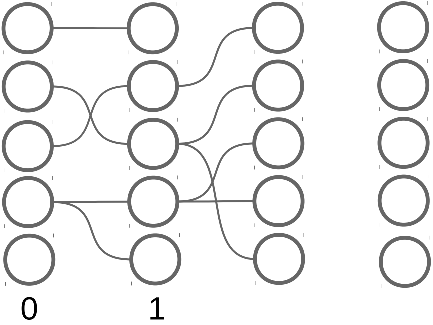


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random “multinomial sampling”

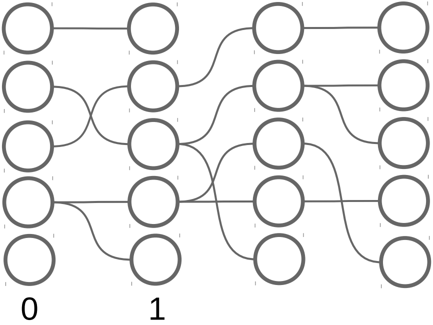


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

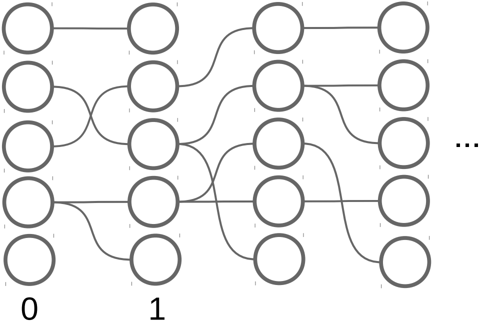


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N (= 5)$

Choose 1 parent uniformly at random   “multinomial sampling”

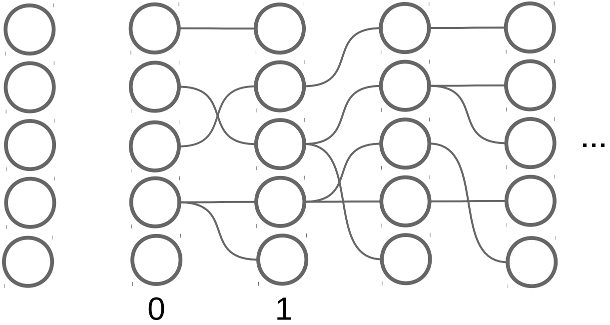


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random “multinomial sampling”

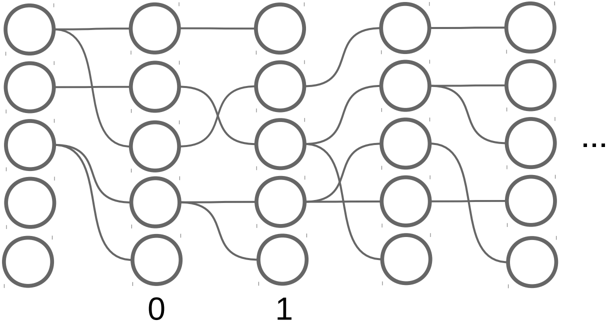


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

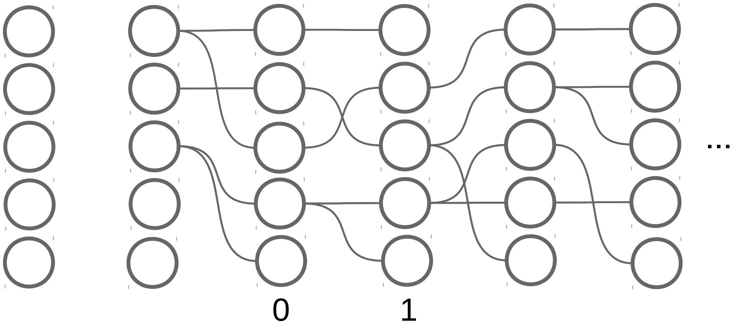


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N (= 5)$

Choose 1 parent uniformly at random "multinomial sampling"

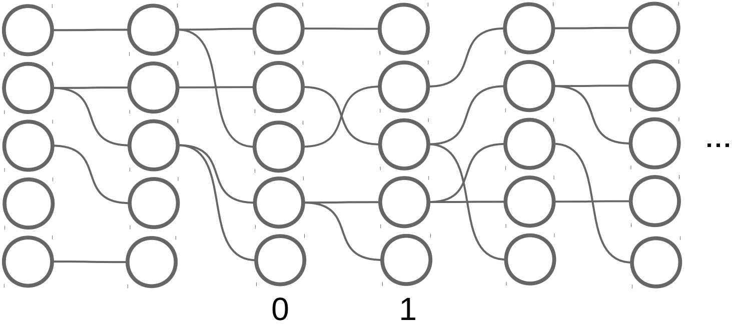


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(=5)$

Choose 1 parent uniformly at random "multinomial sampling"

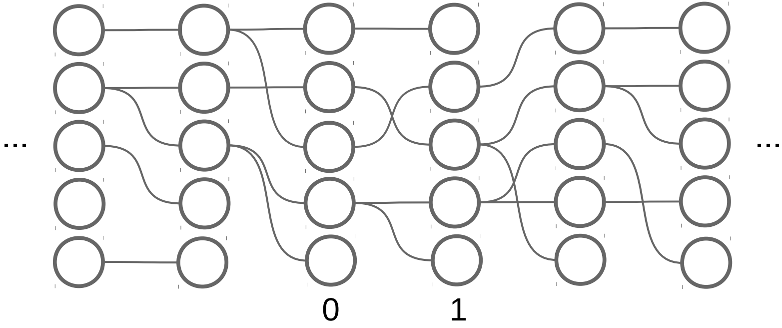


Ancestral graph

Background: Classical Wright-Fisher model

Set-up: haploid population, discrete generations, fixed population size $N(= 5)$

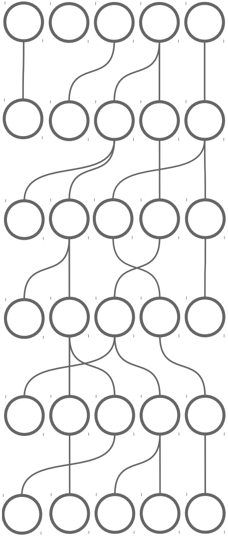
Choose 1 parent uniformly at random "multinomial sampling"



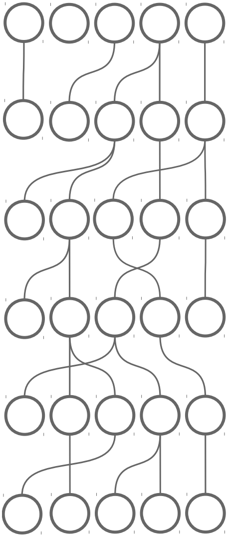
Ancestral graph

Forward – time – Backward

Types: {●, ○} → **Fraction of ●**



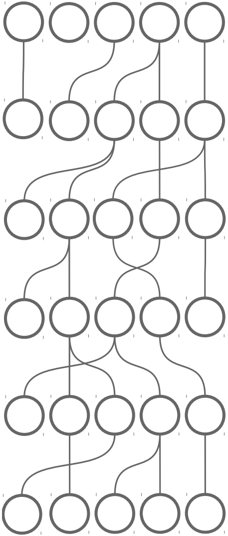
Genealogy of sample

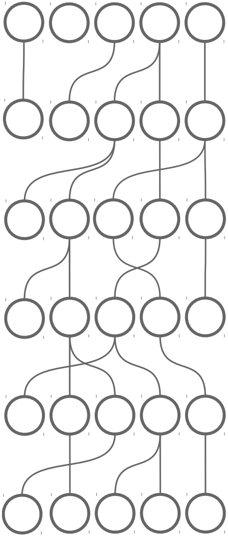


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**

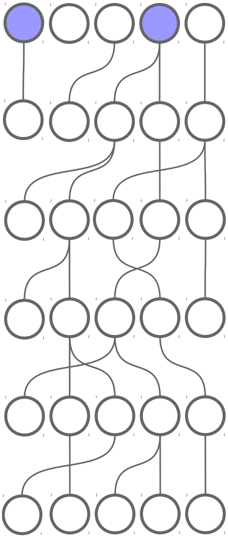
Genealogy of sample



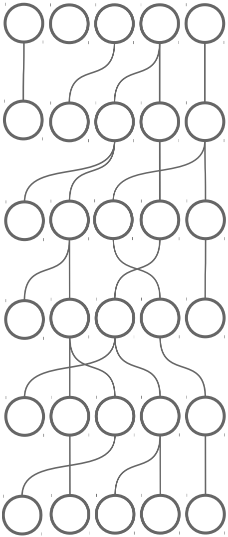


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**

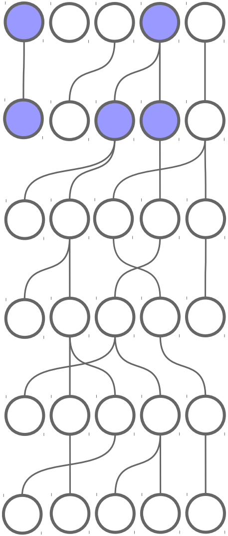


Genealogy of sample

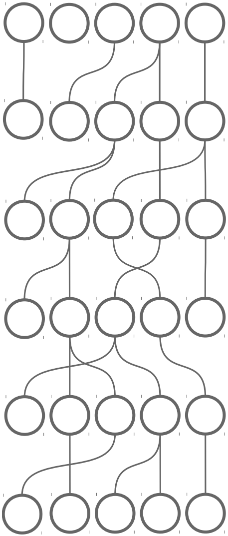


Forward – time – Backward

Types: $\{$●, ○$\}$ → **Fraction of ●**

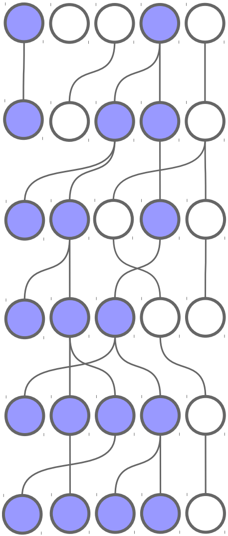


Genealogy of sample

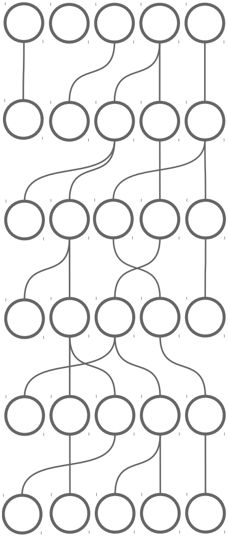


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**

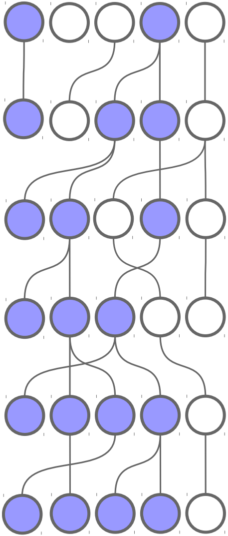


Genealogy of sample

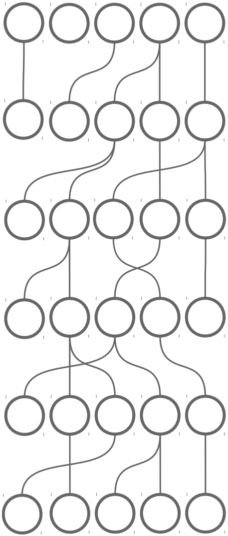


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**



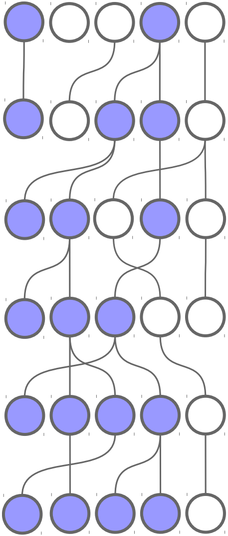
Genealogy of sample

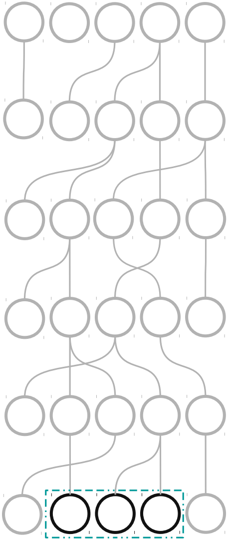


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**

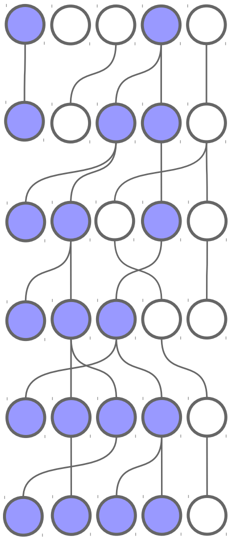
Genealogy of sample



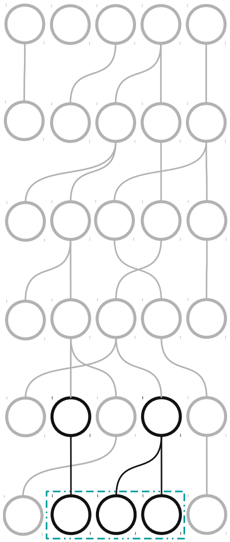


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**

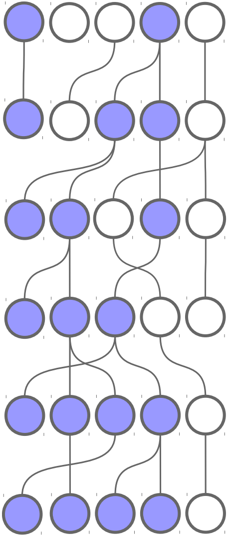


Genealogy of sample

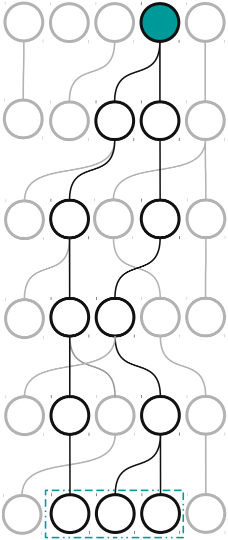


Forward – time – Backward

Types: {●, ○} → **Fraction of ●**



Genealogy of sample



Forward – time – Backward

Fraction X^N of ●

$$X_r^N := \frac{1}{N}\sum_{i\in[\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is } \bullet\}}$$

Theorem 1.1 (Wright '31, Feller '51,...)

$(X^N_{\lfloor Nt \rfloor})_{t\geq 0} \Rightarrow (X_t)_{t\geq 0}$
for $N \to \infty$ on $D_{[0,\infty)}([0,1])$, where

$$dX_t = \sqrt{X_t(1-X_t)}dB_t,$$

Wright-Fisher diffusion

Genealogy Π^N of sample

Theorem 1.2 (Kingman '82)

$(\Pi^{(N,k)}_{\lfloor Nt \rfloor})_{t\geq 0} \Rightarrow (\Pi^k_t)_{t\geq 0}$

Forward – time – Backward

Fraction X^N of ●

$$X^N_r := \frac{1}{N} \sum_{i \in [\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is } ●\}}$$

Theorem 1.1 (Wright '31, Feller '51,...)

$(X^N_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (X_t)_{t \geq 0}$
for $N \to \infty$ on $D_{[0,\infty)}([0,1])$, where

$$dX_t = \sqrt{X_t(1 - X_t)}\, dB_t,$$

Wright-Fisher diffusion

Genealogy Π^N of sample

Theorem 1.2 (Kingman '82)

$(\Pi^{(N,k)}_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (\Pi^k_t)_{t \geq 0}$

Forward – time – Backward

Fraction X^N of ●

$$X_r^N := \frac{1}{N} \sum_{i \in [\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is } \bullet\}}$$

Theorem 1.1 (Wright '31, Feller '51,...)

$(X^N_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (X_t)_{t \geq 0}$
for $N \to \infty$ *on* $D_{[0,\infty)}([0,1])$, *where*

$$dX_t = \sqrt{X_t(1 - X_t)}\,dB_t,$$

Wright-Fisher diffusion

Genealogy Π^N of sample

Theorem 1.2 (Kingman '82)

$$(\Pi^{(N,k)}_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (\Pi^k_t)_{t \geq 0}$$

Forward – time – Backward

Fraction X^N of ●

$$X_r^N := \frac{1}{N}\sum_{i\in[\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is } \bullet\}}$$

Theorem 1.1 (Wright '31, Feller '51,...)

$(X^N_{\lfloor Nt\rfloor})_{t\geq 0} \Rightarrow (X_t)_{t\geq 0}$
for $N \to \infty$ on $D_{[0,\infty)}([0,1])$, where

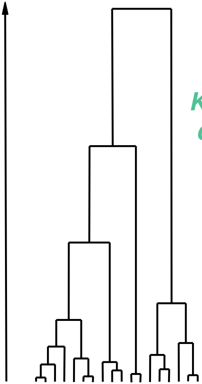
$$dX_t = \sqrt{X_t(1-X_t)}\,dB_t,$$

Wright-Fisher diffusion

Genealogy Π^N of sample

Theorem 1.2 (Kingman '82)

$(\Pi^{(N,k)}_{\lfloor Nt\rfloor})_{t\geq 0} \Rightarrow (\Pi^k_t)_{t\geq 0}$



Kingman coalescent

Moment Duality

Fraction $\{X_t\}$ of ● Wright-Fisher diffusion,

Genealogy $\{\Pi_t\}$ of sample Kingman coalescent,

Block counting process $\{N_t\}$, given by $N_t = |\Pi_t|, t \geq 0$.

Theorem 1.3 (Moment Duality)

$$\mathbb{E}^n[x^{N(t)}] = \mathbb{E}_x[X(t)^n].$$

Here $H(x,n) := x^n$ is the “probability to sample n ● individuals if fraction is x”

Moment Duality

Fraction $\{X_t\}$ of ● Wright-Fisher diffusion,

Genealogy $\{\Pi_t\}$ of sample Kingman coalescent,

Block counting process $\{N_t\}$, given by $N_t = |\Pi_t|, t \geq 0$.

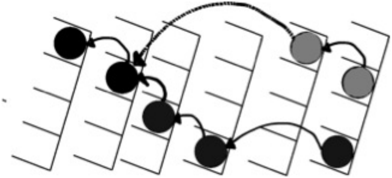
Theorem 1.3 (Moment Duality)

$$\mathbb{E}^n[x^{N(t)}] = \mathbb{E}_x[X(t)^n].$$

Here $H(x,n) := x^n$ *is the "probability to sample* n ● *individuals if fraction is* x*"*

Seed bank models with ancestral jumps

[Kaj, Krone & Lascoux 2001]: One idea to introduce seed banks into a Wright-Fisher model is to allow individuals to pick parents from the $m \in \mathbb{N}$ *past generations* (and not only from previous generation):



Picture: Pop size $N = 4$. Generations $\overset{\uparrow}{-2}$ $\overset{\uparrow}{-1}$ $\overset{\uparrow}{0}$

Known results, I

- [Kaj, Krone & Lascoux 2001] show that if each individual independently picks a random $B \in \{1, \ldots, m\}$, and then its parent from B generations ago, then the ancestral process converges under the usual rescaling to a *time-changed Kingman coalescent*, where coalescence rates are multiplied by
$$\frac{1}{\mathbb{E}[B]^2} \le 1.$$
- This *decelerates* the coalescent, leading to an increase in the effective population size. Note that B and m do not scale with N.
- Since the overall coalescent tree structure is retained, this leaves the normalized site frequency spectrum (nSFS) in the IMS model unchanged. Hence, we speak of a *'weak' seed bank effect*.

Known results, I

- [KAJ, KRONE & LASCOUX 2001] show that if each individual independently picks a random $B \in \{1, \ldots, m\}$, and then its parent from B generations ago, then the ancestral process converges under the usual rescaling to a *time-changed Kingman coalescent*, where coalescence rates are multiplied by
$$\frac{1}{\mathbb{E}[B]^2} \leq 1.$$
- This *decelerates* the coalescent, leading to an increase in the effective population size. Note that B and m do not scale with N.
- Since the overall coalescent tree structure is retained, this leaves the normalized site frequency spectrum (nSFS) in the IMS model unchanged. Hence, we speak of a *'weak' seed bank effect*.

Known results, I

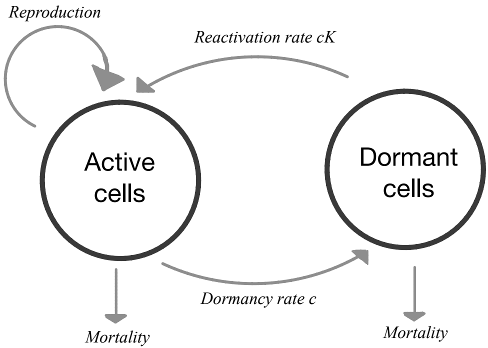
- [Kaj, Krone & Lascoux 2001] show that if each individual independently picks a random $B \in \{1, \ldots, m\}$, and then its parent from B generations ago, then the ancestral process converges under the usual rescaling to a *time-changed Kingman coalescent*, where coalescence rates are multiplied by

$$\frac{1}{\mathbb{E}[B]^2} \leq 1.$$

- This *decelerates* the coalescent, leading to an increase in the effective population size. Note that B and m do not scale with N.
- Since the overall coalescent tree structure is retained, this leaves the normalized site frequency spectrum (nSFS) in the IMS model unchanged. Hence, we speak of a *'weak' seed bank effect*.

Seed banks in bacterial communities

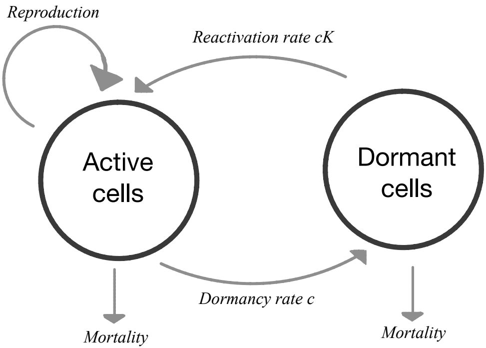
Dormancy: Initiaton and resuscitation à la [Lennon & Jones 2011]



Spontaneous switching - *migration* between active and dormant population!

Seed banks in bacterial communities

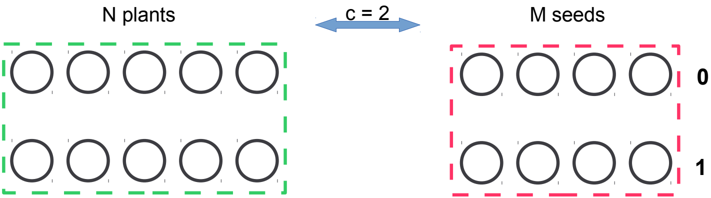
Dormancy: Initiaton and resuscitation à la [Lennon & Jones 2011]



Spontaneous switching - *migration* between active and dormant population!

A Wrigth-Fisher model with seed bank

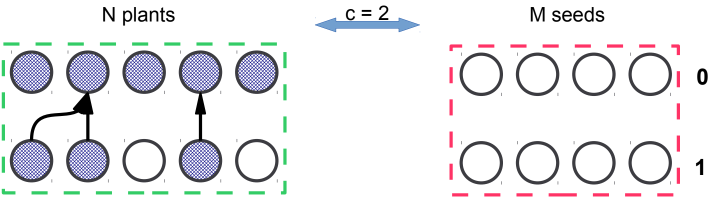
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

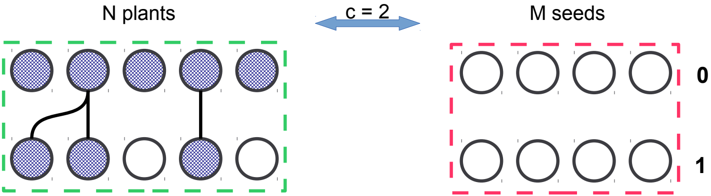
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

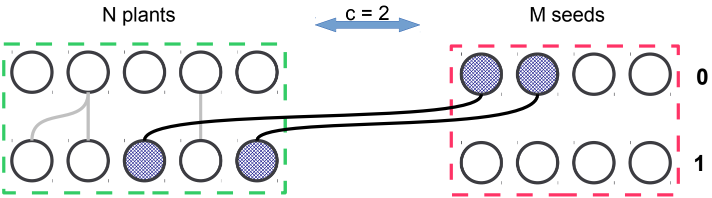
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

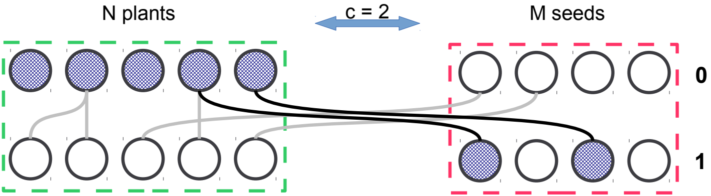
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

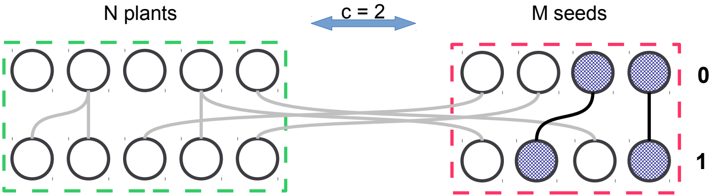
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

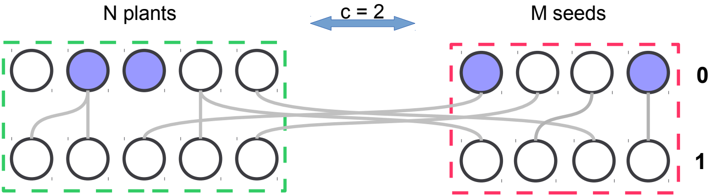
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

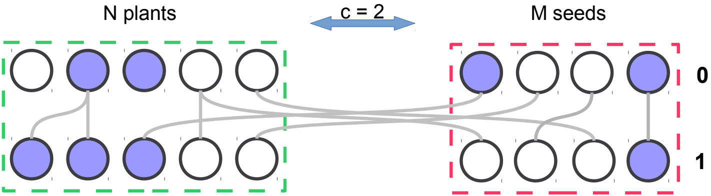
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

A Wrigth-Fisher model with seed bank

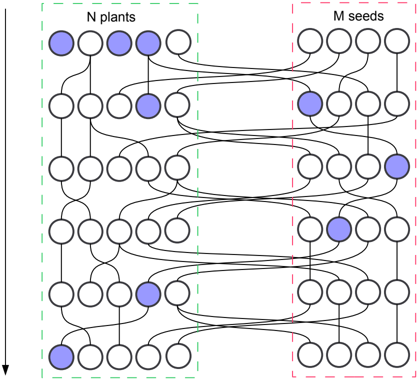
Set-up: WF-type model, active (size N) and dormant (size M) subpopulations



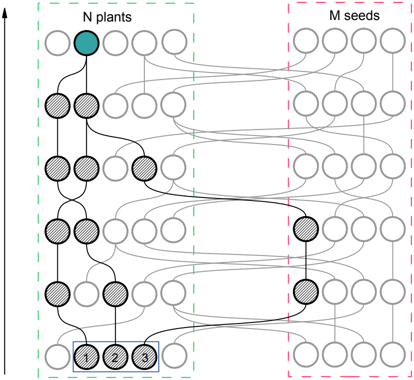
- N plants produce $N - c$ plants by multinomial sampling
- c seeds are selected uniformly to germinate: one plant each
- N plants produce c seeds by multinomial sampling
- $N - c$ seeds stay in the seed bank
- Offspring inherit the type of their parents

Forward and backward in time

Forward: trace fractions of purple type



Backward: trace sample to MRCA



Scaling limits

Forward: trace fractions of purple type

$$X_r^N := \frac{1}{N} \sum_{i \in [\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is purple}\}}$$

$$Y_r^M := \frac{1}{M} \sum_{j \in [\![M]\!]} \mathbf{1}_{\{j \text{ in generation } r \text{ is purple}\}}$$

Theorem 1.4 (BGKW '15)

$c, K > 0$, $M = M(N) = \lfloor KN \rfloor$

$$(X^N_{\lfloor Nt \rfloor}, Y^{M(N)}_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (X_t, Y_t)_{t \geq 0}$$

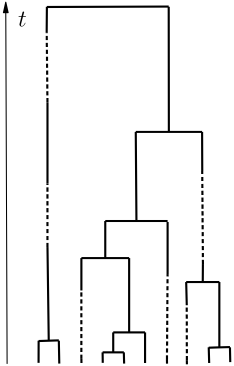
for $N \to \infty$ *on* $D_{[0,\infty)}([0,1]^2)$, *where*

$$\begin{cases} dX_t &= c(Y_t - X_t)dt + \sqrt{X_t(1-X_t)}dB_t, \\ dY_t &= cK(X_t - Y_t)dt. \end{cases}$$

Backward: trace sample to MRCA

Theorem 1.5 (BGKW '15)

$$(\Pi^{(N,k)}_{\lfloor Nt \rfloor})_{t \geq 0} \Rightarrow (\Pi^k_t)_{t \geq 0}$$



Adding bi-allelic mutation

Seed bank diffusion

$$\begin{cases} dX_t &= c(Y_t - X_t)dt + \sqrt{X_t(1-X_t)}dB_t, \\ dY_t &= cK(X_t - Y_t)dt. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[KERMANY, ZHOU & HICKEY 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Adding bi-allelic mutation

Seed bank diffusion with mutation

$$\begin{cases} \mathrm{d}X_t &= \left(-u_1 X_t + u_2(1-X_t) + c(Y_t - X_t)\right)\mathrm{d}t + \sqrt{X_t(1-X_t)}\,\mathrm{d}B_t, \\ \mathrm{d}Y_t &= \left(-u_1' Y_t + u_2'(1-Y_t) + cK(X_t - Y_t)\right)\mathrm{d}t. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[Kermany, Zhou & Hickey 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Adding bi-allelic mutation

"Two-island diffusion"

$$\begin{cases} \mathrm{d}X_t &= \left(-u_1 X_t + u_2(1-X_t) + c(Y_t - X_t)\right)\mathrm{d}t + \alpha\sqrt{X_t(1-X_t)}\mathrm{d}B_t, \\ \mathrm{d}Y_t &= \left(-u_1' Y_t + u_2'(1-Y_t) + c'(X_t - Y_t)\right)\mathrm{d}t + \alpha'\sqrt{Y_t(1-Y_t)}\mathrm{d}B_t'. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[Kermany, Zhou & Hickey 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Adding bi-allelic mutation

$$\begin{cases} \mathrm{d}X_t &= \left(-u_1 X_t + u_2(1-X_t) + c(Y_t - X_t)\right)\mathrm{d}t + \alpha\sqrt{X_t(1-X_t)}\mathrm{d}B_t, \\ \mathrm{d}Y_t &= \left(-u_1' Y_t + u_2'(1-Y_t) + c'(X_t - Y_t)\right)\mathrm{d}t + \alpha'\sqrt{Y_t(1-Y_t)}\mathrm{d}B_t'. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[Kermany, Zhou & Hickey 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Adding bi-allelic mutation

$$\begin{cases} \mathrm{d}X_t &= \left(-u_1 X_t + u_2(1-X_t) + c(Y_t - X_t)\right)\mathrm{d}t + \alpha\sqrt{X_t(1-X_t)}\mathrm{d}B_t, \\ \mathrm{d}Y_t &= \left(-u_1' Y_t + u_2'(1-Y_t) + c'(X_t - Y_t)\right)\mathrm{d}t + \alpha'\sqrt{Y_t(1-Y_t)}\mathrm{d}B_t'. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[Kermany, Zhou & Hickey 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Adding bi-allelic mutation

$$\begin{cases} \mathrm{d}X_t &= \left(-u_1 X_t + u_2(1-X_t) + c(Y_t - X_t)\right)\mathrm{d}t + \alpha\sqrt{X_t(1-X_t)}\mathrm{d}B_t, \\ \mathrm{d}Y_t &= \left(-u_1' Y_t + u_2'(1-Y_t) + c'(X_t - Y_t)\right)\mathrm{d}t + \alpha'\sqrt{Y_t(1-Y_t)}\mathrm{d}B_t'. \end{cases} \tag{1}$$

u_1, u_1': purple → white

u_2, u_2': white → purple

Seed bank: $\alpha' = 0$, $c' = cK$,

Two-island: $\alpha' > 0$

We are now interested in the *boundary behaviour* of these two-dim diffusions.

[KERMANY, ZHOU & HICKEY 2008] investigate the two island diffusion model, but whether boundaries are hit remained open.

Boundary behaviour, known results

For the one-dimensional system (without seed bank),

$$\mathrm{d}X_t = \left[-u_1 X(t) + u_2(1 - X(t))\right]\mathrm{d}t + \sqrt{X_t(1 - X_t)}\,\mathrm{d}B_t, \tag{2}$$

classical *Feller boundary theory* (based on speed masure, scale funtion) shows that the boundary 1 (resp. 0) is hit *iff* $2u_1 < 1$ (resp. $2u_2 < 1$).

However, this method does not work in higher dimensions.

Comparison to the one-dim system (choosing $Y_t \equiv 0$ resp. $Y_t \equiv 1$) leaves open what happens in intervals of type $(u_1, u_1 + c)$.

Boundary behaviour, known results

For the one-dimensional system (without seed bank),

$$\mathrm{d}X_t = \left[-u_1X(t) + u_2(1-X(t))\right]\mathrm{d}t + \sqrt{X_t(1-X_t)}\mathrm{d}B_t, \tag{2}$$

classical *Feller boundary theory* (based on speed masure, scale funtion) shows that the boundary 1 (resp. 0) is hit *iff* $2u_1 < 1$ (resp. $2u_2 < 1$).

However, this method does not work in higher dimensions.

Comparison to the one-dim system (choosing $Y_t \equiv 0$ resp. $Y_t \equiv 1$) leaves open what happens in intervals of type $(u_1, u_1 + c)$.

Boundary behaviour, known results

For the one-dimensional system (without seed bank),

$$\mathrm{d}X_t = \left[- u_1 X(t) + u_2(1 - X(t))\right]\mathrm{d}t + \sqrt{X_t(1 - X_t)}\,\mathrm{d}B_t, \tag{2}$$

classical *Feller boundary theory* (based on speed masure, scale funtion) shows that the boundary 1 (resp. 0) is hit *iff* $2u_1 < 1$ (resp. $2u_2 < 1$).

However, this method does not work in higher dimensions.

Comparison to the one-dim system (choosing $Y_t \equiv 0$ resp. $Y_t \equiv 1$) leaves open what happens in intervals of type $(u_1, u_1 + c)$.

Boundary classification for the general model

Define

$$\tau_0^X := \inf\{t \geq 0 \mid X(t) = 0\} \qquad \text{(and likewise } \tau_1^X, \tau_0^Y, \tau_1^Y.\text{)}$$

Theorem 2.1 (BBGW17+)

Let $c, c' > 0$. For every initial distribution $\mu_0 = \mathcal{L}(X(0), Y(0))$ with $\mu_0((0,1)^2) = 1$

$$\mathbb{P}^{\mu_0}\left(\tau_0^X < \infty\right) = 0 \qquad \textit{iff} \qquad 2u_2 \geq \alpha^2,$$

i.e. X will never hit 0 from the interior if and only if $2u_2 \geq \alpha^2$.

Analogous results hold for all other boundaries.

Boundary classification for the general model

Define

$$\tau_0^X := \inf\{t \geq 0 \mid X(t) = 0\} \qquad \text{(and likewise } \tau_1^X, \tau_0^Y, \tau_1^Y.\text{)}$$

Theorem 2.1 (BBGW17+)

Let $c, c' > 0$. For every initial distribution $\mu_0 = \mathcal{L}(X(0), Y(0))$ with $\mu_0((0,1)^2) = 1$

$$\mathbb{P}^{\mu_0}\left(\tau_0^X < \infty\right) = 0 \qquad \textit{iff} \qquad 2u_2 \geq \alpha^2,$$

i.e. X will never hit 0 from the interior if and only if $2u_2 \geq \alpha^2$.

Analogous results hold for all other boundaries.

Boundary classification for the general model

Define

$$\tau_0^X := \inf\{t \geq 0 \mid X(t) = 0\} \qquad \text{(and likewise } \tau_1^X,\ \tau_0^Y,\ \tau_1^Y\text{.)}$$

Theorem 2.1 (BBGW17+)

Let $c, c' > 0$. *For every initial distribution* $\mu_0 = \mathcal{L}(X(0), Y(0))$ *with* $\mu_0((0,1)^2) = 1$

$$\mathbb{P}^{\mu_0}\left(\tau_0^X < \infty\right) = 0 \qquad \textit{iff} \qquad 2u_2 \geq \alpha^2,$$

i.e. X *will never hit 0 from the interior if and only if* $2u_2 \geq \alpha^2$.

Analogous results hold for all other boundaries.

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [FILIPOVIC & LARSSON 2016] and [LARSSON & PULIDO 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0, t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [FILIPOVIC & LARSSON 2016] and [LARSSON & PULIDO 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0,t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [FILIPOVIC & LARSSON 2016] and [LARSSON & PULIDO 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0, t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [FILIPOVIC & LARSSON 2016] and [LARSSON & PULIDO 2017]).

Strategy of proof:
"$\Rightarrow$" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"$\Leftarrow$" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0, t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [Filipovic & Larsson 2016] and [Larsson & Pulido 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0,t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s. "*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [Filipovic & Larsson 2016] and [Larsson & Pulido 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0,t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [Filipovic & Larsson 2016] and [Larsson & Pulido 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0,t]$.

Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.

"*McKean's argument*"

Boundary classification for the general model

The proof relies on recent results for *polynomial diffusions* (see [Filipovic & Larsson 2016] and [Larsson & Pulido 2017]).

Strategy of proof:
"⇒" Filipović and Larsson *Polynomial diffusions and applications in finance (2016)*

"⇐" Let $\mathcal{P} := \{x, 1-x, y, 1-y\}$, then $[0,1]^2 = \{(x,y) \in \mathbb{R}^2 \mid \forall p \in \mathcal{P} : p(x,y) \geq 0\}$.

Pick $p(x,y) = x$. Check via Ito's formula that for $u_2 \geq \frac{1}{2}\alpha^2$

$$M(s) := \log p(X(s), Y(s)) + \kappa_p s$$

is local submartingale on $[0, \tau_0^X[$ for some suitable constant $\kappa_p \geq 0$,

bounded from above on bounded intervals $[0, \tau_0^X[\cap [0,t]$.

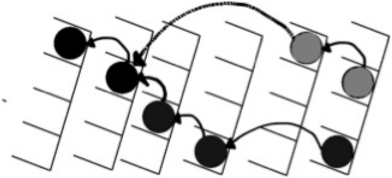
Convergence Theorem: For every fixed t, we have $\lim_{s \to \tau_0^X} M(s \wedge t) \in \mathbb{R}$.

But there is explosion at τ_0^X, hence $\tau_0^X > t$ for every t and thus $\tau_0^X = \infty$ a.s.
"*McKean's argument*"

Weak vs strong seed banks: Delay representation

In the model of [KAJ, KRONE & LASCOUX 2001], there was no explicit seed bank (in form of an "island").

Instead, types of individuals have been directly chosen from the past "active" population:

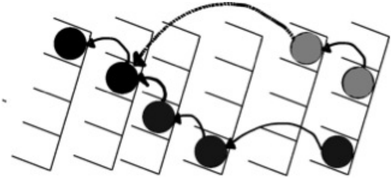


Perhaps surprisingly, there is a similar interpretation for the seed bank diffusion model.

Weak vs strong seed banks: Delay representation

In the model of [Kaj, Krone & Lascoux 2001], there was no explicit seed bank (in form of an “island”).

Instead, types of individuals have been directly chosen from the past “active” population:

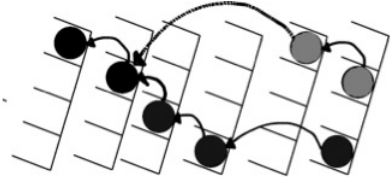


Perhaps surprisingly, there is a similar interpretation for the seed bank diffusion model.

Weak vs strong seed banks: Delay representation

In the model of [Kaj, Krone & Lascoux 2001], there was no explicit seed bank (in form of an “island”).

Instead, types of individuals have been directly chosen from the past “active” population:



Perhaps surprisingly, there is a similar interpretation for the seed bank diffusion model.

Weak vs strong seed banks

An application of the integration by parts rule for semimartingales yields

Proposition 2.2 (BBGW18)

The seed bank diffusion solving (2) *with* $c = K = 1$*, started in* $X_0 = x = y = Y_0$*, agrees with the unique strong solution of the stochastic delay differential equations*

$$\begin{aligned} dX(t) &= \Big(xe^{-t} + \int_0^t e^{-(t-s)} X(s)\,ds - X(t)\Big)dt + \sqrt{X(t)(1-X(t))}\,dB(t), \\ dY(t) &= \Big(-ye^{-t} - \int_0^t e^{-(t-s)} X(s)\,ds + X(t)\Big)dt. \end{aligned} \tag{3}$$

Interpretation: The *type* of any “infinitesimal” resuscitated individual is determined by choosing an ancestor from the *active* pop an exp-distributed time ago (on the “evolutionary timescale”, with cutoff at 0).

Weak vs strong seed banks

An application of the integration by parts rule for semimartingales yields

Proposition 2.2 (BBGW18)

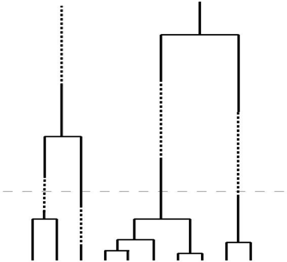
The seed bank diffusion solving (2) *with* $c = K = 1$*, started in* $X_0 = x = y = Y_0$*, agrees with the unique strong solution of the stochastic delay differential equations*

$$\begin{aligned} dX(t) &= \Big(xe^{-t} + \int_0^t e^{-(t-s)} X(s)\,ds - X(t)\Big)dt + \sqrt{X(t)(1-X(t))}\,dB(t), \\ dY(t) &= \Big(-ye^{-t} - \int_0^t e^{-(t-s)} X(s)\,ds + X(t)\Big)dt. \end{aligned} \tag{3}$$

Interpretation: The *type* of any “infinitesimal” resuscitated individual is determined by choosing an ancestor from the *active* pop an exp-distributed time ago (on the “evolutionary timescale”, with cutoff at 0).

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

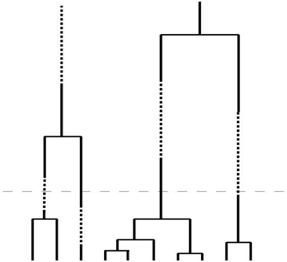


Transitions: If there are currently n active and m dormant lines, then

- Each pair of active lines merges at rate 1 (overall rate $\binom{n}{2}$)
- Each active line becomes dormant with rate c (overall rate cn)
- Each dormant lines becomes active with rate cK (overall rate cKm)

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

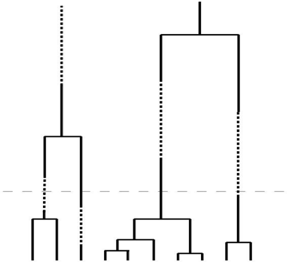


Transitions: If there are currently n active and m dormant lines, then

- Each pair of active lines merges at rate 1 (overall rate $\binom{n}{2}$)
- Each active line becomes dormant with rate c (overall rate cn)
- Each dormant lines becomes active with rate cK (overall rate cKm)

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

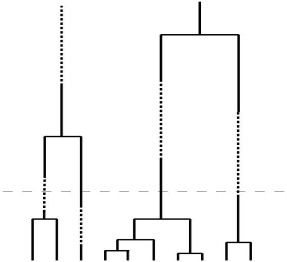


Transitions: If there are currently n active and m dormant lines, then

- Each pair of active lines merges at rate 1 (overall rate $\binom{n}{2}$)
- Each active line becomes dormant with rate c (overall rate cn)
- Each dormant lines becomes active with rate cK (overall rate cKm)

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

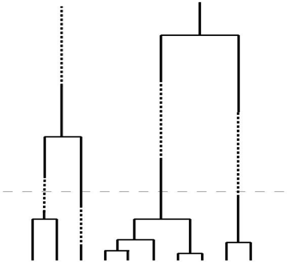


Transitions: If there are currently n active and m dormant lines, then

- Each pair of active lines merges at rate 1 (overall rate $\binom{n}{2}$)
- Each active line becomes dormant with rate c (overall rate cn)
- Each dormant lines becomes active with rate cK (overall rate cKm)

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

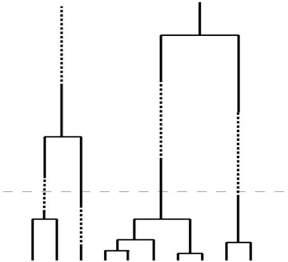


Its *block-counting process* (N_t, M_t) (active lines, dormant lines) satisfies the classical moment (sampling) duality with the seed bank diffusion (X_t, Y_t):

$$\mathbb{E}_{x,y}[X_t^n, Y_t^m] = \mathbb{E}^{n,m}[x^{N_t} y^{M_t}], \quad t \geq 0.$$

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

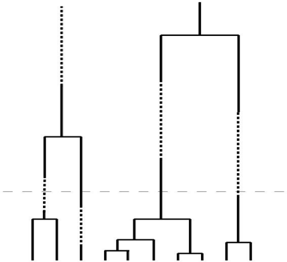


Its *block-counting process* (N_t, M_t) (active lines, dormant lines) satisfies the classical moment (sampling) duality with the seed bank diffusion (X_t, Y_t):

$$\mathbb{E}_{x,y}[X_t^n, Y_t^m] = \mathbb{E}^{n,m}[x^{N_t} y^{M_t}], \quad t \geq 0.$$

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

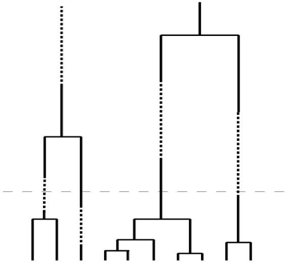


Some properties:

- Does not come down from infinity.
- Elevated T_{MRCA}: $\mathbb{E}[T(n)] \asymp \log\log n$.

The seed bank coalescent

In the *seed bank coalescent* $\Pi_t, t \geq 0$, lines can be "*active*" (black lines) or "*dormant*" (dotted lines).

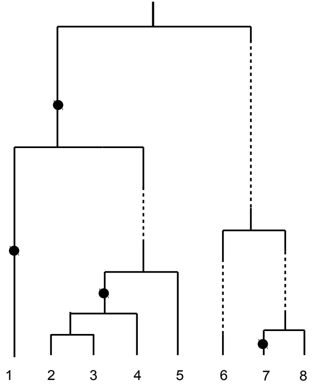


Some properties:

- Does not come down from infinity.
- Elevated T_{MRCA}: $\mathbb{E}[T(n)] \asymp \log\log n$.

The seed bank coalescent with mutation

Mutations can be distributed on active branches, say with rate $u_1/2$.



Modeling choice whether mutations on dormant branches should be possible (perhaps at reduced rate $u_2/2$).

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, “weak seed bank effect”’.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, “strong seed bank effect”.
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson’s estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, "weak seed bank effect"'.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, "strong seed bank effect".
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson's estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, “weak seed bank effect”'.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, “strong seed bank effect”.
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson’s estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, “weak seed bank effect”'.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, “strong seed bank effect”.
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson’s estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, “weak seed bank effect”’.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, “strong seed bank effect”.
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson’s estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

So far, we have encountered the following models:

- `K` *Kingman coalescent*. The standard model of genetic ancestry in the absence of a seed bank.
- `W` *Delayed Kingman coalescent*. Arises in [KKL 2001] model, “weak seed bank effect”’.
- `S` *Seed bank coalescent*, dual to seed bank diffusion, “strong seed bank effect”.
- `TI` *Structured coalescent*, dual to two island diffusion.

Populations under `K` and `W` have same stationary distribution, tree topology identical. Estimate for mutation rates (Watterson’s estimate) will differ.

Populations under `S` and `TI` should produce unique patterns of variability.

Inference and model selection

- Under all models, classical quantities such as expected *total tree length* (segregating sites), *external branch lengths* (singletons), *nSFS* can easily be computed/simulated.
- According to [SHOEMAKER AND LENNON, 2018], seed banks should stratify a population. Indeed, *Wright's* F_{ST} is elevated under `S`, though less pronounced than for `TI` (with otherwise identical model parameters).
- *Identity by descent (IBD)* has been investigated for `S` on a torus by [DEN HOLLANDER, PEDERZANI 2017].
- *Full likelihoods* of data (say under IMS model) can in principle be computed (recursively or by simulation).

Inference and model selection

- Under all models, classical quantities such as expected *total tree length* (segregating sites), *external branch lengths* (singletons), *nSFS* can easily be computed/simulated.
- According to [SHOEMAKER AND LENNON, 2018], seed banks should stratify a population. Indeed, *Wright's* F_{ST} is elevated under `S`, though less pronounced than for `TI` (with otherwise identical model parameters).
- *Identity by descent (IBD)* has been investigated for `S` on a torus by [DEN HOLLANDER, PEDERZANI 2017].
- *Full likelihoods* of data (say under IMS model) can in principle be computed (recursively or by simulation).

Inference and model selection

- Under all models, classical quantities such as expected *total tree length* (segregating sites), *external branch lengths* (singletons), *nSFS* can easily be computed/simulated.
- According to [SHOEMAKER AND LENNON, 2018], seed banks should stratify a population. Indeed, *Wright's* F_{ST} is elevated under `S`, though less pronounced than for `TI` (with otherwise identical model parameters).
- *Identity by descent (IBD)* has been investigated for `S` on a torus by [DEN HOLLANDER, PEDERZANI 2017].
- *Full likelihoods* of data (say under IMS model) can in principle be computed (recursively or by simulation).

Inference and model selection

- Under all models, classical quantities such as expected *total tree length* (segregating sites), *external branch lengths* (singletons), *nSFS* can easily be computed/simulated.
- According to [SHOEMAKER AND LENNON, 2018], seed banks should stratify a population. Indeed, *Wright's* F_{ST} is elevated under `S`, though less pronounced than for `TI` (with otherwise identical model parameters).
- *Identity by descent (IBD)* has been investigated for `S` on a torus by [DEN HOLLANDER, PEDERZANI 2017].
- *Full likelihoods* of data (say under IMS model) can in principle be computed (recursively or by simulation).

Distinguishing the scenarios K, W, S, and TI

Apply Metropolis-Hastings type algorithm of [Andrieu & Roberts, 2009] for model selection and parameter inference simultaneously among $I = \{\mathtt{K}, \mathtt{S}, \mathtt{TI}\}$ in the IMS model. Target distributions is posterior π_I on model space I (and parameter space, omitted for simplicity).

Specifics of the method will be available in [BBKW18+] (in progress). For moderate sample size ($n = 100$) and $u = 10$, and flat prior on model space, we get

True model	$\pi_I(\mathtt{K}\vert\mathbf{t},\mathbf{n})$	$\pi_I(\mathtt{S}\vert\mathbf{t},\mathbf{n})$	$\pi_I(\mathtt{TI}\vert\mathbf{t},\mathbf{n})$
K	0.98	0.00	0.02
S	0.00	1.00	0.00
TI	0.09	0.12	0.79

True model can be recovered from a moderate amount of data, particularly in the case of K and S. Data generated under TI is also somewhat adequately explained by either of the other two model classes.

Distinguishing the scenarios `K`, `W`, `S`, and `TI`

Apply Metropolis-Hastings type algorithm of [ANDRIEU & ROBERTS, 2009] for model selection and parameter inference simultaneously among $I = \{\mathtt{K}, \mathtt{S}, \mathtt{TI}\}$ in the IMS model. Target distributions is posterior π_I on model space I (and parameter space, omitted for simplicity).

Specifics of the method will be available in [BBKW18+] (in progress). For moderate sample size ($n = 100$) and $u = 10$, and flat prior on model space, we get

True model	$\pi_I(\mathtt{K}\vert\mathbf{t},\mathbf{n})$	$\pi_I(\mathtt{S}\vert\mathbf{t},\mathbf{n})$	$\pi_I(\mathtt{TI}\vert\mathbf{t},\mathbf{n})$
`K`	0.98	0.00	0.02
`S`	0.00	1.00	0.00
`TI`	0.09	0.12	0.79

True model can be recovered from a moderate amount of data, particularly in the case of `K` and `S`. Data generated under `TI` is also somewhat adequately explained by either of the other two model classes.

Scaling limits and separation of timescales

Useful application of duality: Identify *degenerate* diffusion limits in certain scaling scenarios.

For example, for the seed bank diffusion (2), consider the following '*rare resuscitation regime*':

- Let the migration rate $c \to 0$,
- speed up time by a factor $1/c \to \infty$.

While exchange of ancestral lineages between active and dormant states thus becomes *rare* in the original timescale, in the new timescale, migration will happen at rate 1, while coalescences in the active population now occur "*instantaneously*" (with rate ∞).

Scaling limits and separation of timescales

Useful application of duality: Identify *degenerate* diffusion limits in certain scaling scenarios.

For example, for the seed bank diffusion (2), consider the following '*rare resuscitation regime*':

- Let the migration rate $c \to 0$,
- speed up time by a factor $1/c \to \infty$.

While exchange of ancestral lineages between active and dormant states thus becomes *rare* in the original timescale, in the new timescale, migration will happen at rate 1, while coalescences in the active population now occur " *instantaneously*" (with rate ∞).

Scaling limits and separation of timescales

Useful application of duality: Identify *degenerate* diffusion limits in certain scaling scenarios.

For example, for the seed bank diffusion (2), consider the following '*rare resuscitation regime*':

- Let the migration rate $c \to 0$,
- speed up time by a factor $1/c \to \infty$.

While exchange of ancestral lineages between active and dormant states thus becomes *rare* in the original timescale, in the new timescale, migration will happen at rate 1, while coalescences in the active population now occur "*instantaneously*" (with rate ∞).

Scaling limits and separation of timescales

Rare resuscitation regime, on slow timescale (let $K = 1$ for simplicity):

$$\begin{cases} \mathrm{d}X_t & = c(Y_t - X_t)\mathrm{d}t + \sqrt{X_t(1 - X_t)}\mathrm{d}B_t \\ \mathrm{d}Y_t & = c(X_t - Y_t)\mathrm{d}t \end{cases} \tag{4}$$

Seed bank diffusion

Scaling limits and separation of timescales

Rare resuscitation regime, on slow timescale (let $K = 1$ for simplicity):

$$\begin{cases} \mathrm{d}X_t & = 1(Y_t - X_t)\mathrm{d}t + \frac{1}{c}\sqrt{X_t(1 - X_t)}\mathrm{d}B_t \\ \mathrm{d}Y_t & = 1(X_t - Y_t)\mathrm{d}t \end{cases} \tag{4}$$

Seed bank diffusion, *sped up time by* $\frac{1}{c}$

Scaling limits and separation of timescales

Rare resuscitation regime, on slow timescale (let $K = 1$ for simplicity):

$$\begin{cases} \mathrm{d}X_t &= (Y_t - X_t)\mathrm{d}t + \infty\sqrt{X_t(1-X_t)}\mathrm{d}B_t \\ \mathrm{d}Y_t &= (X_t - Y_t)\mathrm{d}t \end{cases} \tag{4}$$

Seed bank diffusion, sped up time by $\frac{1}{c}$, limit $c \to 0$, existence of scaling limit?

Scaling limits of the seed bank coalescent

To understand the limit, we work with the *dual* seed bank coalescent.

Rescaled dual seed bank coalescent as $c \to 0$, with time sped up by $\frac{1}{c}$:

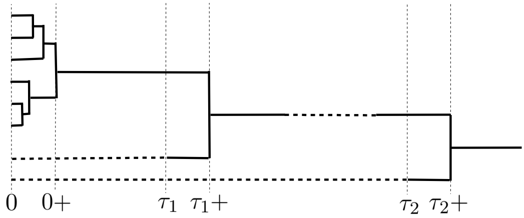


Figure: In the limit, at time $0+$ all lines in the active population have coalesced. At times τ_1 and τ_2 (with rate 1) a seed awakens and *immediately* coalesces.

Limit can be computed explicitly and has transition semigroup $\Pi(t) := Pe^{tG}, t > 0$, where P is a projection. P and G can easily be determined, see e.g. [Möhle & Notohara, 2016].

Scaling limits of the seed bank coalescent

To understand the limit, we work with the *dual* seed bank coalescent.

Rescaled dual seed bank coalescent as $c \to 0$, with time sped up by $\frac{1}{c}$:

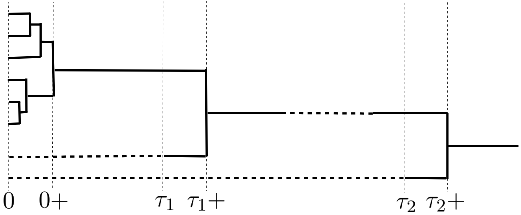


Figure: In the limit, at time $0+$ all lines in the active population have coalesced. At times τ_1 and τ_2 (with rate 1) a seed awakens and *immediately* coalesces.

Limit can be computed explicitly and has transition semigroup $\Pi(t) := Pe^{tG}, t > 0$, where P is a projection. P and G can easily be determined, see e.g. [MÖHLE & NOTOHARA, 2016].

Scaling limits of the seed bank coalescent

To understand the limit, we work with the *dual* seed bank coalescent.

Rescaled dual seed bank coalescent as $c \to 0$, with time sped up by $\frac{1}{c}$:

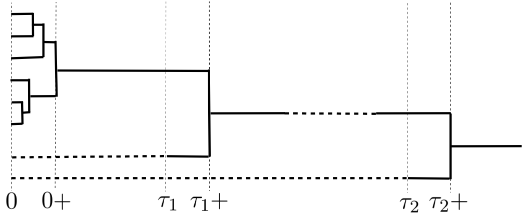


Figure: In the limit, at time $0+$ all lines in the active population have coalesced. At times τ_1 and τ_2 (with rate 1) a seed awakens and *immediately* coalesces.

Limit can be computed explicitly and has transition semigroup $\Pi(t) := Pe^{tG}, t > 0$, where P is a projection. P and G can easily be determined, see e.g. [MÖHLE & NOTOHARA, 2016].

Scaling limits of the seed bank diffusion

Convergence of the dual seed bank coalescent gives us convergence of the *finite dimensional distributions* of the corresponding seed bank diffusion:

$$\mathbb{E}_{x,y}\big[X_t^n, Y_t^m\big] = \mathbb{E}^{n,m}\big[x^{N_t} y^{M_t}\big] \xrightarrow{c\to 0} \mathbb{E}^{n,m}\Big[x^{N_t^\infty} y^{M_t^\infty}\Big].$$

Moment dual uniquely characterizes original diffusion (*moment problem* on $[01]^2$ has unique solution).

Together with Markov property, this yields fdd-convergence.

Scaling limits of the seed bank diffusion

Convergence of the dual seed bank coalescent gives us convergence of the *finite dimensional distributions* of the corresponding seed bank diffusion:

$$\mathbb{E}_{x,y}\left[X_t^n, Y_t^m\right] = \mathbb{E}^{n,m}\left[x^{N_t} y^{M_t}\right] \xrightarrow{c \to 0} \mathbb{E}^{n,m}\left[x^{N_t^\infty} y^{M_t^\infty}\right].$$

Moment dual uniquely characterizes original diffusion (*moment problem* on $[01]^2$ has unique solution).

Together with Markov property, this yields fdd-convergence.

Scaling limits of the seed bank diffusion

Convergence of the dual seed bank coalescent gives us convergence of the *finite dimensional distributions* of the corresponding seed bank diffusion:

$$\mathbb{E}_{x,y}\big[X_t^n, Y_t^m\big] = \mathbb{E}^{n,m}\big[x^{N_t}y^{M_t}\big] \xrightarrow{c\to 0} \mathbb{E}^{n,m}\Big[x^{N_t^\infty}y^{M_t^\infty}\Big].$$

Moment dual uniquely characterizes original diffusion (*moment problem* on $[01]^2$ has unique solution).

Together with Markov property, this yields fdd-convergence.

Identification of scaling limit?

Heuristically, the limiting process should be a *jump-diffusion* (X_t^∞, Y_t^∞) with dynamics as follows:

- X_t^∞ is a *pure jump process* between the states 0 and 1 with the seed-bank dependent rates Y_t^∞ for jumps from 0 to 1, and $(1 - Y_t^\infty)$ for jumps from 1 to 0.
- Y_t^∞ still follows the migration equation $dY_t^\infty = (X_t^\infty - Y_t^\infty)dt$.

Indeed, this is the correct limit, as we can verify via duality.

Identification of scaling limit?

Heuristically, the limiting process should be a *jump-diffusion* (X_t^∞, Y_t^∞) with dynamics as follows:

- X_t^∞ is a *pure jump process* between the states 0 and 1 with the seed-bank dependent rates Y_t^∞ for jumps from 0 to 1, and $(1 - Y_t^\infty)$ for jumps from 1 to 0.
- Y_t^∞ still follows the migration equation $dY_t^\infty = (X_t^\infty - Y_t^\infty)dt$.

Indeed, this is the correct limit, as we can verify via duality.

Identification of scaling limit?

Heuristically, the limiting process should be a *jump-diffusion* (X_t^∞, Y_t^∞) with dynamics as follows:

- X_t^∞ is a *pure jump process* between the states 0 and 1 with the seed-bank dependent rates Y_t^∞ for jumps from 0 to 1, and $(1 - Y_t^\infty)$ for jumps from 1 to 0.
- Y_t^∞ still follows the migration equation $dY_t^\infty = (X_t^\infty - Y_t^\infty)dt$.

Indeed, this is the correct limit, as we can verify via duality.

Scaling limits of the seed bank diffusion

Let

$$f(x,y)(n,m) := x^n y^m.$$

The generator of the candidate process is

$$\begin{aligned} Af(x,y)(n,m) &= (x-y)\frac{\partial f}{\partial y}(x,y)(n,m) + y(f(1,y)(n,m) - f(0,y)(n,m))1_{x=0} \\ &\quad + (1-y)(f(0,y)(n,m) - f(1,y)(n,m))1_{x=1}. \end{aligned}$$

Can be simplified to

$$Af(x,y)(n,m) = y^{m-1}mx^{n+1} - y^m(mx^n + nx) + ny^{m+1}.$$

Scaling limits of the seed bank diffusion

We need to check that this generator satisfies a duality relation with the limiting block counting process (N_t^∞, M_t^∞). Here, we choose the reduced form that already lives in the subspace $\{0,1\} \times \mathbb{N}_0$, with (effective) jump rates

$$\begin{gathered}(0,m) \to (1,m-1) \text{ rate } m,\\ (1,m) \to (1,m-1) \text{ rate } m,\\ (1,m) \to (0,m+1) \text{ rate } 1,\end{gathered}$$

We get

$$\begin{aligned} A^\infty f(x,y)(n,m) &= m(x^1y^{m-1} - x^0y^m)1_{n=0} + m(x^1y^{m-1} - x^1y^m)1_{n=1} \\ &\quad + (x^0y^{m+1} - x^1y^m)1_{n=1} \\ &= m(1-n)(xy^{m-1} - y^m) + mn(xy^{m-1} - xy^m) \\ &\quad + n(y^{m+1} - xy^m) \\ &= y^{m-1}mx - y^m(m(1-n+nx) + nx) + ny^{m+1} = Af(x,y)(n,m). \end{aligned}$$

Scaling limits of the seed bank diffusion

We need to check that this generator satisfies a duality relation with the limiting block counting process (N_t^∞, M_t^∞). Here, we choose the reduced form that already lives in the subspace $\{0,1\} \times \mathbb{N}_0$, with (effective) jump rates

$$\begin{aligned} (0,m) &\to (1,m-1) \text{ rate } m,\\ (1,m) &\to (1,m-1) \text{ rate } m,\\ (1,m) &\to (0,m+1) \text{ rate } 1, \end{aligned}$$

We get

$$\begin{aligned} A^\infty f(x,y)(n,m) &= m(x^1y^{m-1} - x^0y^m)1_{n=0} + m(x^1y^{m-1} - x^1y^m)1_{n=1}\\ &\quad +(x^0y^{m+1} - x^1y^m)1_{n=1}\\ &= m(1-n)(xy^{m-1} - y^m) + mn(xy^{m-1} - xy^m)\\ &\quad +n(y^{m+1} - xy^m)\\ &= y^{m-1}mx - y^m(m(1-n+nx)+nx) + ny^{m+1} = Af(x,y)(n,m). \end{aligned}$$

Technicalities

- Note that from duality we can only infer *fdd convergence* of the diffusion.
- The limit is a jump diffusion, hence we *cannot* have weak convergence wrt Skorohod topology (since this would yield continuity of paths of the limit - continuous paths from a closed subset of càdlàg paths).
- Claim: We have convergence in the *Meyer-Zheng topology*.
- Strictly speaking, we have checked duality of the limits only in *reduced statespaces* (but all mass concentrates there).

If available, duality seems to be a good tool to characterize limits in scenarios with “degenerate diffusion limits”.

Last part: Seed bank models with responsive switching

- So far, we have modeled our seed banks according to the *spontaneous switching regime* which amounts to continuous “migration” between resevoirs.
- However, dormancy seems to be intiated by to exogenous *trigger events* (change in temperature, pH, resources, antibiotics treatment etc.)
- Similarly, one could think about large-scale *simultaneous resuscitation* as result of mass extinction events (e.g. forest fires,...)

Hence one may wish to incorporate “discontinuous” large-scale simultaneous migration between active and dormant population.

We sketch this in a simple setup (to keep notation at a minimum). This is also work in progress.

Last part: Seed bank models with responsive switching

- So far, we have modeled our seed banks according to the *spontaneous switching regime* which amounts to continuous “migration” between resevoirs.
- However, dormancy seems to be intiated by to exogenous *trigger events* (change in temperature, pH, resources, antibiotics treatment etc.)
- Similarly, one could think about large-scale *simultaneous resuscitation* as result of mass extinction events (e.g. forest fires,...)

Hence one may wish to incorporate “discontinuous” large-scale simultaneous migration between active and dormant population.

We sketch this in a simple setup (to keep notation at a minimum). This is also work in progress.

Last part: Seed bank models with responsive switching

- So far, we have modeled our seed banks according to the *spontaneous switching regime* which amounts to continuous “migration” between resevoirs.
- However, dormancy seems to be intiated by to exogenous *trigger events* (change in temperature, pH, resources, antibiotics treatment etc.)
- Similarly, one could think about large-scale *simultaneous resuscitation* as result of mass extinction events (e.g. forest fires,...)

Hence one may wish to incorporate “discontinuous” large-scale simultaneous migration between active and dormant population.

We sketch this in a simple setup (to keep notation at a minimum). This is also work in progress.

Last part: Seed bank models with responsive switching

- So far, we have modeled our seed banks according to the *spontaneous switching regime* which amounts to continuous “migration” between resevoirs.
- However, dormancy seems to be intiated by to exogenous *trigger events* (change in temperature, pH, resources, antibiotics treatment etc.)
- Similarly, one could think about large-scale *simultaneous resuscitation* as result of mass extinction events (e.g. forest fires,...)

Hence one may wish to incorporate “discontinuous” large-scale simultaneous migration between active and dormant population.

We sketch this in a simple setup (to keep notation at a minimum). This is also work in progress.

Last part: Seed bank models with responsive switching

- So far, we have modeled our seed banks according to the *spontaneous switching regime* which amounts to continuous “migration” between resevoirs.
- However, dormancy seems to be intiated by to exogenous *trigger events* (change in temperature, pH, resources, antibiotics treatment etc.)
- Similarly, one could think about large-scale *simultaneous resuscitation* as result of mass extinction events (e.g. forest fires,...)

Hence one may wish to incorporate “discontinuous” large-scale simultaneous migration between active and dormant population.

We sketch this in a simple setup (to keep notation at a minimum). This is also work in progress.

Seed bank models with responsive switching

Let $z_1, z_2 \in (0,1)$. Consider a Wright-Fisher model with seed bank as before.

- Assume that for *typical generations* (with prob. $1 - \frac{1}{N}$), reproduction occurs according to our previous mechanism.
- However, with probability $\frac{1}{2N}$, a *large-scale migration event* occurs: Each individual in the active population is replaced independently with probability z_1 by a uniformly chosen individual from the seed bank (with replacement).
- Similarly, again with probability $\frac{1}{2N}$, each indidividual in the seed bank independetly with probability z_2 is replaced by the offspring of an active individual.

Seed bank models with responsive switching

Let $z_1, z_2 \in (0,1)$. Consider a Wright-Fisher model with seed bank as before.

- Assume that for *typical generations* (with prob. $1 - \frac{1}{N}$), reproduction occurs according to our previous mechanism.
- However, with probability $\frac{1}{2N}$, a *large-scale migration event* occurs: Each individual in the active population is replaced independently with probability z_1 by a uniformly chosen individual from the seed bank (with replacement).
- Similarly, again with probability $\frac{1}{2N}$, each indidividual in the seed bank independetly with probability z_2 is replaced by the offspring of an active individual.

Seed bank models with responsive switching

Let $z_1, z_2 \in (0,1)$. Consider a Wright-Fisher model with seed bank as before.

- Assume that for *typical generations* (with prob. $1 - \frac{1}{N}$), reproduction occurs according to our previous mechanism.
- However, with probability $\frac{1}{2N}$, a *large-scale migration event* occurs: Each individual in the active population is replaced independently with probability z_1 by a uniformly chosen individual from the seed bank (with replacement).
- Similarly, again with probability $\frac{1}{2N}$, each indidividual in the seed bank independetly with probability z_2 is replaced by the offspring of an active individual.

Forward in time

Consider two types (purple and white). Forward: again trace fractions of purple type:

$$\bar{X}_r^N := \frac{1}{N}\sum_{i\in[\![N]\!]} \mathbf{1}_{\{i \text{ in generation } r \text{ is purple}\}}, \quad \bar{Y}_r^M := \frac{1}{M}\sum_{j\in[\![M]\!]} \mathbf{1}_{\{j \text{ in generation } r \text{ is purple}\}}$$

Theorem 4.1 (BGKW18+)

$c > 0$, $M = M(N) = N$

$$(\bar{X}^N_{\lfloor Nt \rfloor}, \bar{Y}^{M(N)}_{\lfloor Nt \rfloor})_{t\geq 0} \Rightarrow (\bar{X}_t, \bar{Y}_t)_{t\geq 0}$$

for $N \to \infty$ *on* $D_{[0,\infty)}([0,1]^2)$*, with generator (including jumps!)*

$$\begin{aligned} Af(x,y) =& f(x+z_1(y-x),y) + f(x,y+z_2(x-y)) - 2f(x,y) \\ & + c(y-x)\frac{\partial}{\partial x}f(x,y) + c(x-y)\frac{\partial}{\partial y}f(x,y) + \frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}f(x,y), \end{aligned}$$

Backward in time

Genealogy is gven by a seed bank coalescent with *simultaneous* initiation of / resuscitation from dormancy!

Theorem 4.2 (BGCK '18 (in progress))

We have

$$(\Pi^{(N,k)}_{[Nt]})_{t\geq 0} \Rightarrow (\Pi^k_t)_{t\geq 0}.$$

The corresponding block-counting process $(N_t, M_t)_{t\geq 0}$ *has transitions*

$$(n,m) \mapsto \begin{cases} (n-1, m+1) & \textit{at rate } (c + z_1(1-z_1)^{n-1})n, \\ (n-k, m+k) & \textit{at rate } \binom{n}{k} z_1^k (1-z_1)^{n-k}, 2 \leq k \leq n, \\ (n+1, m-1) & \textit{at rate } (c + z_2(1-z_2)^{m-1})m, \\ (n+l, m-l) & \textit{at rate } \binom{m}{l} z_2^l (1-z_2)^{m-k}, 2 \leq l \leq m, \\ (n-1, m) & \textit{at rate } \binom{n}{2}. \end{cases} \tag{5}$$

Backward in time

Genealogy is gven by a seed bank coalescent with *simultaneous* initiation of / resuscitation from dormancy!

Theorem 4.2 (BGCK '18 (in progress))

We have

$$(\Pi^{(N,k)}_{\lfloor Nt \rfloor})_{t\geq 0} \Rightarrow (\Pi^k_t)_{t\geq 0}.$$

The corresponding block-counting process $(N_t, M_t)_{t\geq 0}$ *has transitions*

$$(n,m) \mapsto \begin{cases} (n-1, m+1) & \textit{at rate } (c + z_1(1-z_1)^{n-1})n, \\ (n-k, m+k) & \textit{at rate } \binom{n}{k} z_1^k (1-z_1)^{n-k}, 2 \leq k \leq n, \\ (n+1, m-1) & \textit{at rate } (c + z_2(1-z_2)^{m-1})m, \\ (n+l, m-l) & \textit{at rate } \binom{m}{l} z_2^l (1-z_2)^{m-k}, 2 \leq l \leq m, \\ (n-1, m) & \textit{at rate } \binom{n}{2}. \end{cases} \tag{5}$$

Moment duality

Theorem 4.3

For every $(x,y) \in [0,1]^2$*, every* $n,m \in \mathbb{N}_0$ *and every* $t \geq 0$

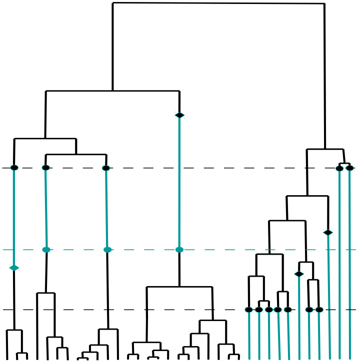
$$\mathbb{E}_{x,y}\left[X_t^n Y_t^m\right] = \mathbb{E}^{n,m}\left[x^{N_t} y^{M_t}\right]. \tag{6}$$

Proof: Let $f(x,y;n,m) := x^n y^m$. We get

$$\begin{aligned} Af(x,y) =& f(x+z_1(y-x),y) + f(x,y+z_2(x-y)) - 2f(x,y) \\ & + c(y-x)\frac{df}{dx}f(x,y) + \frac{1}{2}x(1-x)\frac{d^2f}{dx^2}f(x,y) + c(x-y)\frac{df}{dy}f(x,y) \\ =& \sum_{k=1}^{n}\binom{n}{k}z_1^k(1-z_1)^{n-k}(x^{n-k}y^{m+k} - x^n y^m) + \sum_{l=1}^{n}\binom{m}{l}z_2^l(1-z_2)^{n-l}(x^{m+l}y^{n-l} - x^n y^m) \\ & + cn(x^{n-1}y^{m+1} - x^n y^m) + \binom{n}{2}(x^{n-1}y^m - x^n y^m) + cm(x^{n+1}y^{m-1} - x^n y^m), \end{aligned}$$

as required.

The seed bank coalescent with simultaneous migration



A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

A zoo of on/off coalescents

For people who like to investigate coalescent processes / ancestral influence graphs:

- Kingman coalescent with individual on/off events (seed bank coalescent / peripatric coalescent)
- Kingman coalescent simultaneous on/off events
- Λ-, Ξ-coalescents with individual on/off events
- Λ-, Ξ-coalescents with simultaneous on/off events
- Also: Structured coalescent with simultaneous migration
- All of above combined with selection, recomination, etc: all kinds of on/off graphs...

Finally...

... thank you for your attention!

Talk mostly based on:

- B., ELDON, GONZÁLEZ CASANOVA, KURT, WILKE BERENGUER: Genetic variability under the seed bank coalescent, *Genetics* **200**, 2015
- B., GONZÁLEZ CASANOVA, KURT, WILKE BERENGUER: A new coalescent for seed bank models, *Annals of Applied Probability*, 2016
- B., BUZZONI, GONZÁLEZ CASANOVA, KURT, WILKE BERENGUER: The seed bank diffusion, and its relation to the two-island model, *SPP 1590 (p)reprint* **106**, 2017
- B., BUZZONI, WILKE BERENGUER, KOSKELA: Statistical tools for seed bank detection, *manuscript*, 2018
- B., GONZÁLEZ CASANOVA, KURT: The seed bank coalescent with large migration events, *manuscript*, 2018