An ancestral recombination graph for diploid populations with skewed offspring distribution

Jochen Blath, TU Berlin

Joint work in progress with

Matthias Birkner (Mainz), Bjarki Eldon (Oxford)

Probability, population genetics and evolution, CIRM

June 14, 2012

1) Introduction

Population genetics is concerned with the identification, modelling and interplay of evolutionary forces governing the genetic variability within populations.

Important evolutionary forces are

- mutation
- genetic drift / random reproduction
- selection
- recombination
- migration / spatial structure, etc.

1) Introduction

Population genetics is concerned with the identification, modelling and interplay of evolutionary forces governing the genetic variability within populations.

Important evolutionary forces are

- mutation
- genetic drift / random reproduction
- selection
- recombination
- migration / spatial structure, etc.

Stochastic models in population genetics

- Evolutionary driving forces are inherently stochastic (on microscopic level).
- In this talk: Focus on random genealogies of stochastic populations exhibiting *multiple collisions* of ancestral lineages.
- Such genealogies can be described by Λ- and Ξ-**coalescents**.
- Arise in a variety of scenarios (as result of interplay of evolutionary forces), as we will see later.

General project aim and specific focus of this talk

- General aim of our project: Develop *models* and *inference methods* for population genetics with general coalescents, applicable to real data
  (to be carried out within new DFG Priority Programme SPP 1590 'Probabilistic Structures in Evolution').
- Specific first goal of this talk: Give and introduction to multiple merger genealogies, identify the genealogical process in a diploid multi-locus model with recombination in a Λ-coalescent scenario.
- The latter is work in progress (actually, the SPP hasn't officially started yet).

General project aim and specific focus of this talk

- General aim of our project: Develop *models* and *inference methods* for population genetics with general coalescents, applicable to real data
  (to be carried out within new DFG Priority Programme SPP 1590 'Probabilistic Structures in Evolution').
- Specific first goal of this talk: Give and introduction to multiple merger genealogies, identify the genealogical process in a diploid multi-locus model with recombination in a Λ-coalescent scenario.
- The latter is work in progress (actually, the SPP hasn't officially started yet).

2) Warmup: Classic stochastic population models - haploid case

The **Cannings model** (1974):

- Population of N 'individuals', well mixed.
- Consider fixed non-overlapping discrete generations $r = 0, 1, 2, 3, \ldots$, assume constant population size.
- Let

$$\nu^{(r)} := \left(\nu_1^{(r)}, \ldots, \nu_N^{(r)}\right), \quad \sum_{i=1}^{N} \nu_i^{(r)} = N$$

  denote the random vector of offspring representing the N individuals in the r-th generation, where $\nu_i^{(r)}$ is the number of children of individual $i \in \{1, \ldots, N\}$ in generation r.
- We assume the $\nu^{(r)}$ to be *i.i.d. exchangeable* random vectors.
- Offspring inherit their genetic type from the parent, denote genetic typespace by E.

Example 1: The Wright-Fisher model

This provides a unified framework for many classical models:

For example, in the **Wright-Fisher model** (1930, 1931), the $(\nu_1^{(r)}, \ldots, \nu_N^{(r)})$ are assumed to be symmetric multinomial, i.e.

$$\mathbb{P}\{\nu_1^{(r)} = m_1, \ldots, \nu_N^{(r)} = m_N\} = \frac{n!}{m_1! \cdots m_N!} \left(\frac{1}{N}\right)^N.$$

One may think of this mechanism as each 'offspring choosing its parent' independently and uniformly at random.

Example 2: The Moran model

In the **Moran-model** (1958), the $(\nu_1^{(r)}, \ldots, \nu_N^{(r)})$ are obtained by means of a uniformly chosen permutation of the offspring vector

$$(2, 1, \ldots, 1, 0),$$

that is,

$$\nu^{(r)} = \pi_N^{(r)}((2, 1, \ldots, 1, 0)), \quad \pi_N^{(r)} \in S_N.$$

We can interpret this as follows: In each generation, precisely one individual produces one offspring, and one (distinct) individual dies. All other individuals persist.

Note that this is a *discrete time* version of the Moran model.

Scaling limits

Both models (Moran and Wright-Fisher) are in the domain of attraction the classic **Fleming Viot process** (on a suitable genetic type space E), as $N \to \infty$ and if time is sped up accordingly (for the moment, we ignore mutation). Their genealogy is described by a **Kingman coalescent**.

- Indeed, for the Wright-Fisher model, the probability that two individuals have the same ancestor in the previous generation is $1/N$, hence, to obtain the Fleming-Viot limit, time needs to be measured in units of N.
- For the Moran model, the probability that two individuals have the same ancestor in the previous generation is $O(N^{-2})$, hence time needs to be measured in units of $O(N^2)$.

Scaling limits

Both models (Moran and Wright-Fisher) are in the domain of attraction the classic **Fleming Viot process** (on a suitable genetic type space E), as $N \to \infty$ and if time is sped up accordingly (for the moment, we ignore mutation). Their genealogy is described by a **Kingman coalescent**.

- Indeed, for the Wright-Fisher model, the probability that two individuals have the same ancestor in the previous generation is $1/N$, hence, to obtain the Fleming-Viot limit, time needs to be measured in units of N.
- For the Moran model, the probability that two individuals have the same ancestor in the previous generation is $O(N^{-2})$, hence time needs to be measured in units of $O(N^2)$.

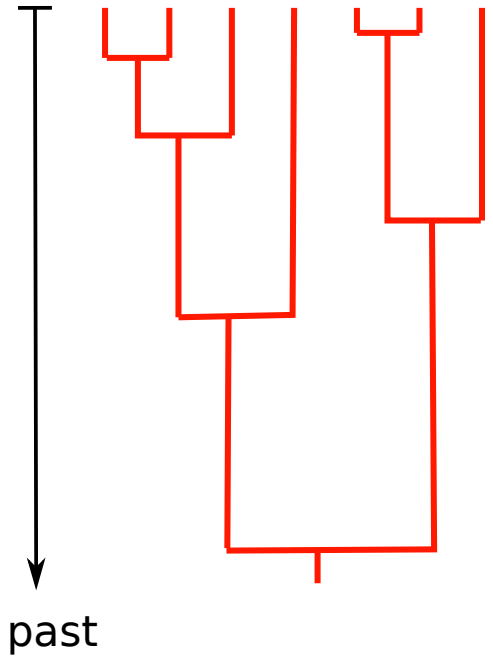
The genealogy of our models

The genealogy of our limting models, and of the Fleming-Viot process, is of course described by the famous **Kingman coalescent** (1982).

It can be embedded into the Fleming-Viot limit on the same probability space with the help of **Donnelly and Kurtz' lookdown construction** (1996).

3) Coalescent processes. Kingman's coalescent

Recall that **Kingman's coalescent** is a partition valued Markov process. A realization could looks as follows:



- Each given pair of blocks merges with rate 1.
- If there are currently n active lineages, the next coalescence happens at rate $\binom{n}{2}$.
- Note that only binary mergers are allowed!

Universality

The Kingman coalescent is a *universal* object.

It describes the law of the genealogy of a sample taken from a Fleming-Viot process (e.g. via the lookdown process), and hence of the Wright-Fisher model and the Moran model, under appropriate scaling.

It is a valid limit of a large class of Cannings models if the variance of the reproduction mechanism 'behaves well' (e.g. stays bounded as $N \to \infty$).

(this is similar to the universality of the Feller diffusion as limit for critical branching processes, and in a sense even to the CLT).

A limit theorem for Cannings models, revisited.

More precisely, for a Cannings-model, the Kingman-coalescent describes the limiting geneology if

$$c_N := \mathbb{P}\{ \text{ two have same parent } \} = \frac{E[\nu_1(\nu_1 - 1)]}{N - 1} \to 0,$$

and (for samples of size three),

$$\frac{E[\nu_1(\nu_1 - 1)(\nu_1 - 2)]}{N^2 c_N} \to 0,$$

as $N \to \infty$, where we have to measure time in units of $1/c_N$. [Full limit theorem due to Möhle & Sagitov (1999)].

Kingman's coalescent - simple applications

In the Kingman scaling regime, it is possible to compute many relevant genetic properties in simple and elegant ways!

Example: Total tree length for a sample of size n satisfies

$$L(n) \stackrel{d}{=} nS(n) + \cdots + 2S(2),$$

where the $S(k)$ are indep. exp r.v. with rate $\binom{k}{2}$. Hence,

$$\mathbb{E}[L(n)] \sim 2 \log n.$$

Mutation models

Examples:

- *Finitely many alleles*, e.g. $E = \{a, A\}$ or $E = \{a, c, g, t\}$, type changes according to some *mutation matrix* P.
- *Infinitely many alleles*: Each mutation produces an entirely new type. Problem: no idea about 'how different' alleles are!
- *Infinitely many sites* (introduced by Watterson, TPB '75): each mutation occurs at a different position on sequence (i.e. affects a different nucleotide). Widely used to analyse real data. Sometimes subtle combinatorics.

Incorporation muations: The infinitely many sites model

Suitable, if sample size n is of the order of the square root of the length of the DNA sequences.

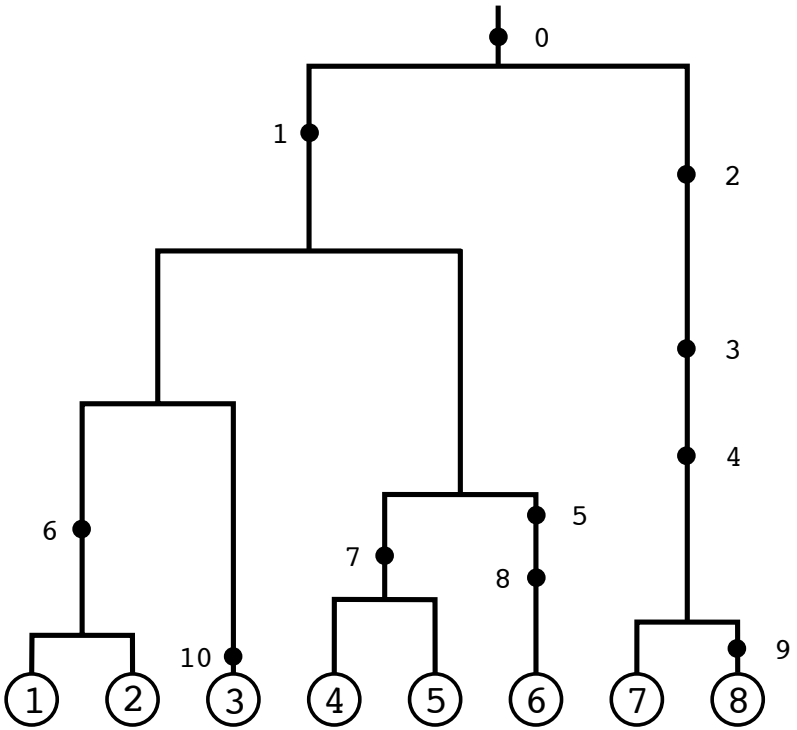
Example: Types given as binary sequences ('1' indicates mutant), segregating sites only:

1 : ...1000010000...
2 : ...1000010000...
3 : ...1000000001...
4 : ...1000001000...
5 : ...1000001000...
6 : ...1000100100...
7 : ...0111000000...
8 : ...0111000010...

Under suitable scaling (e.g. $O(1/N)$ for the Wright-Fisher model), mutations appear as Poisson-process of ancestral lines (with some rate θ, say).

Kingman's coalescent and mutations

One possible binary tree consistent with the **Example** is:



Kingman's coalescent as standard model in population genetics - Aspects of its success

- Scaling limit doesn't depend on precise shape of offspring distribution as long as variance is known (and finite).
- Easy to generate a coalescent tree with mutations forward in time (starting from MRCA) using Hoppe's urn (1984).
- Ewens' (1972) Sampling formula: exact distribution of type probabilities (in inf. alleles model)
- Possible to compute observed type probabilities recursively in inf. sites model (Griffiths & Tavaré 1994).
- Inference of evolutionary parameters: see e.g. Stephens & Donnelly (2000), Hobolth, Uyenoyama & Wiuf (2008) (importance sampling), and others.
- Extensions to incorporate selection (Krone, Neuhauser...), recombination (Hudson, Griffiths, Marjoram...), etc.

Other natural models for genealogies?

Note that many Cannings models are *not* in the domain of attraction of the classical Fleming-Viot process and the Kingman coalescent.

There is a rich class of exchangeable coalescent processes, and there seem to be several relevant evolutionary mechanisms leading to 'exceptional genealogies' outside the Kingman universality class.

More general genealogies: the Λ-coalescent

Pitman (1999), Sagitov (1999) introduce a general class of exchangeable coalescent processes, which allow *multiple*, but not simultaneous multiple collisions of blocks.

These coalescents are known as

Λ-coalescents.

More general genealogies: the Λ-coalescent

A Λ-**coalescent** $\{\prod_t^\Lambda, t \geq 0\}$ is a partition-valued Markov process.

Dynamics: If there are m blocks at present in the partition, each *given* k-tuple merges to a single block at rate

$$\lambda_{m,k} = \int_{[0,1]} x^k (1-x)^{m-k} \frac{1}{x^2} \Lambda(dx),$$

$k \in \{2, \ldots m\}$, where Λ is some finite measure on $[0, 1]$.

These rates agree with natural consistency property

$$\lambda_{m,k} = \lambda_{m+1,k} + \lambda_{m+1,k+1}$$

and uniquely characterize the measure Λ (Hausdorff's moment problem).

A general limit theorem Möhle & Sagitov (2001)

If we consider a Cannings population model on the time scale $1/c_N$, then the limiting genealogy will be governed by a Λ-coalescent iff

- $c_N \to 0$, as $N \to \infty$,
- for all $y \in (0,1)$ with $\Lambda(\{y\}) = 0$, we have

$$\frac{N}{c_N}\mathbb{P}\{\nu_1 > Ny\} \longrightarrow \int_{(y,1]} \frac{1}{x^2}\,\Lambda(dx), \quad \text{as } N \to \infty,$$

- and finally,

$$\frac{\mathbb{E}[\nu_1(\nu_1-1)\,\nu_2(\nu_2-1)]}{N^2 c_N} \longrightarrow 0\,, \quad \text{as } N \to \infty.$$

Note that the third condition guarantees the absence of simultaneous multiple collisions in the genealogy.

Interpretation

Roughly: We will see multiple ancestral collisions in the limit if single individuals are able to produce a positive fraction of the living population in a single reproduction event (high variance indeed).

'highly skewed offspring distribution'

Generalized Λ-Fleming-Viot processes

Donnelly & Kurtz (1999), Bertoin & Le Gall (2003) show that Λ-coalescents are dual to so-called **generalized Λ-Fleming-Viot processes**.

Rather elegantly (and usefully), both processes can simultaneously be embedded in same probability space using the *lookdown construction* by Donnelly & Kurtz (1999).

Largest class of exchangeable coalescents: **Ξ-coalescents**. Allow *simultaneous* multiple mergers (explicit definition omitted) - see e.g. Schweinsberg (2000).

Straightforward (but rather technical) that lookdown-construction can be extended to construct **Ξ-Fleming-Viot processes** [for details and examples see Birkner, Blath, Möhle, Steinrücken & Tams (2008)].

Special cases

Examples:

- $\Lambda = \delta_0$: **Kingman coalescent**.
- $\Lambda = \delta_1$: leads to star-shaped genealogies.
- $\Lambda = \delta_0 + c\psi^2\delta_\psi, \psi \in (0,1]$: two-atom Lambda coalescent (used by Eldon & Wakeley 2006)
- $\Lambda = \mathit{Unif}([0,1])$: **Bolthausen-Sznitman coalescent**.
- Λ the Beta$(2-\alpha, \alpha)$-distribution, $\alpha \in (0,2)$: **Beta-coalescent**, i.e

$$\Lambda(dx) = \frac{\Gamma(2)}{\Gamma(\alpha)\Gamma(2-\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}\, dx.$$

- Ξ-coalescents can be obtained from Lambda-coalescent events via a paintbox procedure.

Example: A two-atom coalescent

Eldon & Wakeley (2006) consider

$$\Lambda = \delta_0 + c\psi^2\delta_\psi, \psi \in (0,1].$$

For the rates of a *given* k-merger, we obtain

$$\begin{aligned}\lambda_{m,k} &= \int_{[0,1]} x^k(1-x)^{m-k}\frac{1}{x^2}\Lambda(dx)\\ &= \int_{[0,1]} x^{k-2}(1-x)^{m-k}(\delta_0(dx) + c\psi^2\delta_\psi(dx))\\ &= 1_{\{k=2\}} + c\psi^k(1-\psi)^{m-k}.\end{aligned}$$

We can interpret this as a Kingman component plus an event at which k lineages merge if they flip successfully an independent 'ψ-coin'.

Example: A two-atom coalescent

Forwards in time, it is easy to see from the limit theorem that this coalescent arises if one considers the following Cannings model: In a population of size N, ν is a (uniform) permutation of

$$(2, 0, 1, \ldots, 1) \quad \text{or of} \quad (\lfloor \psi N \rfloor, \underbrace{0, 0, \ldots, 0}_{\lfloor \psi N \rfloor \text{ times}}, 1, \ldots, 1)$$

with probability $1 - c/N^2$ resp. c/N^2

Interpretation

- Atom in ψ to the right of 0: Each lineage participates independently in the corresponding *coalescence event* with probability ψ.
- For general Λ-coalescent, $x^{-2}\Lambda(dx)$ is the intensity measure with which merger events appear.
- Corresponding *reproduction mechanism*: single parent produces $(x * 100)\%$ of offspring with intensity $x^{-2}\Lambda(dx)$
- Beta-coalescents can be linked with α-stable branching models Donnelly & Kurtz (1999), Hiraba (2000), Schweinsberg (2003), BBCEMSW (2005).

Real populations with exceptional behaviour?

Evolutionary mechanisms in real populations leading to multiple merger genealogies:

- Skewed offspring distributions ('sweepstakes reproduction', type-III-survivorship), see Hedgecock & Pudovkin's overview article (*Bull Marine Sci.* 2011)
- Selective sweeps in combination with genetic hitchhiking (Schweinsberg, Durrett (2005), Coop, Ralph,...).
- Large-scale spatial extinction and recolonisation, see Barton, Etheridge, Véber... (2010 – 2012+).
- Recurrent severe bottlenecks (BBMST 2009)
- ...

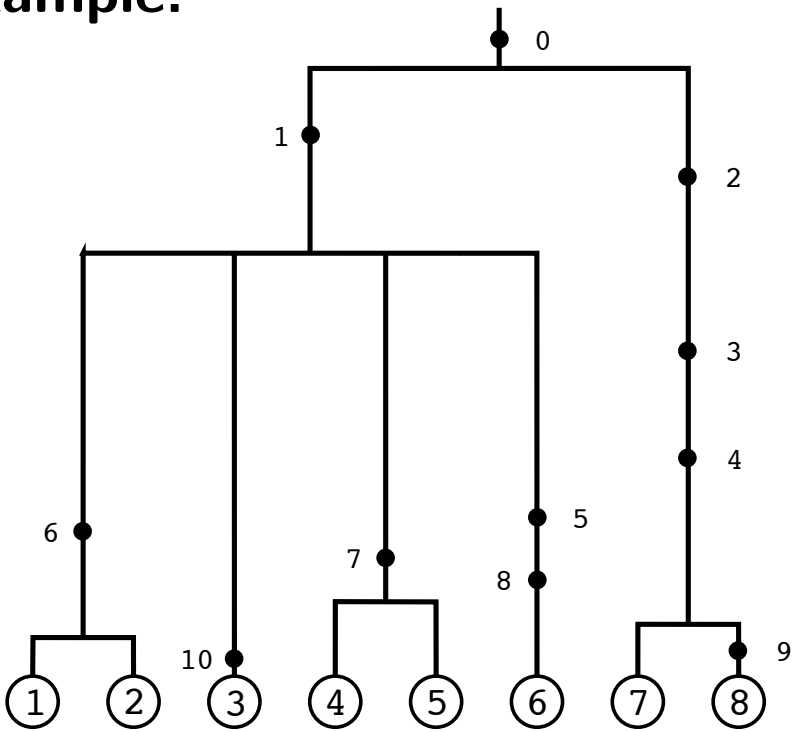
Inference for Lambda coalescents

Eldon & Wakeley (2006) consider Boom et.al.'s *Pacific Oyster* data (RFLPs):

- Assume a special class of population models which allow extreme reproduction events.
  $\Lambda = \delta_0 + c\psi^2\delta_\psi, c > 0, \psi \in (0,1)$.
- Base inference on number of segregating sites and singleton polymorphisms.
- Conclude that their model fits better to data than purely Kingman-based models: at rare reproduction events, about $\hat{\psi} = 8\%$ of the population is replaced by the offspring of a single individual.

Inference: Recursions for rooted tree probabilities

A coalescent tree with multiple collisions consistent with our **Example:**



Again, one can derive recursions for type probabilities (for various coalescent parameters), see Birkner, Blath, JMB (2008).

Inference: Recursions for rooted tree probabilities

This allows us to use a maximum likelihood approach for inference of 'coalescent parameters' (and mutation rates).

- For small complexity of the sample, one can solve the recursion iteratively by 'brute force' directly (complexity is reduced in each transition).
- For moderate sample sizes, one can use Monte Carlo-methods and importance sampling (Birkner, Blath, Steinrücken, TPB 2011)

Genetic variation at the mitochondrial cyt *b* locus of Atlantic cod, see Birkner, Blath, Steinrücken, preprint 2012

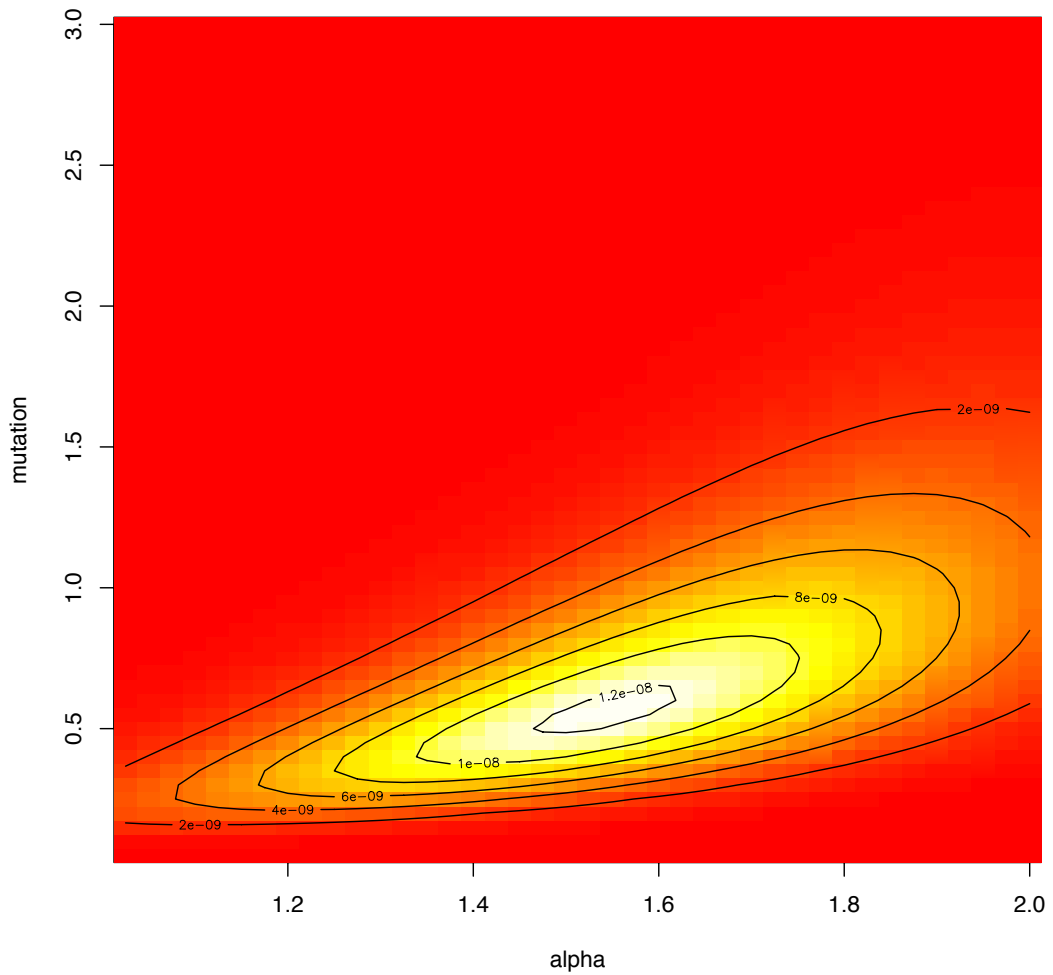


Figure: Likelihood surface for Arnason's Atlantic Cod data

Where do we stand?

- For Cannings models, haploid neutral one-locus theory is mathematically well developed.
- Some tools for estimation in the multiple merger framework available, rather clear evidence for "non-Kingman-ness" in several cases.
- Eldon & Wakeley (2006): *"For many species, the coalescent with multiple mergers might be a better null model than Kingman's coalescent."*
- Hedgecock & Pudovkin (2011) conclude: *"Development of statistical tools to help decide among competing coalescent models and to draw inferences about demographic and genetic parameters of interest is welcome."*

However...

- Statistical properties of estimators in multiple merger scenario? Almost no theoretical results, so far only meager empirical assessments.
- Limited to single-locus haploid datasets (typically mitochondrial DNA)
- Problem of identifiability of classes of coalescents.
- Theoretical limitations (recall the 'tree length example').

Treatment of multi-locus data highly desirable!

However...

- Statistical properties of estimators in multiple merger scenario? Almost no theoretical results, so far only meager empirical assessments.
- Limited to single-locus haploid datasets (typically mitochondrial DNA)
- Problem of identifiability of classes of coalescents.
- Theoretical limitations (recall the 'tree length example').

Treatment of multi-locus data highly desirable!

Diploidy, mulitple loci, and recombination

Next aim: We extend our Moran model to the diploid case (pairs of chromosomes), multiple loci (on the same chromosome) and recombination (between loci).

This extends work of Möhle (1998) and others on scaling limits of diploid populations, and of Hudson (1983), Griffiths (1991), Griffiths & Marjoram (1997) and others on the *ancestral recombination graph*.

For purpose of accessibility, we assume a reproduction mechanism in the domain of attraction of the Eldon & Wakeley (2006) coalescent:

$$\Lambda(dx) = \delta_0 + c\psi^2\delta_\psi, \quad c > 0, \psi \in (0, 1],$$

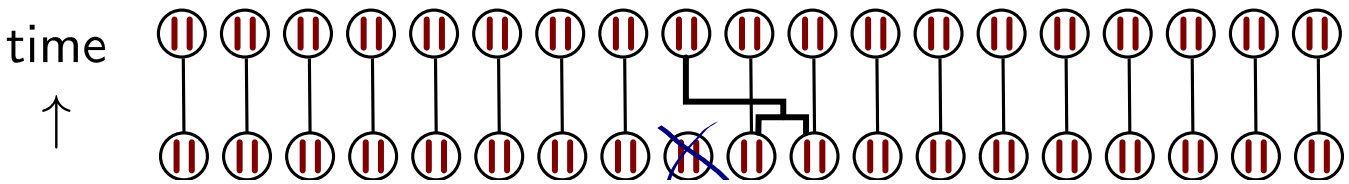
but this can be genralized (messier, though!), see Birkner, Blath, Eldon, Preprint 2012.

A diploid multi-locus Moran model with recombination

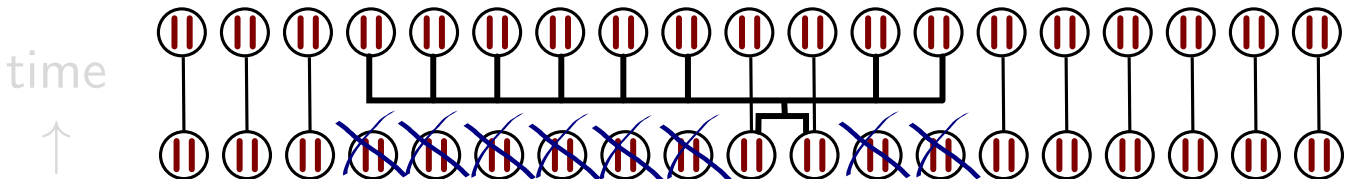
- Assume fixed population size, N individuals, discrete generations.
- Each individuals contains two chromosomes. Each chromosome is structured in $L \in \mathbb{N}$ loci.
- Reproduction comes in two variants: In each generation...
  - ... with probability $1 - \varepsilon_N$, two randomly chosen parents produce one new offspring (small event)...
  - and with probability ε_N, two parents produce $\lfloor \psi N \rfloor$ offspring (large event).
- During reproduction, recombination between locus ℓ and $\ell + 1$ occurs with probability $r_N^{(\ell)}$ (only single crossovers allowed), $1 \leq \ell \leq L - 1$, independently for each offspring.

Illustration of the reproduction mechanism

Small reproduction events (happen most of the time):



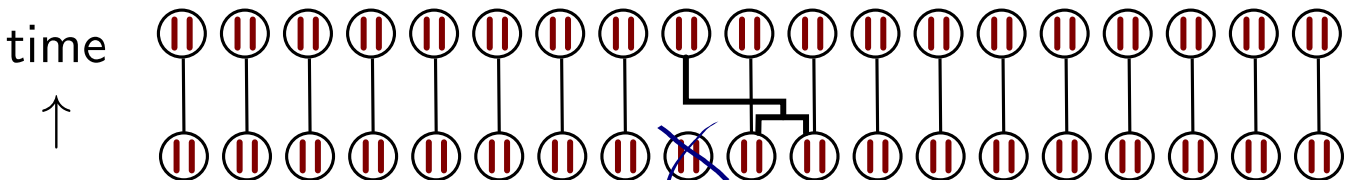
Large reproduction events (probability ε_N):



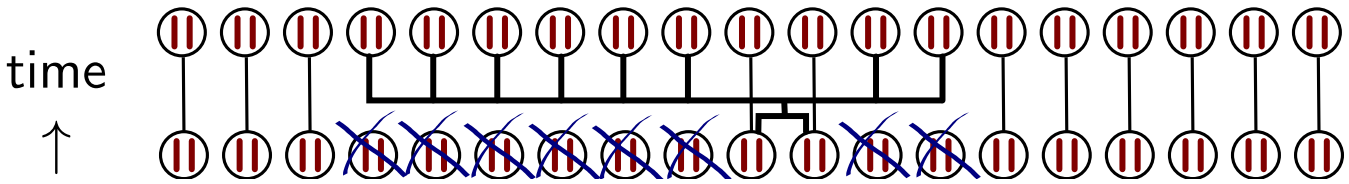
Here, $\lfloor \psi N \rfloor$ individuals are replaced by offspring of the successful parents.

Illustration of the reproduction mechanism

Small reproduction events (happen most of the time):



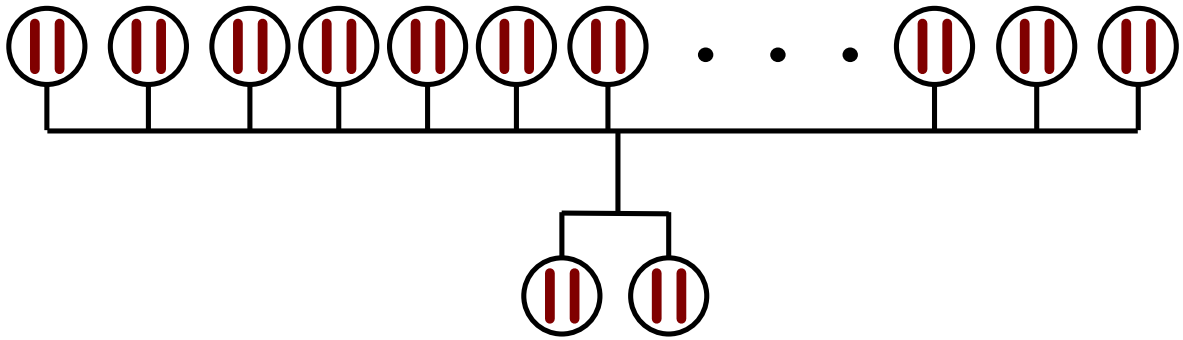
Large reproduction events (probability ε_N):



Here, $\lfloor \psi N \rfloor$ individuals are replaced by offspring of the successful parents.

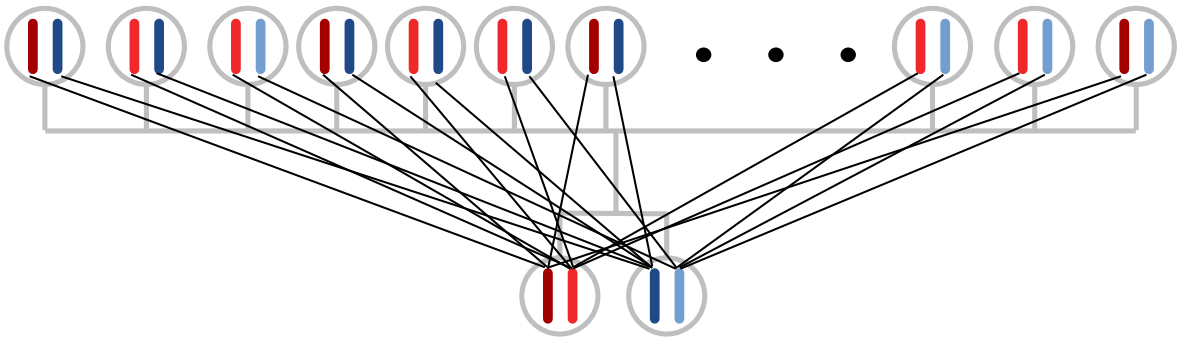
Emergence of Ξ-coalescents with (up to) quadrifold mergers

Important to note that diploidy in combination with 'large events' naturally leads to quadrifold simultaneous multiple mergers:



Emergence of Ξ-coalescents with (up to) quadrifold mergers

Important to note that diploidy in combination with 'large events' naturally leads to quadrifold simultaneous multiple mergers:



Emergence of Ξ-coalescents with (up to) quadrifold mergers

Note that each of the $\lfloor N\psi \rfloor$ offspring uniformly chooses one out of four possible types. This explains the correspondence to simultaneous multiple merger coalescents:

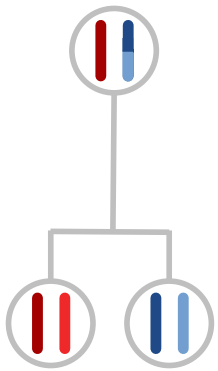
In a large coalescence event according to an atom δ_ψ, each participating lineage (after a successful ψ-coinflip) chooses uniformly one of four possible sub-blocks corresponding to the four possible chromosomal configurations.

For people familiar with Ξ-coalescent notation, this corresponds to a Ξ-atom of the form

$$\delta_{(\psi/4,\psi/4,\psi/4,\psi/4,0,0,\dots)}.$$

Recombination

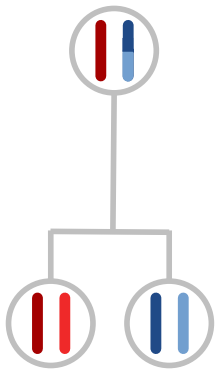
Between our L loci, recombination can occur.
We allow only for single crossover events, that is, at most one of the L loci affected. During a 'small reproduction event', recombination might turn out as follows:



Important to note: 'Large events' in combination with 'recombination' will both be, in our scaling, *rare* events, hence will effectively be negligible for a given individual in the sample (in the scaling limit).

Recombination

Between our L loci, recombination can occur.
We allow only for single crossover events, that is, at most one of the L loci affected. During a 'small reproduction event', recombination might turn out as follows:



Important to note: 'Large events' in combination with 'recombination' will both be, in our scaling, *rare* events, hence will effectively be negligible for a given individual in the sample (in the scaling limit).

Passing to the limit

In order to obatin a non-trivial limit...

- let $N \to \infty$,
- speed up time by $1/N^2$ (the scaling from our Moran model),
- pick $\varepsilon_N = c/N^2$ for the probability of 'large events',
- and for the recombination rate pick $r_N = r/N$, with locus $\ell \in [L]$ affected with probability r_N^ℓ, where

$$r_N = r_N^1 + \cdots + r_N^L.$$

Which events are visible in the limit?

Passing to the limit

In order to obatin a non-trivial limit...

- let $N \to \infty$,
- speed up time by $1/N^2$ (the scaling from our Moran model),
- pick $\varepsilon_N = c/N^2$ for the probability of 'large events',
- and for the recombination rate pick $r_N = r/N$, with locus $\ell \in [L]$ affected with probability r_N^ℓ, where

$$r_N = r_N^1 + \cdots + r_N^L.$$

Which events are visible in the limit?

Effective limiting dynamics for samples

Again, it is useful to assume the retrospective viewpoint, and trace the history of a finite sample under this scaling (still, full detail of notation omitted here, too cumbersome!).

Assume we sample $n << N$ diploid individuals, carrying two chromosomes each.

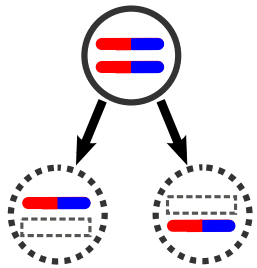
A classification of transitions

We distinguish the following important effective transitions. Their probabilities will lead to a 'separation of time-scales' phenomenon. Convergence can be established by a variant of a Lemma of MÖHLE (1998).

Event 1: A small reproduction event, neither of the parents nor the offspring are in our current sample of the ancestral process. The probability of this event is $O(1)$, but it does not affect our ancestral process.

A classification of transitions

Event 2: A small reproduction event, none of the parents but the offspring in the sample. This happens with probability $O(N^{-1})$, hence, after speeding up time by N^2, will happen 'all the time'. It leads to a complete split of all individuals carrying two active chromosomes into twice as many individuals carrying only one chromosome. Offspring carrying only one active chromosome will not be affected.



One can think of this as a permanent instantaneous projection of the limiting genealogical process on the subset of 'completely split' configurations!

A classification of transitions, cont.

Event 3: A small reproduction event, neither of the parents but one offspring individual with one active chromosome in the sample, on which recombinations acts. This has probability $\approx rN^{-2}$, hence happens with positive rate r in the limit.

Event 4: A small reproduction event, one of the parents and one offspring, both with a single active chromosome (inherited from the parent), no recombination: Again this happens with probability $O(N^{-2})$, and leads to a binary *coalescence* of lineages.

A classification of transitions, cont.

Event 3: A small reproduction event, neither of the parents but one offspring individual with one active chromosome in the sample, on which recombinations acts. This has probability $\approx rN^{-2}$, hence happens with positive rate r in the limit.

Event 4: A small reproduction event, one of the parents and one offspring, both with a single active chromosome (inherited from the parent), no recombination: Again this happens with probability $O(N^{-2})$, and leads to a binary *coalescence* of lineages.

A classification of transitions, cont.

Event 5: A large reproduction event, no parent but (possibly several) offspring in our sample, no recombination. This has probability $O(N^{-2})$, hence happens with positive finite rate in the limit, and leads to a *quadrifold simultaneous multiple merger* (parents distribute their chromosomes uniformly across siblings).

All other events happen with probability $o(N^{-2})$, hence vanish in the limit!

A formal convergence theorem is e.g. via a modification of a result of Möhle (1998).

A classification of transitions, cont.

Event 5: A large reproduction event, no parent but (possibly several) offspring in our sample, no recombination. This has probability $O(N^{-2})$, hence happens with positive finite rate in the limit, and leads to a *quadrifold simultaneous multiple merger* (parents distribute their chromosomes uniformly across siblings).

All other events happen with probability $o(N^{-2})$, hence vanish in the limit!

A formal convergence theorem is e.g. via a modification of a result of Möhle (1998).

A classification of transitions, cont.

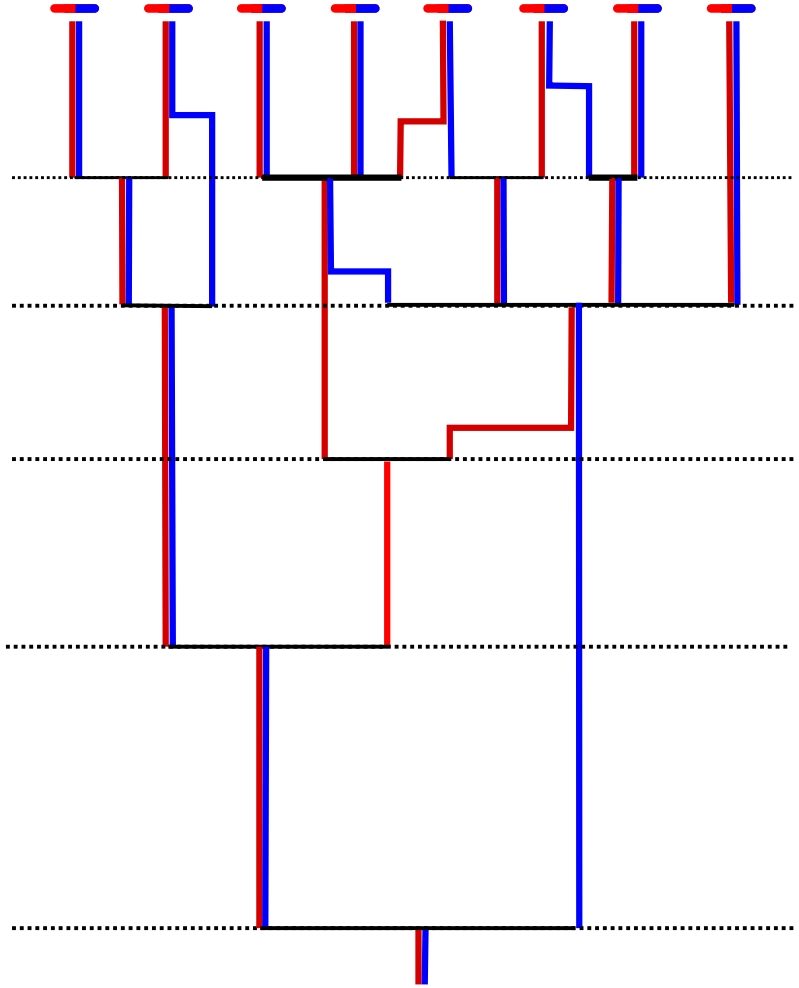
Event 5: A large reproduction event, no parent but (possibly several) offspring in our sample, no recombination. This has probability $O(N^{-2})$, hence happens with positive finite rate in the limit, and leads to a *quadrifold simultaneous multiple merger* (parents distribute their chromosomes uniformly across siblings).

All other events happen with probability $o(N^{-2})$, hence vanish in the limit!

A formal convergence theorem is e.g. via a modification of a result of Möhle (1998).

A visualisation of an ancestry

Start after a complete split, consider only two loci (red, blue):



Remarks

- Coalescences governed by a 'uniform' quadrifold Ξ-coalescent derived from the Eldon & Wakeley coalescent

$$\Lambda(dx) = \delta_0 + c\psi^2\delta_\psi.$$

- Note that the single-locus marginals of the haploid model lead to a genealogy governed by the above Λ-coalescent.
- It is of course possible to modify the reproduction mechanism to obtain general Λ-coalescents (e.g. by randomizing the value of ψ in each step), but notationally too cumbersome for this talk, see the BBE 2012 Preprint.

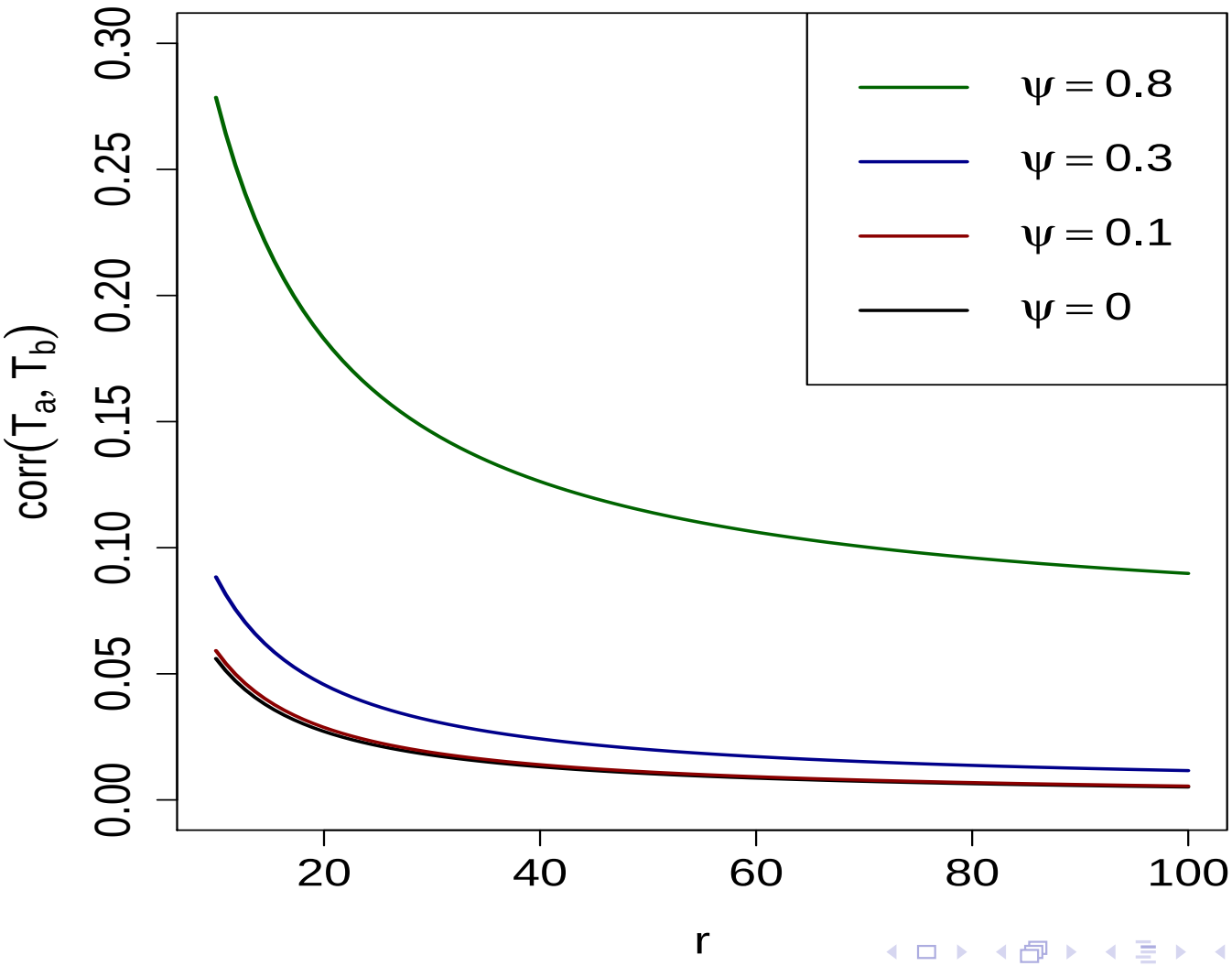
Decorrelation of different loci?

In a classical (Kingman-based) scenario, genealogies of different loci should decorrelate as the recombination rate r approaches ∞.

However, as pointed out by JAY TAYLOR, we do not expect this in a multiple merger scenario.

Behaviour as $r \to \infty$

Here is a simulation of the time to the most recent common ancestro in a two-locus model, which exhibits this effect:



Further extensions

Many further extensions possible (and necessary)...

- random ψ,
- age structure,
- interplay with selection, space...,
- inference of recombination rates...

Ongoing project with Matthias Birkner, Bjarki Eldon.

Collaboration envisaged with Einar Arnason's group, with Anja Sturm, and others within DFG SPP 1590.

Last, but not least...

Thank you for your attention!