Three Lectures about :
“Evolutionary Processes and Patterns of Biodiversity”

Lecture 3/3 : Coevolving genes and species

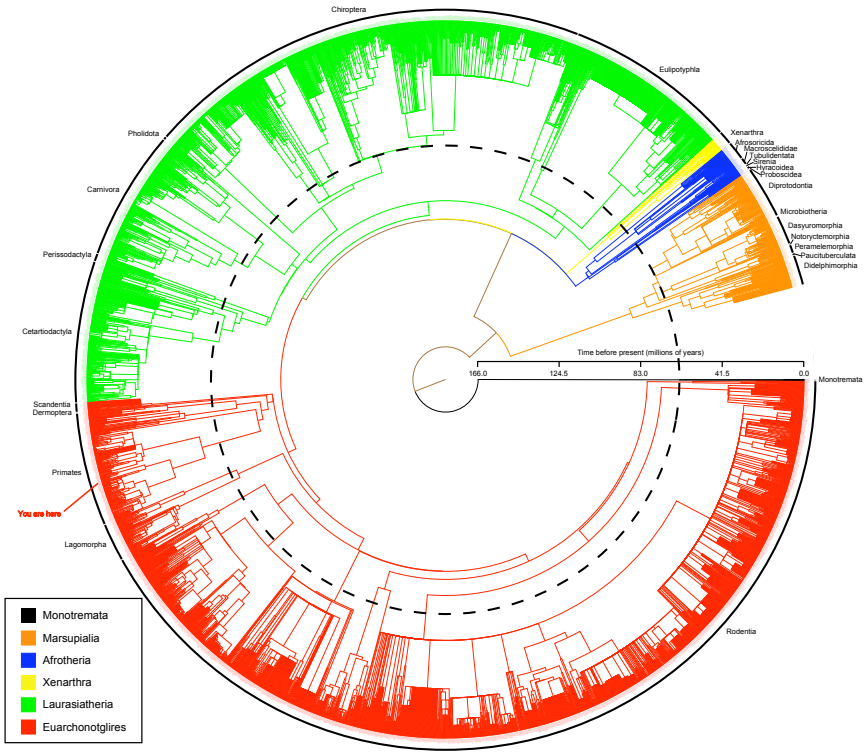
Amaury Lambert

amaury_upmc

IICD & Probability and Society Initiative Joint Seminar Series
Columbia University
October 23, 2020

Outline

Introduction : how to get deeper insight into macro-evolutionary processes

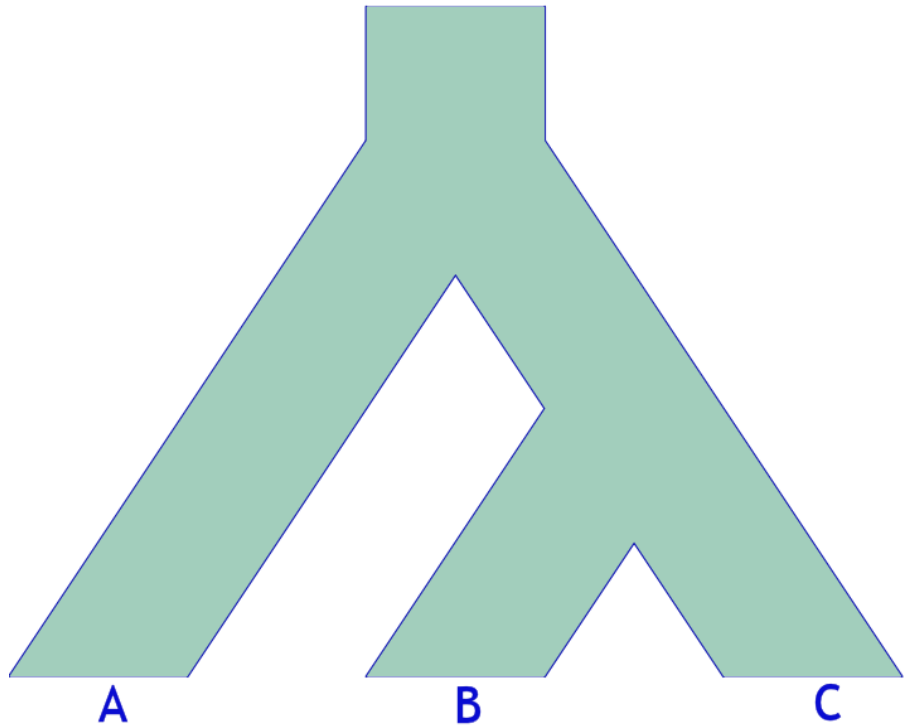


Phylogeny of Mammals
Bininda-Emonds et al *Nature* 2007

- Statistical problems in the phylogenetic approach to diversification (see Lecture 2)
  - False positive associations
    (Rabosky & Goldberg *Syst Biol* 2015)
  - Problems of identifiability
    (Louca & Pennell *Nature* 2020)
- Phylogenies are not big data
- Need to search for additional signal from :
  - Fossils
  - Genomes

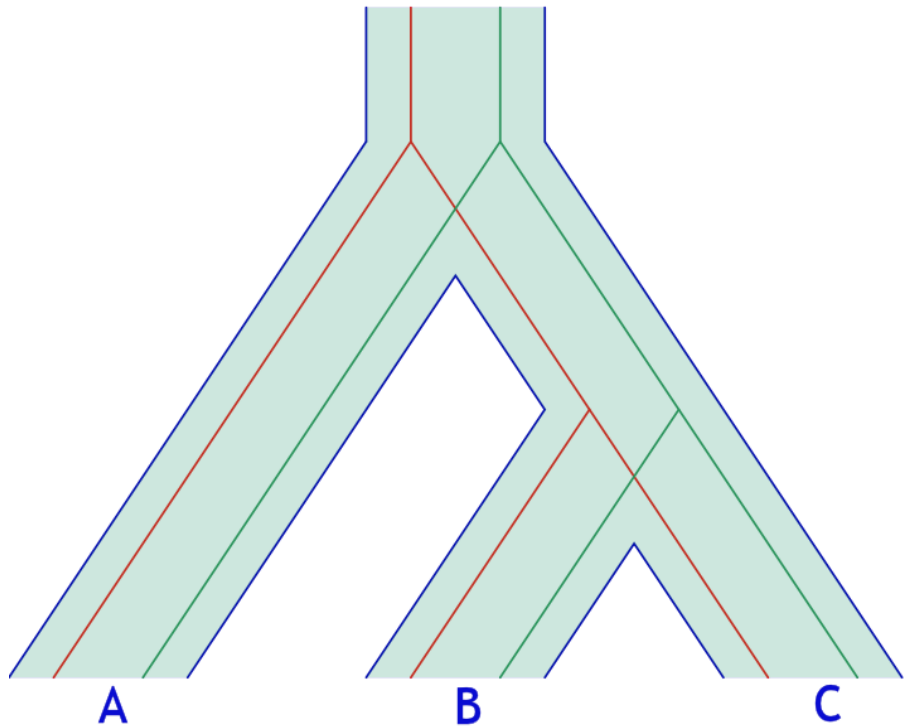
Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences



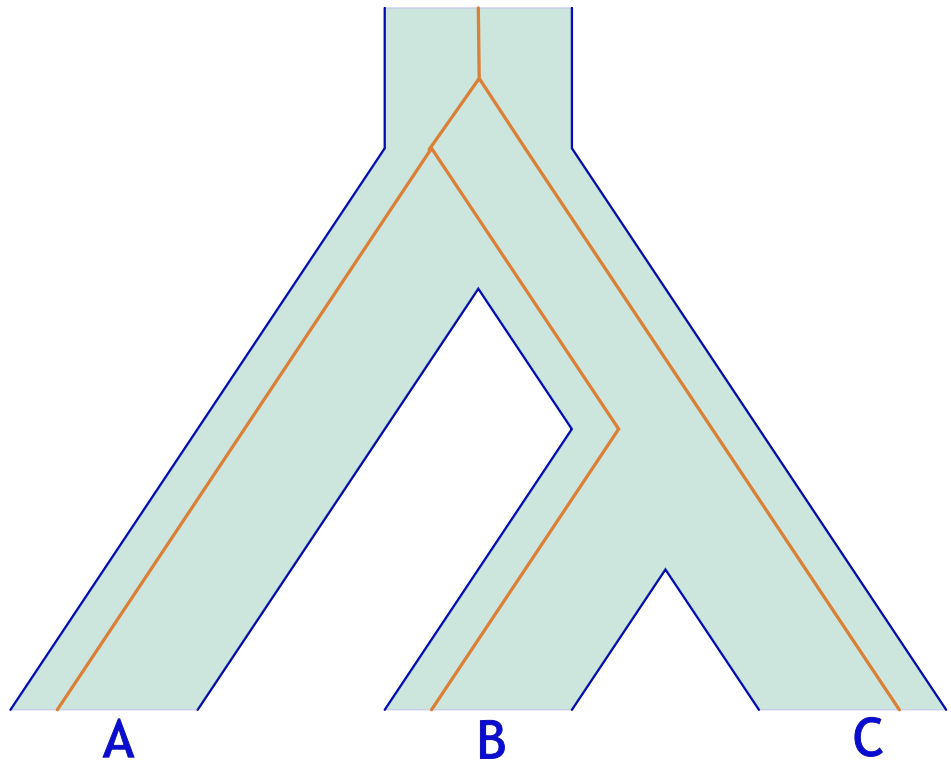
Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree?)



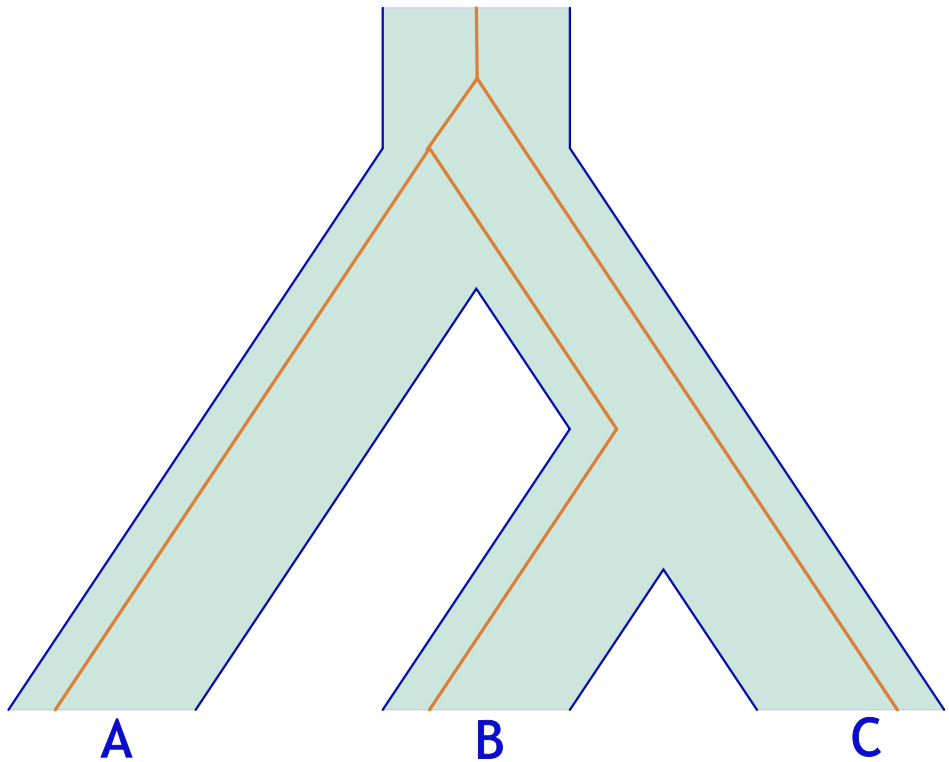
Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree ?)
  - Some do not



Gene trees vs Species tree

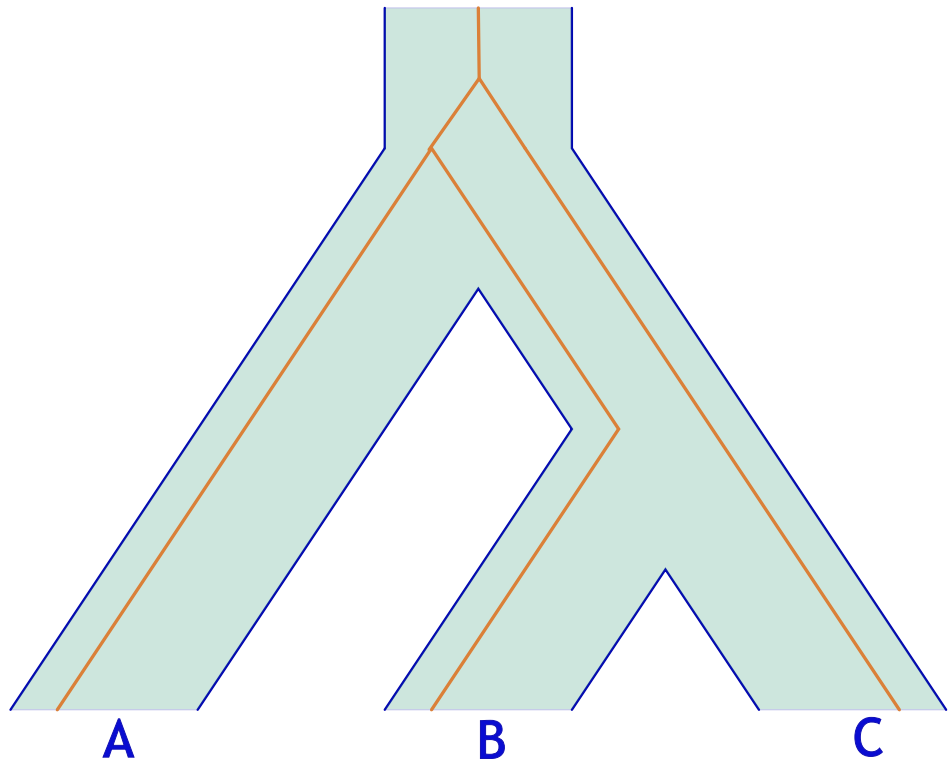
- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree ?)
  - Some do not
- **Q. Can we exploit the signal coming from the diversity of gene trees ?**



Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree ?)
  - Some do not
- **Q. Can we exploit the signal coming from the diversity of gene trees ?**
- **Deep coalescences** due to intrinsic randomness of coalescent or spatial structure

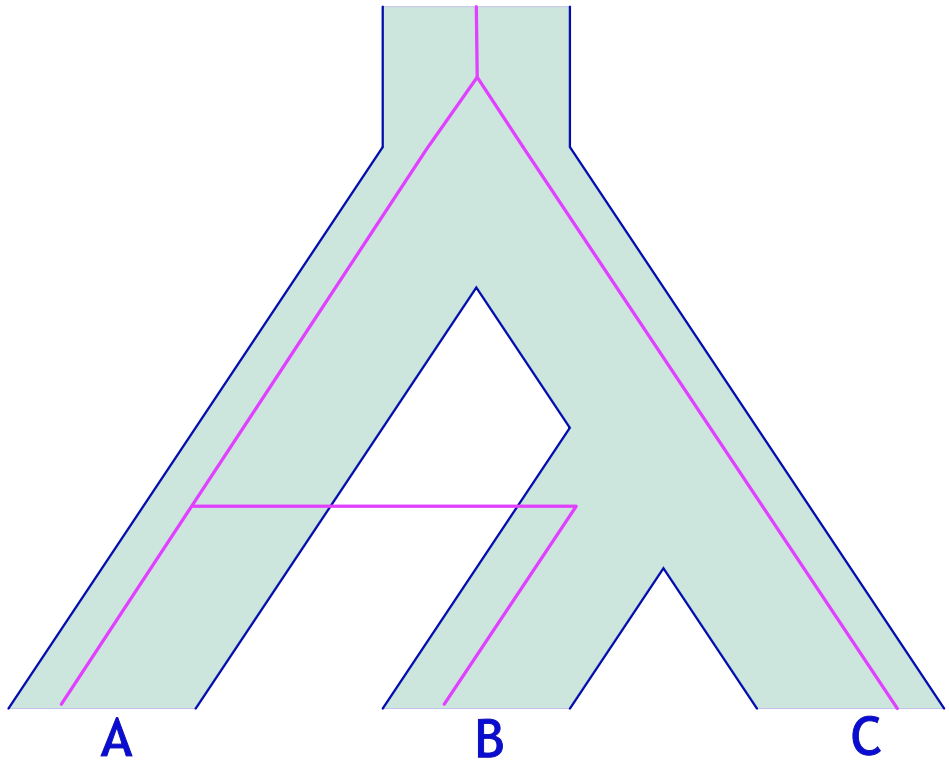
  $\Rightarrow$ Incomplete lineage sorting



Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree ?)
  - Some do not
- **Q. Can we exploit the signal coming from the diversity of gene trees ?**
- **Deep coalescences** due to intrinsic randomness of coalescent or spatial structure

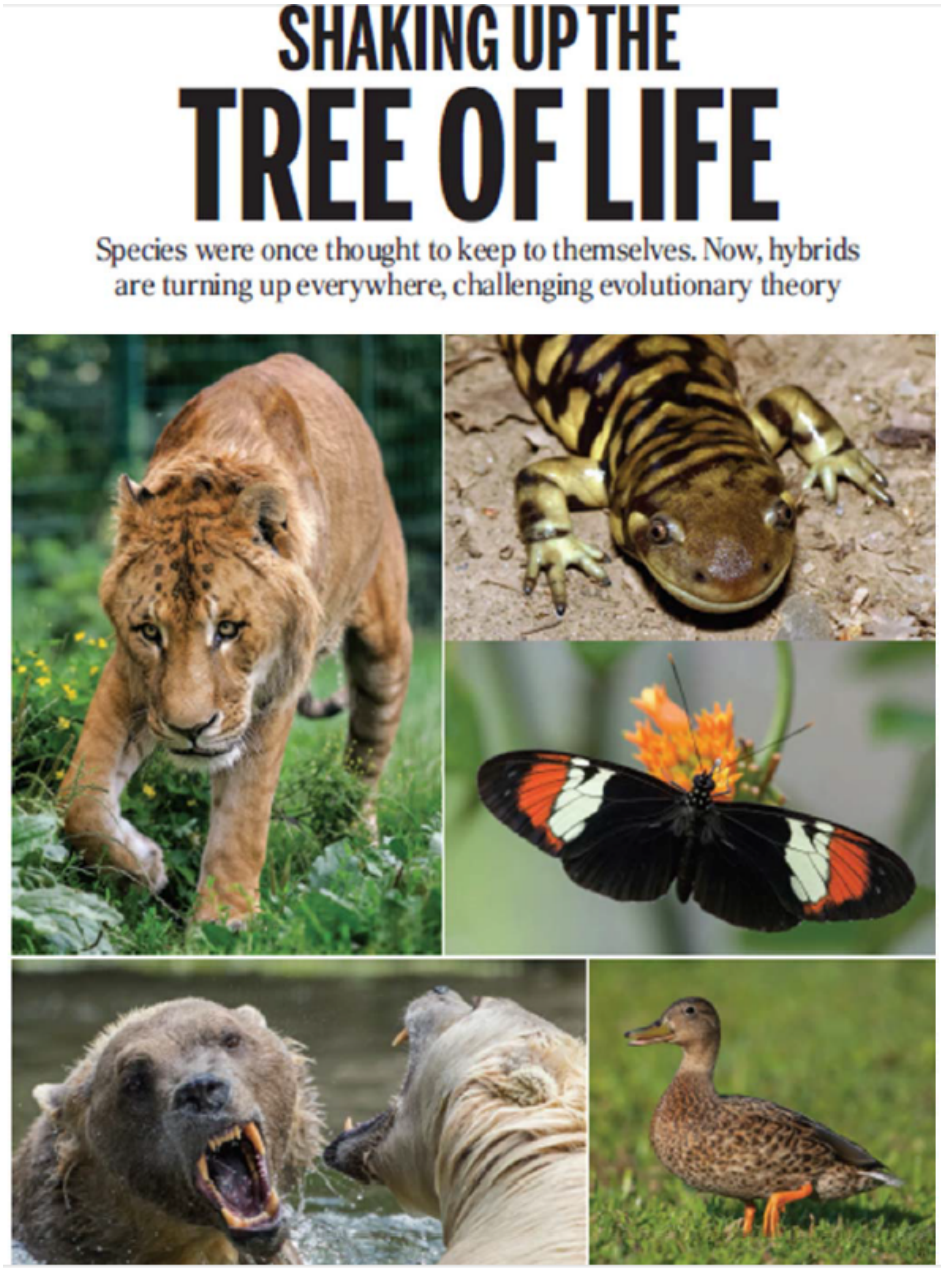
  $\Rightarrow$ Incomplete lineage sorting
- **Shallow coalescences** due to gene flow between species



Gene trees vs Species tree

- Gene trees can be inferred by comparing sequences
  - Some gene trees have the same topology (as putative species tree ?)
  - Some do not
- **Q. Can we exploit the signal coming from the diversity of gene trees ?**
- **Deep coalescences** due to intrinsic randomness of coalescent or spatial structure

  $\Rightarrow$ Incomplete lineage sorting
- **Shallow coalescences** due to gene flow between species



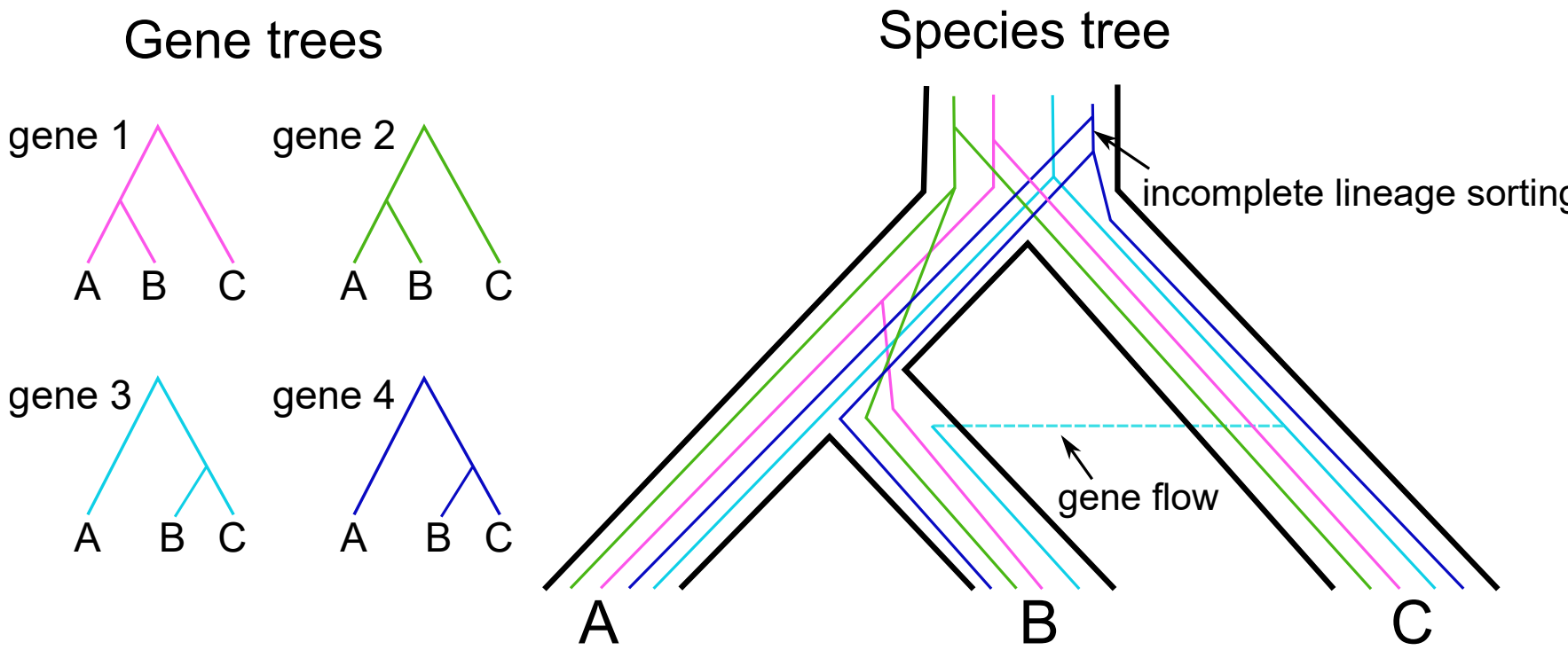
Pennisi *Science* 2016

Outline

- Trees within tree : multispecies coalescent and other nested coalescents
- A model allowing for dynamic spatial structure
- Two models allowing for gene flow between species
- Applications
  - **Q1.** How do nested coalescents come down from infinity ?
  - **Q2.** How to cluster a sample of genomes into putative species

Outline

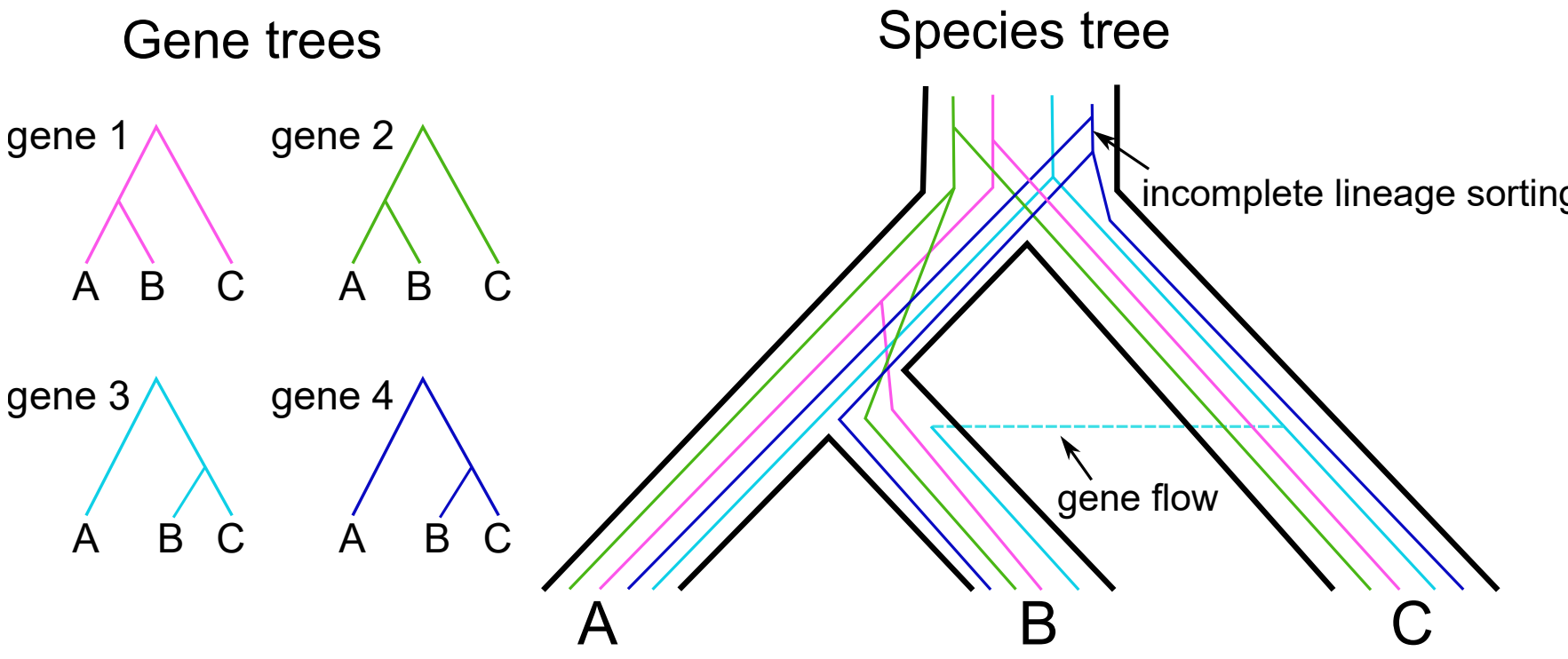
Coupling gene trees and species tree : The Multispecies Coalescent Paradigm



Picture by Julie Marin

- The Multi-Species Coalescent (MSC) : Start from the species tree = fixed or random (e.g., birth-death process), sometimes the unknown to be inferred

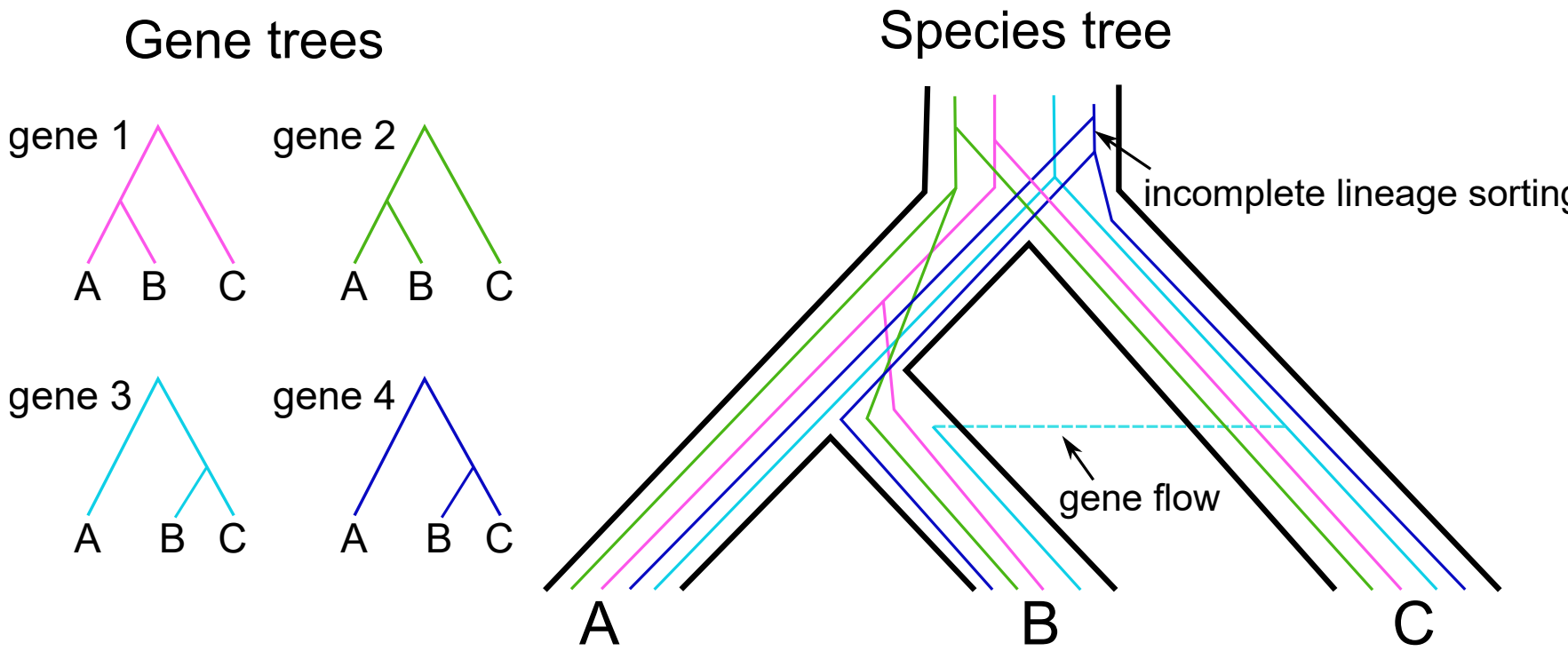
Coupling gene trees and species tree : The Multispecies Coalescent Paradigm



Picture by Julie Marin

- The Multi-Species Coalescent (MSC) : Start from the species tree = fixed or random (e.g., birth-death process), sometimes the unknown to be inferred
- Conditional on the sp tree, gene trees are independent coalescents constrained to reside within sp tree (with occasional gene flow/hybridization events)

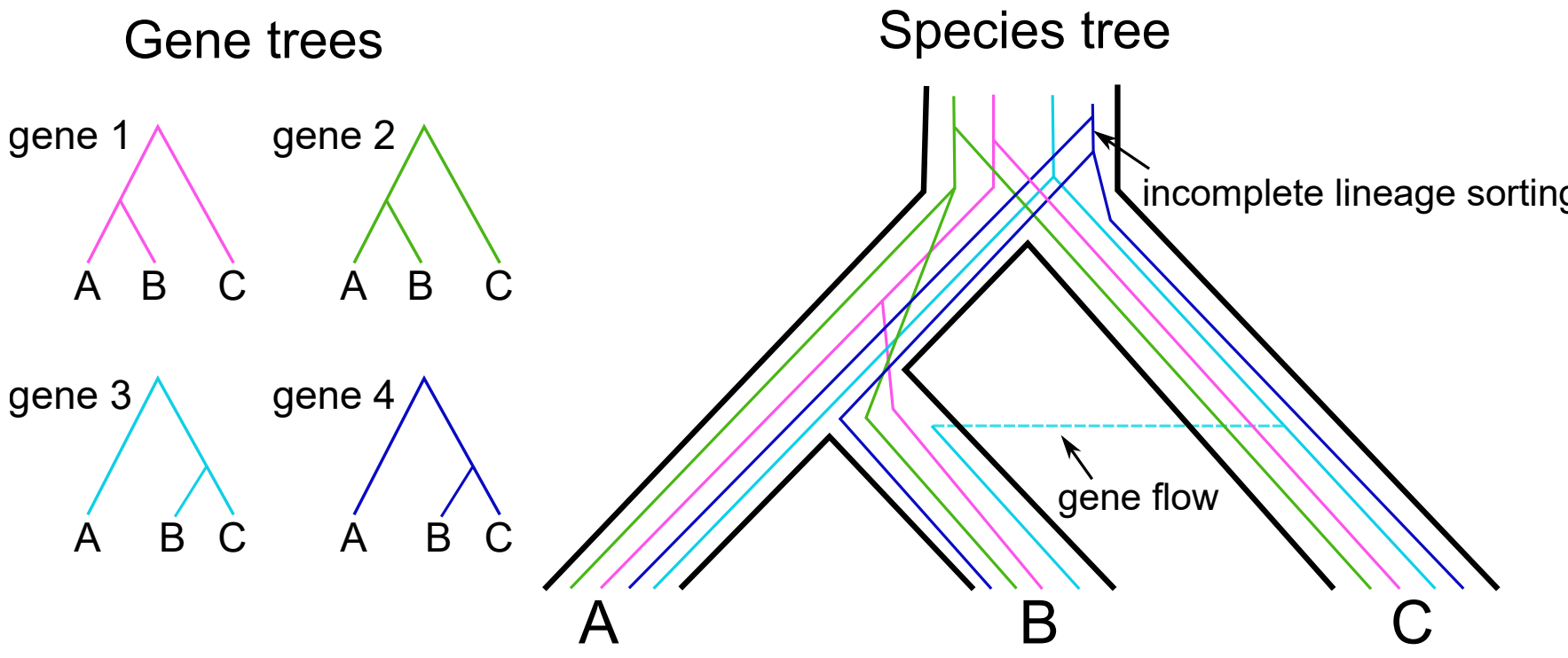
Coupling gene trees and species tree : The Multispecies Coalescent Paradigm



Picture by Julie Marin

- The Multi-Species Coalescent (MSC) : Start from the species tree = fixed or random (e.g., birth-death process), sometimes the unknown to be inferred
- Conditional on the sp tree, gene trees are independent coalescents constrained to reside within sp tree (with occasional gene flow/hybridization events)
- No explicit modeling of physical linkage/recombination

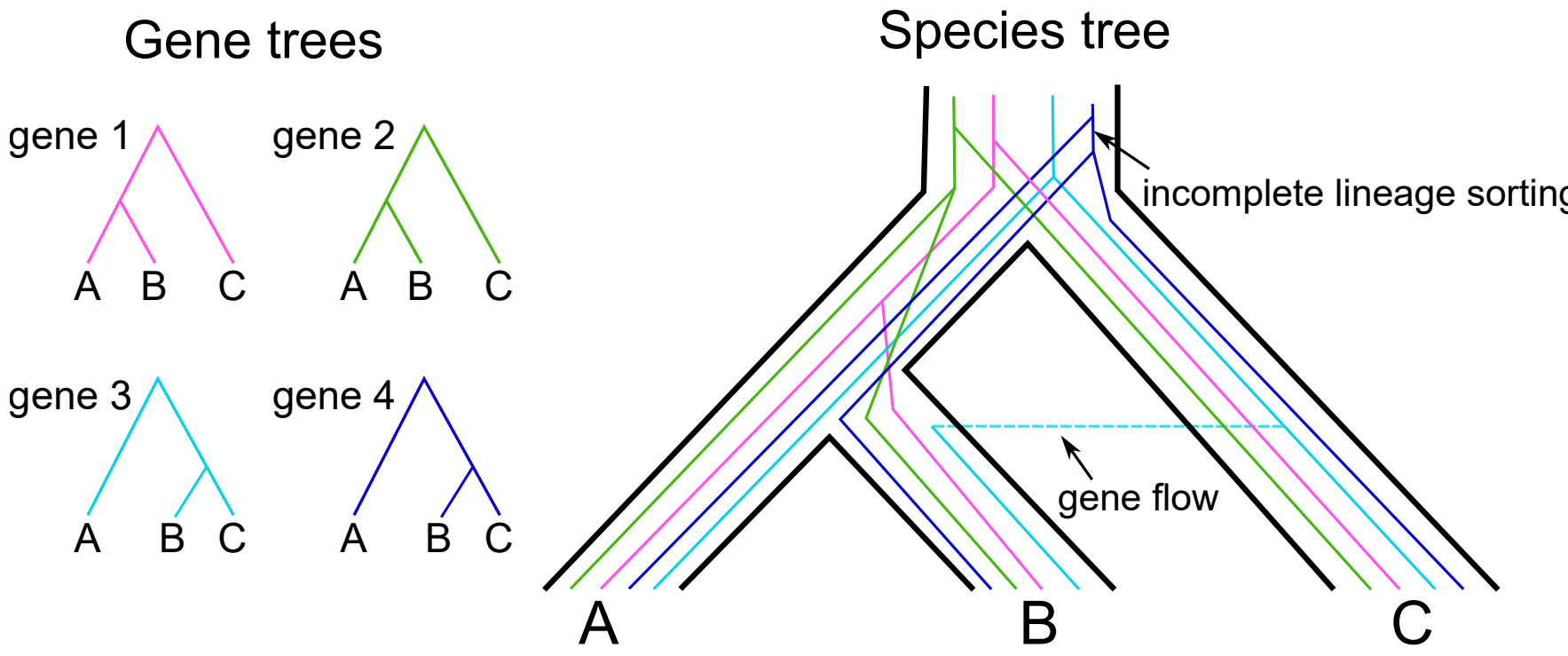
Coupling gene trees and species tree : The Multispecies Coalescent Paradigm



Picture by Julie Marin

- The Multi-Species Coalescent (MSC) : Start from the species tree = fixed or random (e.g., birth-death process), sometimes the unknown to be inferred
- Conditional on the sp tree, gene trees are independent coalescents constrained to reside within sp tree (with occasional gene flow/hybridization events)
- No explicit modeling of physical linkage/recombination
- Disagreements among gene trees mainly due to randomness of coalescent process

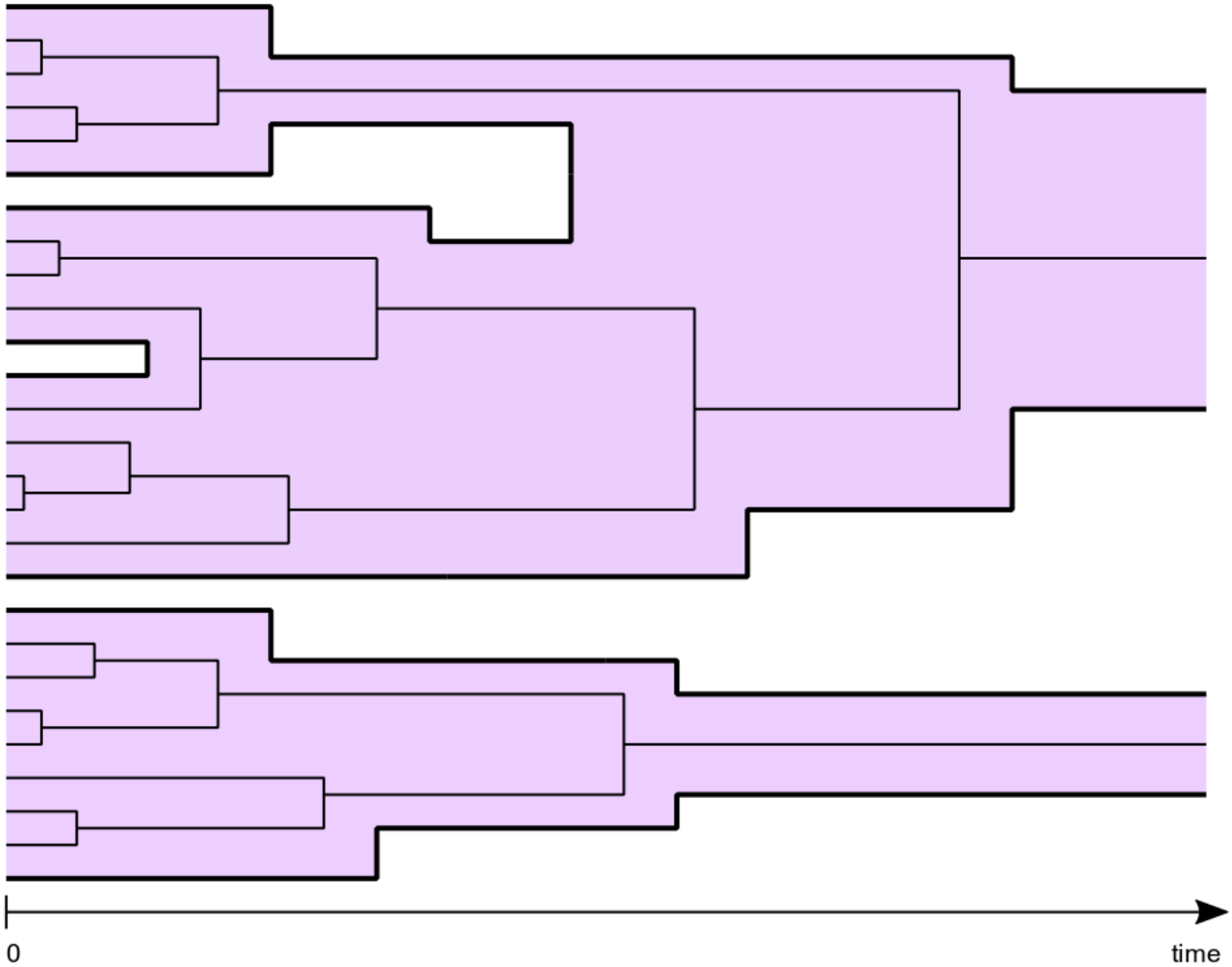
Coupling gene trees and species tree : The Multispecies Coalescent Paradigm



Picture by Julie Marin

- The Multi-Species Coalescent (MSC) : Start from the species tree = fixed or random (e.g., birth-death process), sometimes the unknown to be inferred
- Conditional on the sp tree, gene trees are independent coalescents constrained to reside within sp tree (with occasional gene flow/hybridization events)
- No explicit modeling of physical linkage/recombination
- Disagreements among gene trees mainly due to randomness of coalescent process
- Example : if sp tree given by Kingman's coalescent, the pair gene tree/species tree is called the nested Kingman coalescent, or simply 'Kingman in Kingman'...

Kingman in Kingman



Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \ldots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.
- Consider a Poisson point process (PPP) of intensity $dt\, d\nu$

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \dots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.
- Consider a Poisson point process (PPP) of intensity $dt\, d\nu$
- At each atom (t, x) of the PPP, paint each block of $\Pi_n(t-)$, independently w proba x...

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \dots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.
- Consider a Poisson point process (PPP) of intensity $dt\, d\nu$
- At each atom (t, x) of the PPP, paint each block of $\Pi_n(t-)$, independently w proba x...
- ...And merge all painted blocks into one single block.

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \dots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.
- Consider a Poisson point process (PPP) of intensity $dt\, d\nu$
- At each atom (t, x) of the PPP, paint each block of $\Pi_n(t-)$, independently w proba x...
- ...And merge all painted blocks into one single block.
- A partition into b blocks sees a given subset of $k \geq 2$ blocks merge at rate

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

Λ-coalescents

Sagitov, Möhle, Pitman, Schweinsberg, Berestycki, Bertoin, Le Gall...

- A Λ-coalescent is a Markov process $\Pi_n(t)$ valued in the partitions of $\{1, \dots, n\}$ which is
  - nondecreasing (partitions get coarser and coarser with time)
  - exchangeable, sampling-consistent (coupling all n)
  - one merger at a time.
- Let Λ be a finite measure on $[0, 1]$ and set $\nu(dx) = x^{-2}\Lambda(dx)$.
- Consider a Poisson point process (PPP) of intensity $dt\, d\nu$
- At each atom (t, x) of the PPP, paint each block of $\Pi_n(t-)$, independently w proba x...
- ...And merge all painted blocks into one single block.
- A partition into b blocks sees a given subset of $k \geq 2$ blocks merge at rate

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

- ...corresponding to Kingman coalescent when $\Lambda = \delta_0$.

Nested coalescents

- A nested coalescent describes the coagulation dynamics of genes contained in species
- Sample DNA at the same locus in each individual labelled $1, \ldots, n$ and define $\Pi^g(t)$ and $\Pi^s(t)$ by

  i and j are in same block of $\Pi^g(t) \iff$ the gene lineages of i and j have coalesced before t

  i and j are in same block of $\Pi^s(t) \iff$ the species lineages of i and j have coalesced before t
- In the absence of gene flow, gene lineages can coalesce only if $\in$ the same species

  $$\Rightarrow \Pi^g(t) \textbf{ is finer than } \Pi^s(t).$$

- A nested coalescent is said **simple** if both Π^g and Π^s see only one merger at a time – but possibly the same time.

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*

- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*
  - *For each painted species block i indep, draw a rv $Y_i \in [0,1]$ according to μ*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*
  - *For each painted species block i indep, draw a rv $Y_i \in [0,1]$ according to μ*
  - *Conditional on $Y_i = y_i$, paint each gene block of species i indep with proba y_i*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*
  - *For each painted species block i indep, draw a rv $Y_i \in [0,1]$ according to μ*
  - *Conditional on $Y_i = y_i$, paint each gene block of species i indep with proba y_i*
  - *Merge all painted species blocks together, merge all painted gene blocks together.*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*
  - *For each painted species block i indep, draw a rv $Y_i \in [0,1]$ according to μ*
  - *Conditional on $Y_i = y_i$, paint each gene block of species i indep with proba y_i*
  - *Merge all painted species blocks together, merge all painted gene blocks together.*

Simple nested coalescents

Blancas, Duchamps, Lambert & Siri–Jégousse, "Trees within trees : Simple nested coalescents", *EJP* (2018)

Theorem (Blancas, Duchamps, Lambert & Siri-Jégousse (2018))

A simple nested coalescent can be characterized by the following three types of mechanisms :

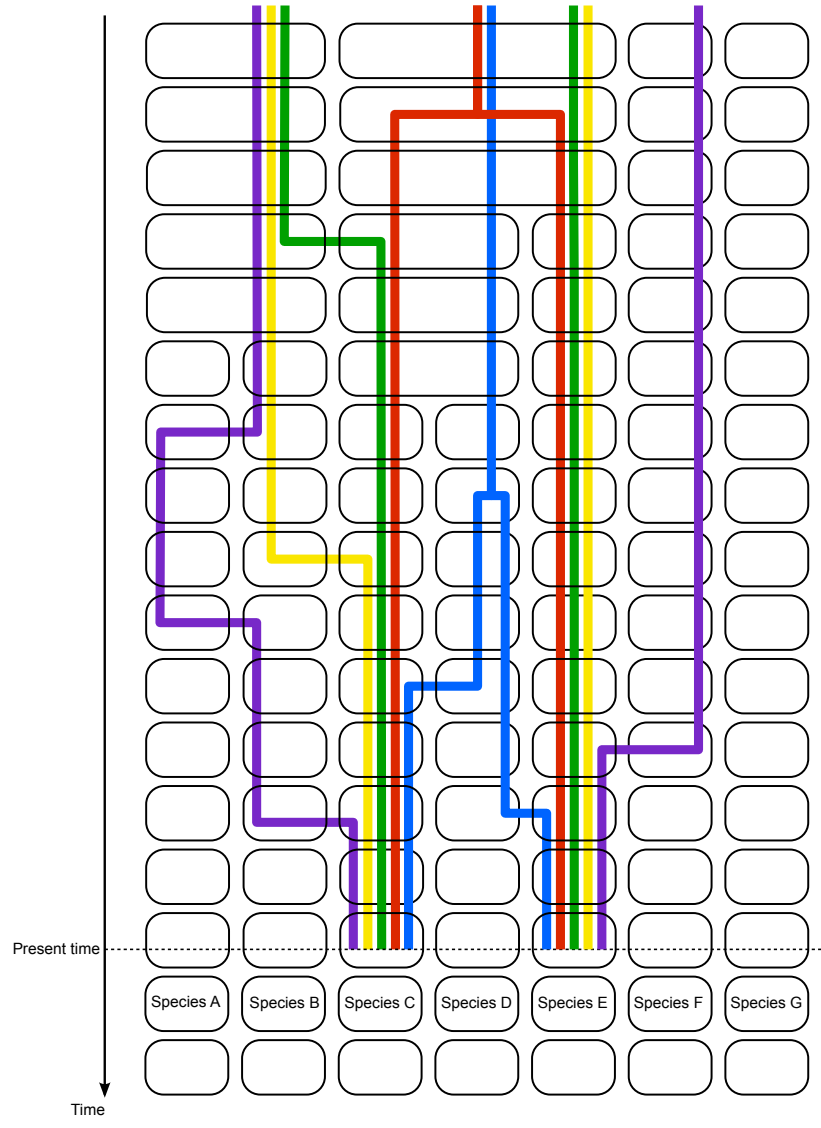
- *There is a measure Λ_s on $[0,1]$ such that species undergo Λ-coalescent-type events according to Λ_s, during which genes do nothing.*
- *There is a measure Λ_g on $[0,1]$ such that in each species block independently, gene blocks contained in this species undergo Λ-coalescent-type events according to Λ_g, during which the species does nothing.*
- *There is a measure ν_{gs} on $(0,1] \times \mathcal{M}_1([0,1])$ driving simultaneous coalescence of genes and species, according to the following rule : at each atom (x, μ) of a PPP with intensity $dt\, d\nu_{gs}$,*
  - *Paint each species block indep with proba x*
  - *For each painted species block i indep, draw a rv $Y_i \in [0,1]$ according to μ*
  - *Conditional on $Y_i = y_i$, paint each gene block of species i indep with proba y_i*
  - *Merge all painted species blocks together, merge all painted gene blocks together.*

For simple nested fragmentations, see :

Duchamps, JJ "Trees within trees II : Nested fragmentations", *Ann Inst H Poincaré Probab Stat* (2020).

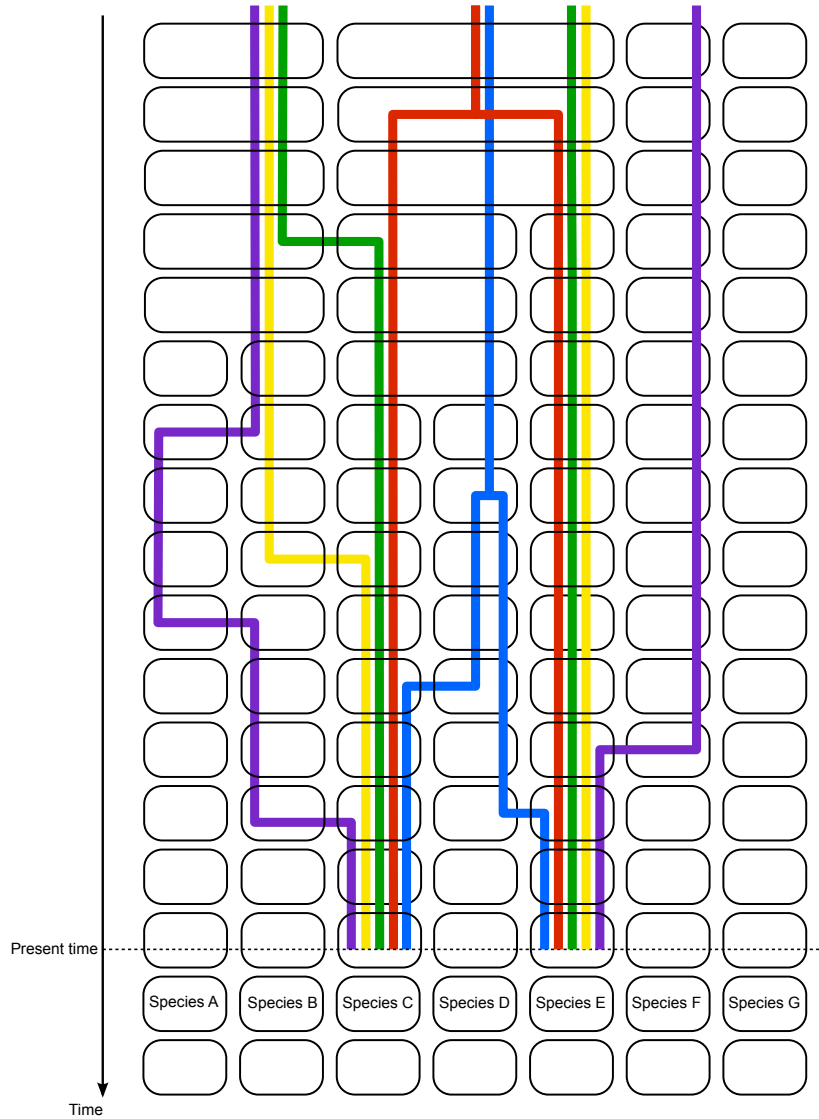
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...



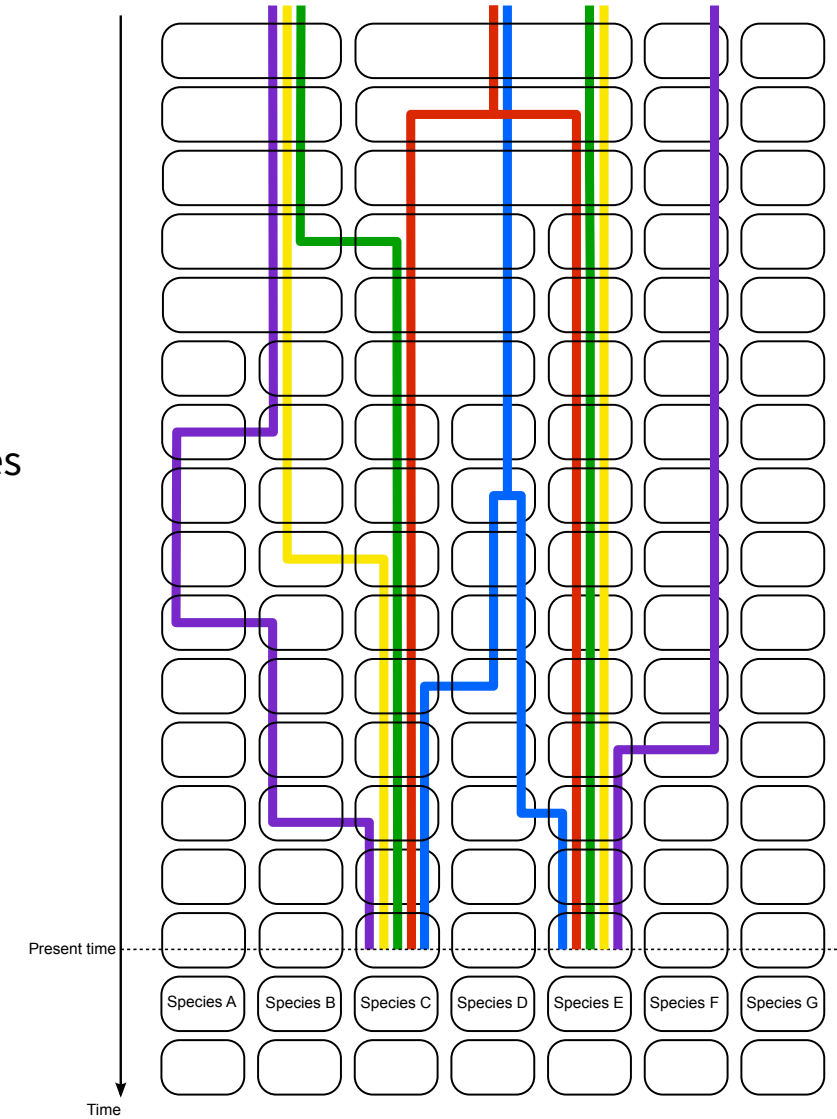
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)



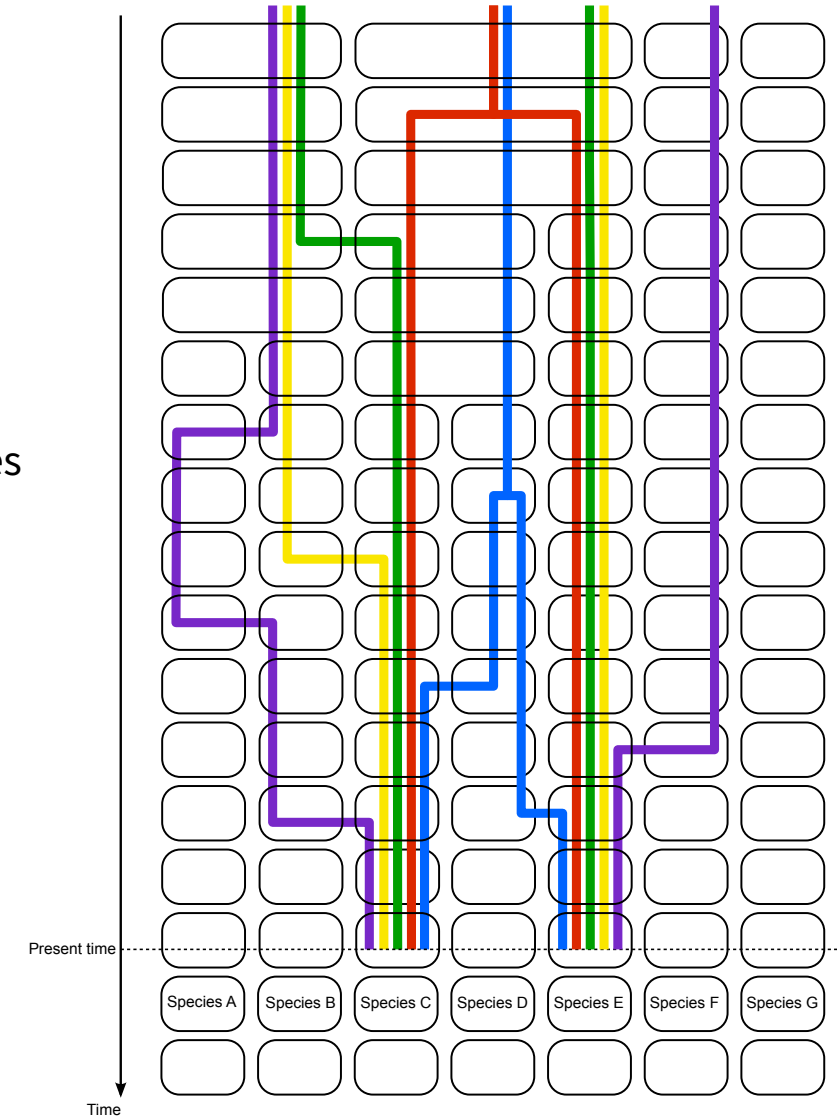
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement



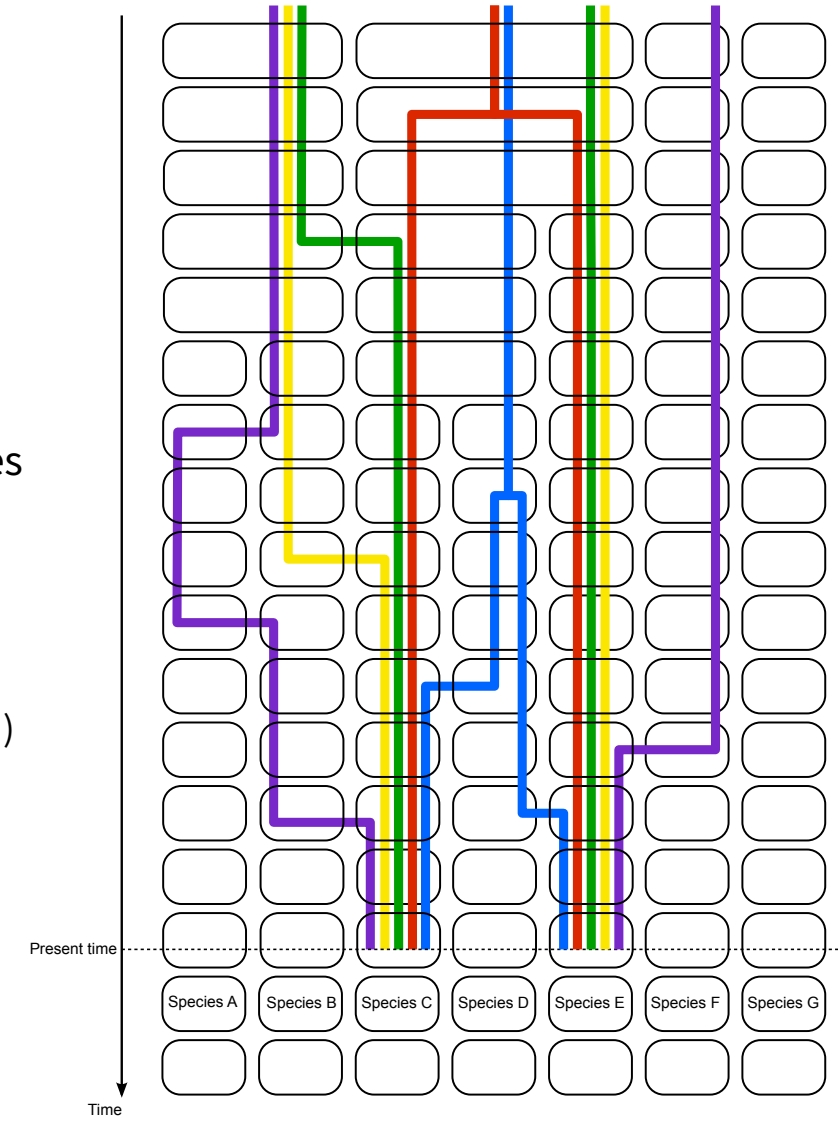
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging



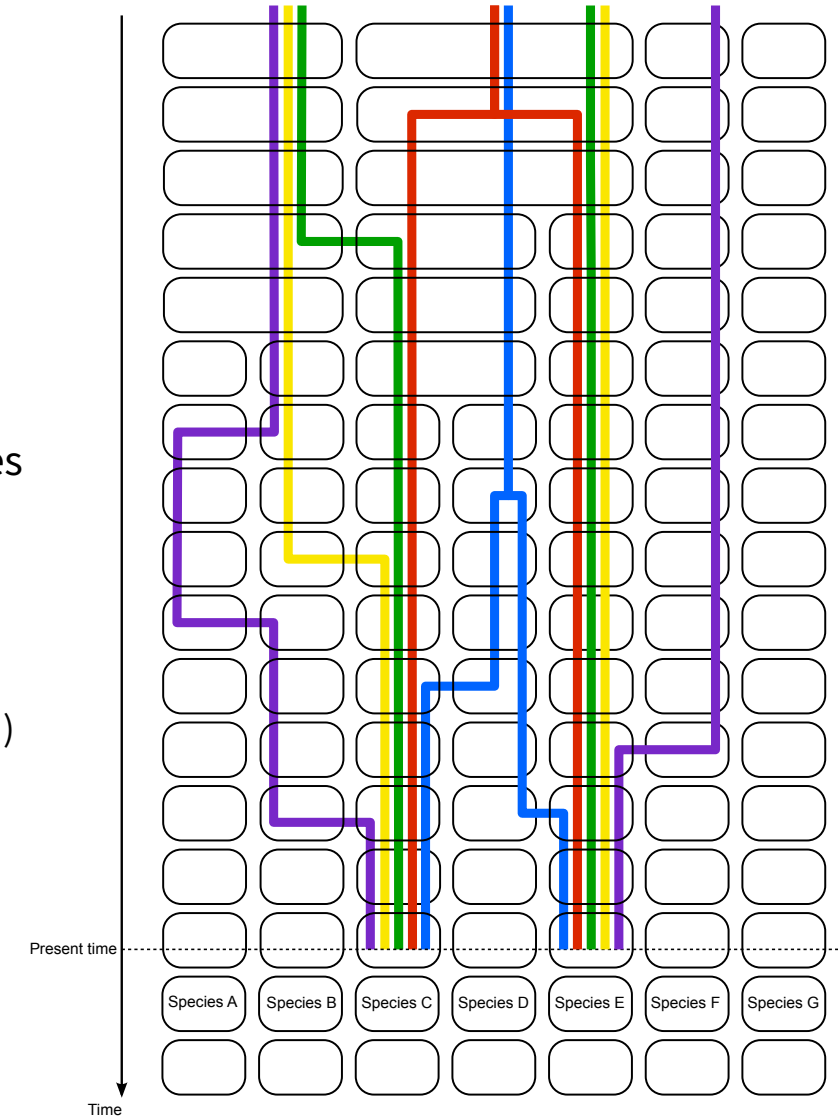
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)



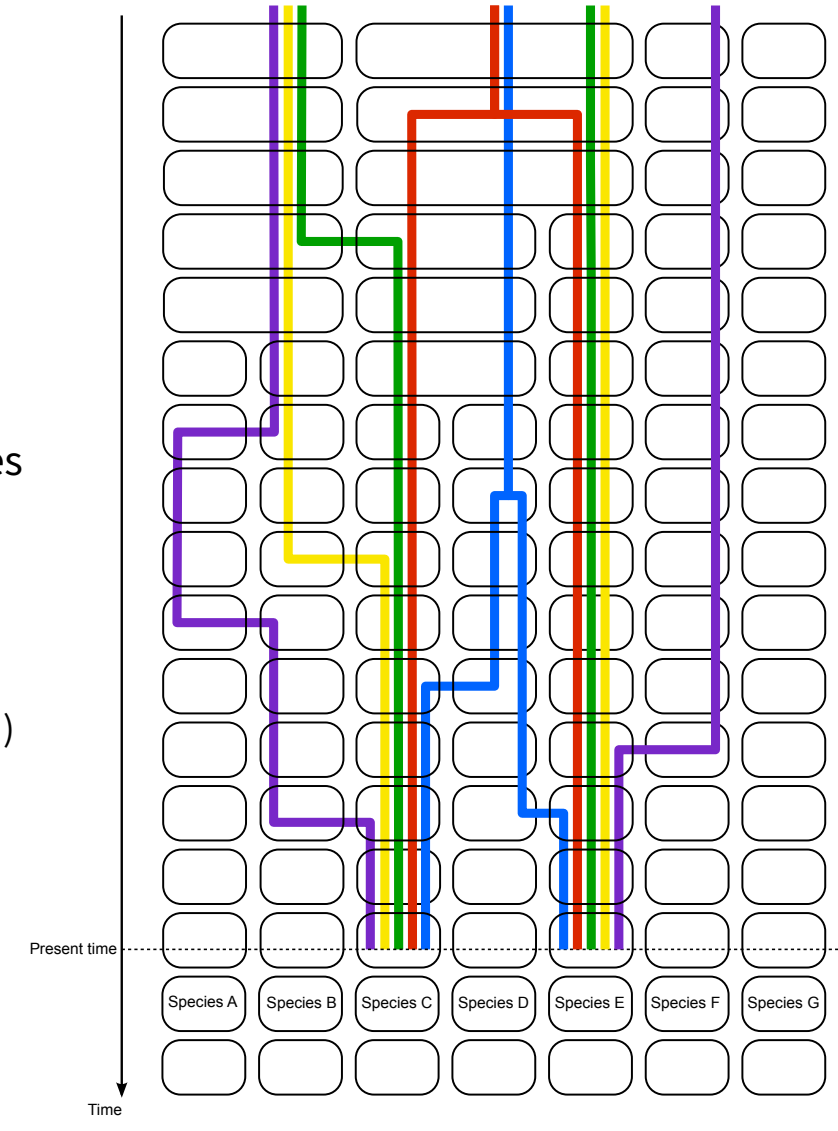
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)
  - Or species do exchange genes continuously (Models 2 and 3)



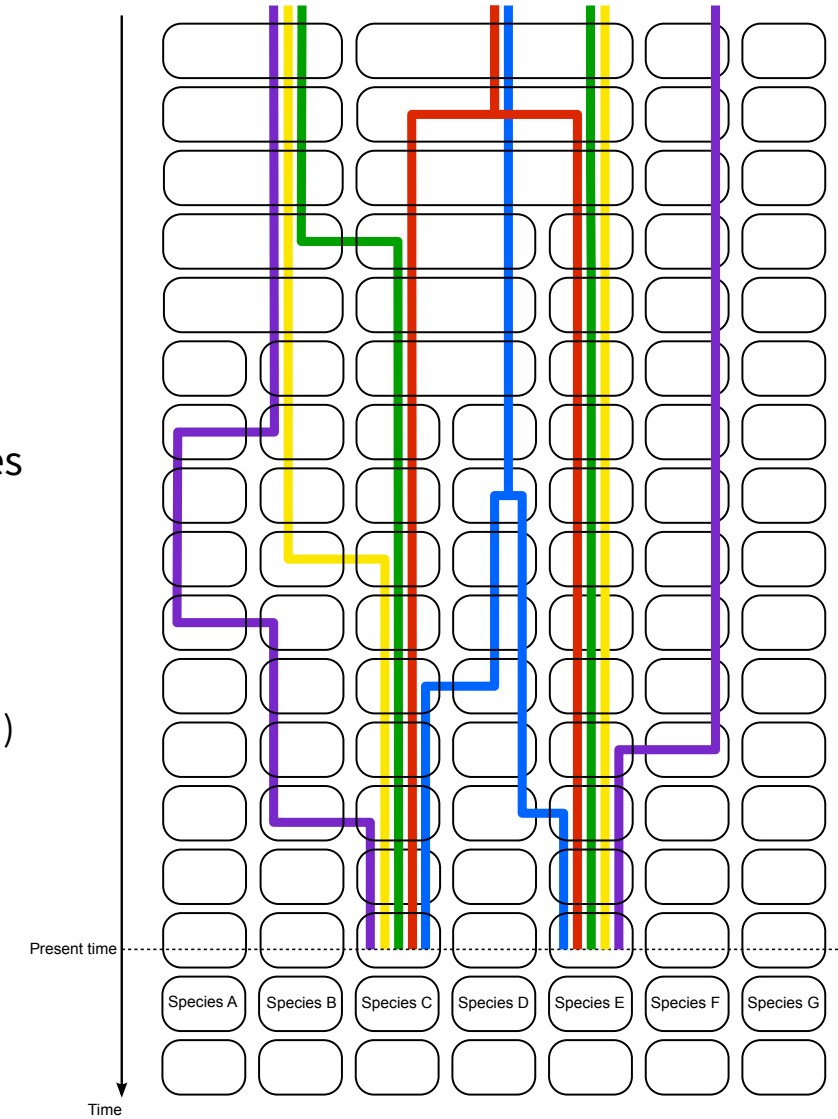
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)
  - Or species do exchange genes continuously (Models 2 and 3)
- Three Models



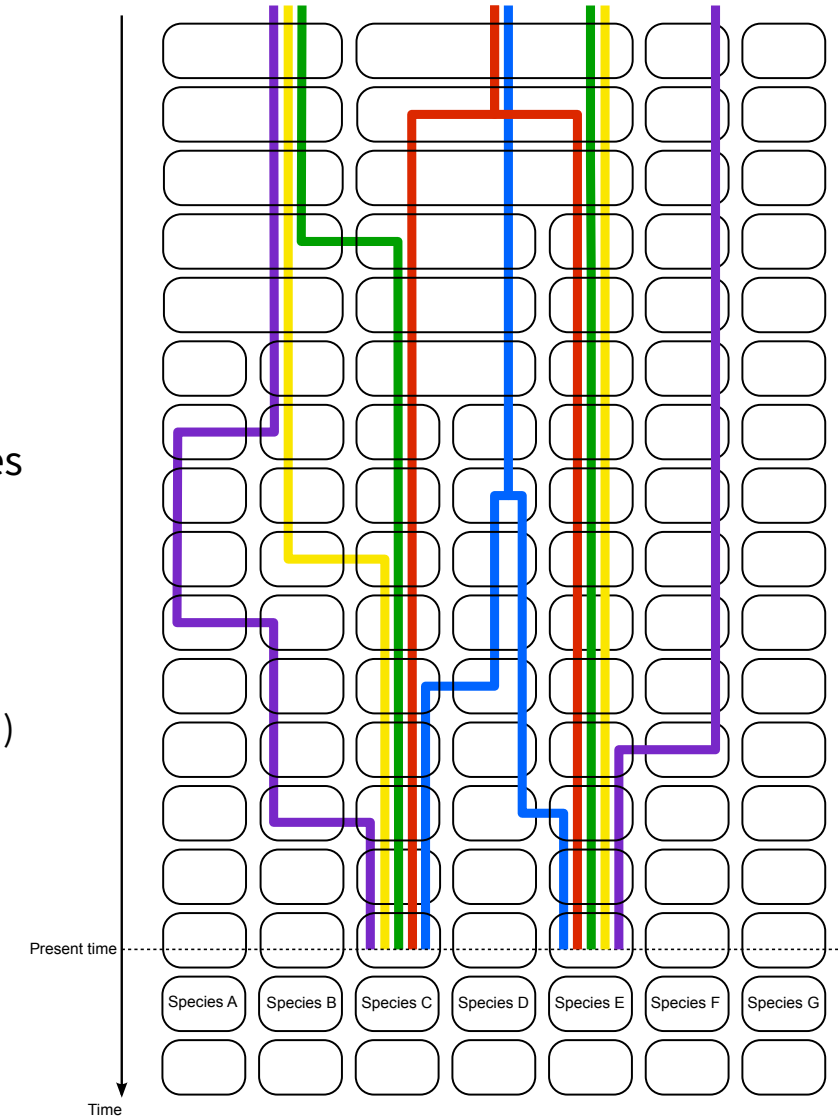
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)
  - Or species do exchange genes continuously (Models 2 and 3)
- Three Models
  - Model 1. One locus, *n* individuals sampled



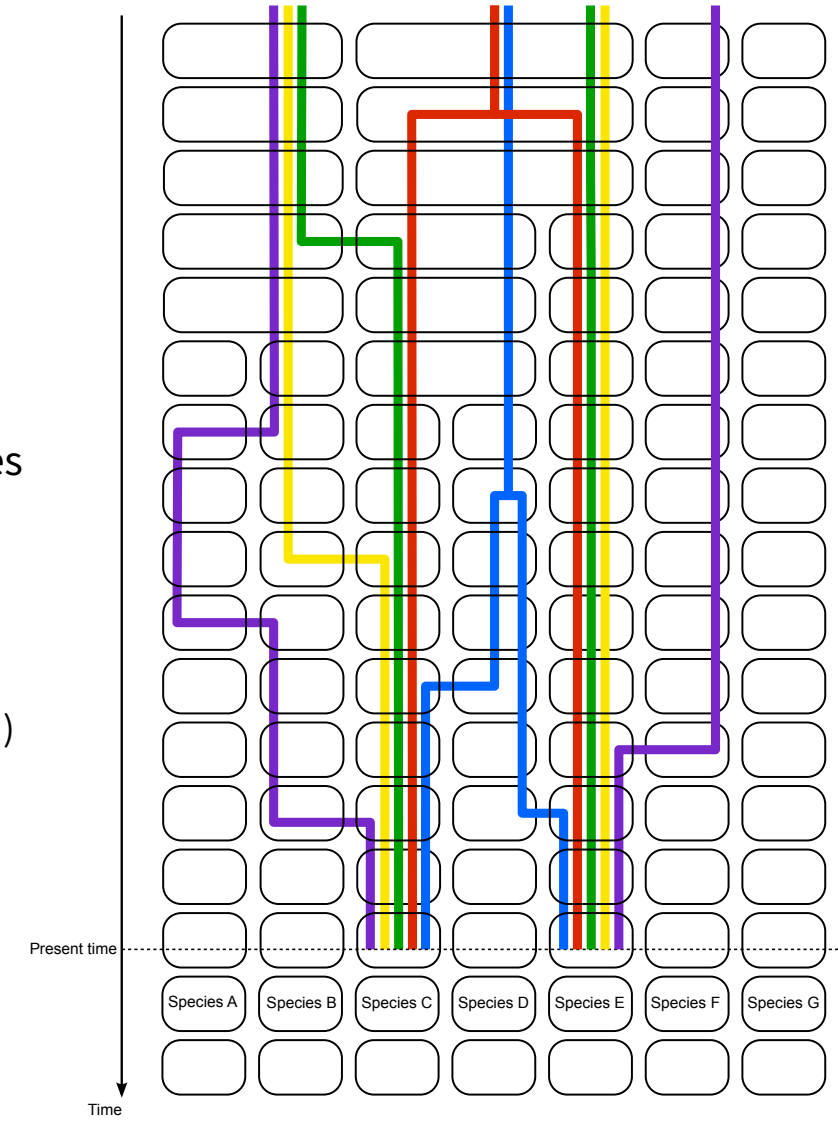
What about several gene trees?

- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)
  - Or species do exchange genes continuously (Models 2 and 3)
- Three Models
  - Model 1. One locus, *n* individuals sampled
  - Model 2. Several loci, one genome sampled.



What about several gene trees?

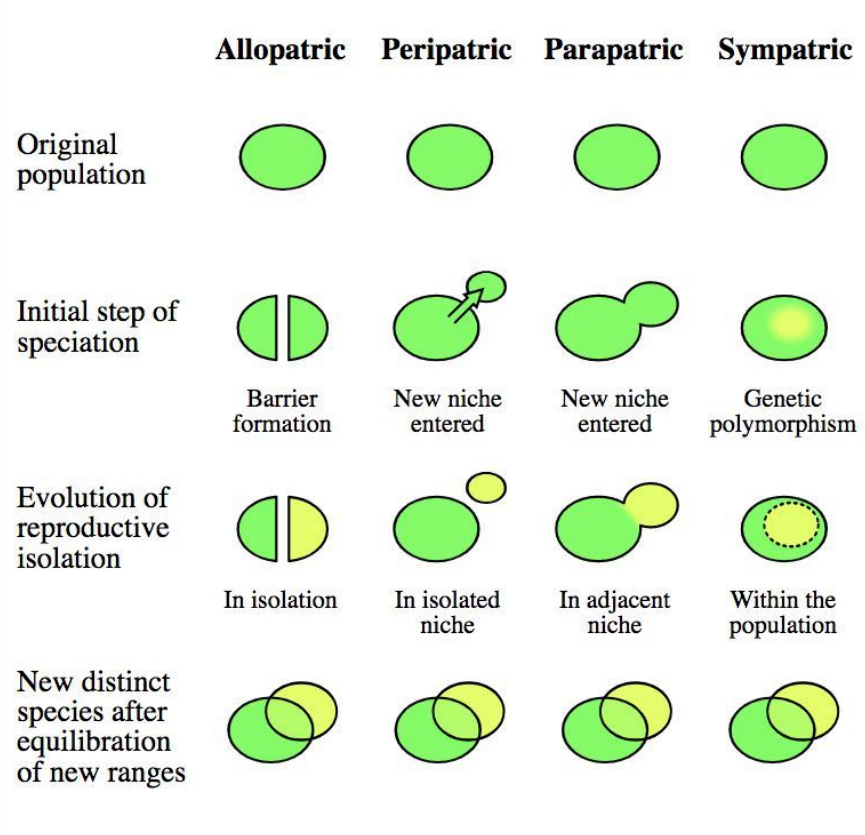
- In the MSC, speciation events are instantaneous and gene trees are shaped by the species tree...
- In reality, speciation takes time, is a by-product of genetic differentiation (see Lecture 2)
- Gene flow and/or spatial structure may be important causes of gene tree disagreement
- We need models of gene trees/species tree acknowledging
  - Speciation takes place in a spatially dynamic context (Model 1)
  - Or species do exchange genes continuously (Models 2 and 3)
- Three Models
  - Model 1. One locus, *n* individuals sampled
  - Model 2. Several loci, one genome sampled.
  - Model 3. Several loci, *n* genomes sampled.



Outline

A dynamic model of peripatric metapopulation

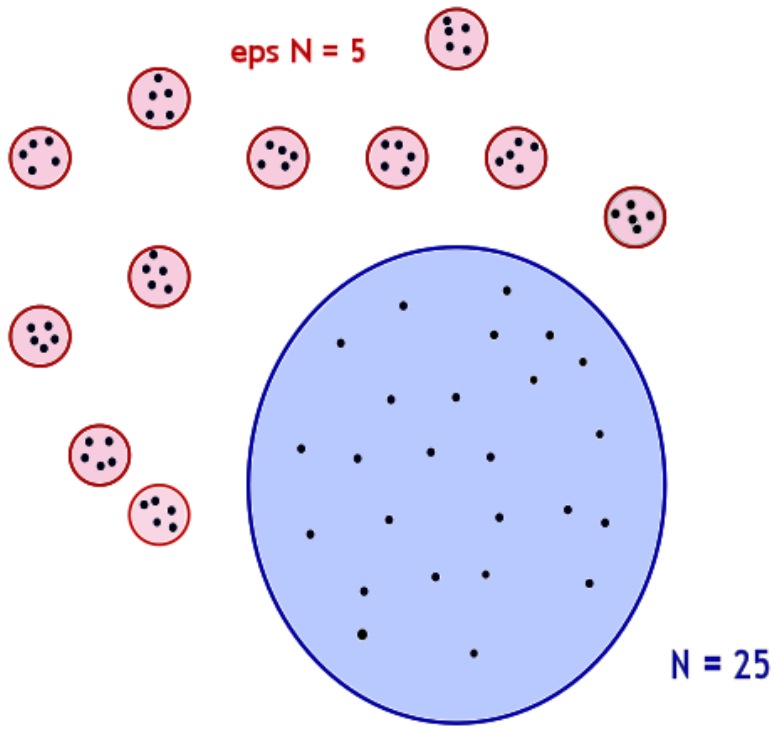
- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)



A dynamic model of peripatric metapopulation

- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)

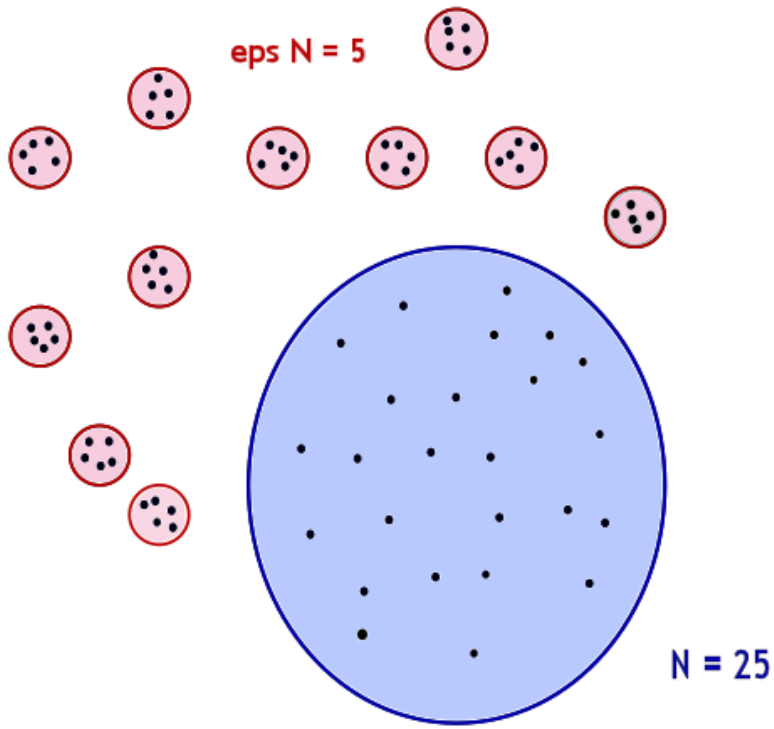
- Assumptions :



A dynamic model of peripatric metapopulation

- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)

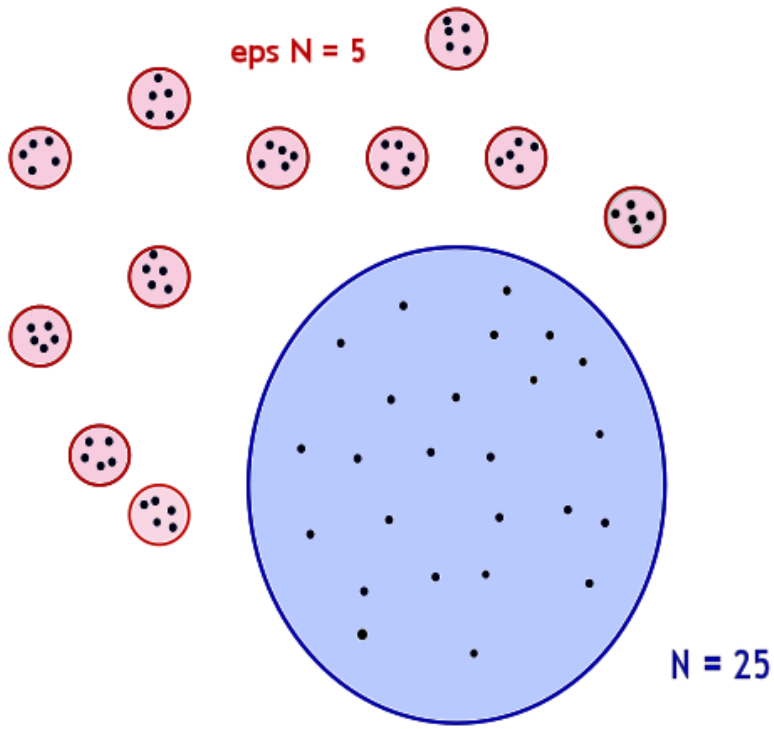
- Assumptions :
  - One **main population** of size *N*



A dynamic model of peripatric metapopulation

- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)

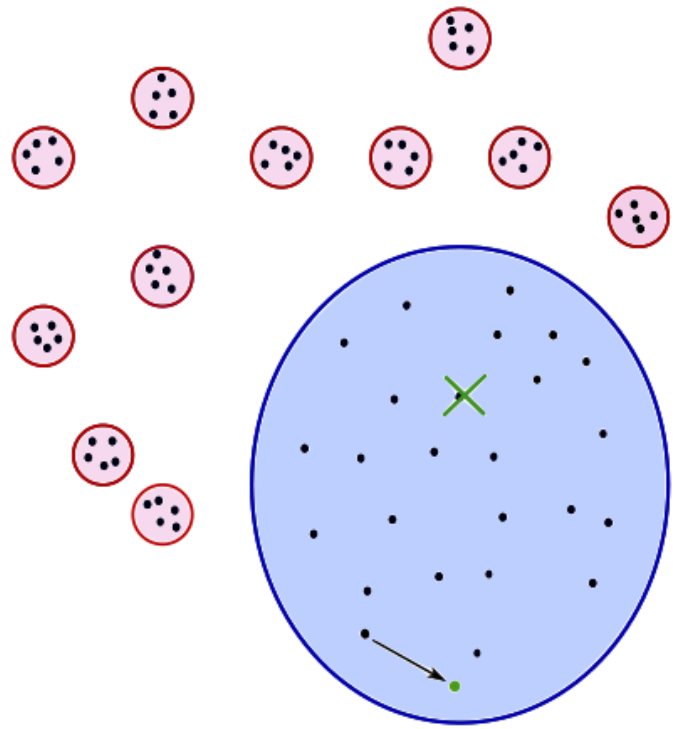
- Assumptions :
  - One **main population** of size N
  - Surrounded by ξ_N (stochastic) **small colonies** of size $\varepsilon_N N$



A dynamic model of peripatric metapopulation

- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)

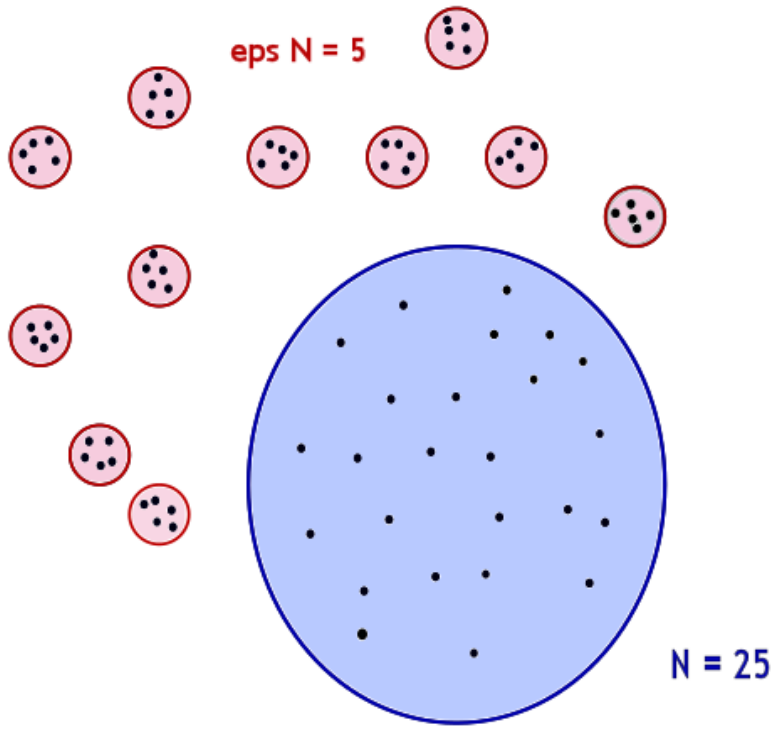
- Assumptions :
  - One **main population** of size N
  - Surrounded by ξ_N (stochastic) **small colonies** of size $\varepsilon_N N$
  - Births/deaths within populations occurring at rate 1



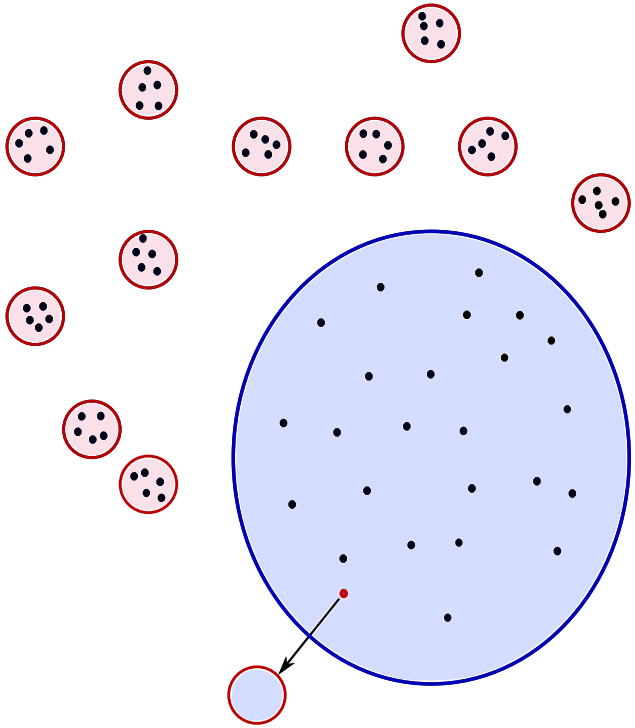
A dynamic model of peripatric metapopulation

- Context of peripatric speciation : one main population surrounded by small, peripheral isolates with marginal environmental conditions (possibly causing speciation)

- Assumptions :
  - One **main population** of size N
  - Surrounded by ξ_N (stochastic) **small colonies** of size $\varepsilon_N N$
  - Births/deaths within populations occurring at rate 1
  - Small colonies emerge from colonization events (fission) and disappear by merging with main pop (fusion)...

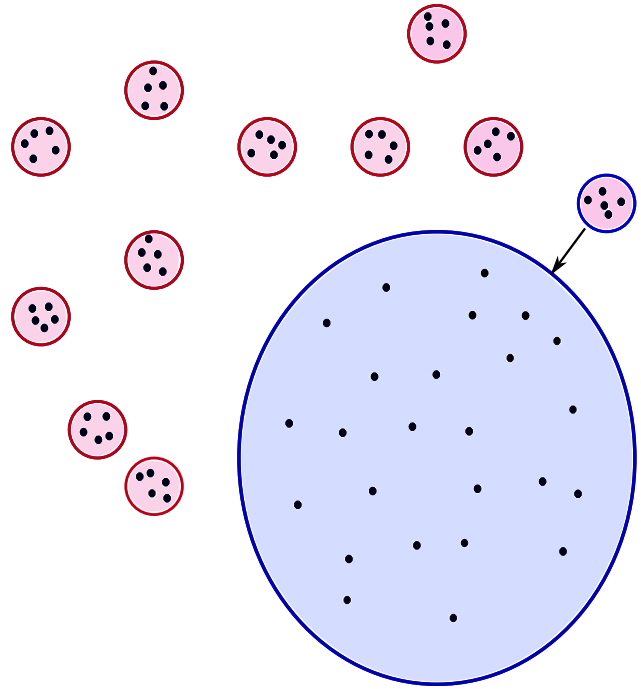


Metapopulation dynamics : Fusions and fissions



- Fission : each individual independently seeds a new colony...
- ...at rate $\theta_N = \theta/\varepsilon_N N^2$ if from main pop
- ...at rate $\beta_N = \beta/\varepsilon_N N^2$ if from a colony
- Total fission rate (gain of colonies) :

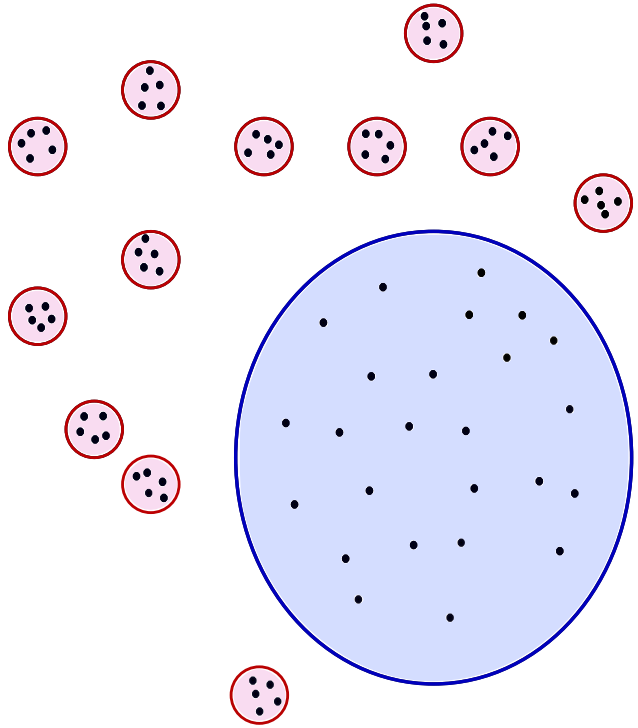
$$= \frac{\theta}{\varepsilon_N N^2} N + \frac{\beta}{\varepsilon_N N^2} (\varepsilon_N N \xi_N) = \frac{\theta + \beta(\varepsilon_N \xi_N)}{\varepsilon_N N}$$



- Fusion : each colony independently merges with main population...
- ...at rate $\gamma_N = \gamma \left(\varepsilon_N \xi_N\right)^{\alpha - 1}/N$
- $\alpha > 1$: density-dependence
- Total fusion rate (loss of colonies) :

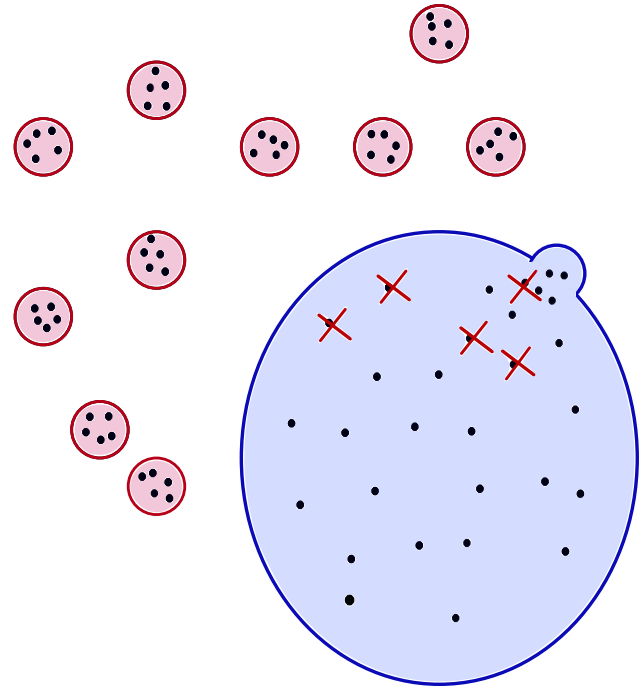
$$= \frac{\gamma \left(\varepsilon_N \xi_N\right)^{\alpha}}{\varepsilon_N N}$$

Metapopulation dynamics : Fusions and fissions



- Fission : each individual independently seeds a new colony...
- ...at rate $\theta_N = \theta/\varepsilon_N N^2$ if from main pop
- ...at rate $\beta_N = \beta/\varepsilon_N N^2$ if from a colony
- Total fission rate (gain of colonies) :

$$= \frac{\theta}{\varepsilon_N N^2} N + \frac{\beta}{\varepsilon_N N^2} (\varepsilon_N N \xi_N) = \frac{\theta + \beta(\varepsilon_N \xi_N)}{\varepsilon_N N}$$



- Fusion : each colony independently merges with main population...
- ...at rate $\gamma_N = \gamma \left(\varepsilon_N \xi_N\right)^{\alpha - 1}/N$
- $\alpha > 1$: density-dependence
- Total fusion rate (loss of colonies) :

$$= \frac{\gamma \left(\varepsilon_N \xi_N\right)^{\alpha}}{\varepsilon_N N}$$

Large populations

Lambert & Ma, "The coalescent in peripatric metapopulations", *J Appl Prob* (2015)

Proposition (Lambert & Ma (2015))

As $N \to \infty$, $\varepsilon_N \to 0$ *and* $\varepsilon_N N \to \infty$,

$$\varepsilon_N \xi_N(t) \longrightarrow x(t), \quad \text{where} \quad \dot{x} = \theta + \beta x - \gamma x^\alpha,$$

As $t \to \infty$,

$$x(t) \longrightarrow \kappa = \kappa(\alpha, \beta, \gamma, \theta), \quad \text{where} \quad \theta + \beta\kappa - \gamma\kappa^\alpha = 0.$$

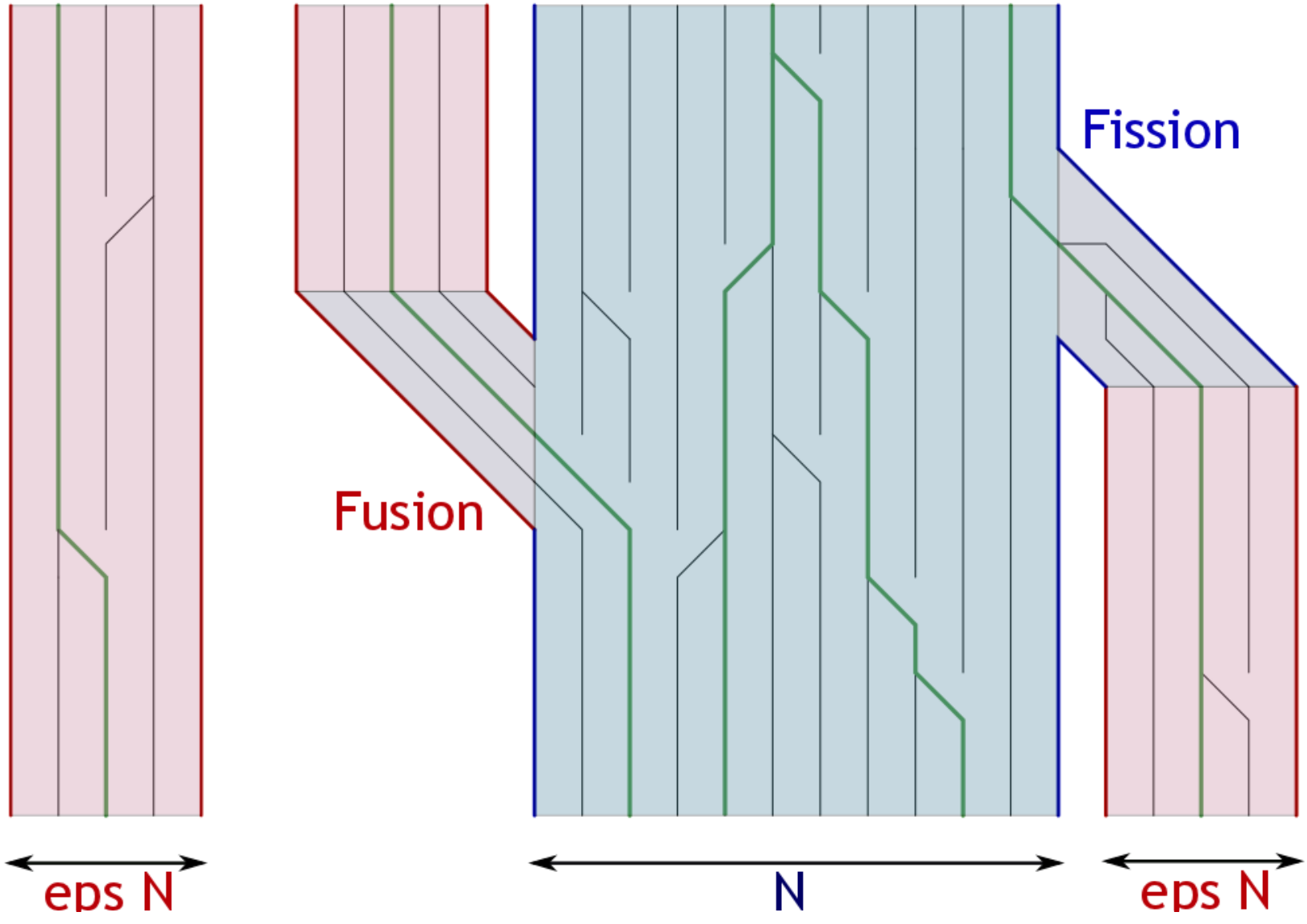
- Total number of colonies at eq $= \varepsilon_N \xi_N \approx \kappa/\varepsilon_N \ll \kappa N$
- Total number of individuals in colonies at eq $= \varepsilon_N N \xi_N \approx \kappa N$
- Fraction of biomass in main pop *eq* $= N/(N + (\varepsilon_N N)\xi_N) \approx 1/(1+\kappa)$

$$\text{Total fission rate (gain of colonies)} \approx \frac{1}{N} \times \frac{\theta + \beta\kappa}{\varepsilon_N}$$

$$\text{Total fusion rate (loss of colonies)} \approx \frac{1}{N} \times \frac{\gamma\kappa^\alpha}{\varepsilon_N}$$

(equal at equilibrium)

Genealogy of a sample in a dynamic metapopulation



The in/out coalescent

Lambert & Ma, "The coalescent in peripatric metapopulations", *J Appl Prob* (2015)

Theorem (Lambert & Ma (2015))

As $N \to \infty$, $\varepsilon_N \to 0$ and $\varepsilon_N N \to \infty$, the genealogy of a sample at equilibrium only depends on which lineages are

- *'inner' = in main pop, or*
- *'outer' = in some colony*

The in/out coalescent

Lambert & Ma, "The coalescent in peripatric metapopulations", *J Appl Prob* (2015)

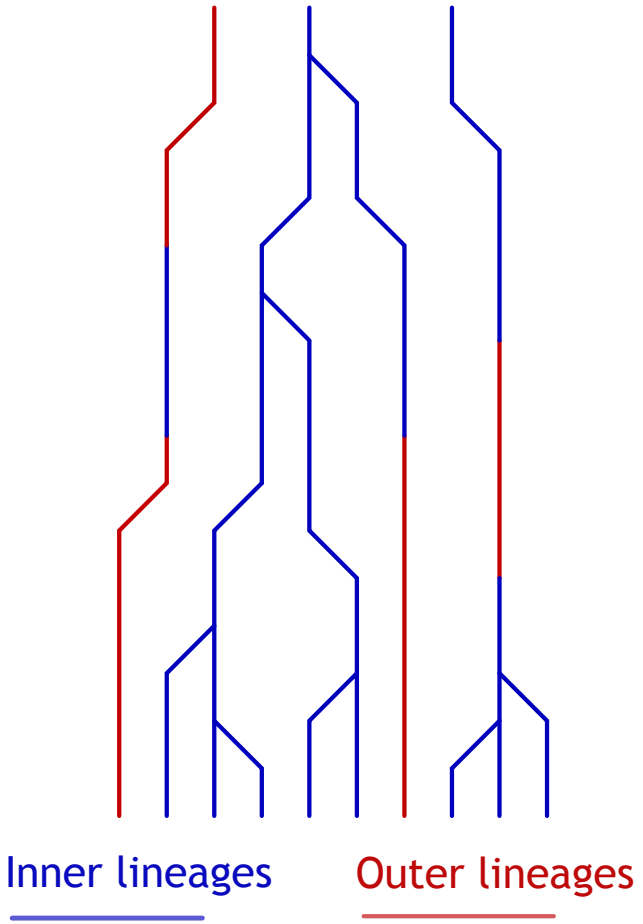
Theorem (Lambert & Ma (2015))

As $N \to \infty$, $\varepsilon_N \to 0$ and $\varepsilon_N N \to \infty$, the genealogy of a sample at equilibrium only depends on which lineages are

- *'inner' = in main pop, or*
- *'outer' = in some colony*

In backward time accelerated by a factor N,

- *Each pair of inner lineages coalesces at rate 1*



The in/out coalescent

Lambert & Ma, "The coalescent in peripatric metapopulations", *J Appl Prob* (2015)

Theorem (Lambert & Ma (2015))

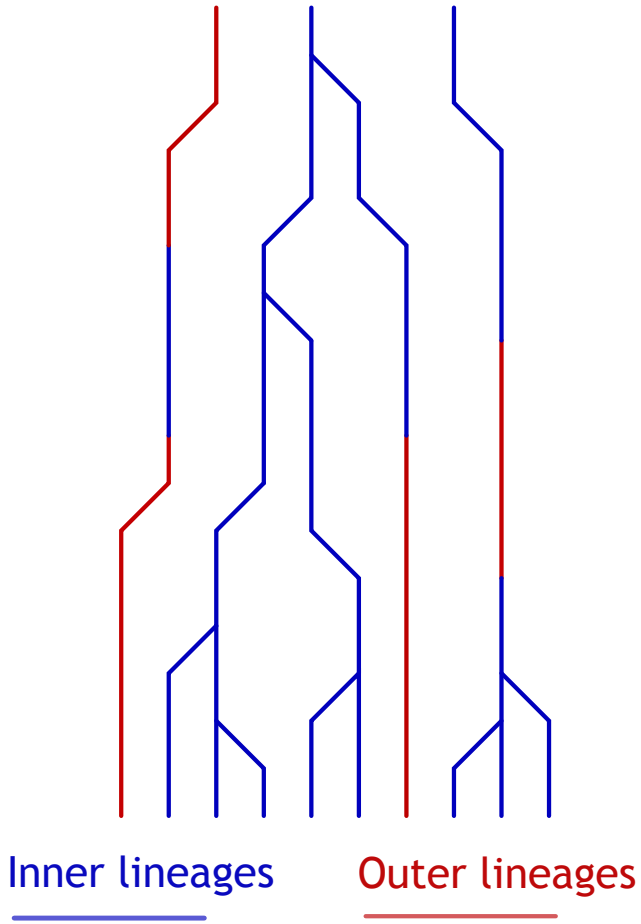
As $N \to \infty$, $\varepsilon_N \to 0$ and $\varepsilon_N N \to \infty$, the genealogy of a sample at equilibrium only depends on which lineages are

- *'inner' = in main pop, or*
- *'outer' = in some colony*

In backward time accelerated by a factor N,

- *Each pair of inner lineages coalesces at rate 1*
- *Each inner lineage turns outer at rate*

$$\text{Rate}\,[\,\text{In} \longrightarrow \text{Out}\,] \approx \frac{\gamma\kappa^\alpha}{\varepsilon_N} \times \frac{\varepsilon_N N}{N + \varepsilon_N N} \approx \gamma\kappa^\alpha$$



The in/out coalescent

Lambert & Ma, "The coalescent in peripatric metapopulations", *J Appl Prob* (2015)

Theorem (Lambert & Ma (2015))

As $N \to \infty$, $\varepsilon_N \to 0$ and $\varepsilon_N N \to \infty$, the genealogy of a sample at equilibrium only depends on which lineages are

- *'inner' = in main pop, or*
- *'outer' = in some colony*

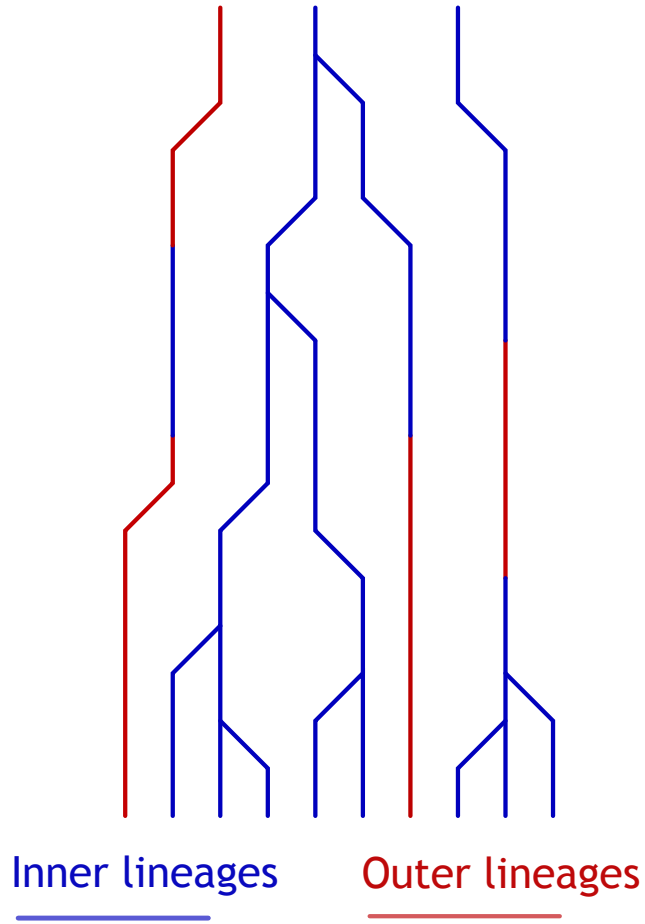
In backward time accelerated by a factor N,

- *Each pair of inner lineages coalesces at rate 1*
- *Each inner lineage turns outer at rate*

$$\text{Rate}\,[\,\text{In} \longrightarrow \text{Out}\,] \approx \frac{\gamma\kappa^\alpha}{\varepsilon_N} \times \frac{\varepsilon_N N}{N + \varepsilon_N N} \approx \gamma\kappa^\alpha$$

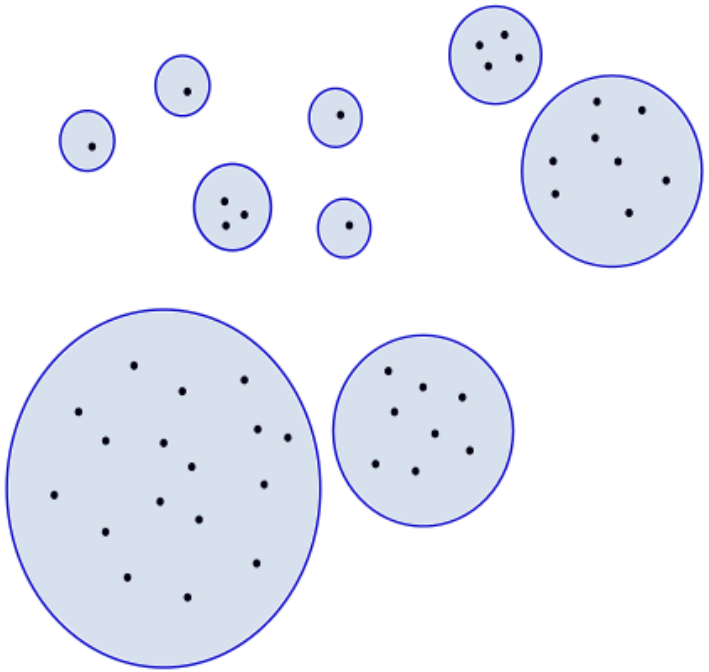
- *Each outer lineage turns inner at rate*

$$\text{Rate}\,[\,\text{Out} \longrightarrow \text{In}\,] \approx \frac{\theta}{\varepsilon_N} \times \frac{1}{\kappa/\varepsilon_N} \approx \frac{\theta}{\kappa}$$



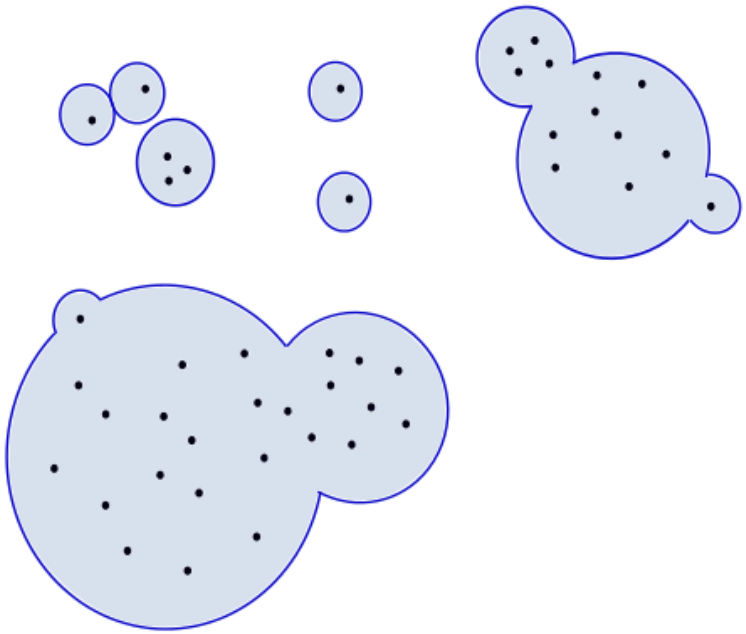
Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population



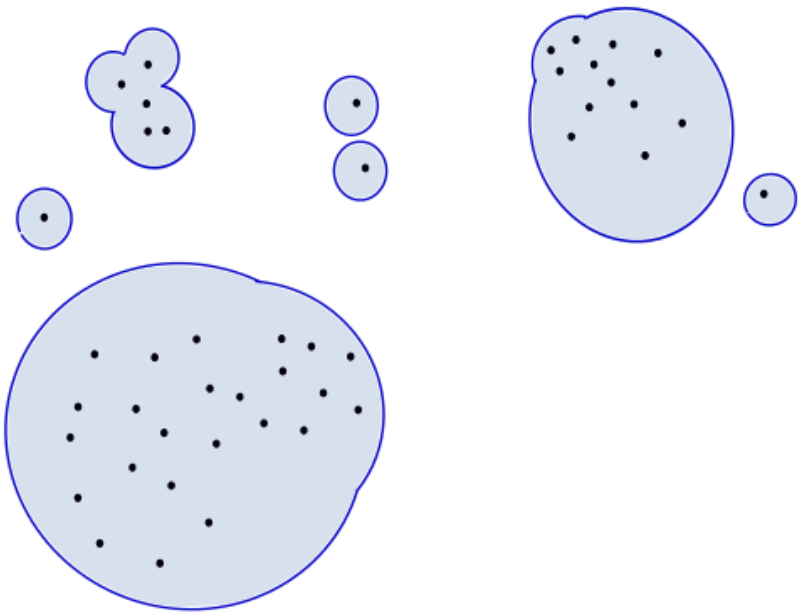
Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population
- A simple two-state coalescent



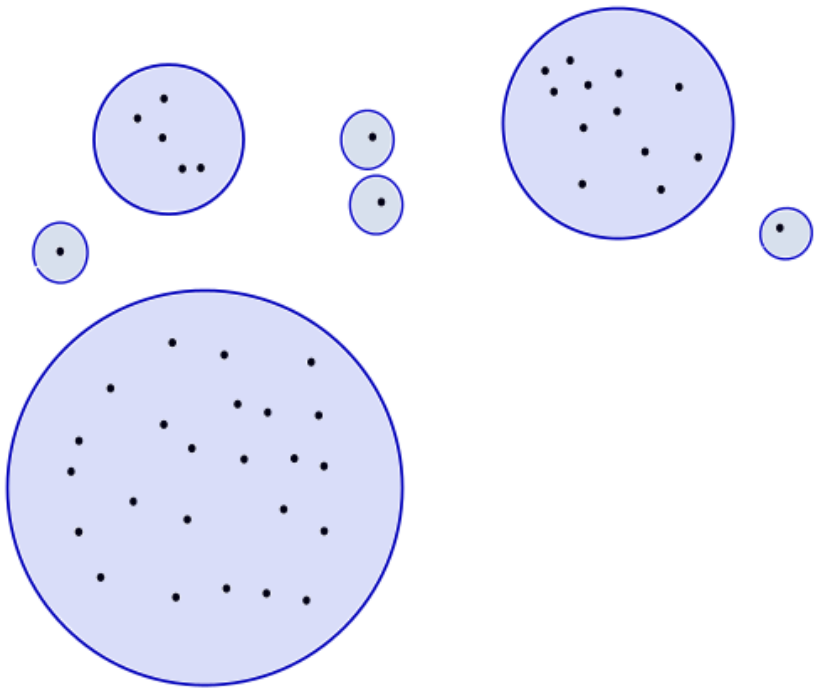
Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population
- A simple two-state coalescent
- Perspective : embed in a speciation framework where each species consists in one main pop surrounded by colonies and new species/main pop's spontaneously emerge



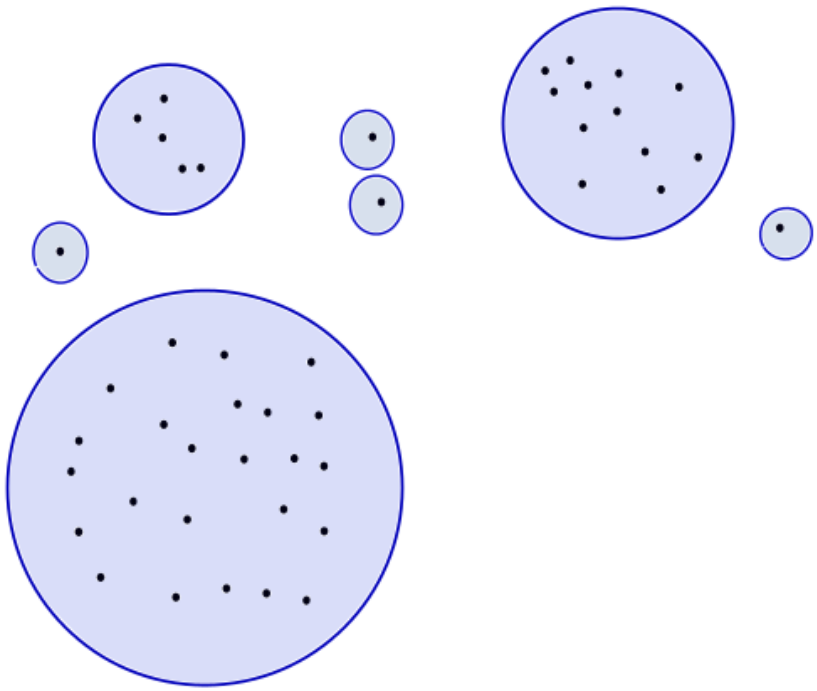
Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population
- A simple two-state coalescent
- Perspective : embed in a speciation framework where each species consists in one main pop surrounded by colonies and new species/main pop's spontaneously emerge
- The active/inactive coalescent also arises in models of seed banks.



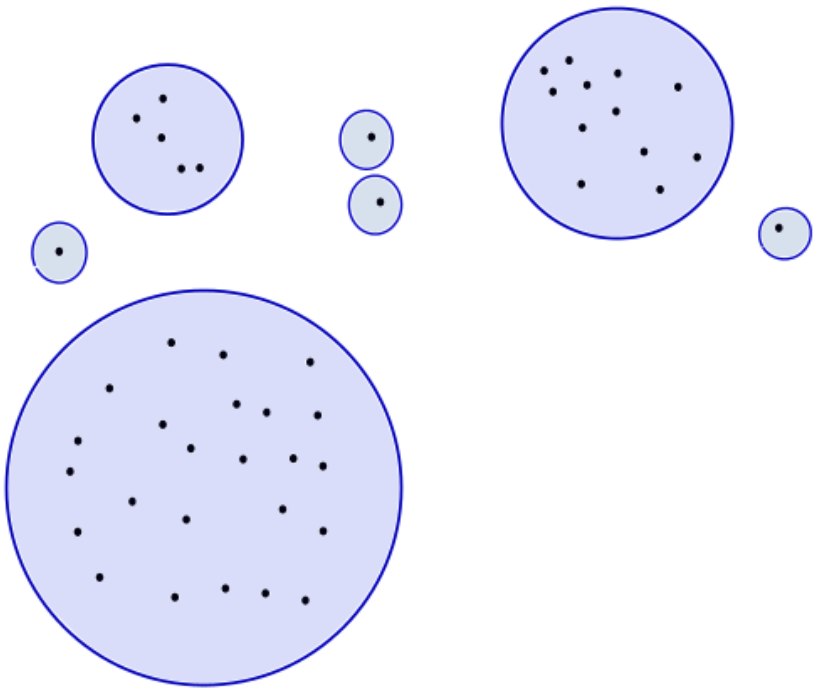
Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population
- A simple two-state coalescent
- Perspective : embed in a speciation framework where each species consists in one main pop surrounded by colonies and new species/main pop's spontaneously emerge
- The active/inactive coalescent also arises in models of seed banks.



Conclusion on Model 1

- A spatially dynamic model, where small populations bud from and merge with main population
- A simple two-state coalescent
- Perspective : embed in a speciation framework where each species consists in one main pop surrounded by colonies and new species/main pop's spontaneously emerge
- The active/inactive coalescent also arises in models of seed banks.



See for example Blath, González Casanova, Kurt & Wilke-Berenguer, "A new coalescent for seedbank models" *Ann Appl Prob* (2016)
... who show for example that for a sample of n, $T_{MRCA} \asymp \ln \ln(n)$...
$\Rightarrow$ The active/inactive coalescent doesn't come down from infinity but ALMOST :)

Outline

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree
- As in ARG, if one sampled genome with n genes, we can define $\Pi_n(t)$ the partition of $\{1, \ldots, n\}$ induced by :

  i and $j \in$ same block of $\Pi_n(t) \iff$ Lineages ancestral to genes i and j at t bkwd lie in same sp

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree
- As in ARG, if one sampled genome with n genes, we can define $\Pi_n(t)$ the partition of $\{1, \ldots, n\}$ induced by :

  i and $j \in$ same block of $\Pi_n(t) \iff$ Lineages ancestral to genes i and j at t bkwd lie in same sp
- By sampling consistency, $\exists\, \Pi_\infty(t)$ exchangeable partition of $\mathbb{N}$ such that $\Pi_n(t) \overset{\mathcal{L}}{=} \Pi_\infty(t)$ restricted to $\{1, \ldots, n\}$

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree
- As in ARG, if one sampled genome with n genes, we can define $\Pi_n(t)$ the partition of $\{1, \ldots, n\}$ induced by :

  i and $j \in$ same block of $\Pi_n(t) \iff$ Lineages ancestral to genes i and j at t bkwd lie in same sp
- By sampling consistency, $\exists\, \Pi_\infty(t)$ exchangeable partition of $\mathbb{N}$ such that $\Pi_n(t) \overset{\mathcal{L}}{=} \Pi_\infty(t)$ restricted to $\{1, \ldots, n\}$
- By the Kingman-De Finetti Theorem, an exchangeable partition of $\mathbb{N}$ can be obtained by 'painting' each integer in color $j \geq 1$ with proba β_j ('frequency of block j'), or in a distinct color with proba $d = 1 - \sum_{j \geq 1} \beta_j$ (singletons, or 'dust').

Models with gene flow

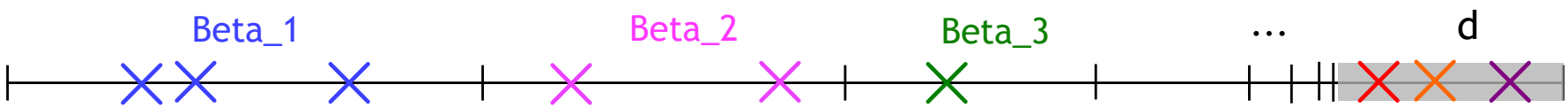
- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree
- As in ARG, if one sampled genome with n genes, we can define $\Pi_n(t)$ the partition of $\{1, \ldots, n\}$ induced by :

  i and $j \in$ same block of $\Pi_n(t) \iff$ Lineages ancestral to genes i and j at t bkwd lie in same sp
- By sampling consistency, $\exists\, \Pi_\infty(t)$ exchangeable partition of $\mathbb{N}$ such that $\Pi_n(t) \overset{\mathcal{L}}{=} \Pi_\infty(t)$ restricted to $\{1, \ldots, n\}$
- By the Kingman-De Finetti Theorem, an exchangeable partition of $\mathbb{N}$ can be obtained by 'painting' each integer in color $j \geq 1$ with proba β_j ('frequency of block j'), or in a distinct color with proba $d = 1 - \sum_{j \geq 1} \beta_j$ (singletons, or 'dust').

Models with gene flow

- The MSC yields a nice way of coupling gene trees thanks to the species tree.
- In models with gene flow, $\exists$ several species which are ancestral $=$ carry a subset of ancestral gene lineages
- No mother-daughter species relationship $\Rightarrow$ no species tree
- As in ARG, if one sampled genome with n genes, we can define $\Pi_n(t)$ the partition of $\{1, \ldots, n\}$ induced by :

  i and $j \in$ same block of $\Pi_n(t) \iff$ Lineages ancestral to genes i and j at t bkwd lie in same sp
- By sampling consistency, $\exists\, \Pi_\infty(t)$ exchangeable partition of $\mathbb{N}$ such that $\Pi_n(t) \overset{\mathcal{L}}{=} \Pi_\infty(t)$ restricted to $\{1, \ldots, n\}$
- By the Kingman-De Finetti Theorem, an exchangeable partition of $\mathbb{N}$ can be obtained by 'painting' each integer in color $j \geq 1$ with proba β_j ('frequency of block j'), or in a distinct color with proba $d = 1 - \sum_{j \geq 1} \beta_j$ (singletons, or 'dust').

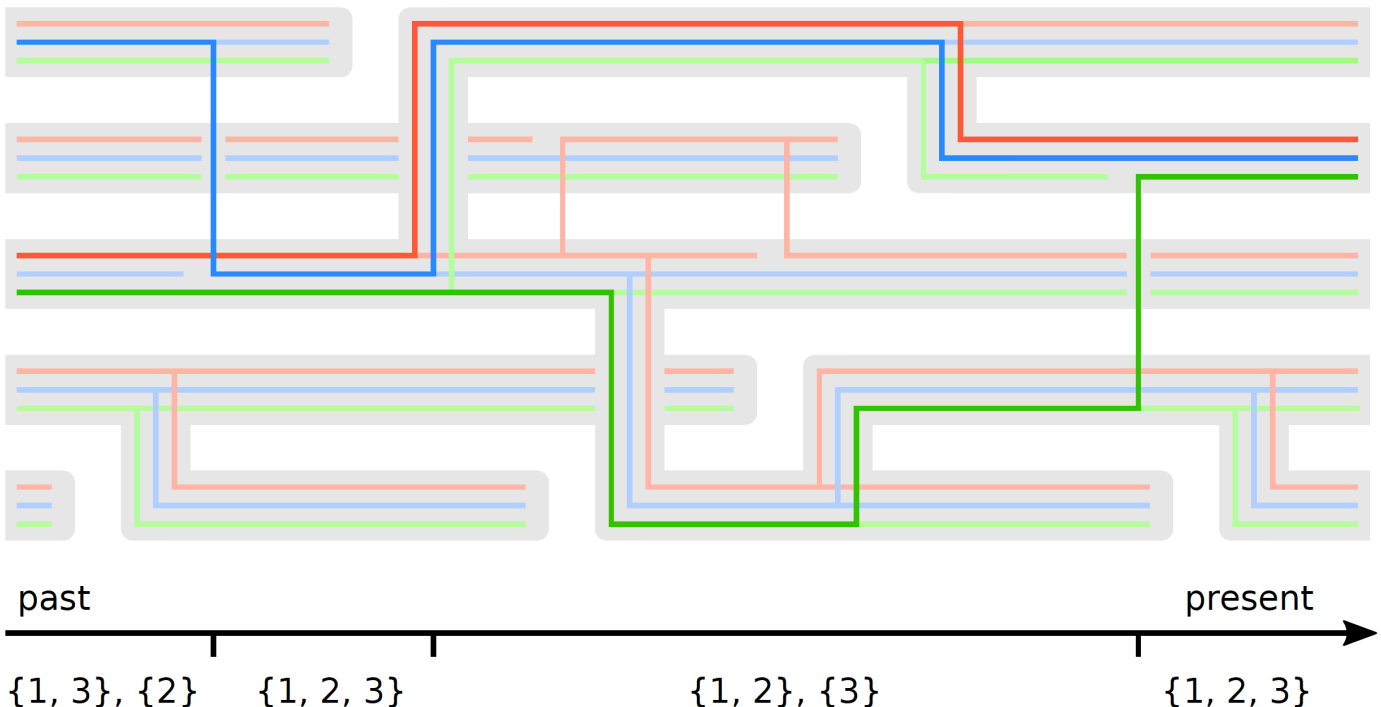


Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$

(B) Backward time : ancestry of 1 sampled genome

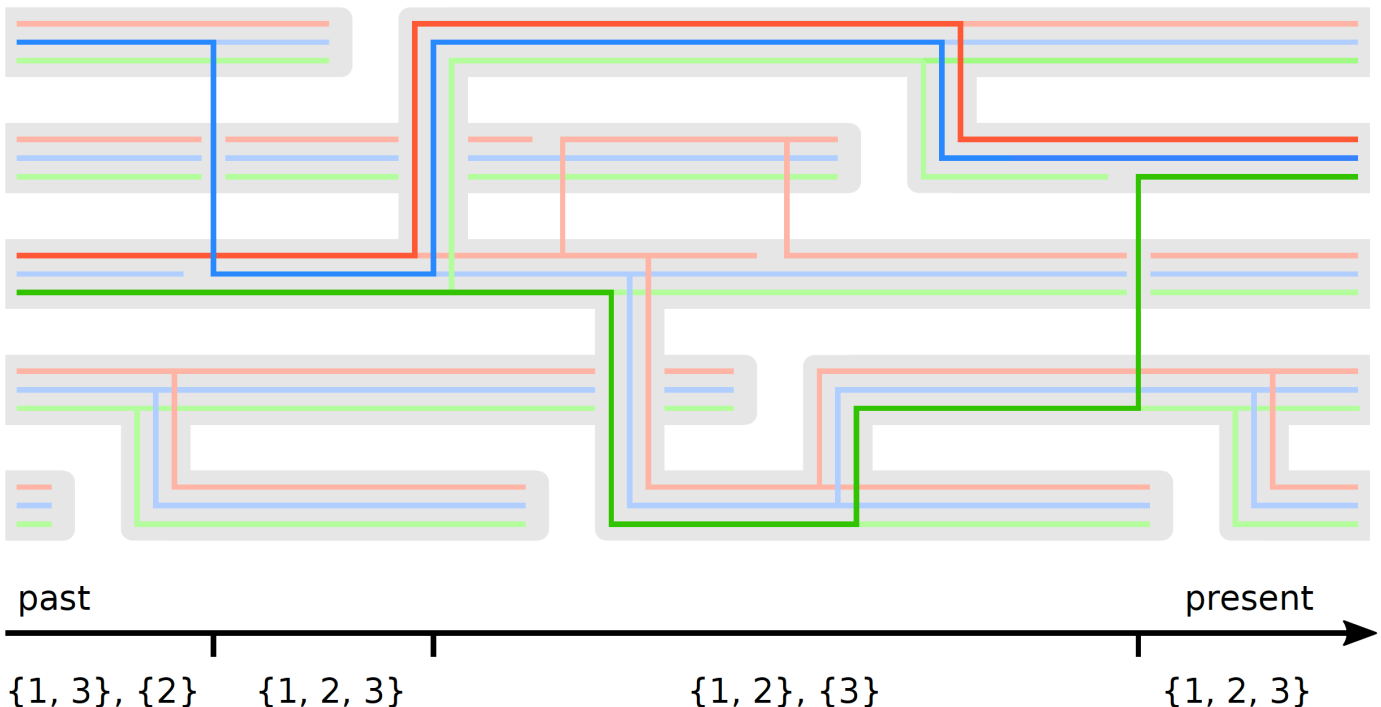


Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization

(B) Backward time : ancestry of 1 sampled genome

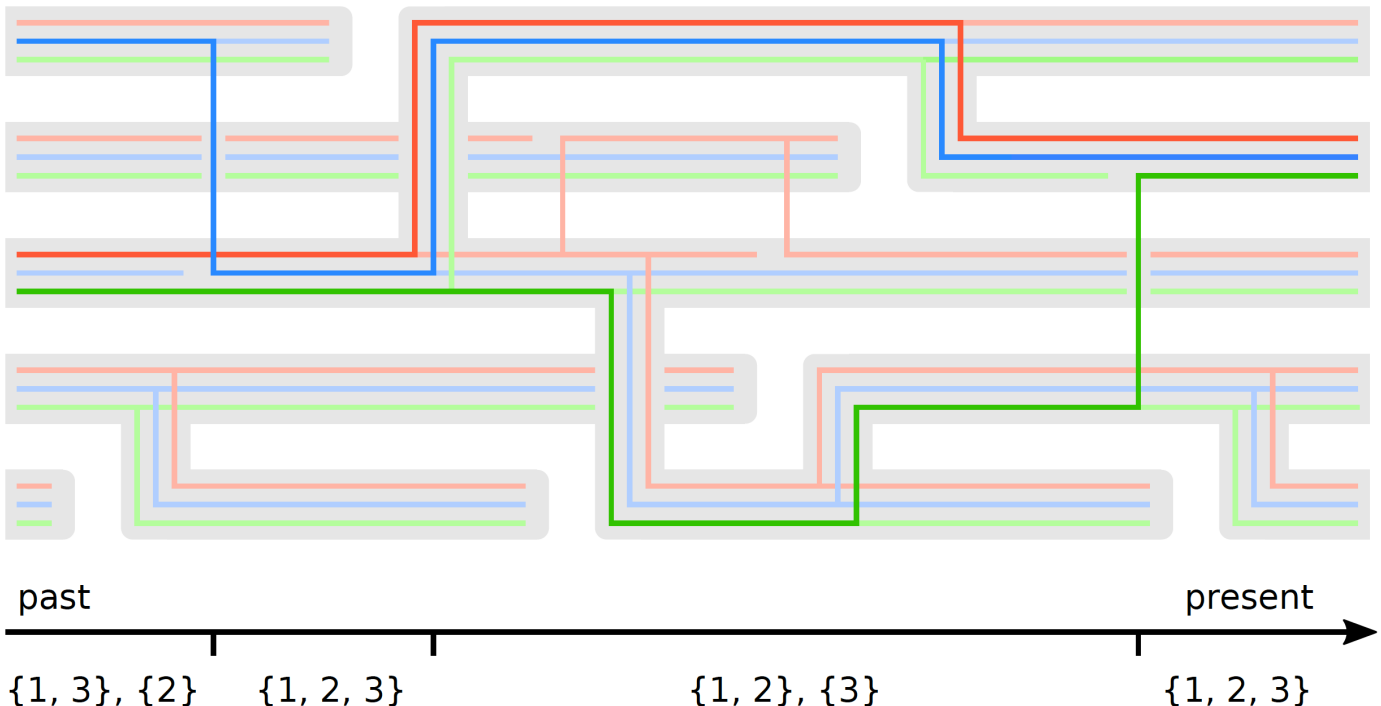


Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization
- **Introgression** : each gene lineage, at rate c can replicate and move from one species to another and then replace its homologous copy in the target species.

(B) Backward time : ancestry of 1 sampled genome



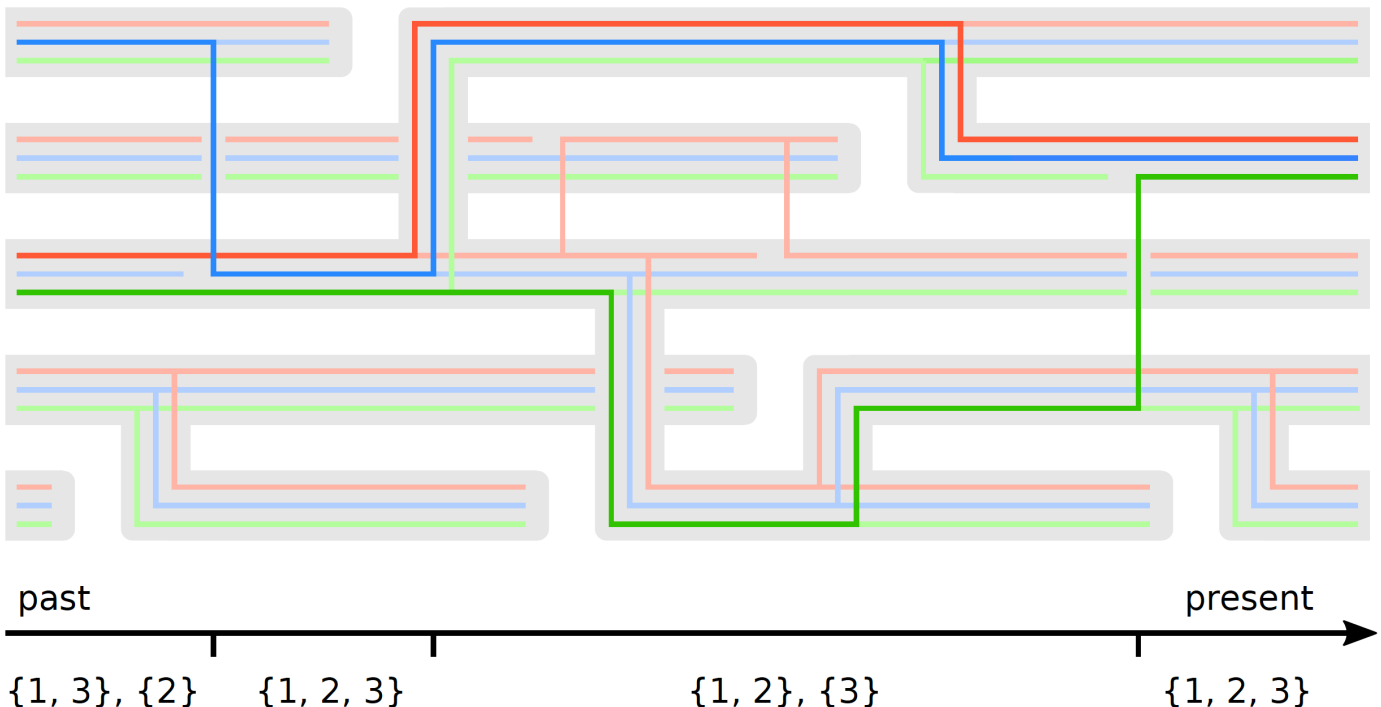
Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization
- **Introgression** : each gene lineage, at rate c can replicate and move from one species to another and then replace its homologous copy in the target species.

(B) Backward time : ancestry of 1 sampled genome

- As $N \to \infty$: Markov process $\Pi_n(t)$ valued in partitions of $\{1, \ldots, n\}$ induced by the relation 'lying in the same species'



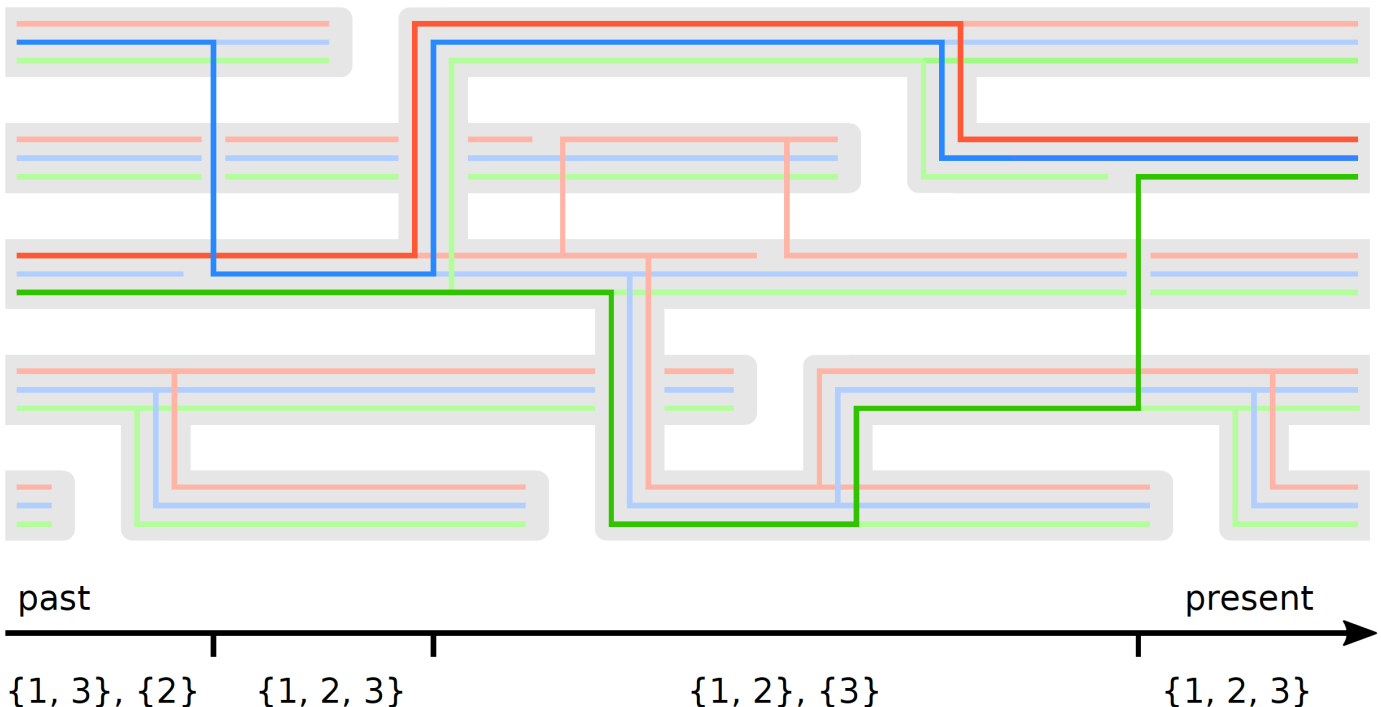
Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization
- **Introgression** : each gene lineage, at rate c can replicate and move from one species to another and then replace its homologous copy in the target species.

(B) Backward time : ancestry of 1 sampled genome

- As $N \to \infty$: Markov process $\Pi_n(t)$ valued in partitions of $\{1, \ldots, n\}$ induced by the relation 'lying in the same species'
- Coalescent : Each pair of blocks coalesces at rate 1



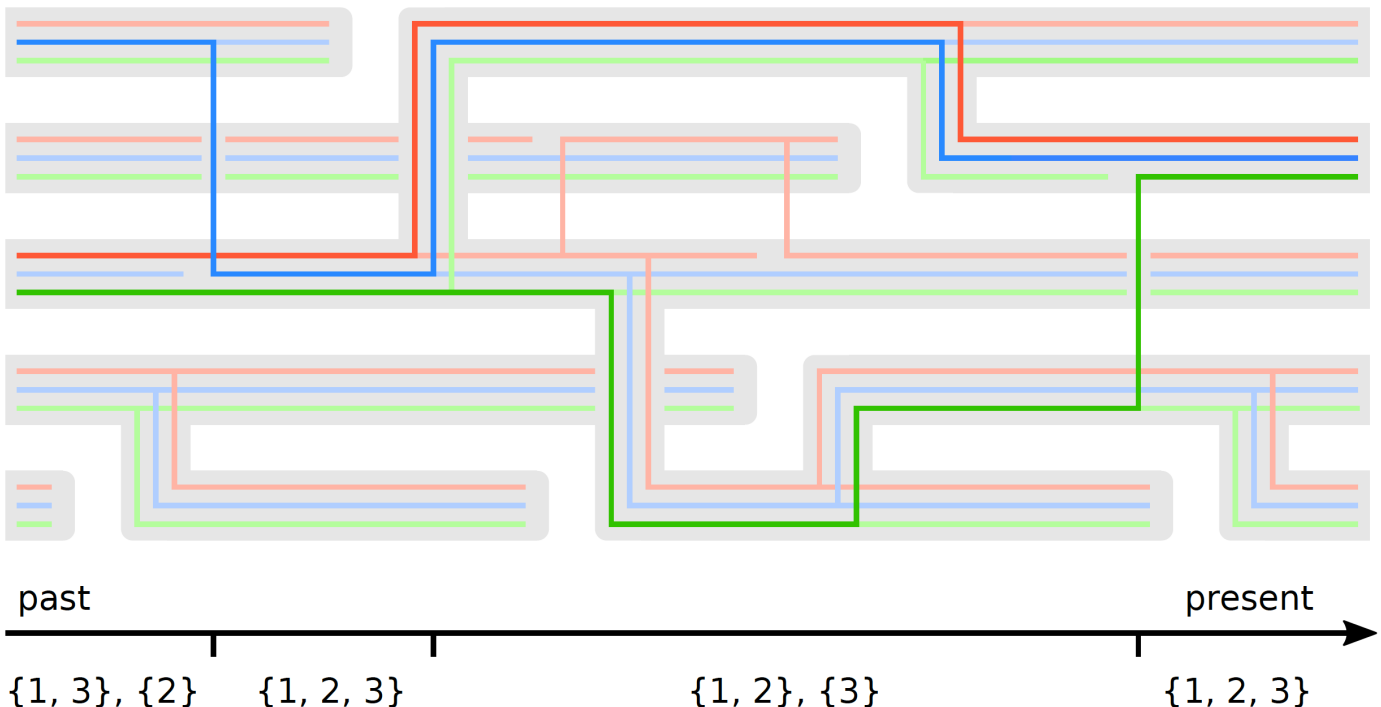
Model 2. Speciation with introgression

(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization
- **Introgression** : each gene lineage, at rate c can replicate and move from one species to another and then replace its homologous copy in the target species.

(B) Backward time : ancestry of 1 sampled genome

- As $N \to \infty$: Markov process $\Pi_n(t)$ valued in partitions of $\{1, \ldots, n\}$ induced by the relation 'lying in the same species'
- Coalescent : Each pair of blocks coalesces at rate 1
- **Erosion** : Each lineage leaves its block at rate c and forms a new singleton.



Model 2. Speciation with introgression

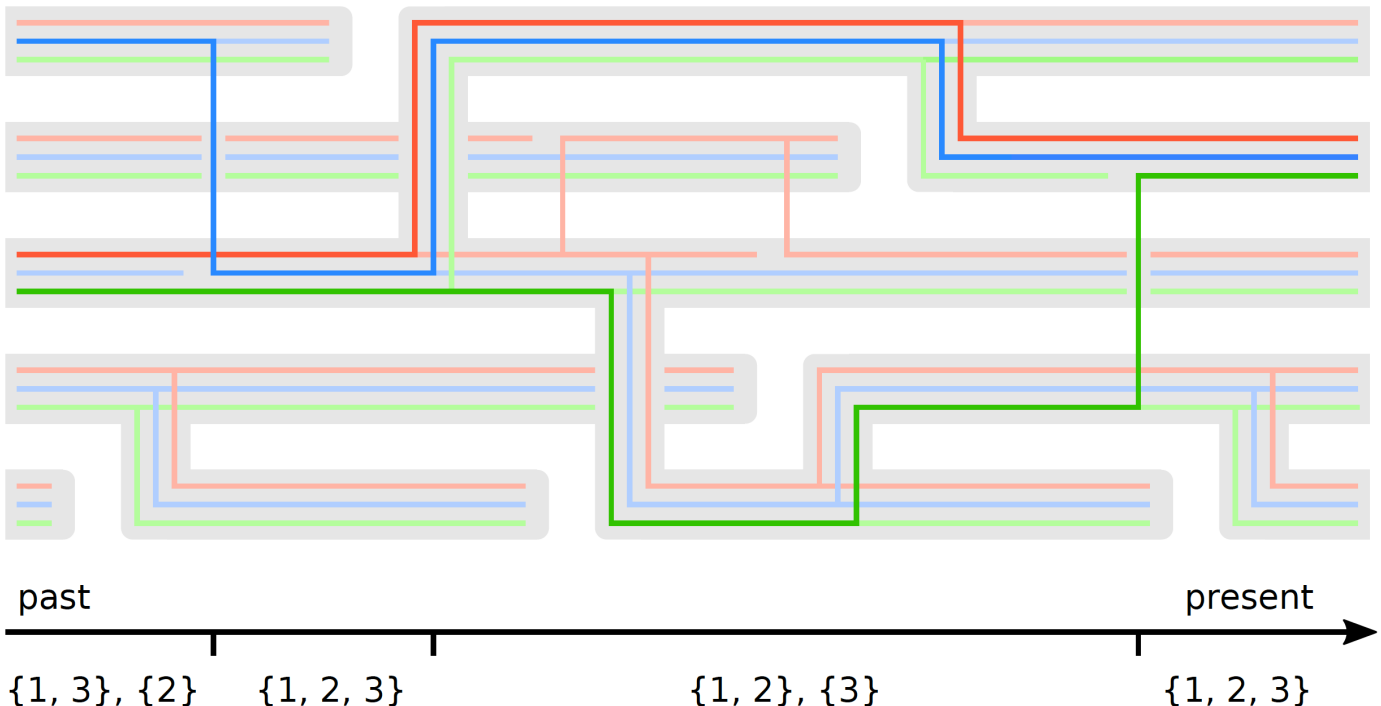
(A) Forward time model with N species

- N monomorphic species, each harboring one genome of n genes indexed by $\{1, \ldots, n\}$
- Extinction-recolonization as in Moran model, pairwise rate 1, genome replicated faithfully upon colonization
- **Introgression** : each gene lineage, at rate c can replicate and move from one species to another and then replace its homologous copy in the target species.

(B) Backward time : ancestry of 1 sampled genome

- As $N \to \infty$: Markov process $\Pi_n(t)$ valued in partitions of $\{1, \ldots, n\}$ induced by the relation 'lying in the same species'
- Coalescent : Each pair of blocks coalesces at rate 1
- **Erosion** : Each lineage leaves its block at rate c and forms a new singleton.

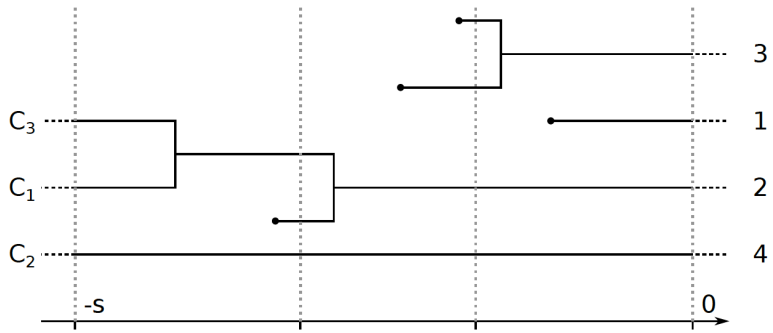
$\Rightarrow$ $\Pi_n(t) =$ Kingman coalescent with erosion



The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago

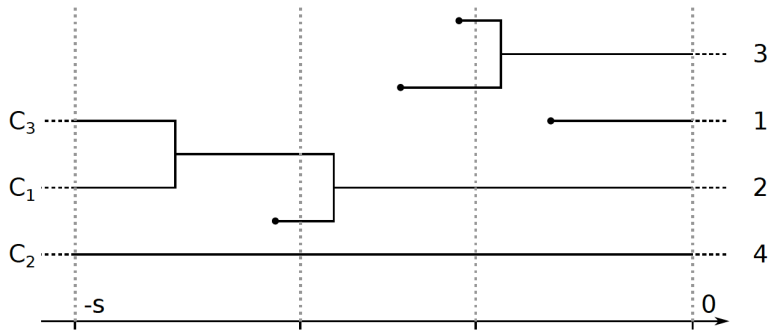


The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of critical birth-death process

The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago
- $\#\Pi_n(t)$ goes from $k \to k-1$ at rate $k(k-1)/2$ and from $k \to k+1$ at rate cn (modulo singletons)

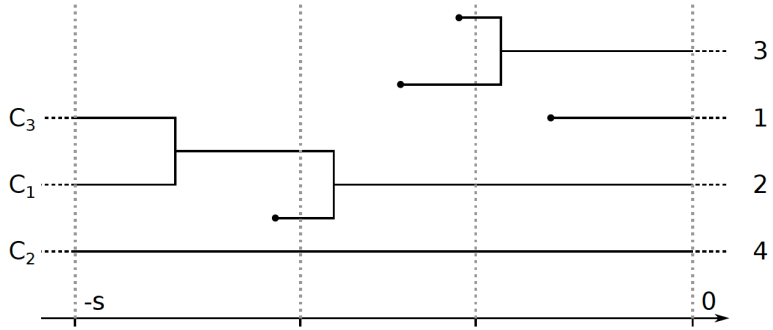


The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of a critical birth-death process

The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago
- $\#\Pi_n(t)$ goes from $k \to k-1$ at rate $k(k-1)/2$ and from $k \to k+1$ at rate cn (modulo singletons)
- As $t \to \infty$, $\#\Pi_n(t) \approx \sqrt{2cn}$

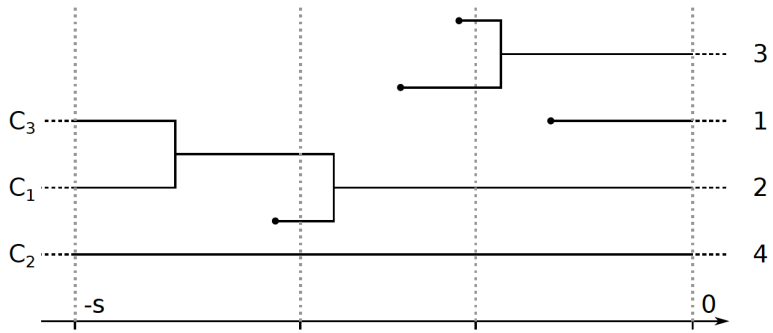


The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of a critical birth-death process

The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago
- $\#\Pi_n(t)$ goes from $k \to k-1$ at rate $k(k-1)/2$ and from $k \to k+1$ at rate cn (modulo singletons)
- As $t \to \infty$, $\#\Pi_n(t) \approx \sqrt{2cn}$
- Reversing time, when k particles are present, each particle gives birth at rate $cn/k \approx \sqrt{cn/2}$ and dies at rate $(k-1)/2 \approx \sqrt{cn/2}$

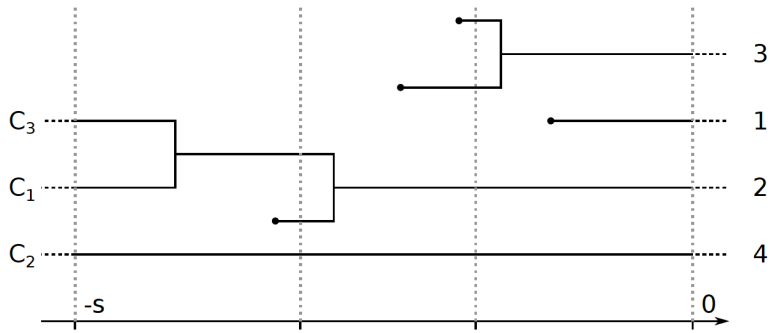


The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of critical birth-death process

The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago
- $\#\Pi_n(t)$ goes from $k \to k-1$ at rate $k(k-1)/2$ and from $k \to k+1$ at rate cn (modulo singletons)
- As $t \to \infty$, $\#\Pi_n(t) \approx \sqrt{2cn}$
- Reversing time, when k particles are present, each particle gives birth at rate $cn/k \approx \sqrt{cn/2}$ and dies at rate $(k-1)/2 \approx \sqrt{cn/2}$

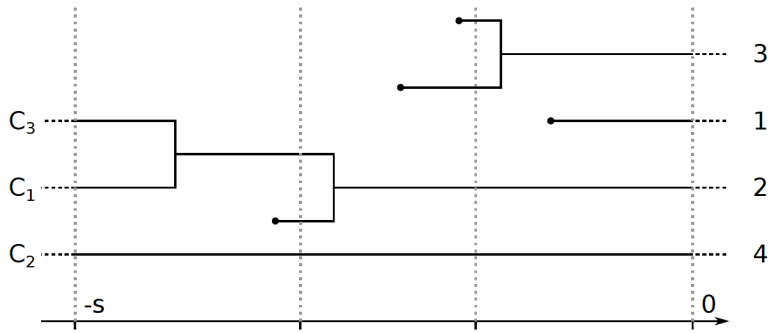


The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of critical birth-death process

The Kingman coalescent with erosion

Foutel–Rodier, Lambert & Schertzer, "Kingman's coalescent with erosion", *EJP* (2020)

- Nb of blocks $\#\Pi_n(t) =$ nb of ancestral species to the sampled genome t time units ago
- $\#\Pi_n(t)$ goes from $k \to k-1$ at rate $k(k-1)/2$ and from $k \to k+1$ at rate cn (modulo singletons)
- As $t \to \infty$, $\#\Pi_n(t) \approx \sqrt{2cn}$
- Reversing time, when k particles are present, each particle gives birth at rate $cn/k \approx \sqrt{cn/2}$ and dies at rate $(k-1)/2 \approx \sqrt{cn/2}$



The origin of a block can be traced back to its singletons
Reversing time, the size of a block is the total progeny of critical birth-death process

Theorem (Foutel–Rodier, Lambert & Schertzer (2020))

Set $M_n(k) :=$ nb of blocks of size k in $\Pi_n(t)$ as $t \to \infty$ and $M_n := \#\Pi_n(\infty) = \sum_k M_n(k)$.
Then as $n \to \infty$, $M_n/\sqrt{n} \to \sqrt{2c}$ and

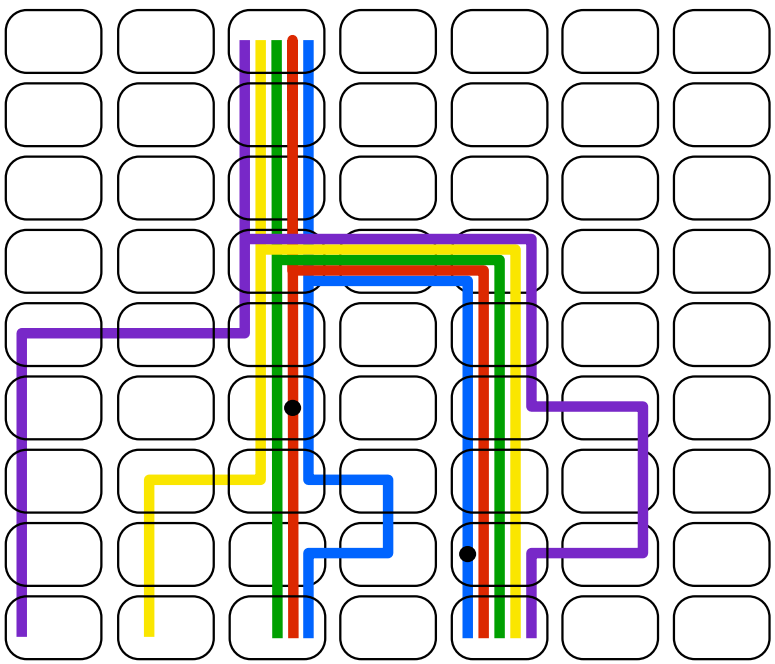
$$\lim_{n\to\infty} \frac{M_n(k)}{\sqrt{2cn}} = \mathbb{P}(J = k),$$

where J is the total progeny of a critical birth-death process. Also $\Pi_\infty(t)$ has no dust ($d = 0$) and as $t \to \infty$, its block frequencies (β_j) can be explicitly characterized.

See also Berestycki, J (2004).

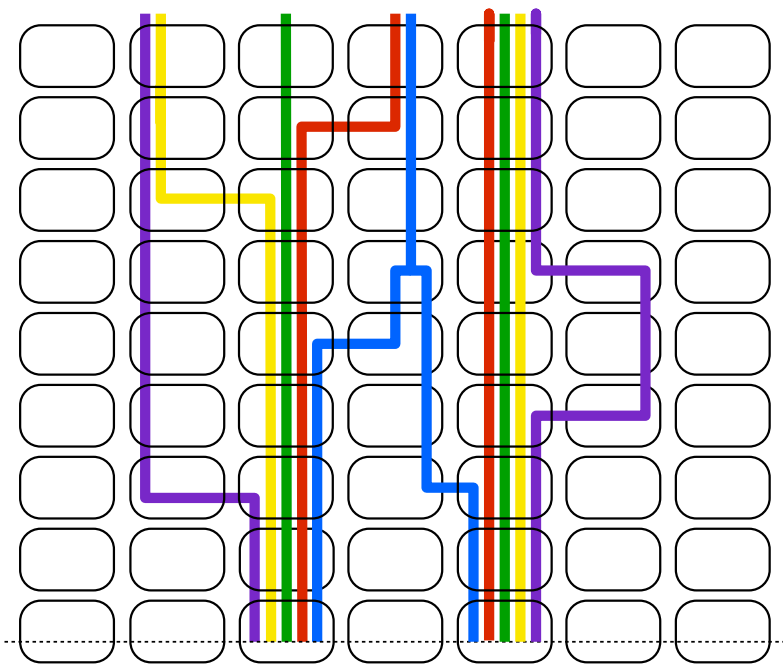
Model 3. Gene-based diversification and the notion of co-adaptation

- In **forward** time, gene lineages accumulate mutations giving rise to new alleles
- A new allele is only **co-adapted** with alleles present at other loci of the same genome at that time
- Separately evolving populations progressively lose co-adaptation of their genomes.



Coadaptation in forward time :
New alleles are co-adapted with genomic background

- In **backward** time, a gene lineage is ancestral to one or several genomes of the sample
- Two non-homologous gene lineages are **co-adapted** iff they are both ancestral to one common genome
- In backward time, co-adaptation increases as coalescences occur.

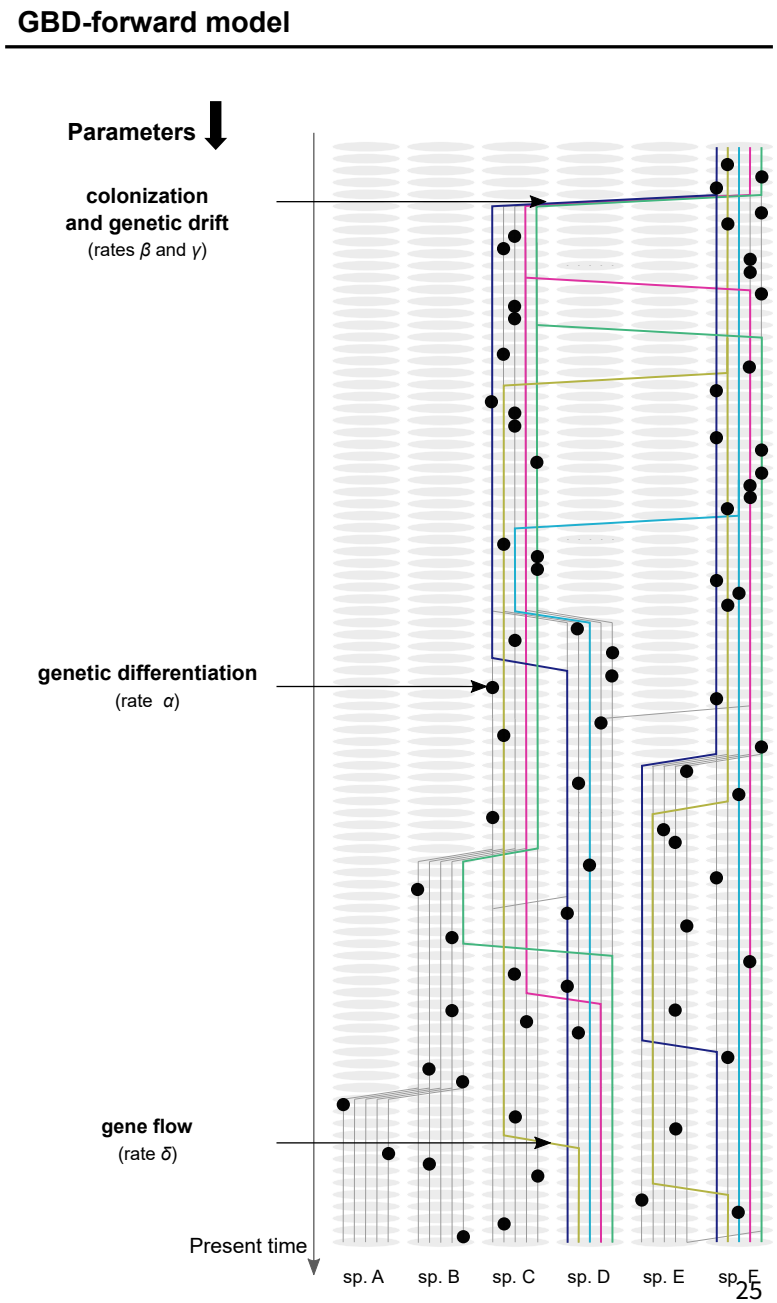


Coadaptation in backward time :
Lineages ancestral to the same genome are co-adapted

Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* 2020

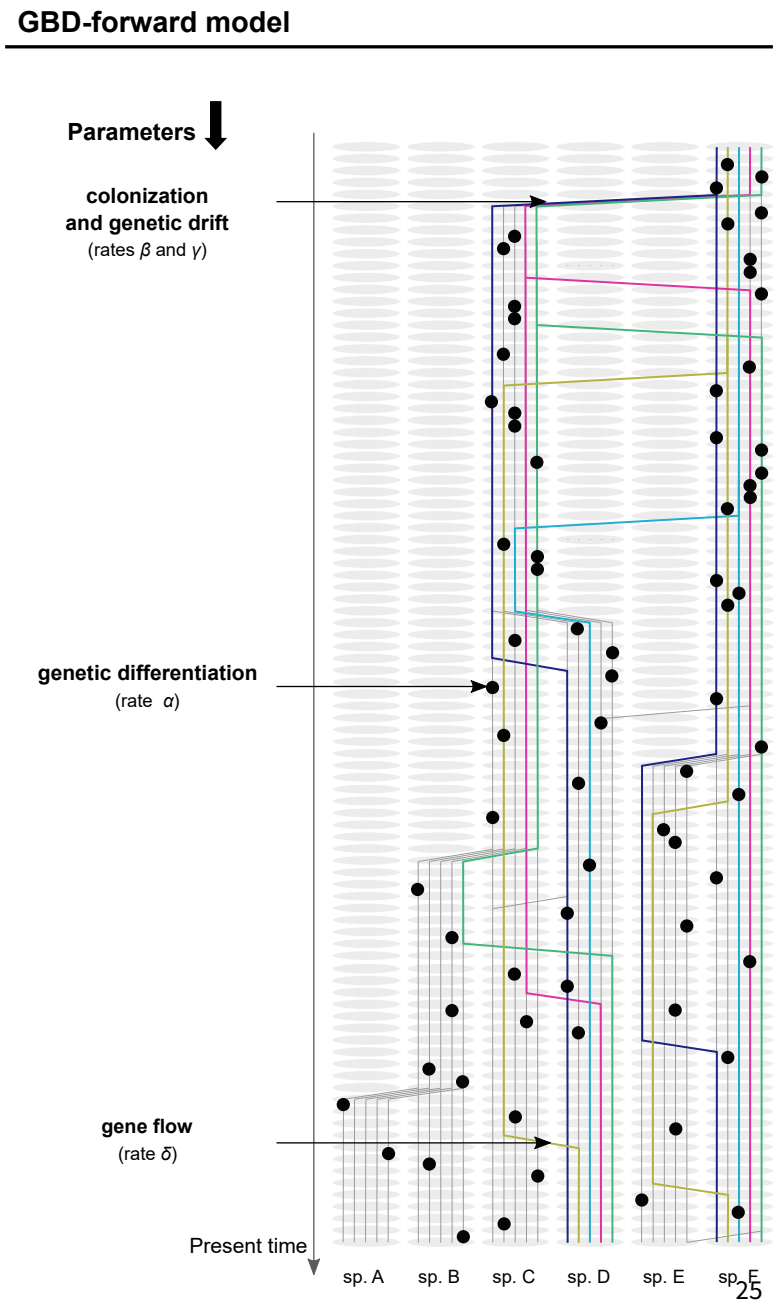
- Genomes accumulate mutations at rate α per lineage



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* 2020

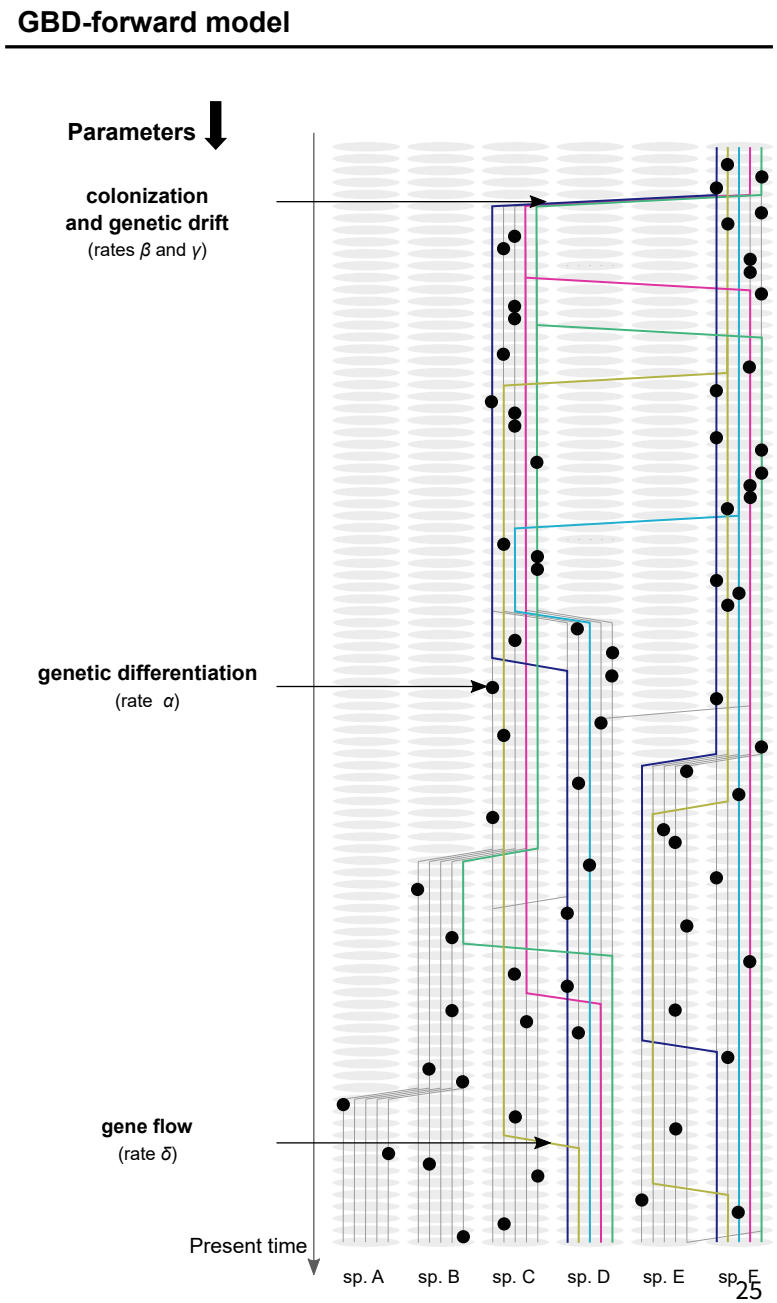
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* 2020

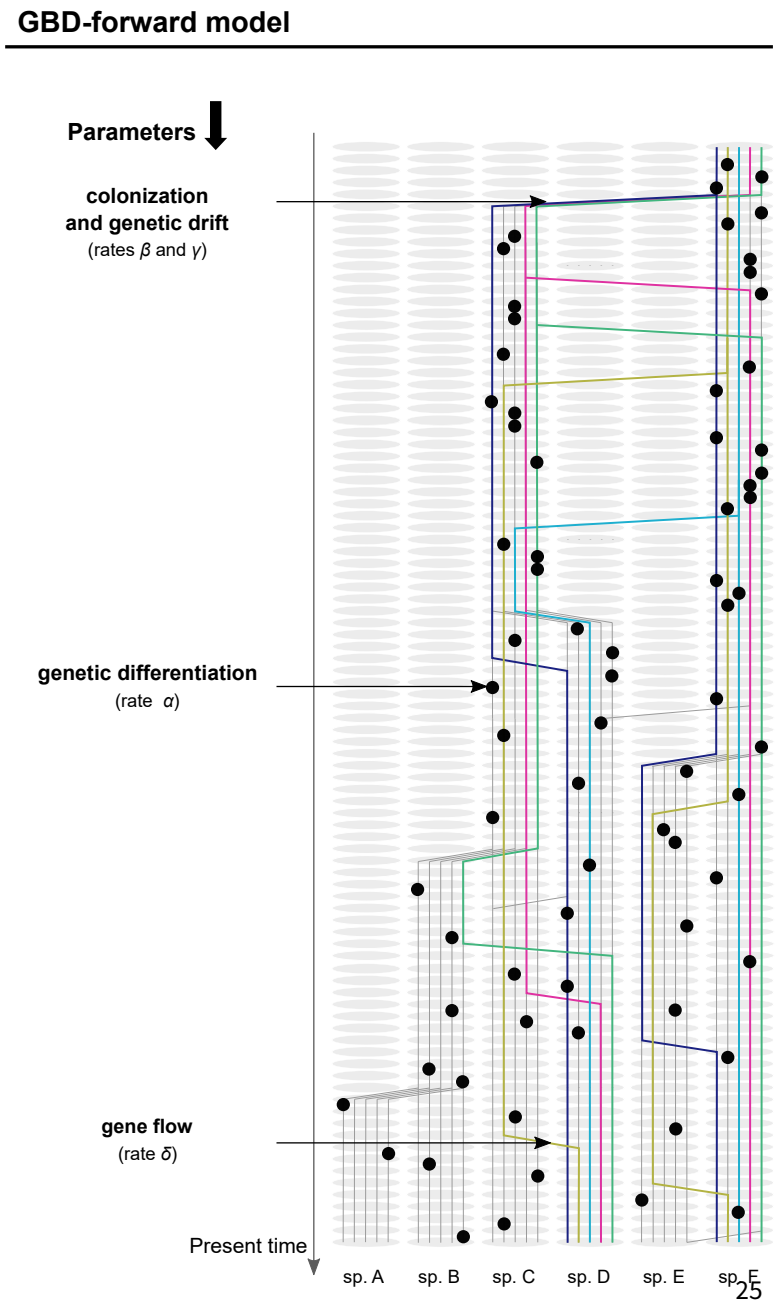
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* 2020

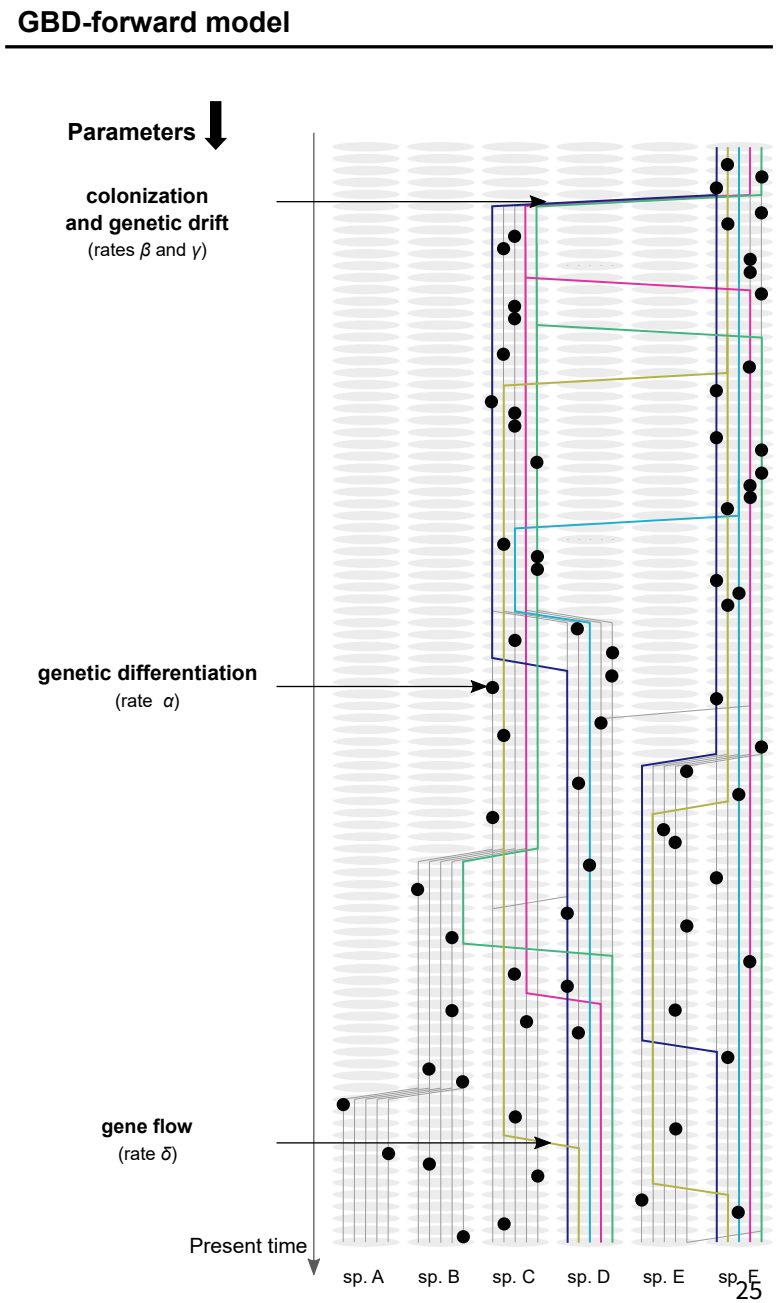
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)
- Populations exchange genes at rate δ per lineage thinned by the **degree of coadaptation** of **introgressed gene** with **target genome**



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* 2020

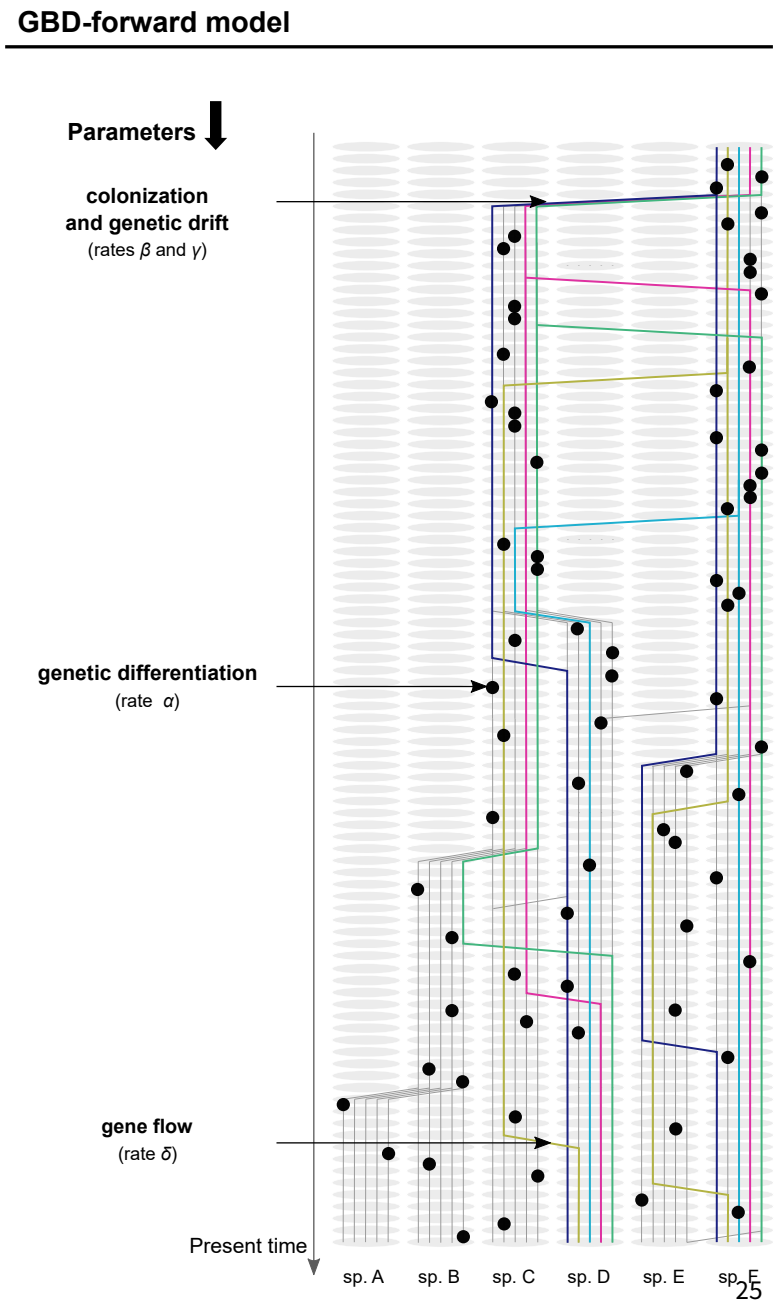
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)
- Populations exchange genes at rate δ per lineage thinned by the **degree of coadaptation** of **introgressed gene** with **target genome**
- 4 parameters



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* 2020

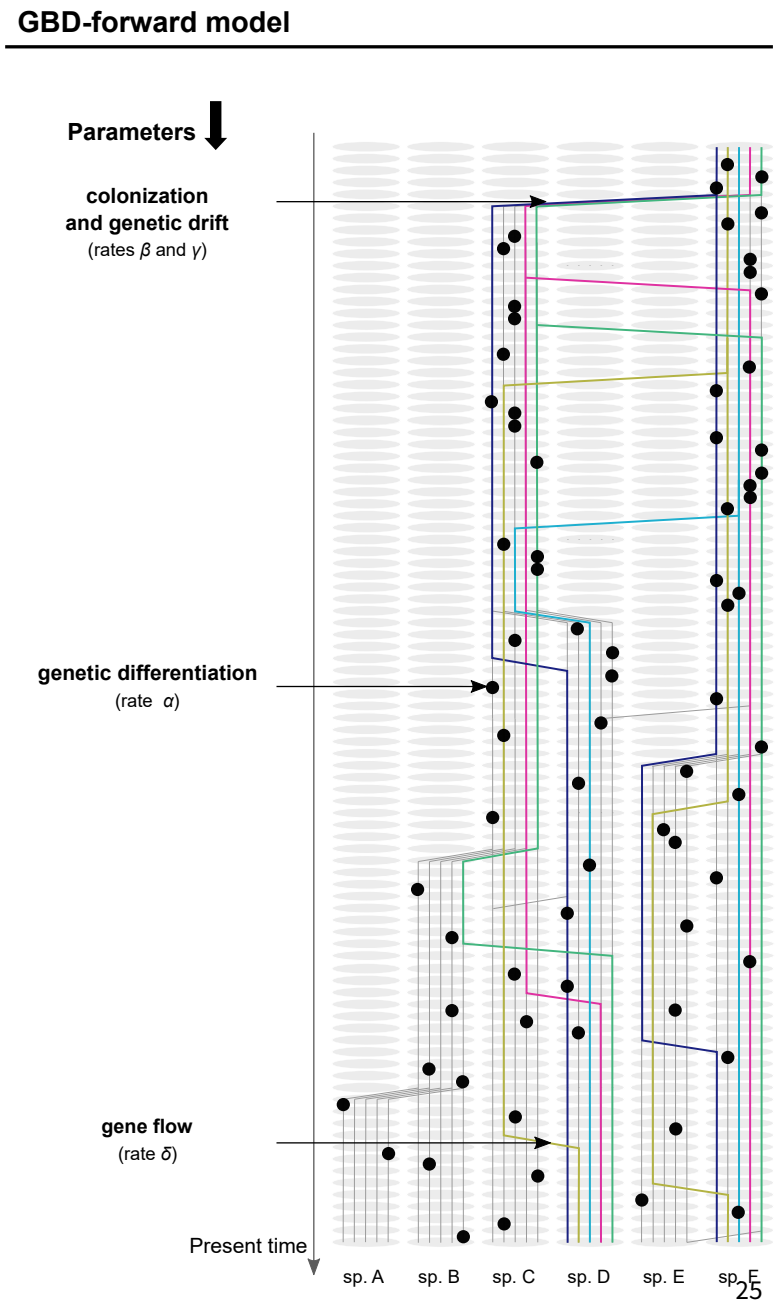
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)
- Populations exchange genes at rate δ per lineage thinned by the **degree of coadaptation** of **introgressed gene** with **target genome**
- 4 parameters
- As in MSC : neutral, no explicit modeling of physical linkage



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* 2020

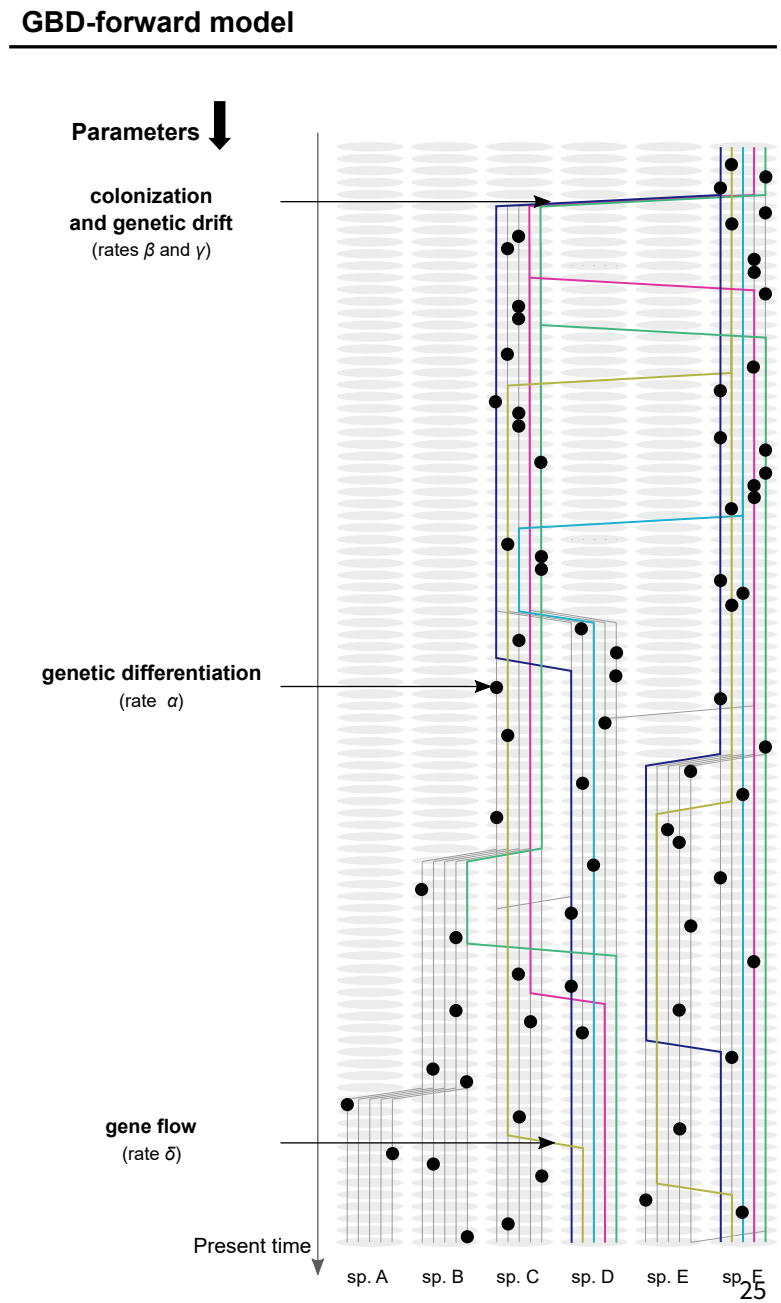
- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)
- Populations exchange genes at rate δ per lineage thinned by the **degree of coadaptation** of **introgressed gene** with **target genome**
- 4 parameters
- As in MSC : neutral, no explicit modeling of physical linkage
- Explicit modeling of the progressive loss of co-adaptation
  = A model coupling gene trees (together and) with progressive emergence of reproductive isolation



Model 3. Gene-based diversification (Forward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* 2020

- Genomes accumulate mutations at rate α per lineage
- Populations/genomes replicate (rate β)
- Each pop is subject to genetic drift (rate γ, not shown)
- Populations exchange genes at rate δ per lineage thinned by the **degree of coadaptation** of **introgressed gene** with **target genome**
- 4 parameters
- As in MSC : neutral, no explicit modeling of physical linkage
- Explicit modeling of the progressive loss of co-adaptation = A model coupling gene trees (together and) with progressive emergence of reproductive isolation
- BUT prohibitively slow for simulation or inference.



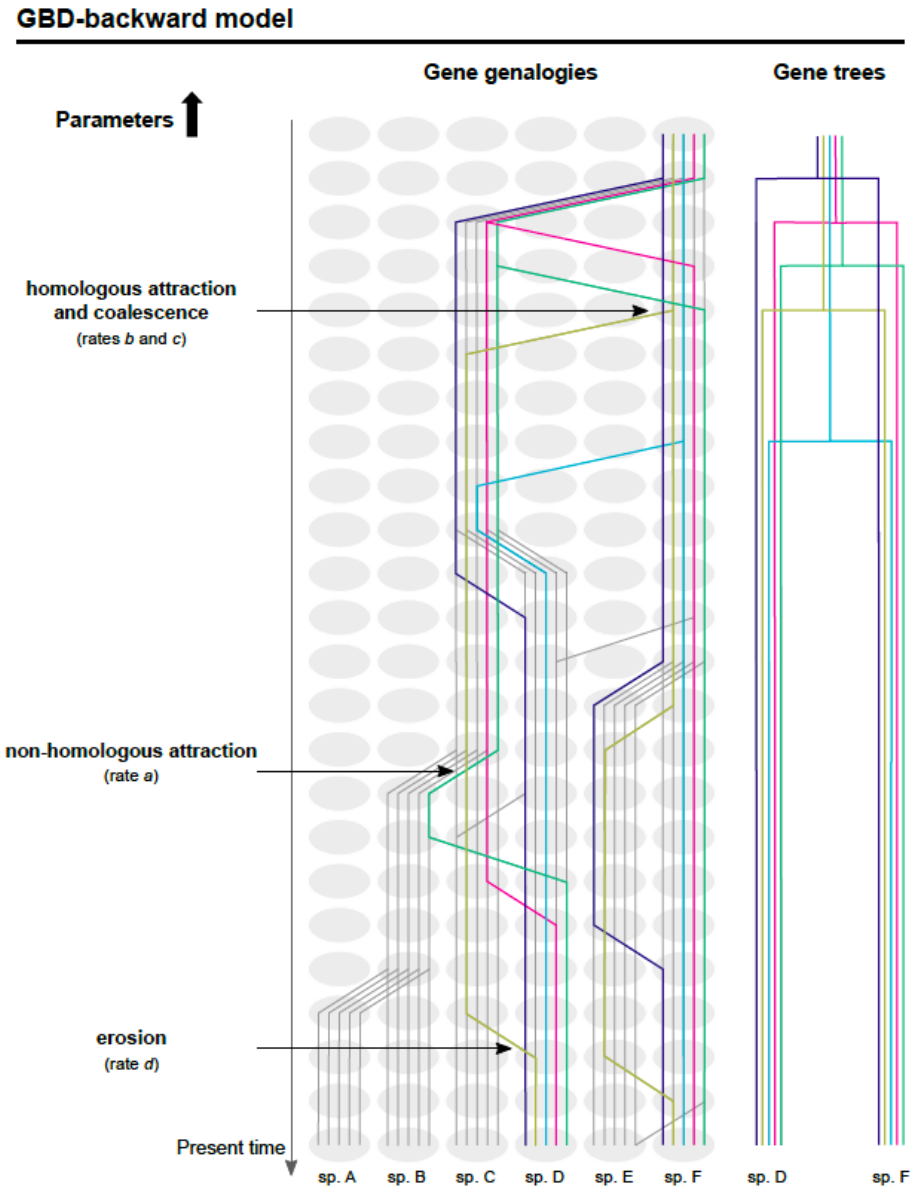
Model 3. Gene-based diversification (Backward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Rate *a* : Non-homologous attraction between **co-adapted** gene lineages

  ...Tunes the timescale of gene flow

  Large *a* ⇒ Most gene lineages concentrated in one or two species



Model 3. Gene-based diversification (Backward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Rate *a* : Non-homologous attraction between **co-adapted** gene lineages

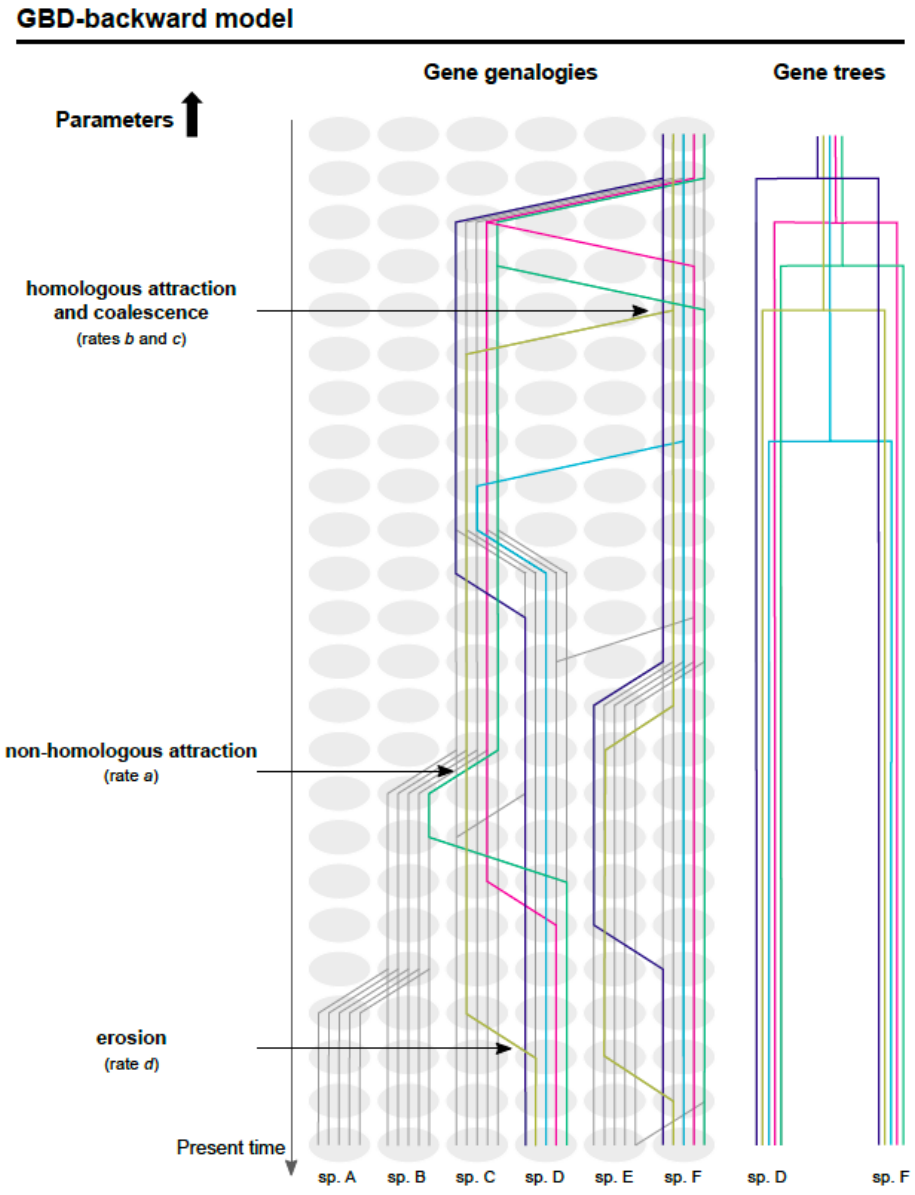
  ...Tunes the timescale of gene flow

  Large *a* ⇒ Most gene lineages concentrated in one or two species

- Rate *b* : Homologous attraction

  ...Tunes the speed of reproductive isolation

  Large *b* ⇒ Short wait between 'speciations'



Model 3. Gene-based diversification (Backward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Rate *a* : Non-homologous attraction between **co-adapted** gene lineages

  ...Tunes the timescale of gene flow

  Large *a* ⇒ Most gene lineages concentrated in one or two species

- Rate *b* : Homologous attraction

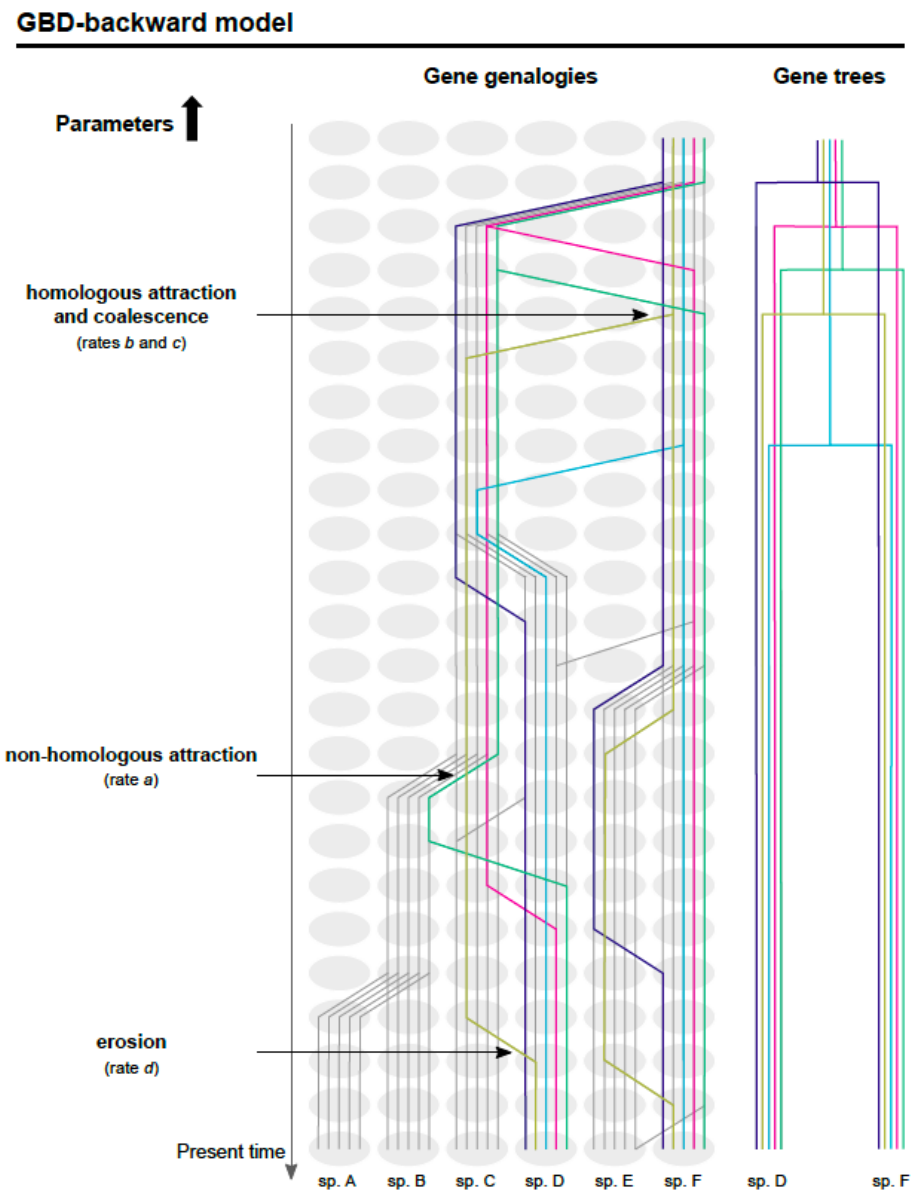
  ...Tunes the speed of reproductive isolation

  Large *b* ⇒ Short wait between 'speciations'

- Rate *c* : Coalescence

  ...Tunes the degree of polymorphism

  Large *c* ⇒ No incomplete lineage sorting



Model 3. Gene-based diversification (Backward)

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Rate *a* : Non-homologous attraction between **co-adapted** gene lineages

  ...Tunes the timescale of gene flow

  Large *a* ⇒ Most gene lineages concentrated in one or two species

- Rate *b* : Homologous attraction

  ...Tunes the speed of reproductive isolation

  Large *b* ⇒ Short wait between 'speciations'

- Rate *c* : Coalescence

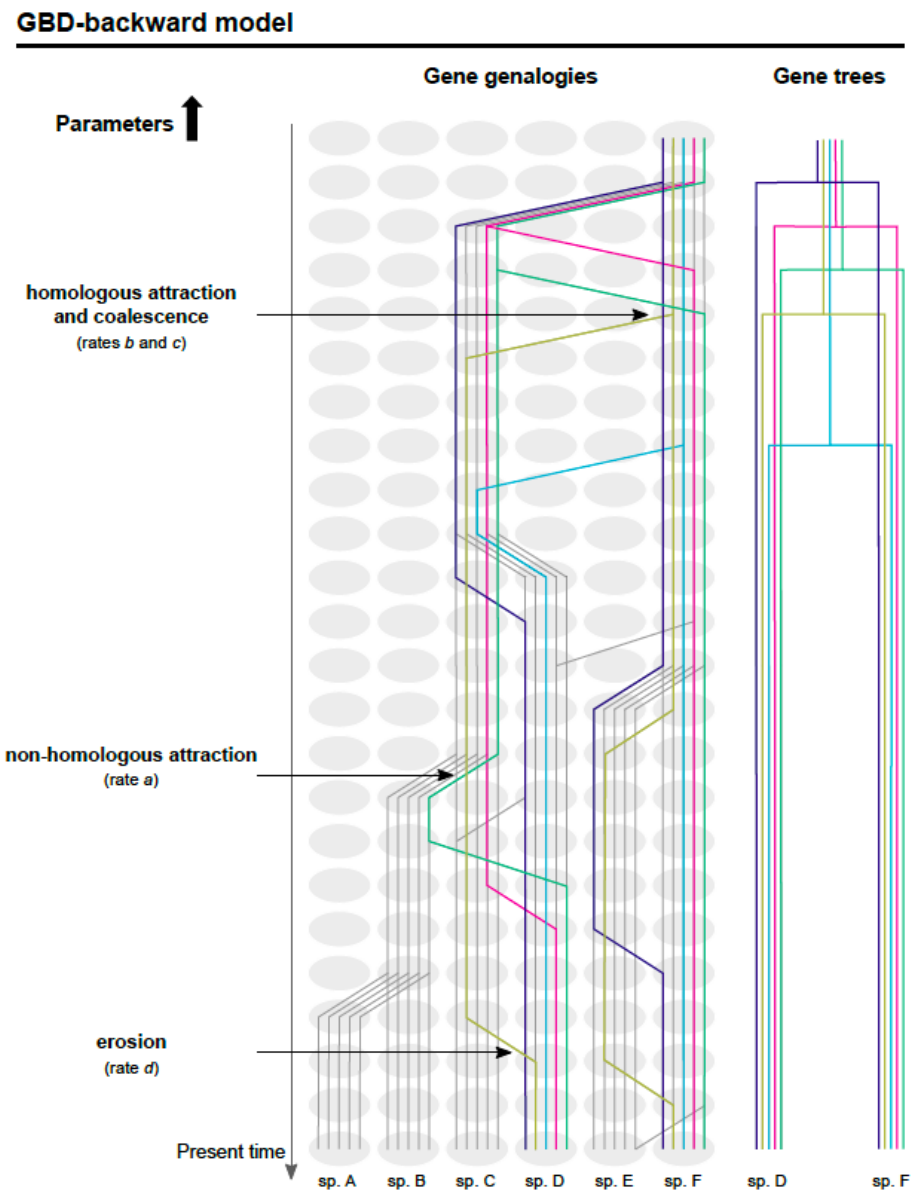
  ...Tunes the degree of polymorphism

  Large *c* ⇒ No incomplete lineage sorting

- Rate *d* : Erosion

  ...Tunes the number of pops exchanging genes

  Large *d* ⇒ Many ancestral species



Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j, k) \to (j-1, k+1)$ at rate ajk

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j, k) \to (j-1, k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1, 1)$ at rate dj

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j, k) \to (j-1, k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1, 1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :
  - Nb of ancestral species to n genes $\#\Pi_n(\infty) \sim \theta \ln(n)$

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :
  - Nb of ancestral species to n genes $\#\Pi_n(\infty) \sim \theta \ln(n)$
  - Few ancestral species containing k genes $\sim \frac{\theta}{k}$

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :
  - Nb of ancestral species to n genes $\#\Pi_n(\infty) \sim \theta \ln(n)$
  - Few ancestral species containing k genes $\sim \frac{\theta}{k}$
  - No dust + block frequencies given by the Poisson–Dirichlet distribution obtained by re-ordering $(\beta_j)_{j\geq 1}$ defined as

$$\beta_j := Z_j \prod_{i=1}^{j-1}(1-Z_i),$$

where the (Z_i) are i.i.d. with density $\theta(1-z)^{\theta-1}$.

Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :
  - Nb of ancestral species to n genes $\#\Pi_n(\infty) \sim \theta \ln(n)$
  - Few ancestral species containing k genes $\sim \frac{\theta}{k}$
  - No dust + block frequencies given by the Poisson–Dirichlet distribution obtained by re-ordering $(\beta_j)_{j\geq 1}$ defined as

$$\beta_j := Z_j \prod_{i=1}^{j-1}(1-Z_i),$$

where the (Z_i) are i.i.d. with density $\theta(1-z)^{\theta-1}$.

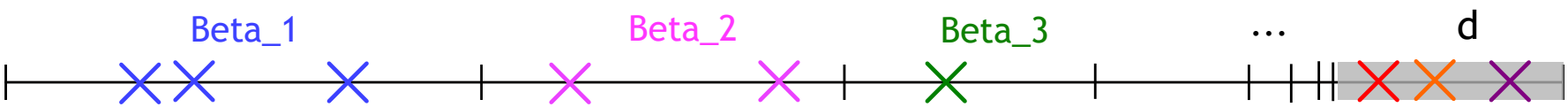
Properties of the GBD backward model

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Coalescent-based, fast to simulate
- Plastic : can incorporate time-inhomogeneity, gene-specific rates...
- Sampling-consistent
- Ancestral configuration of one single genome $\Pi_n(t)$
  - Non-homologous attraction = pairs of blocks with masses $(j,k) \to (j-1,k+1)$ at rate ajk
  - Erosion = block with mass $j \to (j-1,1)$ at rate dj
  - = Moran model with resampling rate a and mutation rate d.
- As $t \to \infty$, $\Pi_n(t)$ is given by Ewens' sampling formula with $\theta = d/a$ (see Lecture 1) :
  - Nb of ancestral species to n genes $\#\Pi_n(\infty) \sim \theta \ln(n)$
  - Few ancestral species containing k genes $\sim \frac{\theta}{k}$
  - No dust + block frequencies given by the Poisson–Dirichlet distribution obtained by re-ordering $(\beta_j)_{j\geq 1}$ defined as

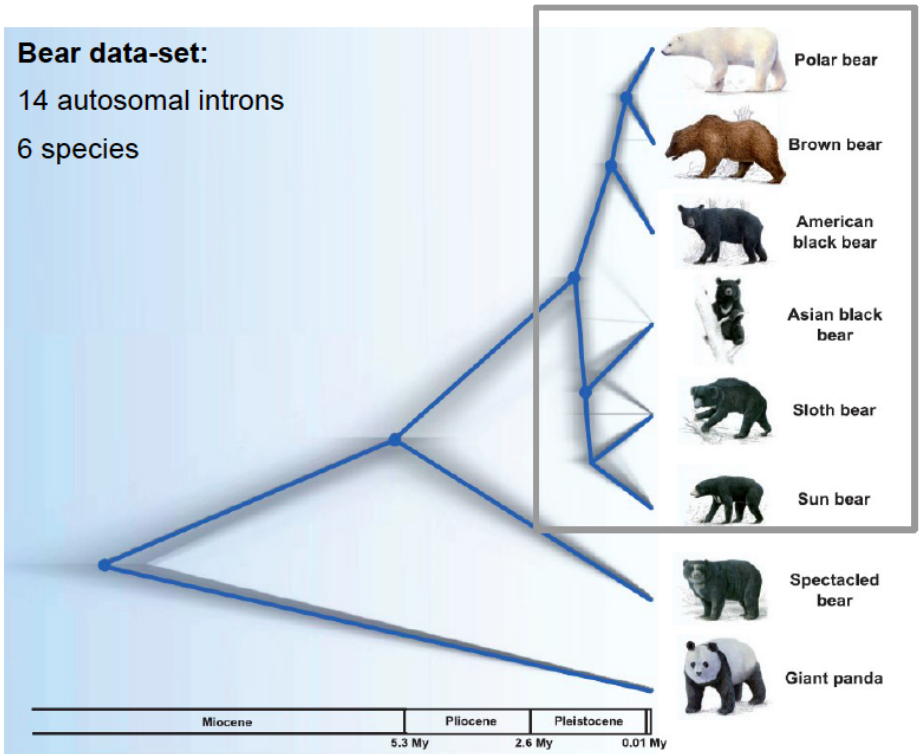
$$\beta_j := Z_j \prod_{i=1}^{j-1}(1-Z_i),$$

where the (Z_i) are i.i.d. with density $\theta(1-z)^{\theta-1}$.

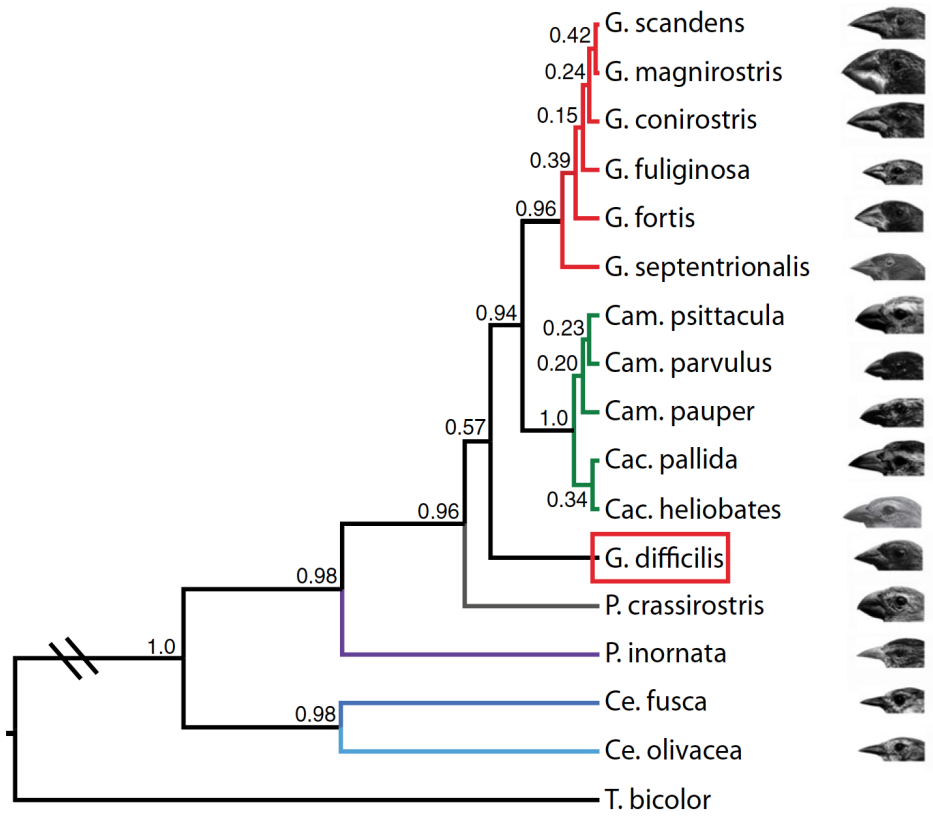


Bear and Finch datasets

Kutschera et al 14, Kumar et al 17, Farrington et al 14



Kutschera et al *MBE* 14



Farrington et al *Evolution* 14

Bears = 6 species, 14 loci

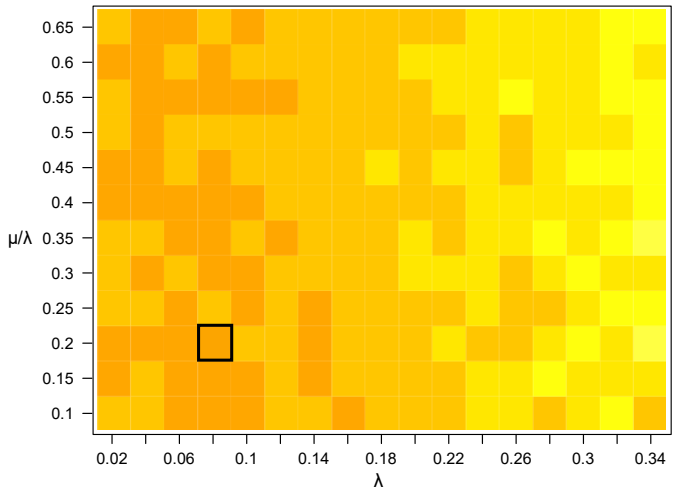
Finches = 6 species, 7 loci

ABC on Bear and Finch datasets

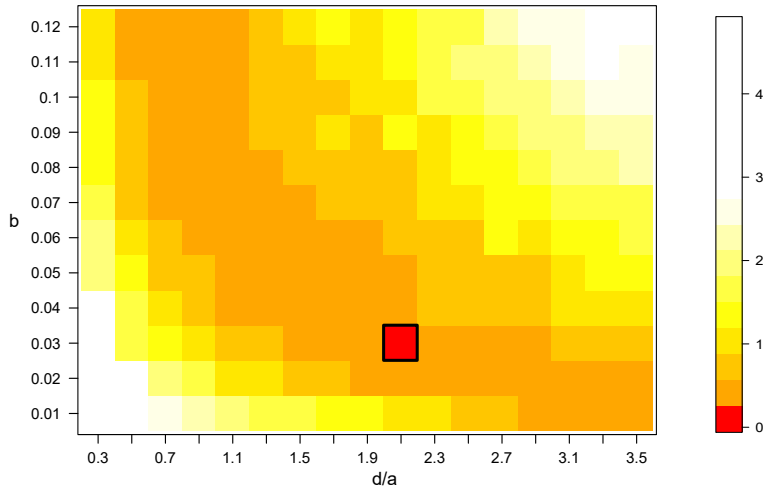
Marin, Achaz, Crombach & Lambert “The genomic view of diversification”, *J Evol Biol* (2020)

Bear data-set

Multi-species coalescent (MSC) model

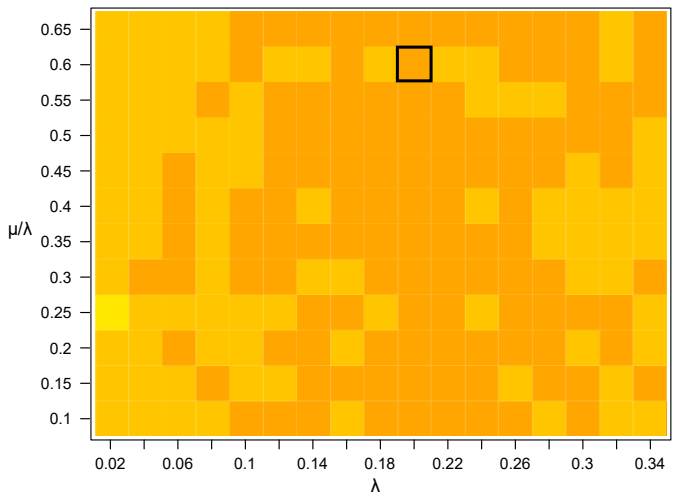


Gene-based diversification (GBD) model

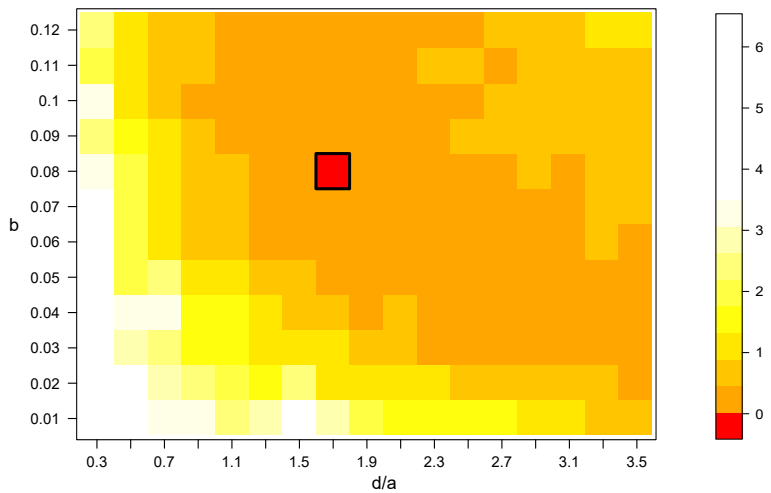


Finch data-set

Multi-species coalescent (MSC) model



Gene-based diversification (GBD) model

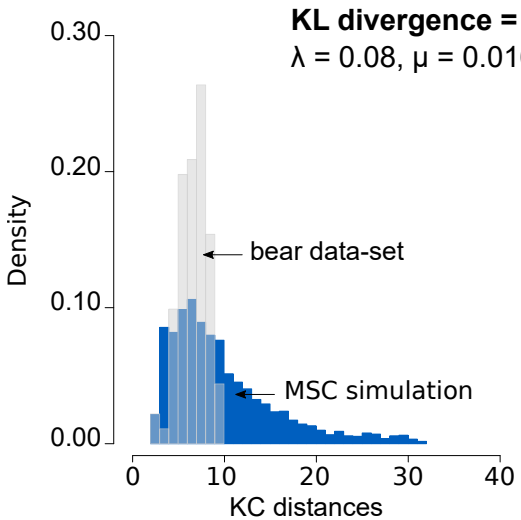


ABC on Bear and Finch datasets

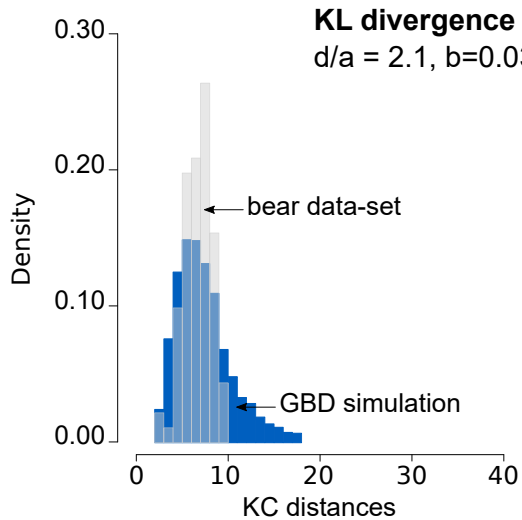
Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

Bear data-set

Multi-species coalescent (MSC) model

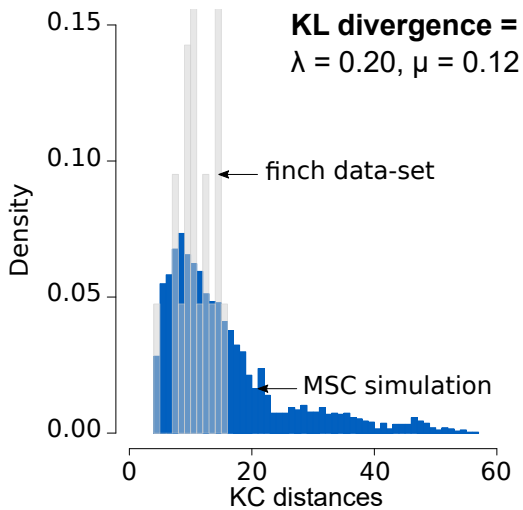


Gene-based diversification (GBD) model

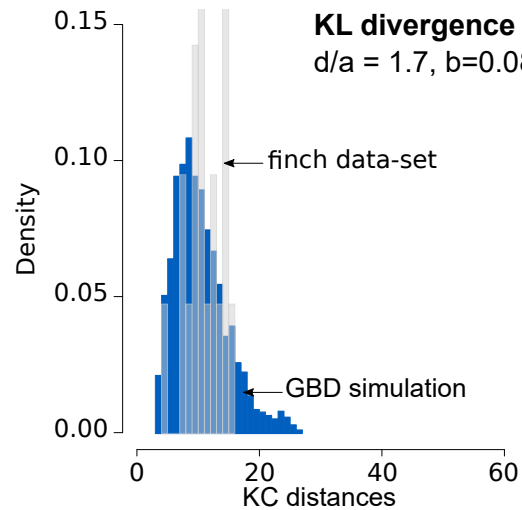


Finch data-set

Multi-species coalescent (MSC) model



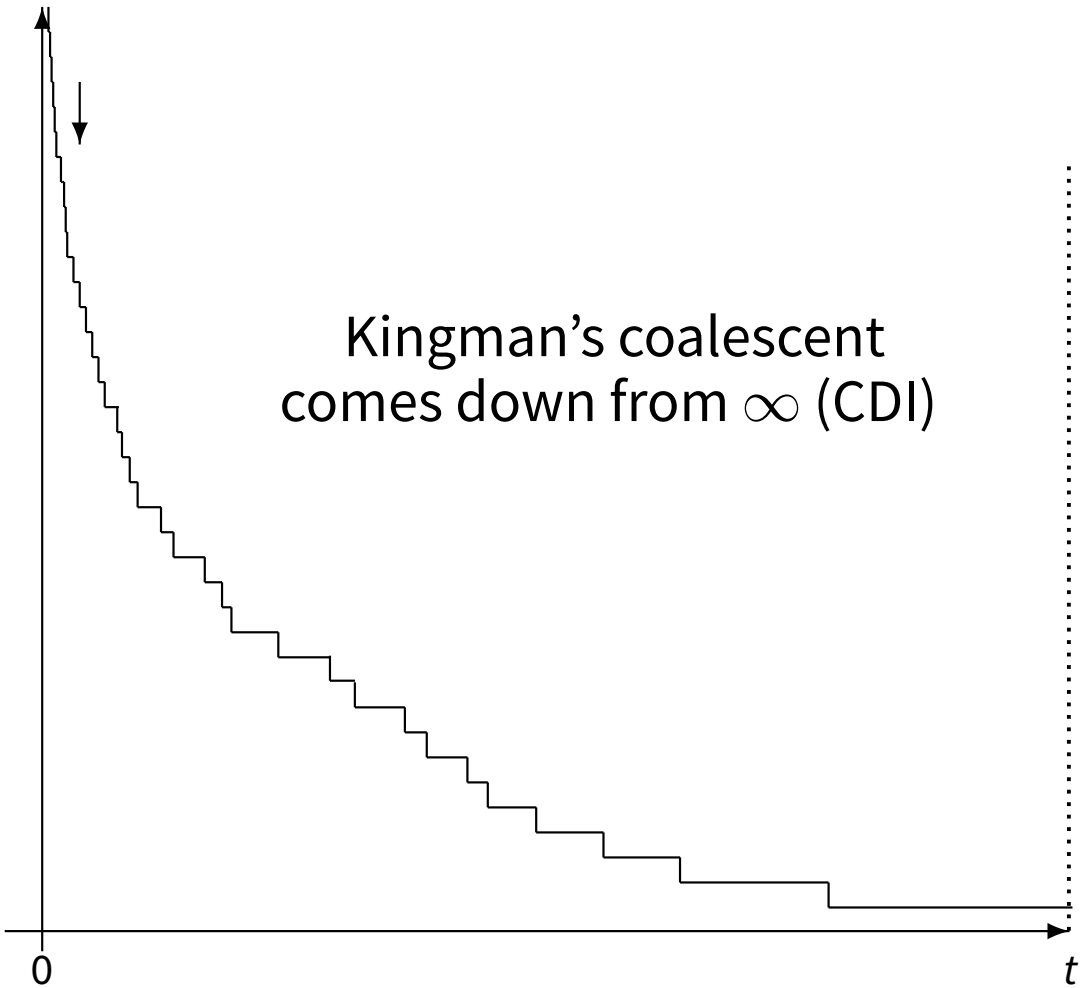
Gene-based diversification (GBD) model



Outline

How does Kingman's coalescent CDI ?

CDI = come down from infinity



- The process K_t counting the number of blocks in Kingman's coalescent is a pure-death process going from k to $k-1$ at rate $\binom{k}{2}$

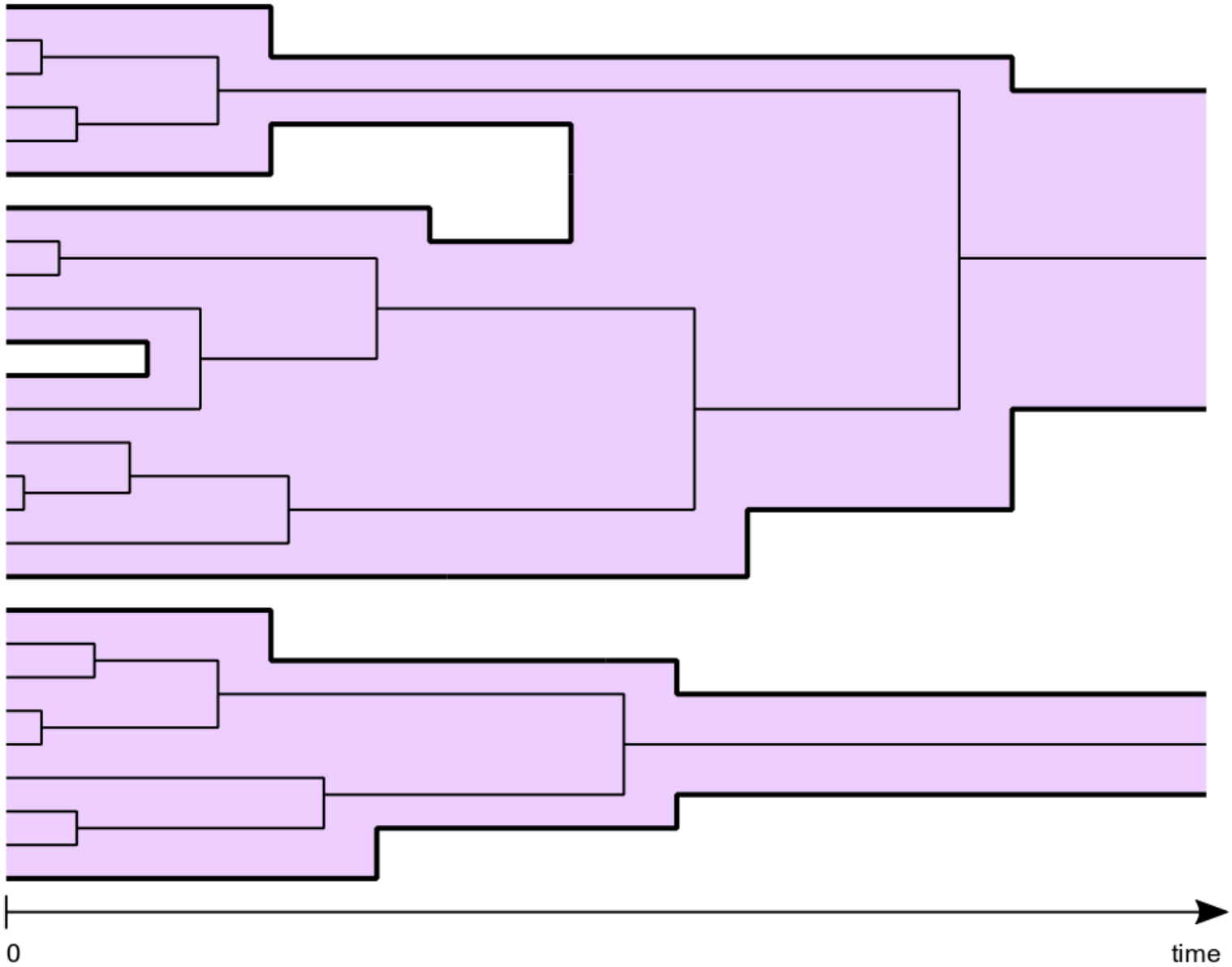
- Under $\mathbb{P}_\infty$, K_t behaves like the solution to $\dot{x} = -\frac{x^2}{2}$ starting at ∞, equal to

$$x(t) = \frac{2}{t}$$

i.e.

$$\mathbb{P}_\infty(tK_t \to 2) = 1.$$

Kingman in Kingman



How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1

How do nested coalescents come down from infinity ? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c

How do nested coalescents come down from infinity ? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

- Then the Λ-coalescent CDI iff

$$\text{[Grey's condition]} \qquad \int_{\lambda_0}^{\infty} \frac{dx}{\psi(x)} < \infty.$$

and then it behaves at time 0^+ like $\dot{x} = -\psi(x)$ starting at ∞.

Schweinsberg (2000), Berestycki, Berestycki & Limic (2010)

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

- Then the Λ-coalescent CDI iff

$$\text{[Grey's condition]} \qquad \int_{\lambda_0}^{\infty} \frac{dx}{\psi(x)} < \infty.$$

  and then it behaves at time 0^+ like $\dot{x} = -\psi(x)$ starting at ∞.

  Schweinsberg (2000), Berestycki, Berestycki & Limic (2010)

- Do nested coalescents CDI? Yes : Domination by coal's started at ∞.

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

- Then the Λ-coalescent CDI iff

$$\text{[Grey's condition]} \qquad \int_{\lambda_0}^{\infty} \frac{dx}{\psi(x)} < \infty.$$

and then it behaves at time 0^+ like $\dot{x} = -\psi(x)$ starting at ∞.

Schweinsberg (2000), Berestycki, Berestycki & Limic (2010)

- Do nested coalescents CDI? Yes : Domination by coal's started at ∞.
- How do nested coalescents CDI?

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

- Then the Λ-coalescent CDI iff

$$\text{[Grey's condition]} \qquad \int_{\lambda_0}^{\infty} \frac{dx}{\psi(x)} < \infty.$$

and then it behaves at time 0^+ like $\dot{x} = -\psi(x)$ starting at ∞.

Schweinsberg (2000), Berestycki, Berestycki & Limic (2010)

- Do nested coalescents CDI? Yes : Domination by coal's started at ∞.
- How do nested coalescents CDI?
  1. Simple coalescents : speed is determined by a simple ODE with ∞-initial condition.

How do nested coalescents come down from infinity? (CDI)

- Here, pairs of species coalesce at rate 1
- Genes : provided they lie within the same species,
  - Pairs of genes coalesce at rate c
  - And mutiple mergers according to a Λ-coalescent (CDI)
- Set

$$(\Lambda) \qquad \psi(\lambda) := \frac{c\lambda^2}{2} + \int_{(0,1]} (e^{-\lambda r} - 1 + \lambda r)\, r^{-2} \Lambda(dr)$$

- Then the Λ-coalescent CDI iff

$$\text{[Grey's condition]} \qquad \int_{\lambda_0}^{\infty} \frac{dx}{\psi(x)} < \infty.$$

and then it behaves at time 0^+ like $\dot{x} = -\psi(x)$ starting at ∞.

Schweinsberg (2000), Berestycki, Berestycki & Limic (2010)

- Do nested coalescents CDI? Yes : Domination by coal's started at ∞.
- How do nested coalescents CDI?
  1. Simple coalescents : speed is determined by a simple ODE with ∞-initial condition.
  2. Nested coalescents : speed is determined by a degenerate PDE.

The coagulation-transport PDE

- The **transport equation** with velocity $-\psi$

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles with mass x at time t, where particles **spontaneously lose mass** as in the ODE $\dot{x} = -\psi(x)$.

The coagulation-transport PDE

- The **transport equation** with velocity $-\psi$

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \qquad t,x \geq 0$$

  describes the evolution of the concentration $d(t,x)$ of particles with mass x at time t, where particles **spontaneously lose mass** as in the ODE $\dot{x} = -\psi(x)$.

- Here, $\psi(0) = 0$, $\psi \nearrow$ and $\int_{\lambda_0}^{\infty} dx/\psi(x) < \infty$ (Grey's cond)

The coagulation-transport PDE

- The **transport equation** with velocity $-\psi$

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles with mass x at time t, where particles **spontaneously lose mass** as in the ODE $\dot{x} = -\psi(x)$.

- Here, $\psi(0) = 0$, $\psi \nearrow$ and $\int_{\lambda_0}^{\infty} dx/\psi(x) < \infty$ (Grey's cond)

- The **coagulation-transport equation** with velocity $-\psi$ and coagulation rate a

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \;+\; a(t)\,(d \star d(t,x) \;-\; d(t,x)) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles which lose mass as in the ODE $\dot{x} = -\psi(x)$ and **gain mass by pairwise coagulation** at rate a.

The coagulation-transport PDE

- The **transport equation** with velocity $-\psi$

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles with mass x at time t, where particles **spontaneously lose mass** as in the ODE $\dot{x} = -\psi(x)$.

- Here, $\psi(0) = 0$, $\psi \nearrow$ and $\int_{\lambda_0}^{\infty} dx/\psi(x) < \infty$ (Grey's cond)

- The **coagulation-transport equation** with velocity $-\psi$ and coagulation rate a

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \;+\; a(t)\,(d \star d(t,x) \;-\; d(t,x)) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles which lose mass as in the ODE $\dot{x} = -\psi(x)$ and **gain mass by pairwise coagulation** at rate a.

The coagulation-transport PDE

- The **transport equation** with velocity $-\psi$

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles with mass x at time t, where particles **spontaneously lose mass** as in the ODE $\dot{x} = -\psi(x)$.

- Here, $\psi(0) = 0$, $\psi \nearrow$ and $\int_{\lambda_0}^{\infty} dx/\psi(x) < \infty$ (Grey's cond)

- The **coagulation-transport equation** with velocity $-\psi$ and coagulation rate a

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \;+\; a(t)\,(d \star d(t,x) \;-\; d(t,x)) \qquad t,x \geq 0$$

describes the evolution of the concentration $d(t,x)$ of particles which lose mass as in the ODE $\dot{x} = -\psi(x)$ and **gain mass by pairwise coagulation** at rate a.

Note : $d \star d(t,x) = \int_0^x d(t,x-y)d(t,y)\,dy$ denotes the convolution product.

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Start with ∞ many species, each one containing at least one gene.

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Start with ∞ many species, each one containing at least one gene.
- Then $n^{-1}S(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t}$

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Start with ∞ many species, each one containing at least one gene.
- Then $n^{-1}S(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t}$

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Start with ∞ many species, each one containing at least one gene.
- Then $n^{-1}S(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t}$

Theorem (Lambert & Schertzer (2020))

Under these conditions, $(G_n(t/n); t \geq 0)$ *converges to* **the unique solution** *of the* ***coagulation-transport equation***

$$\partial_t d(t,x) = \partial_x(\psi d)(t,x) + a(t)\,(d \star d(t,x) - d(t,x)) \quad t > 0, x \geq 0$$

such that $\limsup_{t\downarrow 0} \int_y^\infty d(t,x)\,dx > 0$ *(for some* y*)* $=:$ *proper solution.*

- Solution starts at δ_∞ and ends at δ_0.

∞ species with at least one gene (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* 2020

- Let $S(t) := \#$ species, $\Pi^i(t) := \#$ genes or **mass** of the i-th species

$$G_n(t) := \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Start with ∞ many species, each one containing at least one gene.
- Then $n^{-1}S(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t}$

Theorem (Lambert & Schertzer (2020))

Under these conditions, $(G_n(t/n); t \geq 0)$ *converges to* **the unique solution** *of the* ***coagulation-transport equation***

$$\partial_t d(t,x) \;=\; \partial_x(\psi d)(t,x) \;+\; a(t)\,(d \star d(t,x) \;-\; d(t,x)) \qquad t > 0, x \geq 0$$

such that $\limsup_{t \downarrow 0} \int_y^\infty d(t,x)\,dx > 0$ *(for some y)* $=:$ *proper solution.*

- Solution starts at δ_∞ and ends at δ_0.
- Uniqueness of solution holds under the sole condition that $\limsup_{t \downarrow 0} \int_y^\infty d(t,x)\,dx > 0$ for some y : starting at ∞ many species does not require tuning initial number of genes : species coalescences dominate gene coalescences so as to generate initial condition.

CDI of the nested Kingman coalescent

see also Blancas, Rogers, Schweinsberg & Siri-Jégousse (2018)

- For $\psi(x) = cx^\gamma$ ($\gamma > 1$) and $a(t) = 1/t$, if Υ is the rv with density $d(1, \cdot)$, then $d(t, \cdot)$ is the density of

$$t^{-\frac{1}{\gamma-1}}\,\Upsilon$$

CDI of the nested Kingman coalescent

see also Blancas, Rogers, Schweinsberg & Siri-Jégousse (2018)

- For $\psi(x) = cx^\gamma$ ($\gamma > 1$) and $a(t) = 1/t$, if Υ is the rv with density $d(1, \cdot)$, then $d(t, \cdot)$ is the density of

$$t^{-\frac{1}{\gamma-1}}\,\Upsilon$$

- Recall

$$G_n(t) = \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

CDI of the nested Kingman coalescent

see also Blancas, Rogers, Schweinsberg & Siri-Jégousse (2018)

- For $\psi(x) = cx^\gamma$ ($\gamma > 1$) and $a(t) = 1/t$, if Υ is the rv with density $d(1, \cdot)$, then $d(t, \cdot)$ is the density of

$$t^{-\frac{1}{\gamma-1}}\,\Upsilon$$

- Recall

$$G_n(t) = \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Let $\rho(t) =$ total # genes, so that for any n

$$\rho(t) = S(t) \int nx\, [G_n(t)](dx).$$

CDI of the nested Kingman coalescent

see also Blancas, Rogers, Schweinsberg & Siri-Jégousse (2018)

- For $\psi(x) = cx^\gamma$ ($\gamma > 1$) and $a(t) = 1/t$, if Υ is the rv with density $d(1, \cdot)$, then $d(t, \cdot)$ is the density of

$$t^{-\frac{1}{\gamma-1}}\,\Upsilon$$

- Recall

$$G_n(t) = \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Let $\rho(t) = \text{total \# genes}$, so that for any n

$$\rho(t) = S(t) \int nx\, [G_n(t)](dx).$$

- So as $n \to \infty$

$$\rho(t/n) = nS(t/n) \int x\, [G_n(t/n)](dx) \sim \frac{2n^2}{t} \int x\, d(t, x)\, dx = \frac{2n^2}{t^2}\, \mathbb{E}(\Upsilon),$$

CDI of the nested Kingman coalescent

see also Blancas, Rogers, Schweinsberg & Siri-Jégousse (2018)

- For $\psi(x) = cx^\gamma$ ($\gamma > 1$) and $a(t) = 1/t$, if Υ is the rv with density $d(1, \cdot)$, then $d(t, \cdot)$ is the density of

$$t^{-\frac{1}{\gamma-1}}\,\Upsilon$$

- Recall

$$G_n(t) = \frac{1}{S(t)} \sum_{i=1}^{S(t)} \delta_{\Pi^i(t)/n}.$$

- Let $\rho(t) =$ total # genes, so that for any n

$$\rho(t) = S(t) \int nx\, [G_n(t)](dx).$$

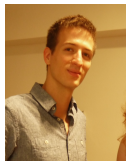
- So as $n \to \infty$

$$\rho(t/n) = nS(t/n) \int x\, [G_n(t/n)](dx) \sim \frac{2n^2}{t} \int x\, d(t, x)\, dx = \frac{2n^2}{t^2}\, \mathbb{E}(\Upsilon),$$

- In other words

$$\lim_{t \downarrow 0} t^2 \rho(t) = 2\mathbb{E}(\Upsilon).$$

Collaborators

Jean-Jil DUCHAMPS*, Airam BLANCAS & Arno SIRI-JÉGOUSSE   

Chunhua MA* (Nankin U) . . . . . 

Félix FOUTEL–RODIER* & Emmanuel SCHERTZER* . . . . .  

Julie MARIN*, Guillaume ACHAZ* & Anton CROMBACH . . . . .   

Emmanuel SCHERTZER* . . . . . 

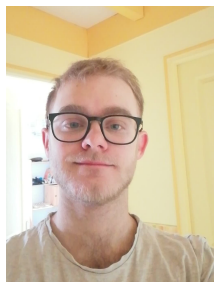
*= members or former members of SMILE.

SMILE : An interdisciplinary group in Paris

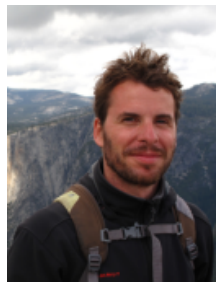
Below : SMILE members in May 2020

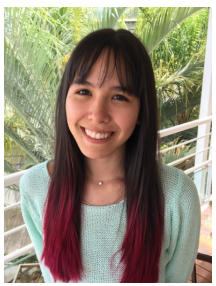

Jasmine 12/01


François 15/01


Guillaume T. 25/02


Pete 29/03


Emmanuel 18/04


Laura 21/05

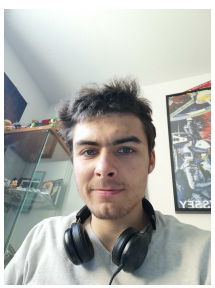

Jean-Jil 24/05


Felix 05/09


Rob 16/09

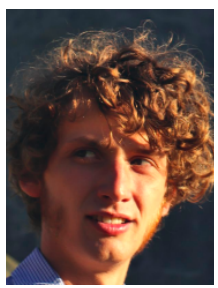

Julie 21/09

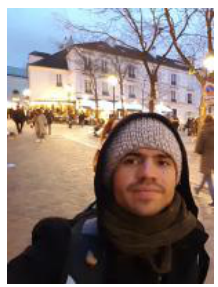

Elise 10/10


Léo 07/11


Guillaume A. 09/12


Amaury 16/12


Philibert 30/12


Alejandro

SMILE = **S**tochastic **M**odels for the **I**nference of **L**ife **E**volution

Outline

Books and monographs

- Gascuel, O. (Ed) (2005) *Mathematics of evolution and phylogeny*
- Steel, M. (2016) *Phylogeny : Discrete and random processes in evolution*
- Scornavacca, C., Delsuc, F., Galtier, N. (Eds) (2019) *Phylogenetics in the genomic era*
- Balding, D., Moltke, I., Marioni, J. (Eds) (2019) *Handbook of statistical genomics*
- Anisimova, M. (Eds) (2019) *Evolutionary genomics – statistical and computational methods*

Papers

- Schweinsberg, “A necessary and sufficient condition for the Λ-coalescent to come down from infinity”, *Elec Comm Prob* (2000)
- Berestycki, “Exchangeable fragmentation-coalescence processes and their equilibrium measures”, *Elec J Prob* (2004)
- Berestycki, Berestycki & Limic, “The Λ-coalescent speed of coming down from infinity”, *Ann Probab* (2010)
- Puillandre, Lambert, Brouillet, Achaz, “ABGD, Automatic Barcode Gap Discovery for primary species delimitation” *Mol Ecol* (2012)
- Lambert & Ma, “The coalescent in peripatric metapopulations”, *J Appl Prob* (2015)
- Blancas, Duchamps, Lambert & Siri–Jégousse, “Trees within trees : Simple nested coalescents”, *Elec J Prob* (2018)
- Blancas, Rogers, Schweinsberg & Siri-Jégousse, “The nested Kingman coalescent : Speed of coming down from infinity”, *Ann Appl Prob* (2019)
- Duchamps, “Trees within trees II : Nested fragmentations”, *Ann Inst H Poincaré Probab Stat* (2020)
- Foutel–Rodier, Lambert & Schertzer, “Kingman’s coalescent with erosion”, *Elec J Prob* (2020)
- Marin, Achaz, Crombach & Lambert, “The genomic view of diversification”, *J Evol Biol* (2020)
- Lambert & Schertzer, “Coagulation-transport equations and the nested coalescents”, *Probab Theory & Related Fields* (2020)

Comparing qualitative behavior of GBD Forward/GBD Backward

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- ⋆ 2 species sampled at present time
- ⋆ 1 genome with 20 homologous genes per species
- ⋆ GBD-forward : $a = 2, d = 6$
- ⋆ GBD backward : $\alpha = 0.1, \delta = 0.06$

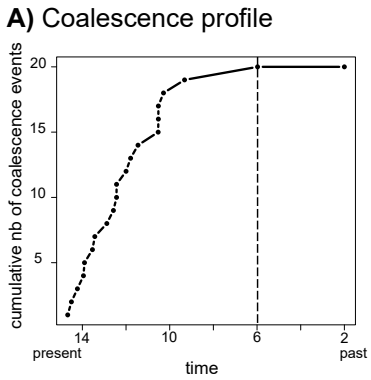
→ As time goes backward, we count...

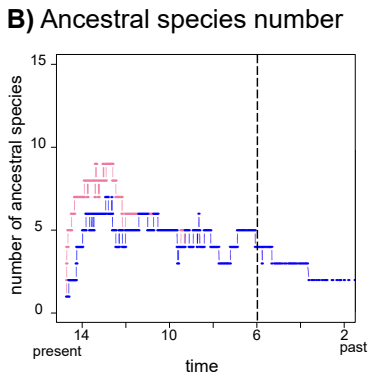
A) # Pairs of homologous gene lineages having coalesced :

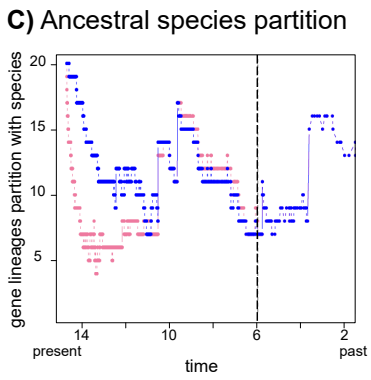
B) # Species ancestral to either present-day species :

C) # Gene lineages in main ancestral species :

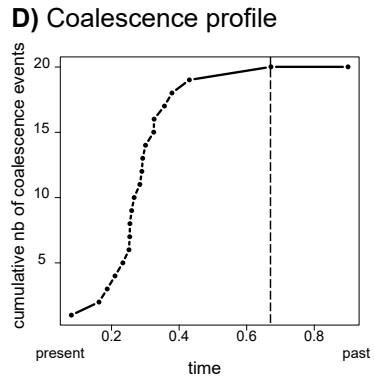
GBD Forward

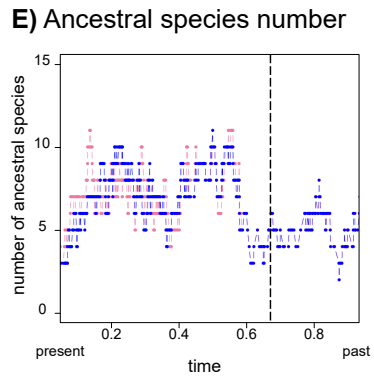


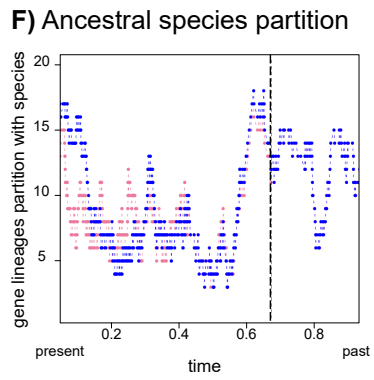




GBD Backward







ABC on distribution of pairwise distances between gene trees

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Simulate or infer gene trees

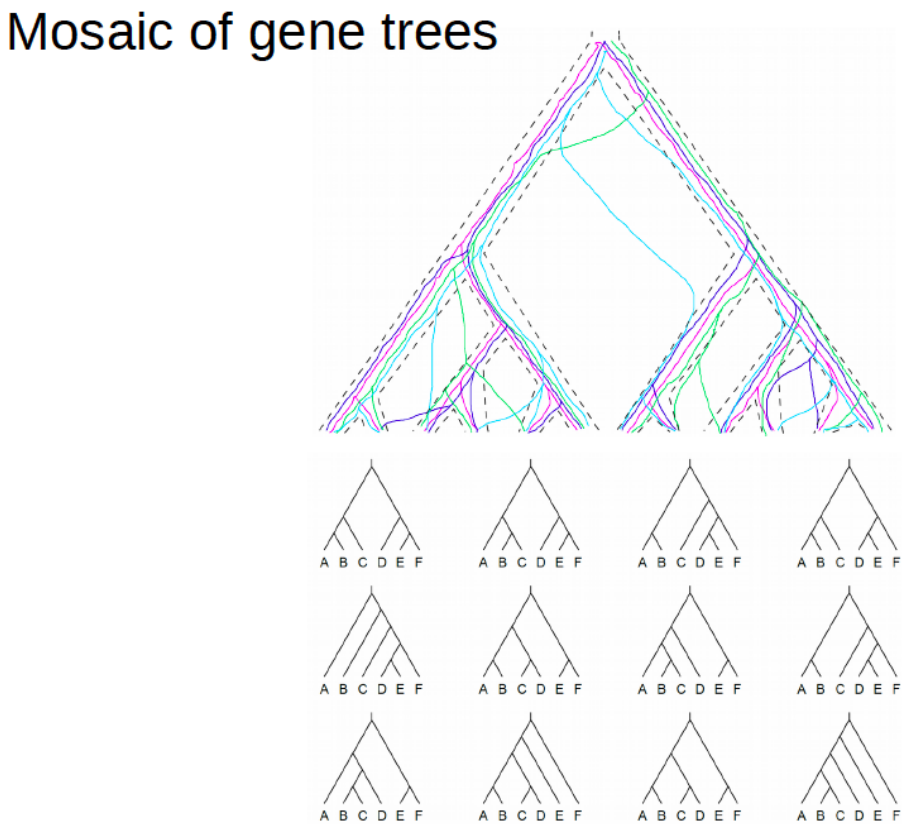
- Compute (BHV or KC) distance between each pair of gene trees

  (Billera-Holmes-Vogtmann 05, Kendall-Colijn 16)

- Compare distribution of pairwise distances via Kullback-Leibler divergence to simulations w parameters (a, b, d)

- Minimize KL divergence in parameter space

- Recall parameters :
  a tunes timescale of gene flow,
  b tunes timescale of reproductive isolation,
  d tunes # ancestral species.

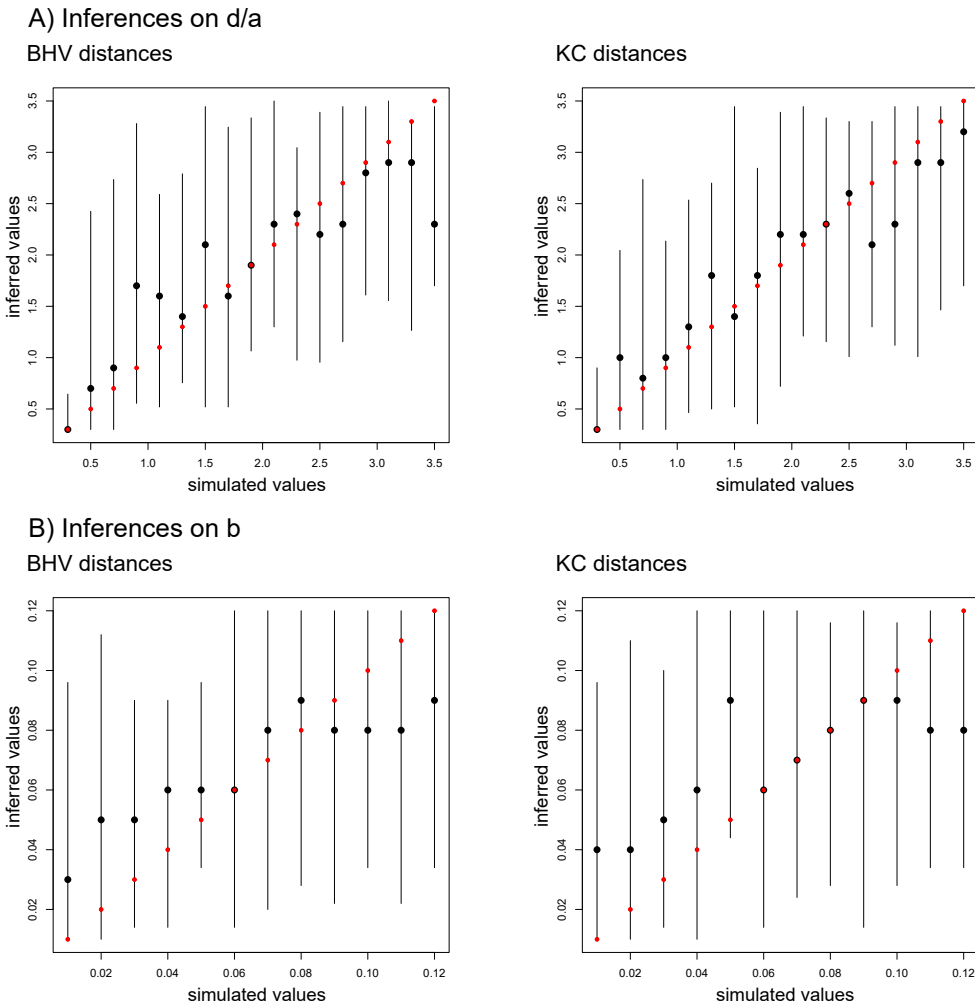


ABC on distribution of pairwise distances between gene trees

Marin, Achaz, Crombach & Lambert "The genomic view of diversification", *J Evol Biol* (2020)

- Simulate or infer gene trees
- Compute (BHV or KC) distance between each pair of gene trees

  (Billera-Holmes-Vogtmann 05, Kendall-Colijn 16)
- Compare distribution of pairwise distances via Kullback-Leibler divergence to simulations w parameters (a, b, d)
- Minimize KL divergence in parameter space
- Recall parameters :
  a tunes timescale of gene flow,
  b tunes timescale of reproductive isolation,
  d tunes # ancestral species.



$O(n)$ species with $O(n)$ genes (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* (2020)

- Let $S_n(t) := \#$ species, $\Pi_n^i(t) := \#$ genes or **mass** of the i-th species,

$$G_n(t) := \frac{1}{S_n(t)} \sum_{i=1}^{S_n(t)} \delta_{\Pi_n^i(t)/n}.$$

$O(n)$ species with $O(n)$ genes (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* (2020)

- Let $S_n(t) := \#$ species, $\Pi_n^i(t) := \#$ genes or **mass** of the i-th species,

$$G_n(t) := \frac{1}{S_n(t)} \sum_{i=1}^{S_n(t)} \delta_{\Pi_n^i(t)/n}.$$

- Start with $r\,n$ species containing genes in iid numbers w law ν^n.

$O(n)$ species with $O(n)$ genes (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* (2020)

- Let $S_n(t) := \#$ species, $\Pi_n^i(t) := \#$ genes or **mass** of the i-th species,

$$G_n(t) := \frac{1}{S_n(t)} \sum_{i=1}^{S_n(t)} \delta_{\Pi_n^i(t)/n}.$$

- Start with $r\,n$ species containing genes in iid numbers w law ν^n.

- Then $n^{-1} S_n(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t+\delta}$ with $\delta = \frac{2}{r}$

$O(n)$ species with $O(n)$ genes (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* (2020)

- Let $S_n(t) := \#$ species, $\Pi_n^i(t) := \#$ genes or **mass** of the i-th species,

$$G_n(t) := \frac{1}{S_n(t)} \sum_{i=1}^{S_n(t)} \delta_{\Pi_n^i(t)/n}.$$

- Start with $r\,n$ species containing genes in iid numbers w law ν^n.

- Then $n^{-1} S_n(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t+\delta}$ with $\delta = \frac{2}{r}$

$O(n)$ species with $O(n)$ genes (each)

Lambert & Schertzer, "Coagulation-transport equations and the nested coalescents", *Probability Theory & Related Fields* (2020)

- Let $S_n(t) := \#$ species, $\Pi_n^i(t) := \#$ genes or **mass** of the i-th species,

$$G_n(t) := \frac{1}{S_n(t)} \sum_{i=1}^{S_n(t)} \delta_{\Pi_n^i(t)/n}.$$

- Start with $r\,n$ species containing genes in iid numbers w law ν^n.

- Then $n^{-1} S_n(t/n) \to 2a(t)$, where $a(t) = \frac{1}{t+\delta}$ with $\delta = \frac{2}{r}$

Theorem (Lambert & Schertzer 2020)

If $G_n(0) \Longrightarrow \nu$, *then* $(G_n(t/n); t \geq 0)$ *converges to* **the unique solution** *of the* ***coagulation-transport equation***

$$(\star) \quad \partial_t d(t,x) = \partial_x(\psi d)(t,x) + a(t)\,(d \star d(t,x) - d(t,x)) \qquad t, x \geq 0$$

with initial condition $d(0,x)\,dx = \nu(dx)$.

weak solutions

Call $f : [0, \infty) \to \mathbb{R}$ a test-function if f is C^1, and f and $\psi f'$ are bounded.
For d solution of

$$(\star) \quad \partial_t d(t,x) = \partial_x(\psi d)(t,x) + a(t)\,(d \star d(t,x) - d(t,x)) \qquad t, x \geq 0$$

Set $\mu_t(dx) = d(t,x)\,dx$ and $\mu_t(f) = \int_{\mathbb{R}^+} f(x)\,\mu_t(dx)$.
Assuming that d is a probability density and integrating by parts, we get

$$\partial_t(\mu_t(f)) = -\mu_t(\psi f') + a(t)(\mu_t \star \mu_t(f) - \mu_t(f)).$$

weak solutions

Call $f : [0, \infty) \to \mathbb{R}$ a test-function if f is C^1, and f and $\psi f'$ are bounded.
For d solution of

$$(\star) \quad \partial_t d(t,x) = \partial_x(\psi d)(t,x) + a(t)\,(d \star d(t,x) - d(t,x)) \qquad t, x \geq 0$$

Set $\mu_t(dx) = d(t,x)\,dx$ and $\mu_t(f) = \int_{\mathbb{R}^+} f(x)\,\mu_t(dx)$.
Assuming that d is a probability density and integrating by parts, we get

$$\partial_t(\mu_t(f)) = -\mu_t(\psi f') + a(t)(\mu_t \star \mu_t(f) - \mu_t(f)).$$

Definition

A probability-valued process $(\mu_t; t \geq 0)$ is a ***weak solution of*** $(\star)$ *with initial condition ν if for every test-function f and every $t \geq 0$:*

$$\mu_t(f) = \nu(f) - \int_0^t \mu_s(\psi f')\,ds + \int_0^t a(s)(\mu_s \star \mu_s(f) - \mu_s(f))\,ds \qquad t \geq 0.$$

Mc-Kean–Vlasov solutions

Definition

A probability-valued process $(\mu_t; t \geq 0)$ *is a* ***McKean–Vlasov (MK–V) solution of*** $(\star)$ *with initial condition* ν *if there exists a process* $(x_t; t \geq 0)$ *such that* $\mu_t = \mathcal{L}(x_t)$ *and*

$$dx_t \;=\; -\psi(x_t)\,dt \;+\; v_t\,\Delta J_t, \qquad \mathcal{L}(x_0) = \nu$$

where J *is an inhomogeneous Poisson process with rate* $a(t)$ *at time* t*, and* $(v_t; t \geq 0)$ *is a sequence of independent rvs with* $\mathcal{L}(v_t) = \mathcal{L}(x_t)$.

Mc-Kean–Vlasov solutions

Definition

A probability-valued process $(\mu_t; t \geq 0)$ *is a* ***McKean–Vlasov (MK–V) solution of*** $(\star)$ *with initial condition* ν *if there exists a process* $(x_t; t \geq 0)$ *such that* $\mu_t = \mathcal{L}(x_t)$ *and*

$$dx_t \;=\; -\psi(x_t)\,dt \;+\; v_t\,\Delta J_t, \qquad \mathcal{L}(x_0) = \nu$$

where J *is an inhomogeneous Poisson process with rate* $a(t)$ *at time* t*, and* $(v_t; t \geq 0)$ *is a sequence of independent rvs with* $\mathcal{L}(v_t) = \mathcal{L}(x_t)$.

Proposition

If $(\mu_t; t \geq 0)$ *is a MK–V solution of* $(\star)$ *with initial condition* ν*, then it is a weak solution of* $(\star)$ *with initial condition* ν.

Proof : For every test-function f, take cond expectation of following eq (Itô) :

$$df(x_t) \;=\; -f'(x_{t-})\psi(x_{t-})\,dt \;+\; \big(f(x_{t-} + v_t) - f(x_{t-})\big)\,\Delta J_t.$$

$$\begin{aligned} \text{Existence of a MK–V solution} &\implies \text{Existence of a weak solution,} \\ \text{Uniqueness of the weak solution} &\implies \text{Uniqueness of the MK–V solution.} \end{aligned}$$

construction of a MK–V solution

- Remark 1. Zooming on (or conditional sampling) the Kingman coalescent gives the Brownian coalescent point process (CPP) :

  Lambert & Schertzer, "Recovering the Brownian coalescent point process from the Kingman coalescent by conditional sampling", *Bernoulli* (2019).

- Remark 2. The Brownian CPP has node depths given by a PPP with intensity $d\ell\, t^{-2} dt$ = leftmost lineage branches at rate $t^{-2} / \int_t^\infty s^{-2} ds = 1/t$

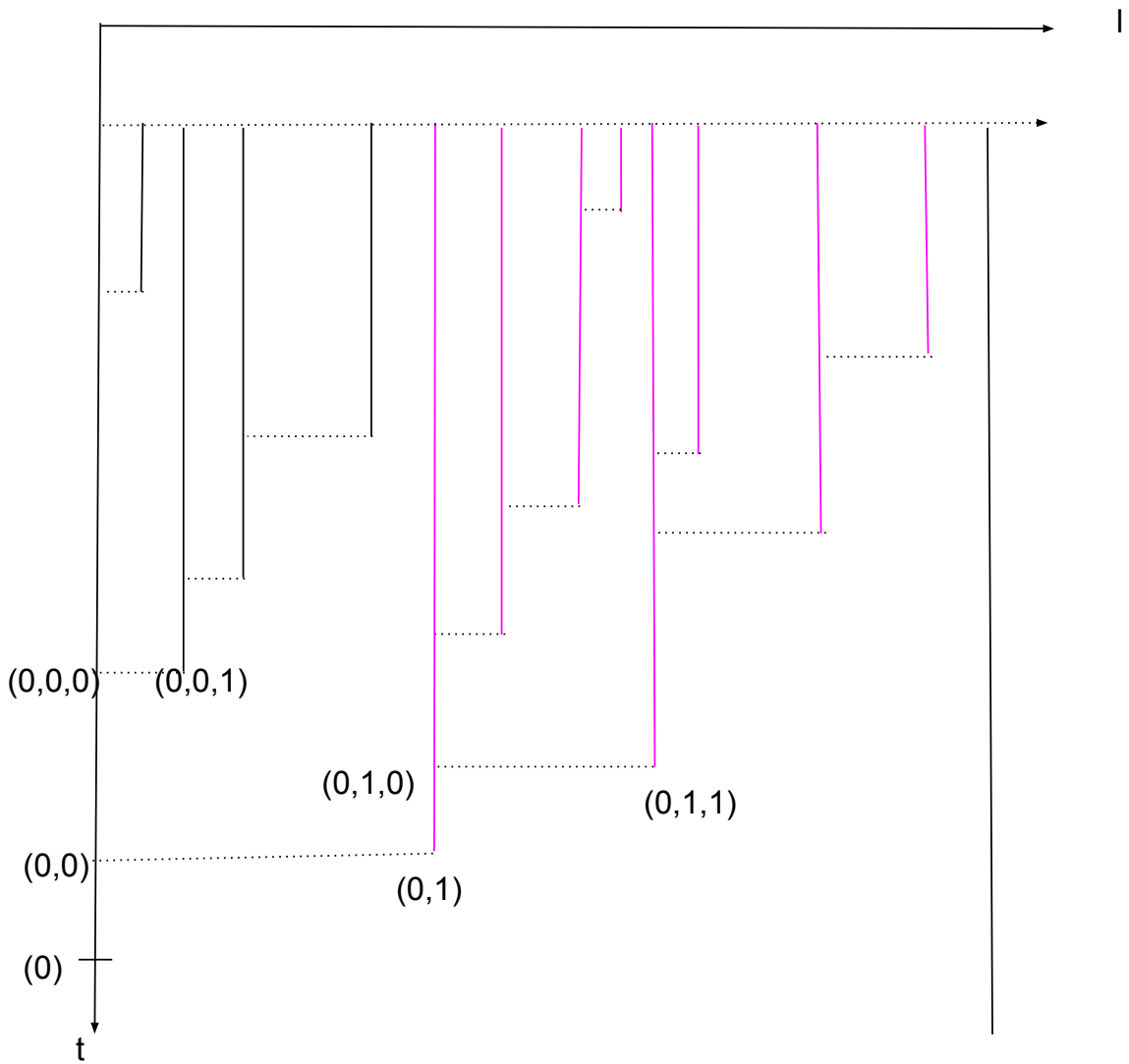
- Recall that $\mu_t = \mathcal{L}(x_t)$ is a MK–V solution of $(\star)$ with initial cond ν if

$$dx_t \;=\; -\psi(x_t)\, dt \;+\; v_t\, \Delta J_t, \qquad \mathcal{L}(x_0) = \nu$$

  where J jumps at rate $1/(t+\delta)$ and $\mathcal{L}(v_t) = \mathcal{L}(x_t)$.

Brownian Coalescent Point Process

Node depths form a Poisson point process with intensity $d\ell\, t^{-2} dt$



construction of a MK–V solution of $(\star)$

- Assume $\psi(0) = 0$ and ψ nondecreasing.
- Let $\mathbb{t}$ be a plane, binary tree embedded in time, leaves at time 0.
- Mark leaf i with a positive real number W_i
- Let each mark decrease with time as $\dot{x} = -\psi(x)$
- When two branches coalesce, add the two marks
- Let $X_t(\mathbb{t}, (W_i)) :=$ value of mark on leftmost branch at depth t.

construction of a MK–V solution of $(\star)$

- Assume $\psi(0) = 0$ and ψ nondecreasing.
- Let $\mathbb{t}$ be a plane, binary tree embedded in time, leaves at time 0.
- Mark leaf i with a positive real number W_i
- Let each mark decrease with time as $\dot{x} = -\psi(x)$
- When two branches coalesce, add the two marks
- Let $X_t(\mathbb{t}, (W_i)) :=$ value of mark on leftmost branch at depth t.

Theorem

If the (W_i) are iid $\sim \nu$ and $\mathbb{t}$ is a CPP with intensity $d\ell\,(t+\delta)^{-2}dt$, then $(X_t(\mathbb{t}, (W_i)); t \geq 0)$ is a MK–V solution to $(\star)$.

construction of a MK–V solution of $(\star)$

- Assume $\psi(0) = 0$ and ψ nondecreasing.
- Let $\mathbb{t}$ be a plane, binary tree embedded in time, leaves at time 0.
- Mark leaf i with a positive real number W_i
- Let each mark decrease with time as $\dot{x} = -\psi(x)$
- When two branches coalesce, add the two marks
- Let $X_t(\mathbb{t}, (W_i)) :=$ value of mark on leftmost branch at depth t.

Theorem

If the (W_i) are iid $\sim \nu$ and $\mathbb{t}$ is a CPP with intensity $d\ell\,(t+\delta)^{-2}dt$, then $(X_t(\mathbb{t}, (W_i)); t \geq 0)$ is a MK–V solution to $(\star)$.

In particular, $\mu_t^{(\delta)} := \mathcal{L}(X_t(\mathbb{t}, (W_i))$ is a weak solution to $(\star)$.

Laplace transform of the weak solution

- Let $(\mu_t; t \geq 0)$ be a weak solution to $(\star)$ and set

$$u(t, \lambda) := \int_{[0,\infty)} e^{-\lambda x} \mu_t(dx).$$

- Whenever ψ is of type (Λ), including $\psi(x) = cx^\gamma$ for $\gamma \in (1, 2]$, u is solution to

$$(L) \qquad \partial_t u = Au + a(t)(u^2 - u),$$

where A is the infinitesimal generator of a Markov process Z.

- The process $(Z_t; t \geq 0)$ is a critical, continuous-state branching process or CSBP with branching mechanism ψ.

- Under Grey's condition, Z is a.s. absorbed at 0 in finite time.

uniqueness of weak solution to $(\star)$ and $(\star\star)$

- Let $\mathbb{t}$ is a pure-birth tree with birth rate $\tilde{a}(t) = a(T - t)$ starting at time 0 and stopped at time T.
- Let Q_x^T be the law of a branching particle system
  - starting with one particle with mass x at time 0
  - each particle follows an edge of $\mathbb{t}$ and splits into two copies of itself at the next branching time
  - particle masses evolve like independent copies of Z between nodes of $\mathbb{t}$
  - $(Z_t^i; 1 \leq i \leq N_t)$ denote the particle masses at time t

uniqueness of weak solution to $(\star)$ and $(\star\star)$

- Let $\mathbb{t}$ is a pure-birth tree with birth rate $\tilde{a}(t) = a(T-t)$ starting at time 0 and stopped at time T.
- Let Q_x^T be the law of a branching particle system
  - starting with one particle with mass x at time 0
  - each particle follows an edge of $\mathbb{t}$ and splits into two copies of itself at the next branching time
  - particle masses evolve like independent copies of Z between nodes of $\mathbb{t}$
  - $(Z_t^i; 1 \leq i \leq N_t)$ denote the particle masses at time t

Theorem

Case $(\star)$ *: The only solution to (L) with* $u(0,\lambda) = \int_{[0,\infty)} e^{-\lambda x}\nu(dx)$ *is given by*

$$u(T,\lambda) = Q_\lambda^T\left(\prod_{i=1}^{N_T} e^{-W_i Z_T^i}\right),$$

where the (W_i) *are iid with law* ν*.*

uniqueness of weak solution to $(\star)$ and $(\star\star)$

- Let $\mathbb{t}$ is a pure-birth tree with birth rate $\tilde{a}(t) = a(T-t)$ starting at time 0 and stopped at time T.
- Let Q_x^T be the law of a branching particle system
  - starting with one particle with mass x at time 0
  - each particle follows an edge of $\mathbb{t}$ and splits into two copies of itself at the next branching time
  - particle masses evolve like independent copies of Z between nodes of $\mathbb{t}$
  - $(Z_t^i; 1 \le i \le N_t)$ denote the particle masses at time t

Theorem

Case $(\star)$ *: The only solution to (L) with* $u(0,\lambda) = \int_{[0,\infty)} e^{-\lambda x}\nu(dx)$ *is given by*

$$u(T,\lambda) = Q_\lambda^T\left(\prod_{i=1}^{N_T} e^{-W_i Z_T^i}\right),$$

where the (W_i) *are iid with law* ν*.*
Case $(\star\star)$ *: If now* $\int_0^\epsilon a(t)\,dt = \infty$*, under Grey's condition, any proper sol to* (L) *equals*

$$u(T,\lambda) = Q_\lambda^T(\mathscr{M}),$$

where $\mathscr{M} := \{$*all masses become extinct before time* $T\}$*.*

uniqueness of weak solution to $(\star)$ and $(\star\star)$: disintegration formula

- Using the branching property of Z, we can invert the Laplace transform

$$u(T,\lambda) = Q_\lambda^T \left(\exp\left(-\sum_{i=1}^{N_T} W_i Z_T^i \right) \right)$$

- In the case $(\star)$, μ_T is the law (wrt pure-birth tree $\mathbb{t}$ and the W_i's) of

$$F(\mathbb{t}, W_i) \;=\; M^{\mathbb{t}} \left(1 - \exp\left(-\sum_{i=1}^{N_T} W_i Z_T^i \right) \right),$$

  where $M^{\mathbb{t}}$ is the 0-entrance law of the branching CSBP.

- In the case $(\star\star)$, μ_T is the law (wrt pure-birth tree $\mathbb{t}$) of

$$F(\mathbb{t}) \;=\; M^{\mathbb{t}} \left(\mathscr{M}^c \right),$$

  where $\mathscr{M} := \{\text{all masses become extinct before time } T\}$.

conclusion

- We exposed a relation between the empirical measure of the nested coalescent at small times and a coagulation-transport equation.
- We constructed a finite pop ($\delta > 0$) McKean–Vlasov solution using the Brownian coalescent point process.
- Using a disintegration formula, we proved that there is at most one ∞-pop ($\delta = 0$) proper solution under Grey's condition, given by the law of

$$F(\mathbb{t}) \;=\; M^{\mathbb{t}}\left(\mathscr{M}^{c}\right).$$

open problems

- Convergence of nested Λ coalescents (not only nested Kingman).
- Uniqueness of the entrance law at ∞.
- What if ψ is not the Laplace exponent of a Lévy process? e.g., $\psi(x) = x^\gamma$ with $\gamma > 2$?
- Characterization of and convergence to dust solutions.