

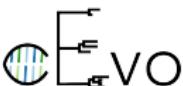
# Computational Biology

Lecturers:  
Tanja Stadler, Tim Vaughan & Carsten Magnus

Teaching Assistants:  
Antoine Zwaans, Adrian Lison,  
James Munday & Marcus Overwater

Computational Evolution  
Department of Biosystems Science and Engineering

HS 2023



- Structured Populations
  - Population structure
  - Spatial structure
  - Non-spatial structure
  - Structure in phylodynamic analyses
  - Structured birth-death models
  - Structured coalescent models
  - Structural dynamics
  - Species trees
  - Species networks
- Phylogenetic Networks
  - Introduction to networks
  - Linkage and phylogeny
  - Wright-Fisher with Recombination
  - Coalescent with Recombination
  - Bayesian inference
  - Population dynamics inference
- References

# Bayesian Inference: Questions

- ?
- Does a Bayesian phylogenetic analysis of the kind described here allow one to directly infer ancestral sequences?  
Why/Why not?

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian Inference: Questions

- not
- ② Does a Bayesian phylogenetic analysis of the kind described here allow one to directly infer ancestral sequences?  
Why/Why not?
  - ② How might we test to see whether a Bayesian MCMC analysis has explored the full state space supported by the posterior?

Run multiple independent checks

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian Inference: Questions

- ② Does a Bayesian phylogenetic analysis of the kind described here allow one to directly infer ancestral sequences?  
Why/Why not?
- ② How might we test to see whether a Bayesian MCMC analysis has explored the full state space supported by the posterior?
- ② Suppose you have conducted a Bayesian phylodynamic analysis and recovered a 95% HPD interval for the birth rate parameter. If you take this result and use it to construct a new prior for this parameter and use this prior to analyze the *same data*, would the resulting second posterior be valid?  
**No, don't use same data twice, like double dating**

## Structured Populations

Population structure

Spatial structure

Non-spatial structure

Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

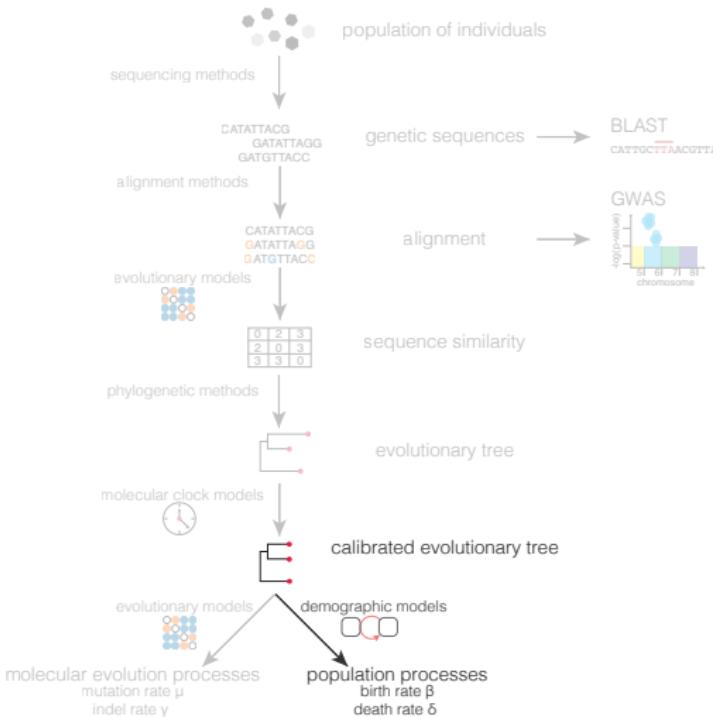
Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Overview



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

## Structured Populations

Population structure

Spatial structure

Non-spatial structure

Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Structured Populations

## Section 1

# Structured Populations

Biological populations often have some internal “structure”:

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

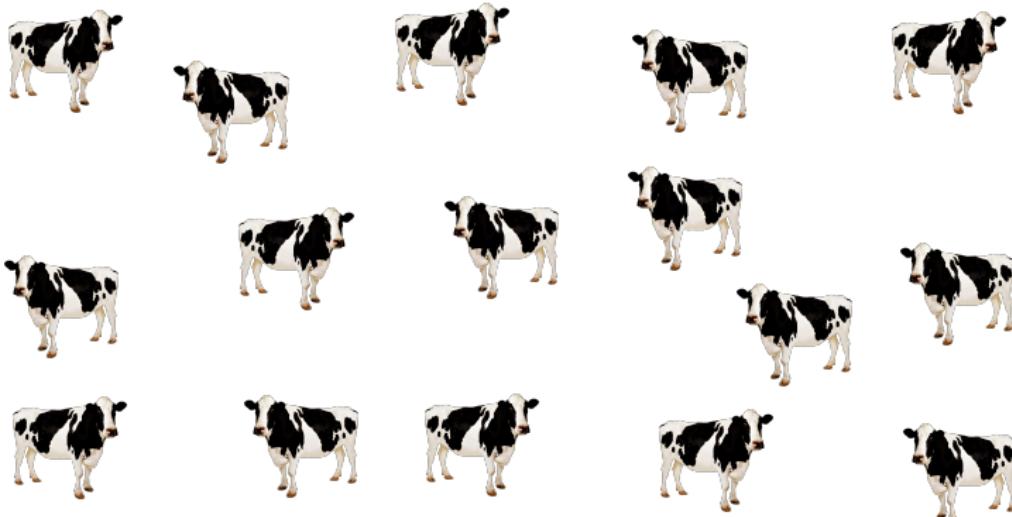
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Populations

Biological populations often have some internal “structure”:



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

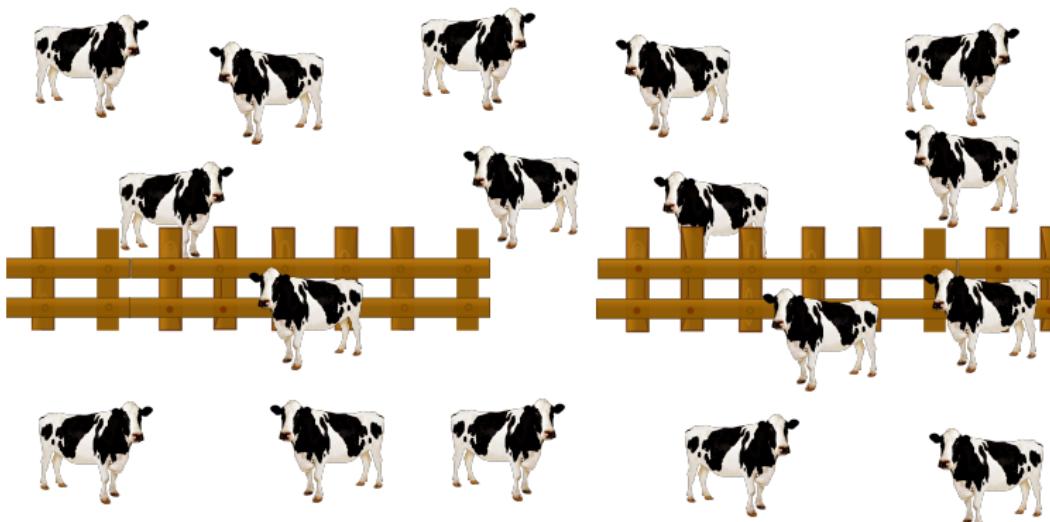
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Populations

Biological populations often have some internal “structure”:



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Populations

What do we really mean by a “structured population”?

## Our Definition

A population is structured if its members possess one or more traits (e.g. location, group membership, ...) that affects their phylodynamic parameters (e.g. birth rate, death rate, sampling rate, coalescence rate).

### Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

### Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

### References

# Geographic/Spatial structuring



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Geographic/Spatial structuring

- ▶ Gene flow limited by spatial/geographic segregation of subpopulations.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Geographic/Spatial structuring

- ▶ Gene flow limited by spatial/geographic segregation of subpopulations.
- ▶ Impact of structure depends on the rate of migration across boundaries relative to the local birth-rate.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Geographic/Spatial structuring

- ▶ Gene flow limited by spatial/geographic segregation of subpopulations.
- ▶ Impact of structure depends on the rate of migration across boundaries relative to the local birth-rate.
- ▶ A population spanning an archipelago is a classic example of a spatially structured population.



## Structured Populations

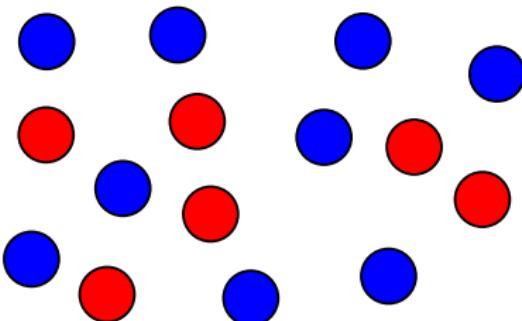
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Non-spatial structuring



Even populations that are spatially mixed can be structured in other ways:

## Structured Populations

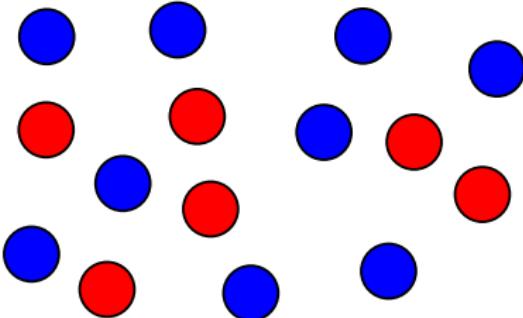
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Non-spatial structuring



Even populations that are spatially mixed can be structured in other ways:

- ▶ pathogen populations are generally composed of many within-host sub-populations,

## Structured Populations

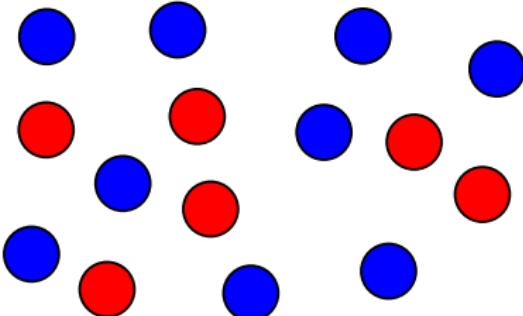
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Non-spatial structuring



Even populations that are spatially mixed can be structured in other ways:

- ▶ pathogen populations are generally composed of many within-host sub-populations,
- ▶ pathogen sub-populations may possess traits (e.g. drug resistance/susceptibility) affecting reproductive success,

## Structured Populations

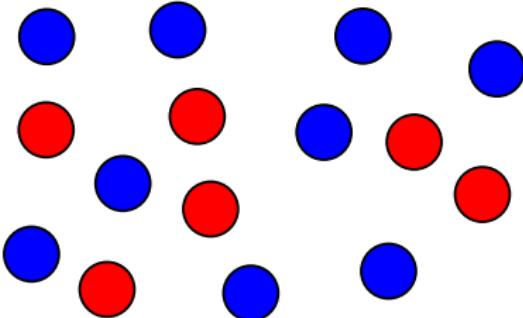
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Non-spatial structuring



Even populations that are spatially mixed can be structured in other ways:

- ▶ pathogen populations are generally composed of many within-host sub-populations,
- ▶ pathogen sub-populations may possess traits (e.g. drug resistance/susceptibility) affecting reproductive success,
- ▶ infected individuals may be in different epidemiological states (e.g. exposed vs infectious),

## Structured Populations

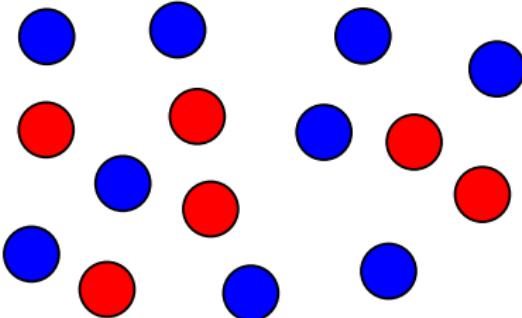
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Non-spatial structuring



Even populations that are spatially mixed can be structured in other ways:

- ▶ pathogen populations are generally composed of many within-host sub-populations,
- ▶ pathogen sub-populations may possess traits (e.g. drug resistance/susceptibility) affecting reproductive success,
- ▶ infected individuals may be in different epidemiological states (e.g. exposed vs infectious),
- ▶ sampled animals may be members of different species, between which there may be (extremely rare) horizontal gene transfer.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

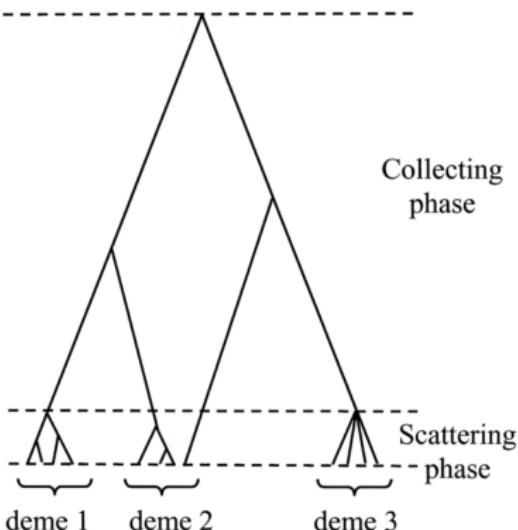
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Importance to phylodynamic analyses

- ▶ Population structure can play an important role in shaping the phylogenetic relationships between samples.

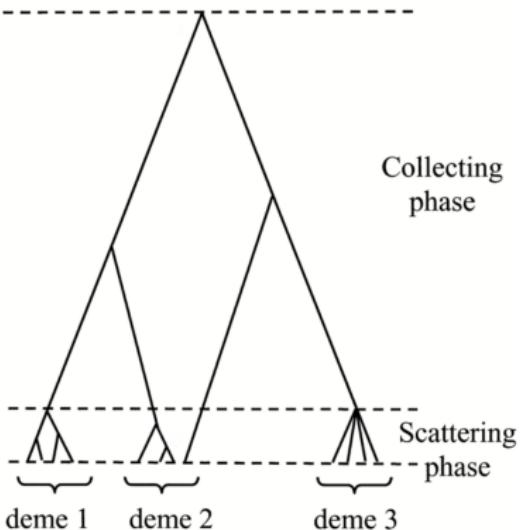


Deme just a lingo for population sample  
Pannell, 2003]

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# Importance to phylodynamic analyses

- ▶ Population structure can play an important role in shaping the phylogenetic relationships between samples.
- ▶ Failing to account for existing structure in phylodynamic analyses can bias results.



[Pannell, 2003]

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

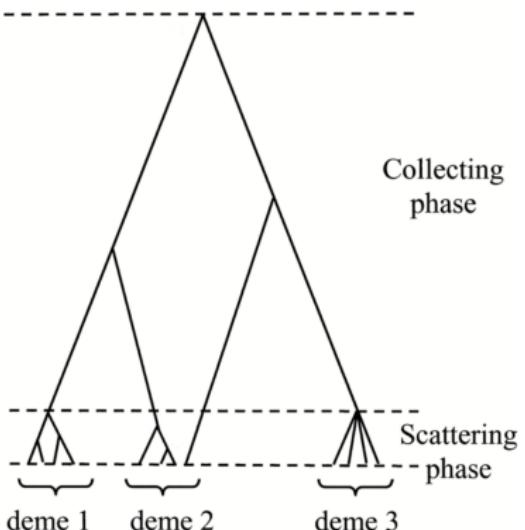
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Importance to phylodynamic analyses

- ▶ Population structure can play an important role in shaping the phylogenetic relationships between samples.
- ▶ Failing to account for existing structure in phylodynamic analyses can bias results.
- ▶ We can also learn about parameters of structured models (e.g. local birth/death rates and sub-population sizes) using structure-aware phylodynamic models.

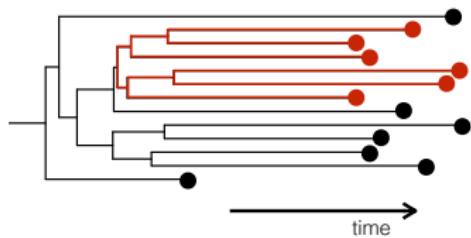


[Pannell, 2003]

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

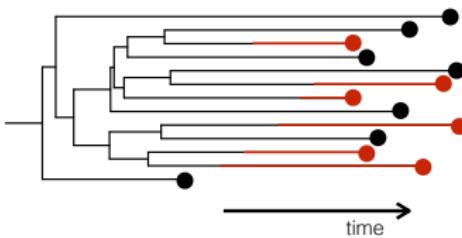
# Phylogenetic trees contain information about population structure

Transmitted drug resistance:



- drug resistant
- drug sensitive

*De novo* drug resistance:



- ▶ Each tip corresponds to one patient. The tree corresponds to the transmission chain.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

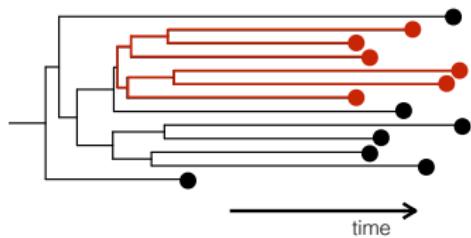
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

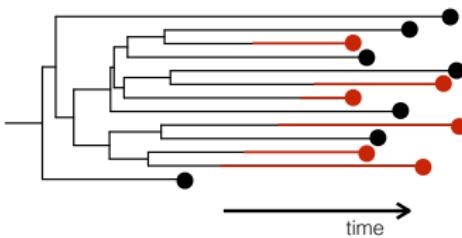
# Phylogenetic trees contain information about population structure

Transmitted drug resistance:



- drug resistant
- drug sensitive

*De novo* drug resistance:



- ▶ Each tip corresponds to one patient. The tree corresponds to the transmission chain.
- ▶ Counting number of sensitive and number of resistant samples through time does not provide insight into which scenario happened (transmitted or *de novo* drug resistance).

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

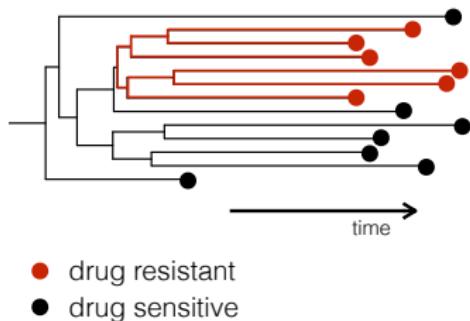
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

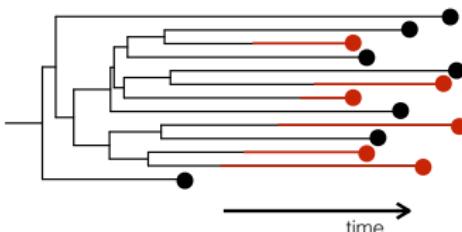
## References

# Phylogenetic trees contain information about population structure

Transmitted drug resistance:



*De novo* drug resistance:



- ▶ Each tip corresponds to one patient. The tree corresponds to the transmission chain.
- ▶ Counting number of sensitive and number of resistant samples through time does not provide insight into which scenario happened (transmitted or *de novo* drug resistance).
- ▶ However, the phylogeny with tip labels contains information about the scenarios...

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

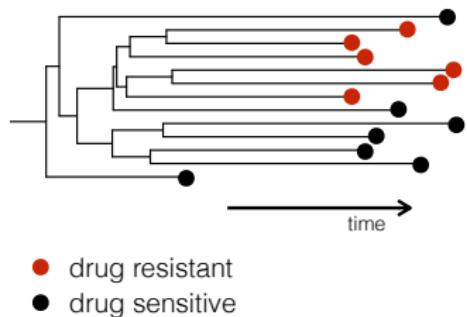
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

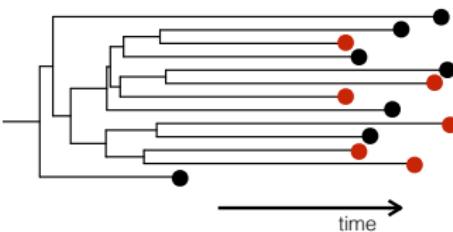
## References

# Phylogenetic trees contain information about population structure

Transmitted drug resistance:



*De novo* drug resistance:



(Note that this information is hidden from traditional epidemiological data such as hospital case records.)

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

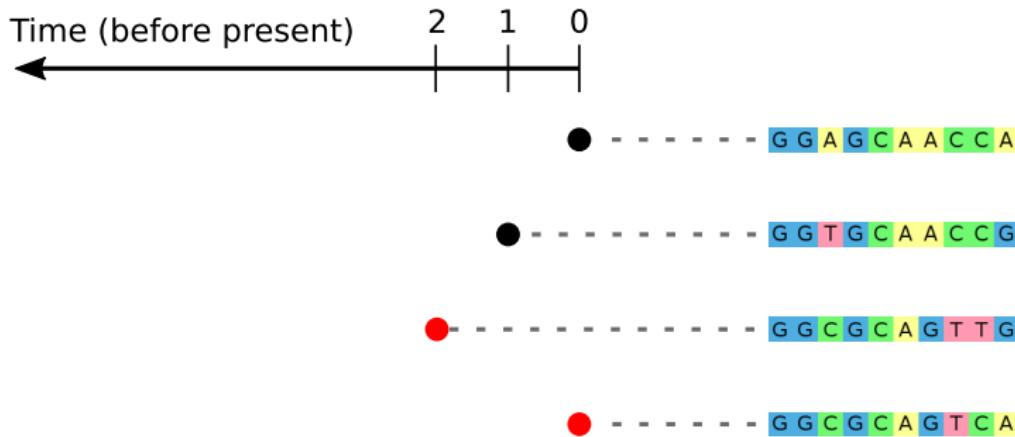
Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Basic structured pop. inference problem



## Structured Populations

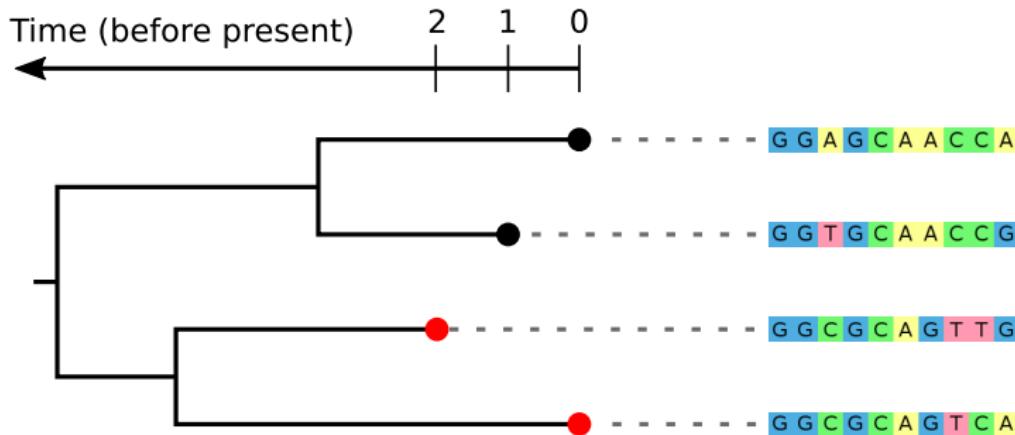
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Basic structured pop. inference problem



## Structured Populations

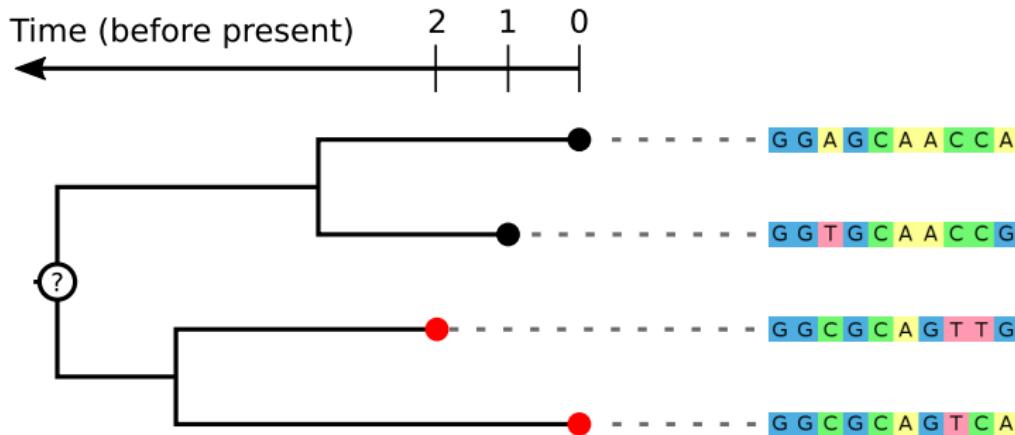
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Basic structured pop. inference problem



## Structured Populations

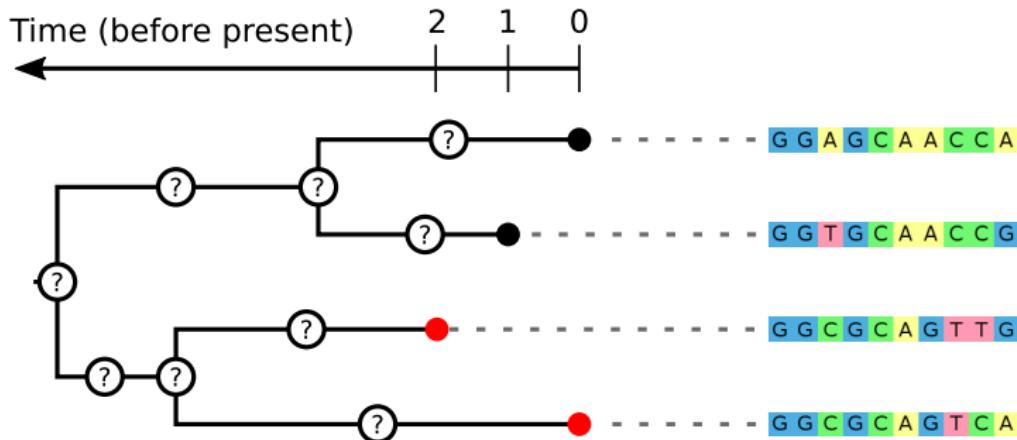
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Basic structured pop. inference problem



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

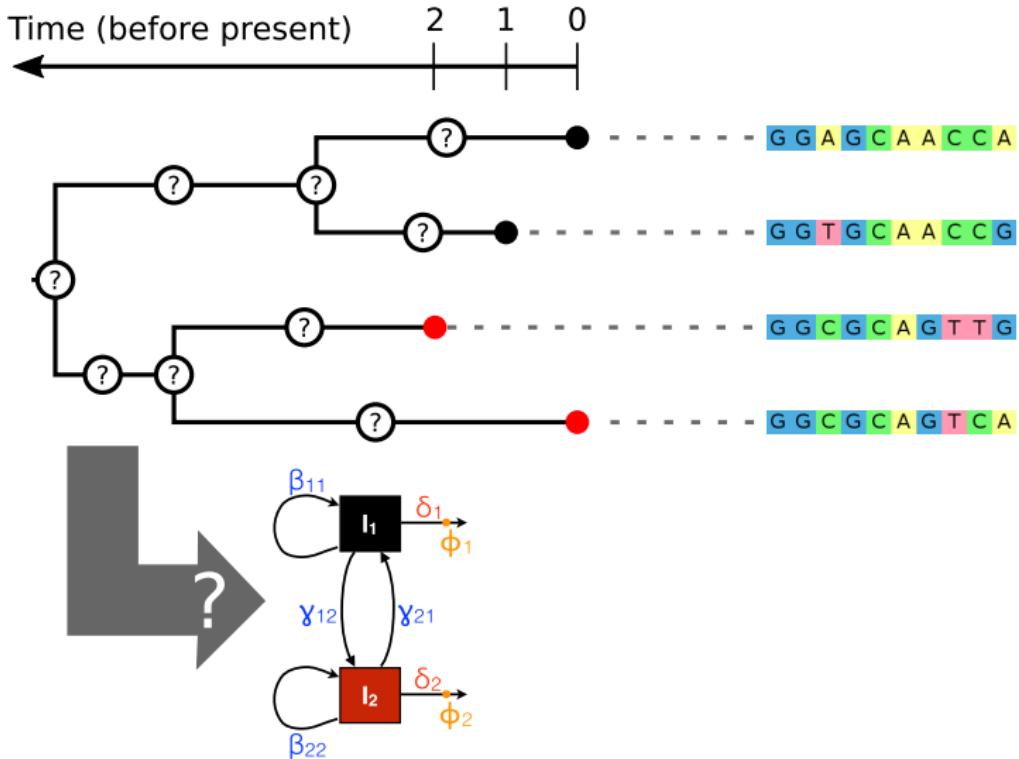
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

## Basic structured pop. inference problem

CB



## Structured Populations

- Population structure
  - Spatial structure
  - Non-spatial structure
  - Structure in phylogenetic analyses**
  - Structured birth-death models
  - Structured coalescent models
  - Structural dynamics
  - Species trees
  - Species networks

## Phylogenetic Networks

- Introduction to networks
  - Linkage and phylogeny
  - Wright-Fisher with Recombination
  - Coalescent with Recombination
  - Bayesian inference
  - Population dynamics inference

## References

# Structured birth-death models

Several models augment the existing birth-death models to include individuals of different types, including:

## Binary State Speciation and Extinction (BiSSE)

Allows for 2 states affecting birth and death rates,  
assumes complete present-day sampling.

[Maddison et al., 2007]

## Multi-State Speciation and Extinction (MuSSE)

Allows for > 2 states, allows for incomplete  
present-day sampling. [FitzJohn et al., 2009]

## Multi-type Birth Death (MTBD)

Allows for > 2 states, allows for both incomplete  
present-day and through-time sampling.  
[Kühnert et al., 2016, Scire et al., ]

### Structured Populations

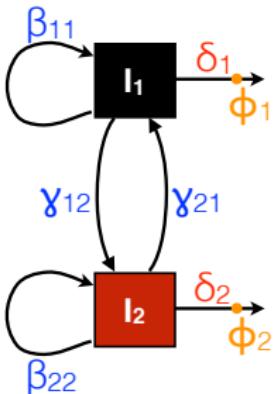
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

### Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

### References

# Simple structured birth-death model



- In this case,  $\beta_{11}$  is transmission (birth) rate of drug sensitive strains,  $\beta_{22}$  is transmission (birth) rate of drug resistant strains,  $\gamma_{12}$  is the rate of resistance evolution.

## Structured Populations

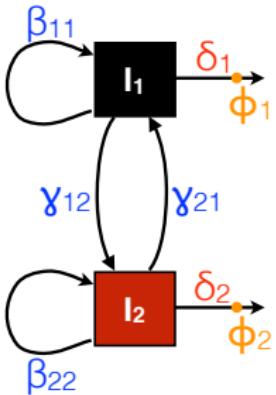
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Simple structured birth-death model



- In this case,  $\beta_{11}$  is transmission (birth) rate of drug sensitive strains,  $\beta_{22}$  is transmission (birth) rate of drug resistant strains,  $\gamma_{12}$  is the rate of resistance evolution.
- In general:
  - Different compartments may represent pathogen strains, geographic locations, host risk groups, etc.

## Structured Populations

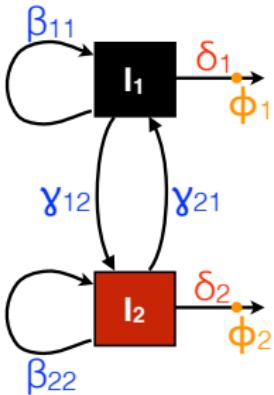
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Simple structured birth-death model



- In this case,  $\beta_{11}$  is transmission (birth) rate of drug sensitive strains,  $\beta_{22}$  is transmission (birth) rate of drug resistant strains,  $\gamma_{12}$  is the rate of resistance evolution.
- In general:
  - Different compartments may represent pathogen strains, geographic locations, host risk groups, etc.
  - The  $\beta_{ij}$  represent the rate at which individuals of type  $i$  **produce** individuals of type  $j$ , while  $\gamma_{ij}$  is the rate at which individuals of type  $i$  **become** type  $j$ .

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Multi-type birth-death phylodynamic likelihood

Very similar to the unstructured case:

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Multi-type birth-death phylodynamic likelihood

Very similar to the unstructured case:

- ▶ Let  $p_i(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past, gives rise to *no* sampled descendants.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Multi-type birth-death phylodynamic likelihood

Very similar to the unstructured case:

- ▶ Let  $p_i(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past, gives rise to *no* sampled descendants.
- ▶ Satisfies the following ODE:

$$\begin{aligned} \frac{d}{dt} p_i(t) = & - \left( \sum_{j=1}^d (\beta_{ij} + \gamma_{ij}) + \delta_i \right) p_i(t) \\ & + \sum_{j=1}^d \beta_{ij} p_i(t) p_j(t) + \sum_{j=1}^d \gamma_{ij} p_j(t) + \delta_i \end{aligned}$$

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Multi-type birth-death phylodynamic likelihood

Very similar to the unstructured case:

- ▶ Let  $p_i(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past, gives rise to *no* sampled descendants.
- ▶ Satisfies the following ODE:

$$\begin{aligned} \frac{d}{dt} p_i(t) = & - \left( \sum_{j=1}^d (\beta_{ij} + \gamma_{ij}) + \delta_i \right) p_i(t) \\ & + \sum_{j=1}^d \beta_{ij} p_i(t) p_j(t) + \sum_{j=1}^d \gamma_{ij} p_j(t) + \delta_i \end{aligned}$$

- ▶ Let  $g_i^e(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past and belonging to sampled tree edge  $e$ , gives rise to the sampled phylogeny below that edge. (This obeys a similar ODE.)

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Multi-type birth-death phylodynamic likelihood

Very similar to the unstructured case:

- ▶ Let  $p_i(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past, gives rise to *no* sampled descendants.
- ▶ Satisfies the following ODE:

$$\begin{aligned} \frac{d}{dt} p_i(t) = & - \left( \sum_{j=1}^d (\beta_{ij} + \gamma_{ij}) + \delta_i \right) p_i(t) \\ & + \sum_{j=1}^d \beta_{ij} p_i(t) p_j(t) + \sum_{j=1}^d \gamma_{ij} p_j(t) + \delta_i \end{aligned}$$

- ▶ Let  $g_i^e(t)$  be the probability that an individual of type  $i$ , alive at time  $t$  in the past and belonging to sampled tree edge  $e$ , gives rise to the sampled phylogeny below that edge. (This obeys a similar ODE.)

Unlike unstructured case, no known analytical solutions to these equations exist: must be solved numerically to compute structured phylodynamic likelihood.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for multi-type birth-death models

The phylogenetic posterior for a multi-type birth-death analysis becomes

$$P(\mathcal{T}, Q, \eta | A, L) = \frac{P(A|Q, \mathcal{T})P(\mathcal{T}, L|\eta)P(Q)P(\eta)}{P(A, L)}$$

- ▶  $\mathcal{T}$  is a phylogenetic tree.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\eta$  are the parameters of the multi-type birth-death model.
- ▶  $P(\mathcal{T}, L|\eta)$  is the structured phylodynamic likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for multi-type birth-death models

The phylogenetic posterior for a multi-type birth-death analysis becomes

$$P(\mathcal{T}, Q, \eta | A, L) = \frac{P(A|Q, \mathcal{T})P(\mathcal{T}, L|\eta)P(Q)P(\eta)}{P(A, L)}$$

- ▶  $\mathcal{T}$  is a phylogenetic tree.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\eta$  are the parameters of the multi-type birth-death model.
- ▶  $P(\mathcal{T}, L|\eta)$  is the structured phylodynamic likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶ Likelihood computed using equations on previous page integrates out (i.e. averages over) ancestral states.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for multi-type birth-death models

The phylogenetic posterior for a multi-type birth-death analysis becomes

$$P(\mathcal{T}, Q, \eta | A, L) = \frac{P(A|Q, \mathcal{T})P(\mathcal{T}, L|\eta)P(Q)P(\eta)}{P(A, L)}$$

- ▶  $\mathcal{T}$  is a phylogenetic tree.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\eta$  are the parameters of the multi-type birth-death model.
- ▶  $P(\mathcal{T}, L|\eta)$  is the structured phylodynamic likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶ Likelihood computed using equations on previous page integrates out (i.e. averages over) ancestral states.
- ▶ Can also derive probability of “coloured” tree with ancestral states marked. In that case, MCMC must be performed on this expanded state space of coloured trees.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for multi-type birth-death models

The phylogenetic posterior for a multi-type birth-death analysis becomes

$$P(\mathcal{T}, Q, \eta | A, L) = \frac{P(A|Q, \mathcal{T})P(\mathcal{T}, L|\eta)P(Q)P(\eta)}{P(A, L)}$$

- ▶  $\mathcal{T}$  is a phylogenetic tree.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\eta$  are the parameters of the multi-type birth-death model.
- ▶  $P(\mathcal{T}, L|\eta)$  is the structured phylodynamic likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶ Likelihood computed using equations on previous page integrates out (i.e. averages over) ancestral states.
- ▶ Can also derive probability of “coloured” tree with ancestral states marked. In that case, MCMC must be performed on this expanded state space of coloured trees.
  - + Posterior distribution for ancestral types directly available.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for multi-type birth-death models

The phylogenetic posterior for a multi-type birth-death analysis becomes

$$P(\mathcal{T}, Q, \eta | A, L) = \frac{P(A|Q, \mathcal{T})P(\mathcal{T}, L|\eta)P(Q)P(\eta)}{P(A, L)}$$

- ▶  $\mathcal{T}$  is a phylogenetic tree.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\eta$  are the parameters of the multi-type birth-death model.
- ▶  $P(\mathcal{T}, L|\eta)$  is the structured phylodynamic likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶ Likelihood computed using equations on previous page integrates out (i.e. averages over) ancestral states.
- ▶ Can also derive probability of “coloured” tree with ancestral states marked. In that case, MCMC must be performed on this expanded state space of coloured trees.
  - + Posterior distribution for ancestral types directly available.
  - Analyses much more computationally demanding.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Example: Seasonal influenza (H3N2)

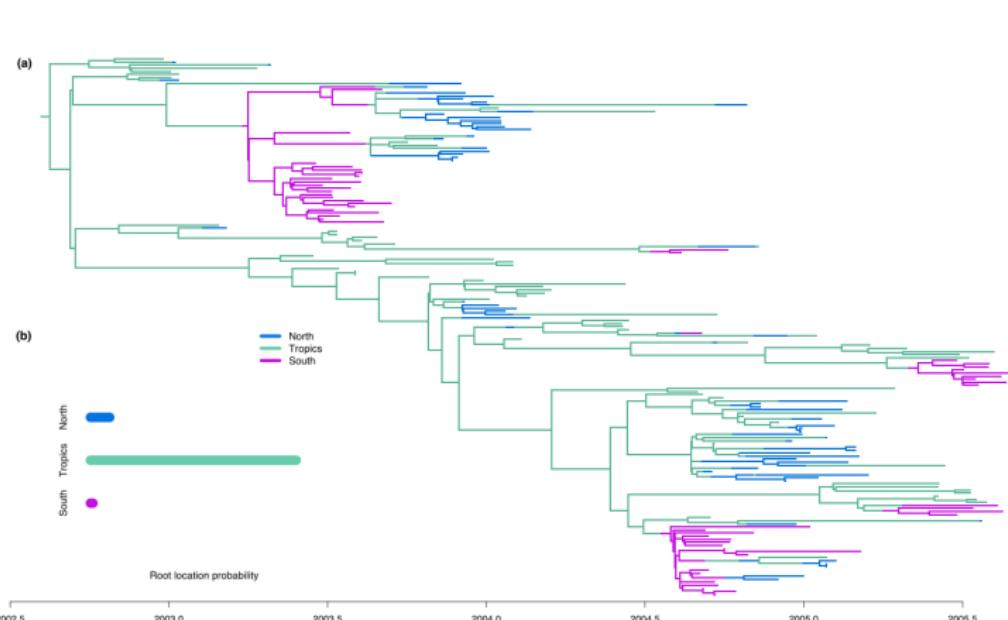


Figure adapted from [Kühnert et al., 2016]

Phylogenetic tree of geographic spread of seasonal influenza inferred under a multi-type birth-death model.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Example: Seasonal influenza (H3N2)

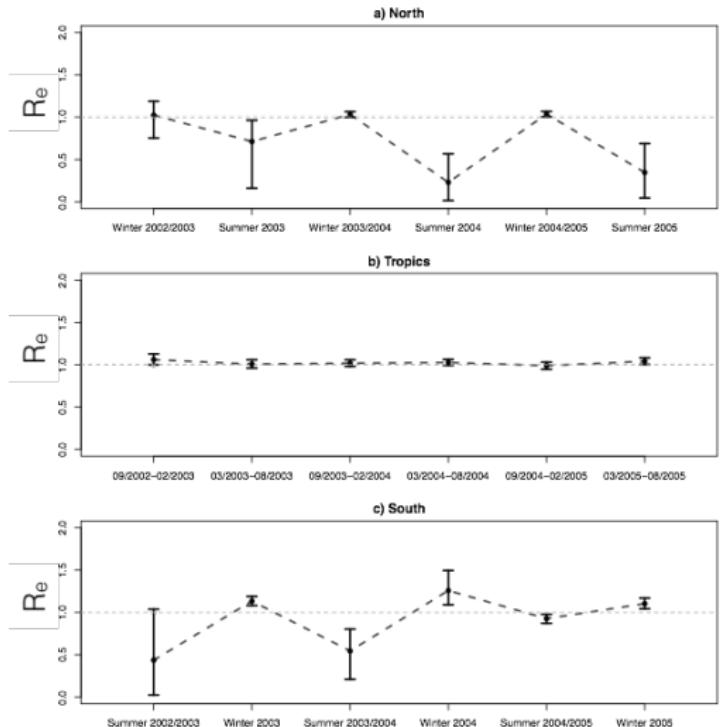


Figure adapted from [Kühnert et al., 2016]

Effective reproductive number of seasonal influenza inferred using the same multi-type birth-death model.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Unknown Leaf Types

- Newer methods are available that address the problem of accounting for unknown leaf types/locations.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Unknown Leaf Types

- ▶ Newer methods are available that address the problem of accounting for unknown leaf types/locations.
- ▶ This problem is particularly challenging when the **number** of types/locations is also unknown.

## Structured Populations

Population structure  
Spatial structure  
Non-spatial structure  
Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks  
Linkage and phylogeny  
Wright-Fisher with Recombination  
Coalescent with Recombination  
Bayesian inference  
Population dynamics inference

## References

# Unknown Leaf Types

- ▶ Newer methods are available that address the problem of accounting for unknown leaf types/locations.
- ▶ This problem is particularly challenging when the **number** of types/locations is also unknown.

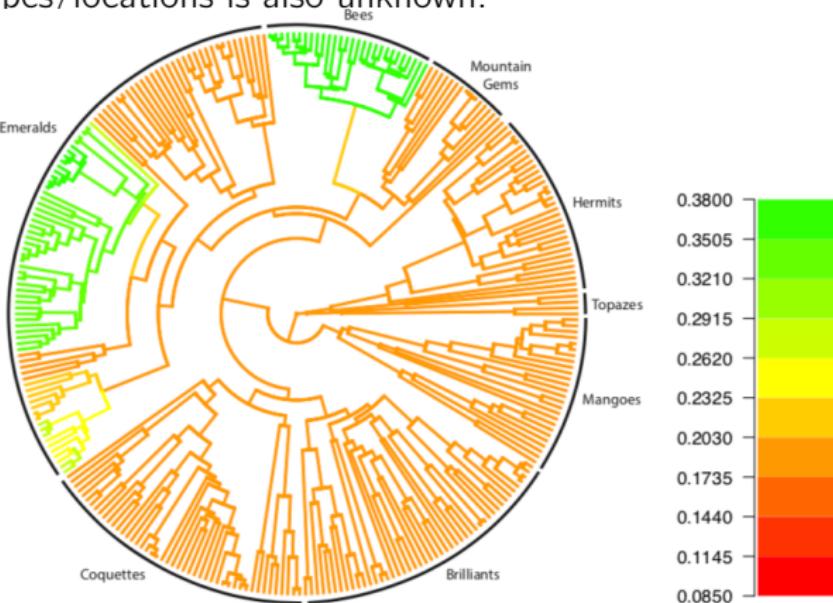


Figure adapted from [Barido-Sottani et al., 2018]  
Speciation rate classes across phylogeny of hummingbird species.

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
<b>Structured birth-death models</b>
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# Structured coalescent models

- ▶ Just as for the birth-death models, we can augment coalescent models to include multiple types.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured coalescent models

- ▶ Just as for the birth-death models, we can augment coalescent models to include multiple types.
- ▶ Instead of type-specific birth, death and sampling rates however, this yields models with type-specific population sizes.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured coalescent models

- ▶ Just as for the birth-death models, we can augment coalescent models to include multiple types.
- ▶ Instead of type-specific birth, death and sampling rates however, this yields models with type-specific population sizes.
- ▶ In a similar way to the unstructured coalescent process, the structured coalescent process can be derived from a variety of distinct forward-time models of population dynamics.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured coalescent models

- ▶ Just as for the birth-death models, we can augment coalescent models to include multiple types.
- ▶ Instead of type-specific birth, death and sampling rates however, this yields models with type-specific population sizes.
- ▶ In a similar way to the unstructured coalescent process, the structured coalescent process can be derived from a variety of distinct forward-time models of population dynamics.
- ▶ We focus here on an extension to the Wright-Fisher model.

## Structured Populations

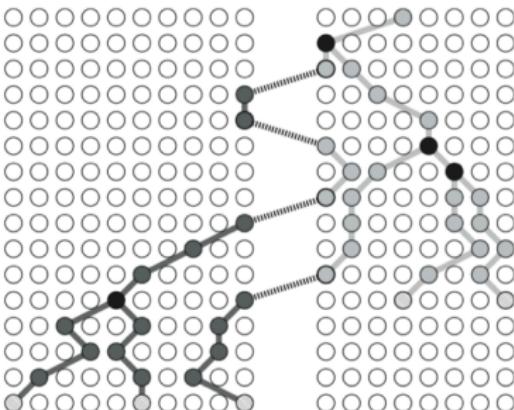
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Wright-Fisher model



- ▶ Introduced by [Notohara, 1990] (and probably others).

## Structured Populations

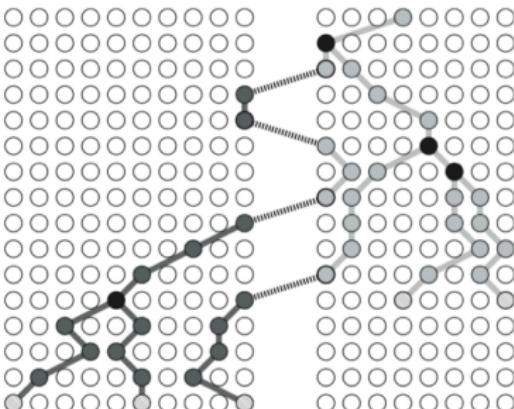
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Wright-Fisher model



- ▶ Introduced by [Notohara, 1990] (and probably others).
- ▶ Assumes a single population is divided into sub-populations (demes) of size  $N_i$  for  $i \in [1, d]$ .

## Structured Populations

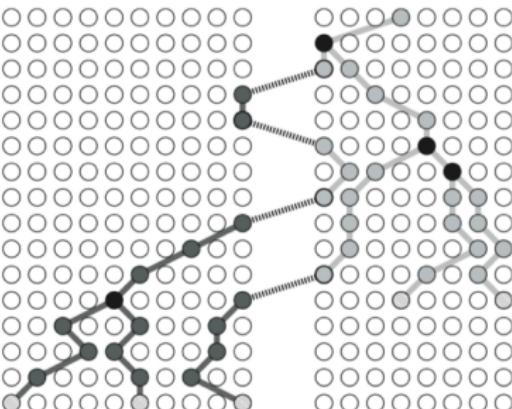
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Wright-Fisher model



- ▶ Introduced by [Notohara, 1990] (and probably others).
- ▶ Assumes a single population is divided into sub-populations (demes) of size  $N_i$  for  $i \in [1, d]$ .
- ▶ Allows for migration between demes at rate  $q_{ij}$ .

## Structured Populations

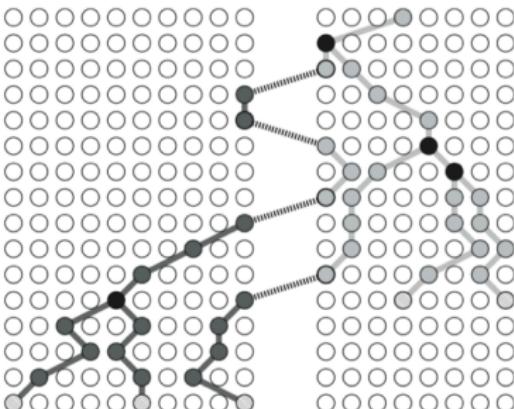
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Wright-Fisher model



- ▶ Introduced by [Notohara, 1990] (and probably others).
- ▶ Assumes a single population is divided into sub-populations (demes) of size  $N_i$  for  $i \in [1, d]$ .
- ▶ Allows for migration between demes at rate  $q_{ij}$ .
- ▶ As in unstructured case, assume a fixed time interval  $g$  between successive generations.

## Structured Populations

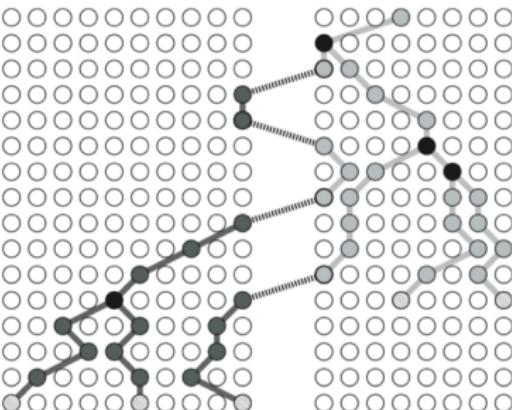
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Structured Wright-Fisher model



- ▶ Introduced by [Notohara, 1990] (and probably others).
- ▶ Assumes a single population is divided into sub-populations (demes) of size  $N_i$  for  $i \in [1, d]$ .
- ▶ Allows for migration between demes at rate  $q_{ij}$ .
- ▶ As in unstructured case, assume a fixed time interval  $g$  between successive generations.
- ▶ Assumes that sub-population sizes are unaffected by migration in the long term.

## Structured Populations

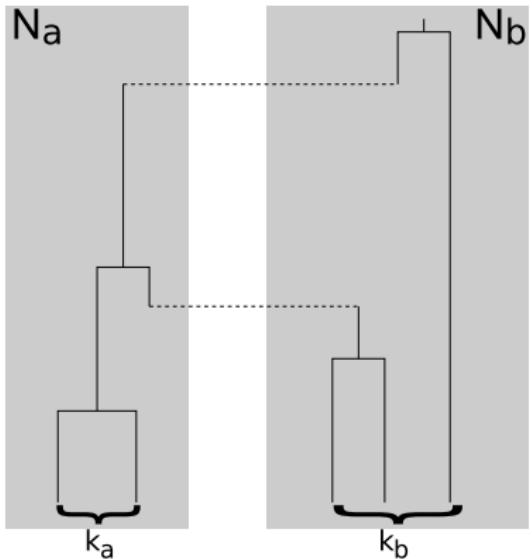
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The structured coalescent



This is a backward in time tree generation process corresponding to the coalescent limit of the structured WF model.

Coalescence rate in deme i:

$$\binom{k_i}{2} \frac{1}{gN_i}$$

Migration rate (backward)  $i \rightarrow j$ :

$$k_i m_{ij}$$

## Structured Populations

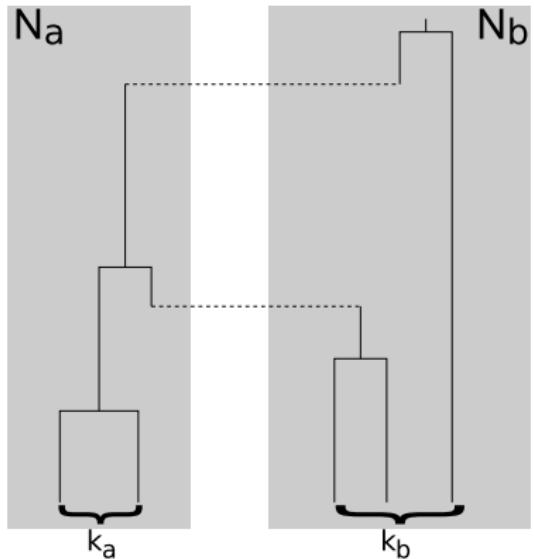
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The structured coalescent



This is a backward in time tree generation process corresponding to the coalescent limit of the structured WF model.

Coalescence rate in deme i:

$$\binom{k_i}{2} \frac{1}{gN_i}$$

Migration rate (backward) i → j:

$$k_i m_{ij}$$

- The backward-time migration rate  $m_{ij}$  (also called the immigration rate) is related to the forward time rate  $q_{ji}$  from the structured WF model by

$$m_{ij} = q_{ji} \frac{N_j}{N_i}$$

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

- ▶ Define  $T_d$  and  $T_s$  as the expected coalescence times when lineages are in different/same demes respectively.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

- ▶ Define  $T_d$  and  $T_s$  as the expected coalescence times when lineages are in different/same demes respectively.
- ▶ Lineages in distinct demes cannot coalesce, so we have:

$$T_d = \frac{1}{2m} + T_s$$

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

- ▶ Define  $T_d$  and  $T_s$  as the expected coalescence times when lineages are in different/same demes respectively.
- ▶ Lineages in distinct demes cannot coalesce, so we have:

$$T_d = \frac{1}{2m} + T_s$$

- ▶ Lineages in the same deme wait for average time  $1/(2m + 1/Ng)$  before either coalescing or migrating:

$$T_s = \frac{1}{2m + 1/Ng} + \frac{1/Ng}{2m + 1/Ng} 0 + \frac{2m}{2m + 1/Ng} T_d$$

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

- ▶ Define  $T_d$  and  $T_s$  as the expected coalescence times when lineages are in different/same demes respectively.
- ▶ Lineages in distinct demes cannot coalesce, so we have:

$$T_d = \frac{1}{2m} + T_s$$

- ▶ Lineages in the same deme wait for average time  $1/(2m + 1/Ng)$  before either coalescing or migrating:

$$T_s = \frac{1}{2m + 1/Ng} + \frac{1/Ng}{2m + 1/Ng} 0 + \frac{2m}{2m + 1/Ng} T_d$$

- ▶ Solving this pair of simultaneous equations yields:  
 $T_s = 2Ng$  and  $T_d = \frac{1}{2m} + 2Ng$ .

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Expected coalescence time

For a symmetric 2 deme model ( $N_1 = N_2 = N$  and  $m_{12} = m_{21} = m$ ) we can derive the *expected* time for two lineages to coalesce:

- ▶ Define  $T_d$  and  $T_s$  as the expected coalescence times when lineages are in different/same demes respectively.
- ▶ Lineages in distinct demes cannot coalesce, so we have:

$$T_d = \frac{1}{2m} + T_s$$

- ▶ Lineages in the same deme wait for average time  $1/(2m + 1/Ng)$  before either coalescing or migrating:

$$T_s = \frac{1}{2m + 1/Ng} + \frac{1/Ng}{2m + 1/Ng} 0 + \frac{2m}{2m + 1/Ng} T_d$$

- ▶ Solving this pair of simultaneous equations yields:

$$T_s = 2Ng \text{ and } T_d = \frac{1}{2m} + 2Ng.$$

Interesting: Expected time to coalesce from the same deme is independent of migration rate!

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for structured coalescent models

The phylogenetic posterior for a structured coalescent analysis becomes

$$P(\mathcal{T}_{\text{col}}, Q, \theta | A, L) = \frac{P(A|Q, \mathcal{T}_{\text{col}})P(\mathcal{T}_{\text{col}}|\theta, L)P(Q)P(\theta)}{P(A, L)}$$

- ▶  $\mathcal{T}_{\text{col}}$  is a phylogenetic tree with ancestral locations marked.
- ▶ L are the locations/types associated with the sequences.
- ▶  $\theta$  are the parameters of the structured coalescent model.
- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is the structured coalescent likelihood.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for structured coalescent models

The phylogenetic posterior for a structured coalescent analysis becomes

$$P(\mathcal{T}_{\text{col}}, Q, \theta | A, L) = \frac{P(A|Q, \mathcal{T}_{\text{col}})P(\mathcal{T}_{\text{col}}|\theta, L)P(Q)P(\theta)}{P(A, L)}$$

- ▶  $\mathcal{T}_{\text{col}}$  is a phylogenetic tree with ancestral locations marked.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\theta$  are the parameters of the structured coalescent model.
- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is the structured coalescent likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for structured coalescent models

The phylogenetic posterior for a structured coalescent analysis becomes

$$P(\mathcal{T}_{\text{col}}, Q, \theta | A, L) = \frac{P(A|Q, \mathcal{T}_{\text{col}})P(\mathcal{T}_{\text{col}}|\theta, L)P(Q)P(\theta)}{P(A, L)}$$

- ▶  $\mathcal{T}_{\text{col}}$  is a phylogenetic tree with ancestral locations marked.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\theta$  are the parameters of the structured coalescent model.
- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is the structured coalescent likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is a simple extension of the unstructured coalescent expression.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for structured coalescent models

The phylogenetic posterior for a structured coalescent analysis becomes

$$P(\mathcal{T}_{\text{col}}, Q, \theta | A, L) = \frac{P(A|Q, \mathcal{T}_{\text{col}})P(\mathcal{T}_{\text{col}}|\theta, L)P(Q)P(\theta)}{P(A, L)}$$

- ▶  $\mathcal{T}_{\text{col}}$  is a phylogenetic tree with ancestral locations marked.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\theta$  are the parameters of the structured coalescent model.
- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is the structured coalescent likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is a simple extension of the unstructured coalescent expression.
- ▶ Unlike the multi-type birth-death case, it is difficult to integrate over ancestral locations.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian inference for structured coalescent models

The phylogenetic posterior for a structured coalescent analysis becomes

$$P(\mathcal{T}_{\text{col}}, Q, \theta | A, L) = \frac{P(A|Q, \mathcal{T}_{\text{col}})P(\mathcal{T}_{\text{col}}|\theta, L)P(Q)P(\theta)}{P(A, L)}$$

- ▶  $\mathcal{T}_{\text{col}}$  is a phylogenetic tree with ancestral locations marked.
- ▶  $L$  are the locations/types associated with the sequences.
- ▶  $\theta$  are the parameters of the structured coalescent model.
- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is the structured coalescent likelihood.

As before,  $A$  represents the sequence alignment,  $Q$  the substitution rate matrix.

- ▶  $P(\mathcal{T}_{\text{col}}|\theta, L)$  is a simple extension of the unstructured coalescent expression.
- ▶ Unlike the multi-type birth-death case, it is difficult to integrate over ancestral locations.
- ▶ Has been done approximately  
[De Maio et al., 2015, Müller et al., 2017].

## Structured Populations

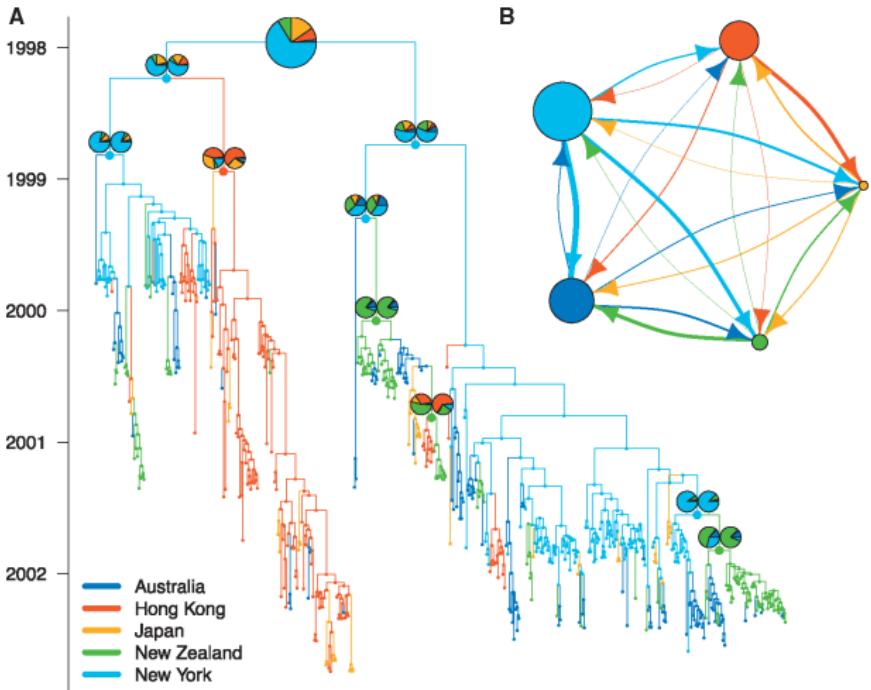
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Example: Seasonal influenza (H3N2)



Tree with ancestral locations inferred from influenza sequences sampled from five locations. Effective sub-population sizes and migration rates are also inferred. [Müller et al., 2018]

- Structured Populations**
  - Population structure
  - Spatial structure
  - Non-spatial structure
  - Structure in phylodynamic analyses
  - Structured birth-death models
  - Structured coalescent models
  - Structural dynamics
  - Species trees
  - Species networks
- Phylogenetic Networks**
  - Introduction to networks
  - Linkage and phylogeny
  - Wright-Fisher with Recombination
  - Coalescent with Recombination
  - Bayesian inference
  - Population dynamics inference
- References

# Populations with dynamic structure

Coalescent models have also been extended to account for **structure** which changes through time.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Populations with dynamic structure

Coalescent models have also been extended to account for **structure** which changes through time.

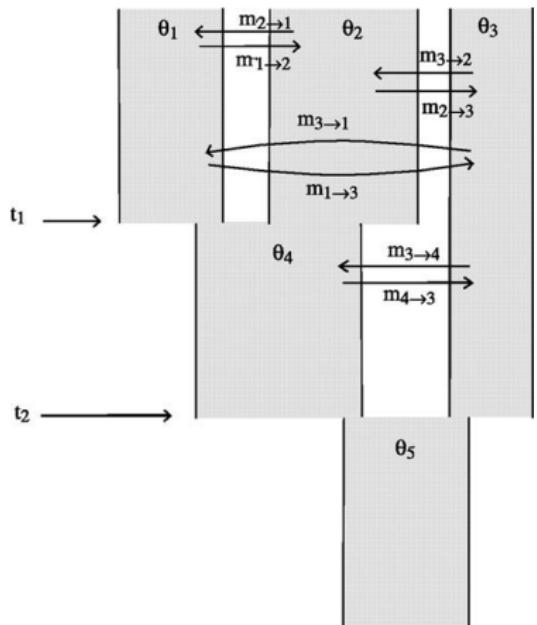


Figure adapted from [Hey, 2009]

Isolation-with-migration model.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Populations with dynamic structure

Coalescent models have also been extended to account for **structure** which changes through time.

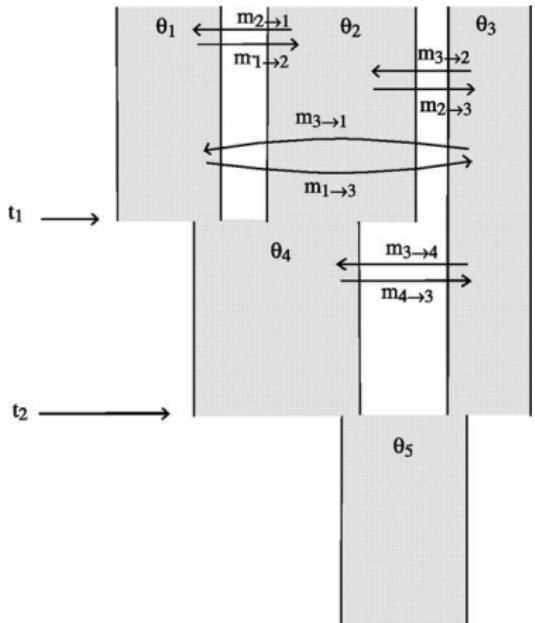


Figure adapted from [Hey, 2009]

Isolation-with-migration model.

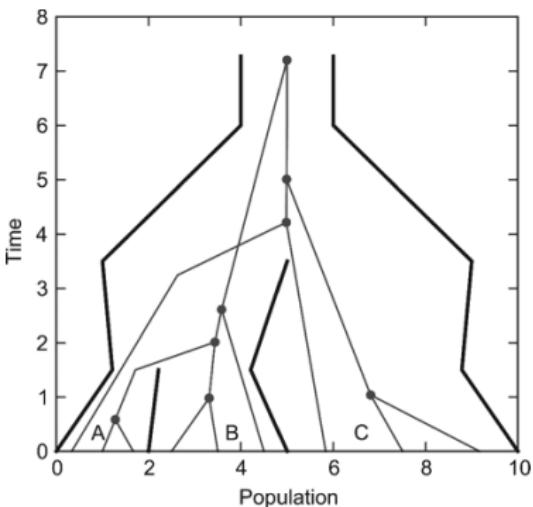


Figure adapted from  
[Heled and Drummond, 2010]

Multi-species coalescent model.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species tree models: Multi-species coalescent

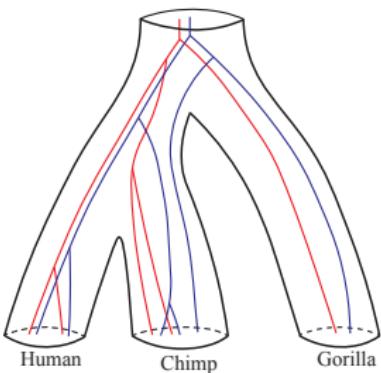


Figure adapted from [Zhang et al., 2017]

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species tree models: Multi-species coalescent

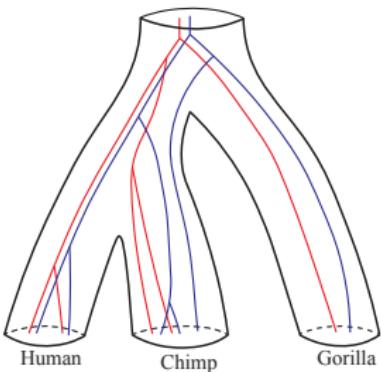


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a species tree (or transmission tree).

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species tree models: Multi-species coalescent

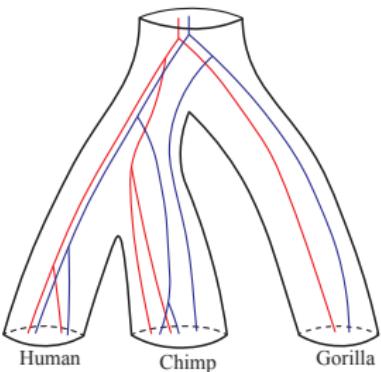


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a species tree (or transmission tree).
- ▶ Gene trees may be different to the species tree as the genes may coalesce any time prior to being in the same population.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species tree models: Multi-species coalescent

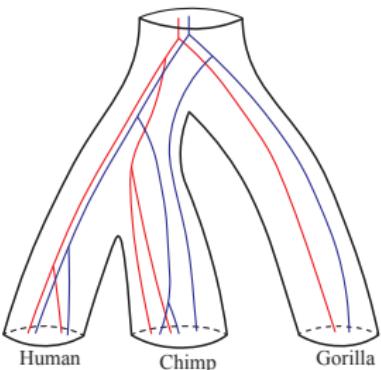


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a species tree (or transmission tree).
- ▶ Gene trees may be different to the species tree as the genes may coalesce any time prior to being in the same population.
- ▶ Failure of two homologous genes within the same species to coalesce during the lifetime of the species is called **incomplete lineage sorting**.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species tree models: Multi-species coalescent

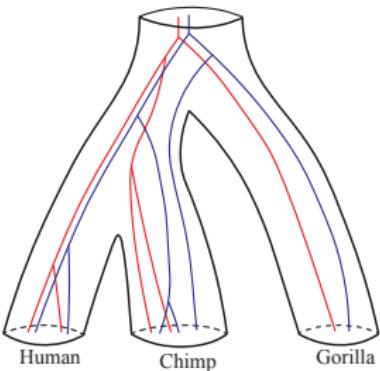


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a species tree (or transmission tree).
- ▶ Gene trees may be different to the species tree as the genes may coalesce any time prior to being in the same population.
- ▶ Failure of two homologous genes within the same species to coalesce during the lifetime of the species is called **incomplete lineage sorting**.
- ▶ Software: \*BEAST, \*BEAST2  
[Heled and Drummond, 2010, Ogilvie et al., 2016].

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species network inference

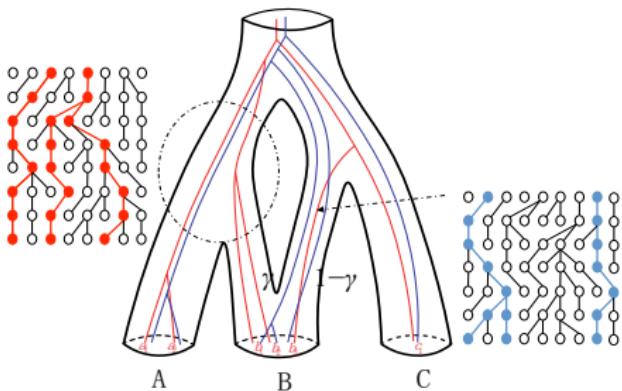


Figure adapted from [Zhang et al., 2017]

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species network inference

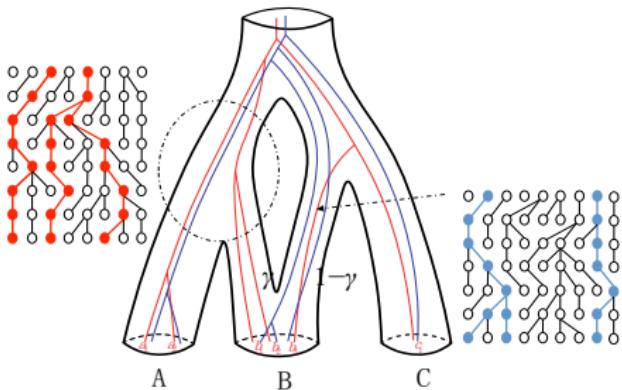


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a hybridization network or a lateral gene transfer network.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species network inference

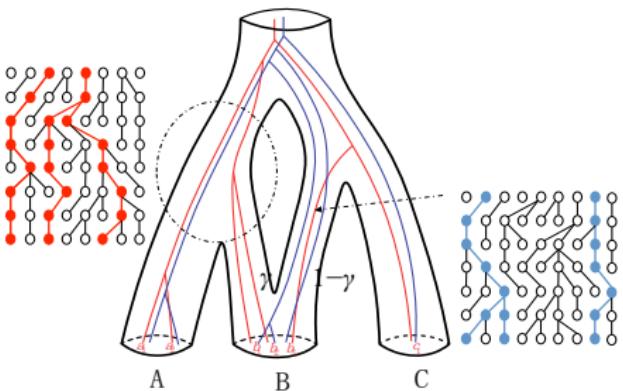


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a hybridization network or a lateral gene transfer network.
- ▶ A hybrid node has two ancestor lineages and one descendant lineage.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Species network inference

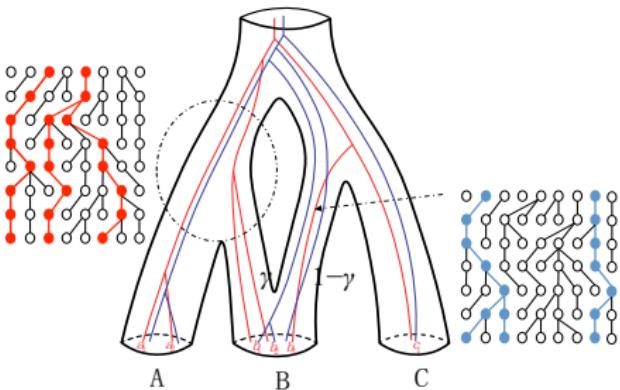


Figure adapted from [Zhang et al., 2017]

- ▶ Each gene has its own gene tree which is embedded within a hybridization network or a lateral gene transfer network.
- ▶ A hybrid node has two ancestor lineages and one descendant lineage.
- ▶ \*BEAST2 has been extended [Zhang et al., 2017] to allow for hybridization, lateral gene transfer.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Questions

- ② Under a structured birth-death model, how do the sub-population sizes vary (if at all) through time?
- ③ Suppose you perform a structured coalescent analysis on sequences collected from a relatively *unstructured* population. Would you expect the posterior migration rate to be very low or very high? Why?

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models

- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

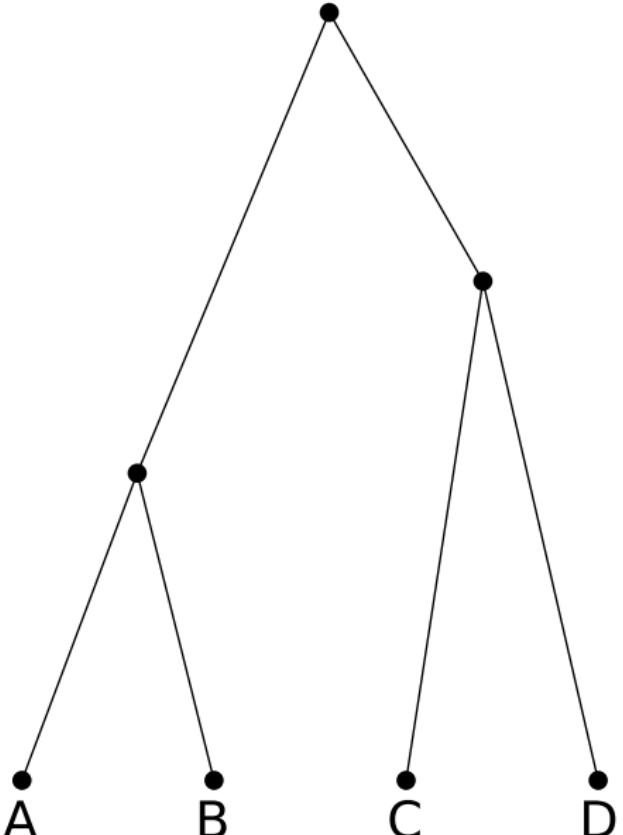
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Section 2

## Phylogenetic Networks

# Phylogenetic Trees



## Structured Populations

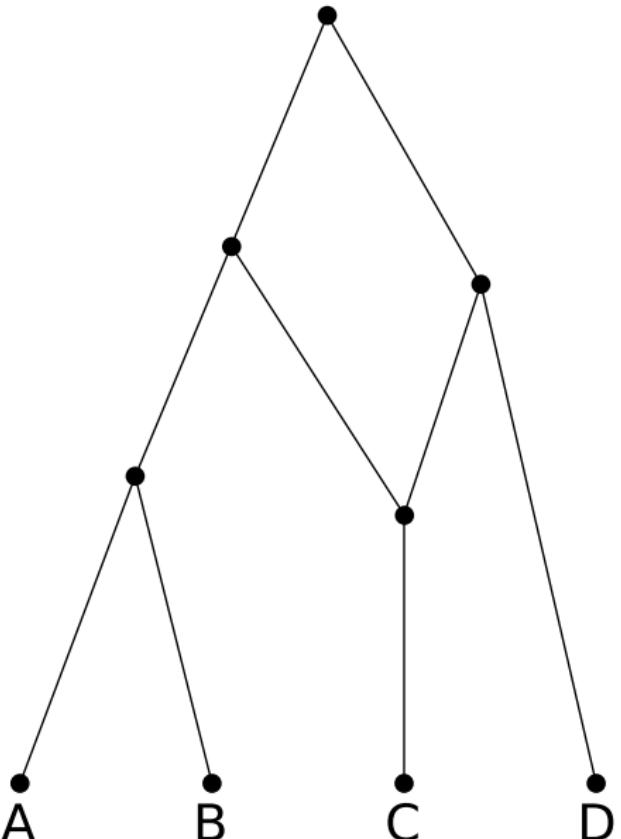
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Phylogenetic Networks



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

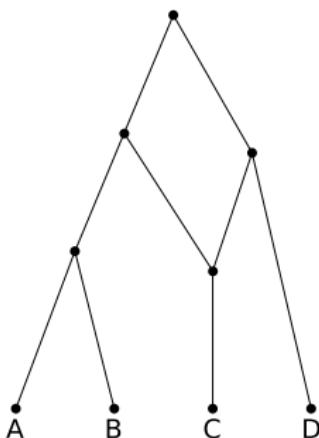
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Phylogenetic Networks

Just as for trees, phylogenetic networks represent a wide range of evolutionary relationships:



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

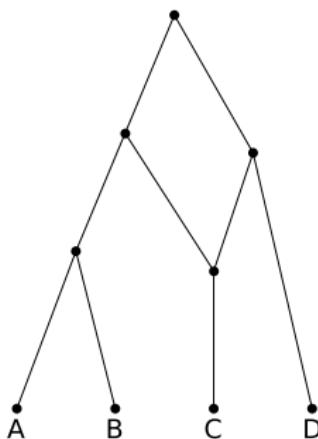
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Phylogenetic Networks

Just as for trees, phylogenetic networks represent a wide range of evolutionary relationships:

- ▶ For **species**, the network represents species ancestry and nodes with multiple parents represent hybridisation or horizontal gene transfer (HGT) events.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

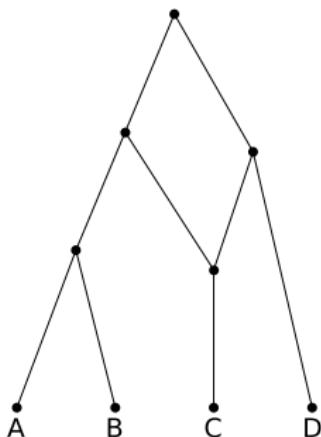
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Phylogenetic Networks

Just as for trees, phylogenetic networks represent a wide range of evolutionary relationships:

- ▶ For **species**, the network represents species ancestry and nodes with multiple parents represent hybridisation or horizontal gene transfer (HGT) events.
- ▶ For **individuals**, the network represents ancestry of individual lineages and nodes with multiple parents represent either hybridization, HGT or simply a node in a pedigree (family tree) of a sexually reproducing organism.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

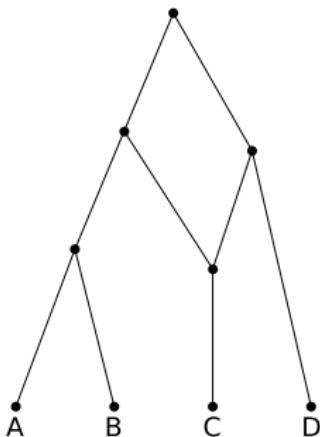
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Phylogenetic Networks

Just as for trees, phylogenetic networks represent a wide range of evolutionary relationships:

- ▶ For **species**, the network represents species ancestry and nodes with multiple parents represent hybridisation or horizontal gene transfer (HGT) events.
- ▶ For **individuals**, the network represents ancestry of individual lineages and nodes with multiple parents represent either hybridization, HGT or simply a node in a pedigree (family tree) of a sexually reproducing organism.
- ▶ For **genes or chromosomes**, the network represents ancestry of sequence data and nodes with multiple parents represent recombination events.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

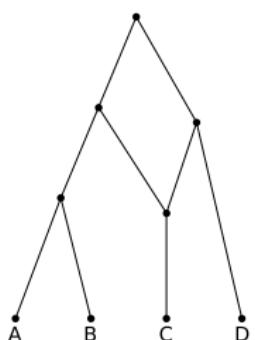
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models

- Structured coalescent models

- Structural dynamics

- Species trees

- Species networks

## Phylogenetic Networks

- Introduction to networks

- Linkage and phylogeny

- Wright-Fisher with Recombination

- Coalescent with Recombination

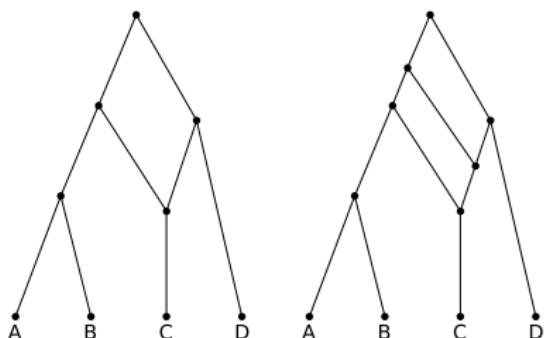
- Bayesian inference

- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

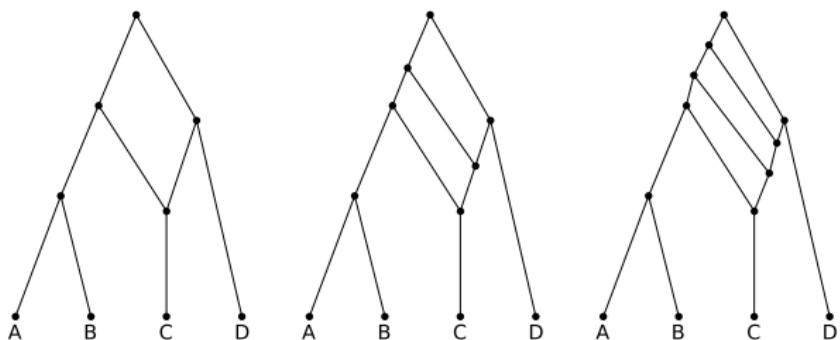
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

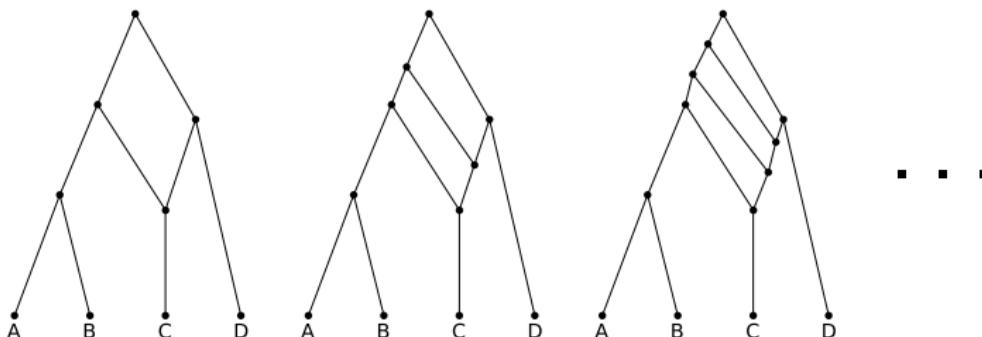
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

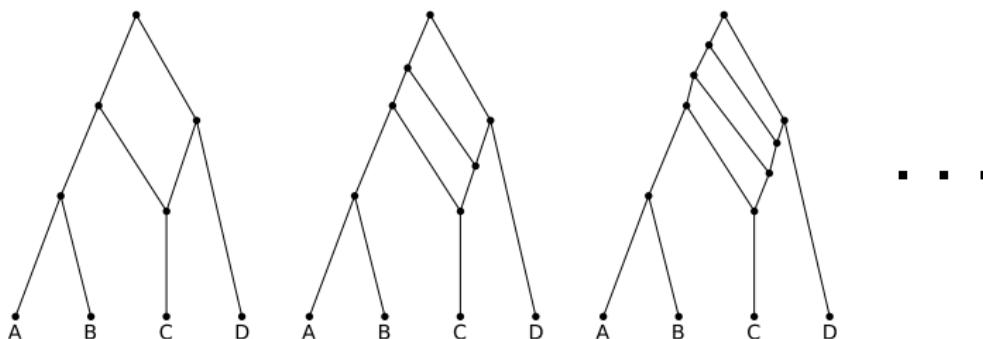
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



- ▶ Recall that there are  $(2n - 3)!!$  rooted trees with  $n$  leaves.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

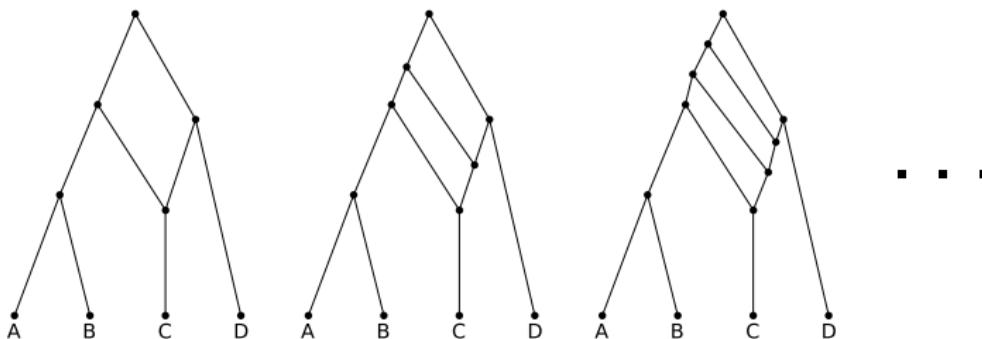
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# How many networks are there?

- ▶ Consider the ancestry of 4 species. How many distinct network topologies are there?



- ▶ Recall that there are  $(2n - 3)!!$  rooted trees with  $n$  leaves.
- ▶ There are an **infinite** number of possible ancestral network topologies.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Sexual reproduction and genetic linkage

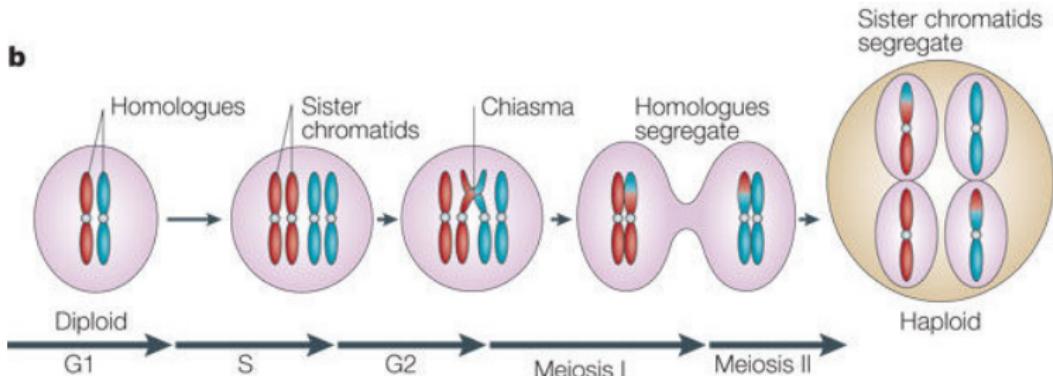
**b**

Figure adapted from [Marston and Amon, 2004]

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Sexual reproduction and genetic linkage

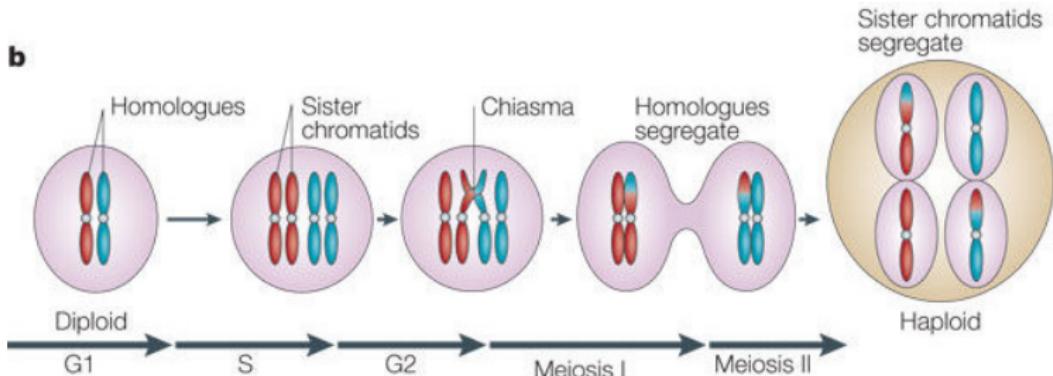
**b**

Figure adapted from [Marston and Amon, 2004]

- Recall: genetic linkage is the tendency for nearby sites to be inherited together.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Sexual reproduction and genetic linkage

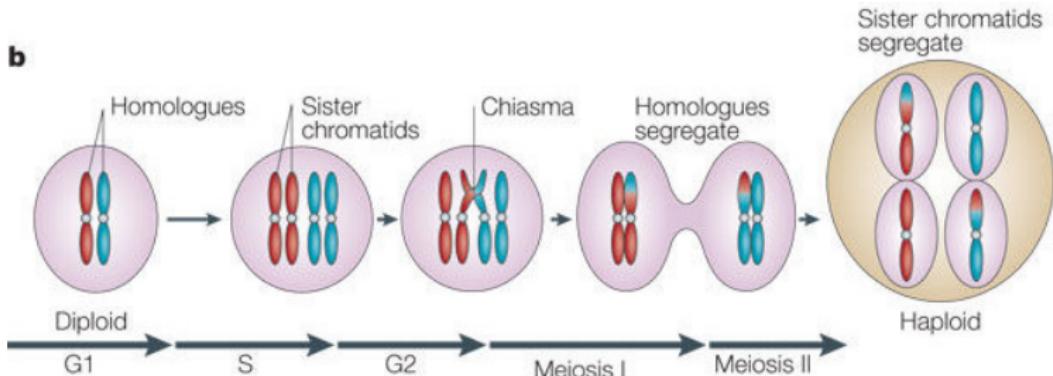
**b**

Figure adapted from [Marston and Amon, 2004]

- ▶ Recall: genetic linkage is the tendency for nearby sites to be inherited together.
- ▶ For sexually reproducing organisms, sites on different chromosomes are completely unlinked.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Sexual reproduction and genetic linkage

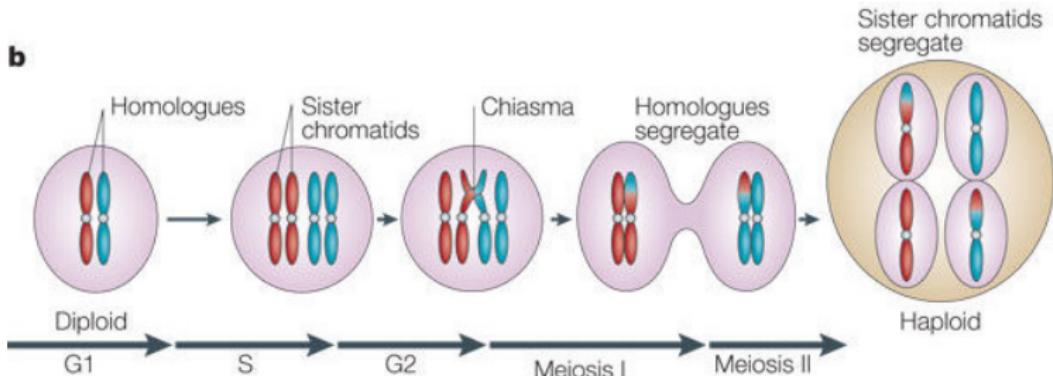
**b**

Figure adapted from [Marston and Amon, 2004]

- ▶ Recall: genetic linkage is the tendency for nearby sites to be inherited together.
- ▶ For sexually reproducing organisms, sites on different chromosomes are completely unlinked.
- ▶ Sites on the same chromosome are inherited together unless a homologous recombination event divides them.

## Structured Populations

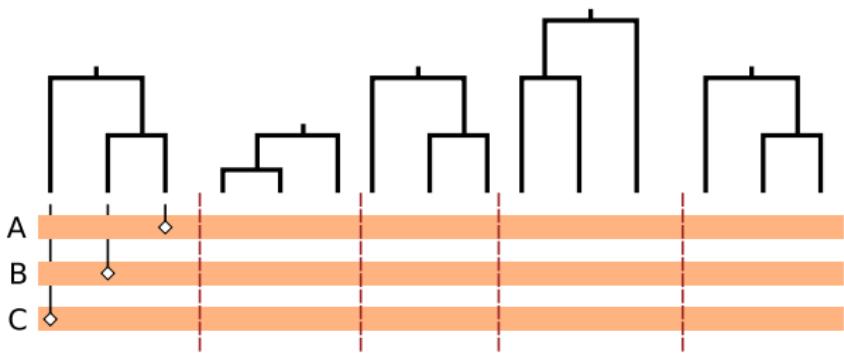
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



## Structured Populations

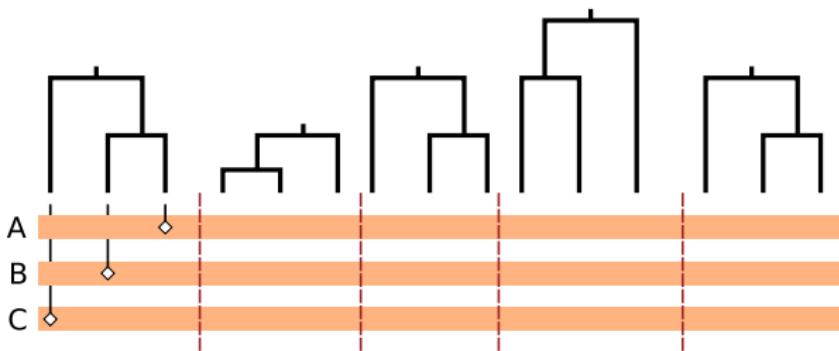
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



- ▶ Different sites correspond to different trees.

## Structured Populations

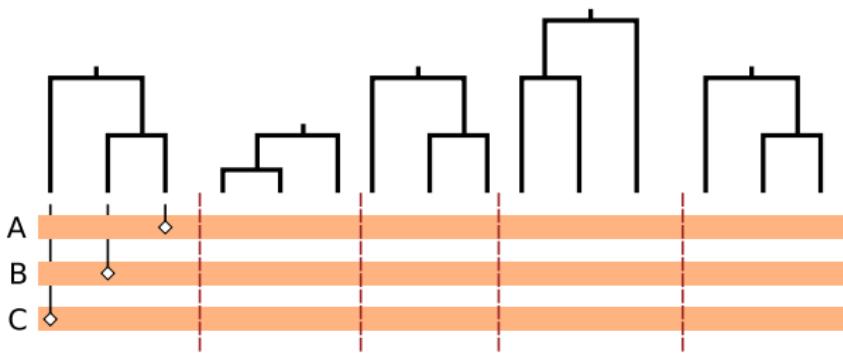
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



- ▶ Different sites correspond to different trees.
- ▶ The further away sites are on the alignment, the more likely they are to possess different ancestry.

## Structured Populations

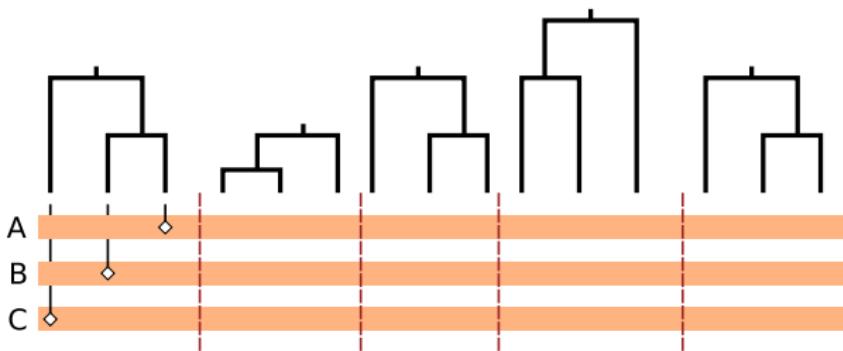
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



- ▶ Different sites correspond to different trees.
- ▶ The further away sites are on the alignment, the more likely they are to possess different ancestry.
- ▶ Single nucleotide polymorphisms (SNPs) are usually widely separated and are thus assumed to be completely unlinked - necessary for the validity of GWAS analyses.

## Structured Populations

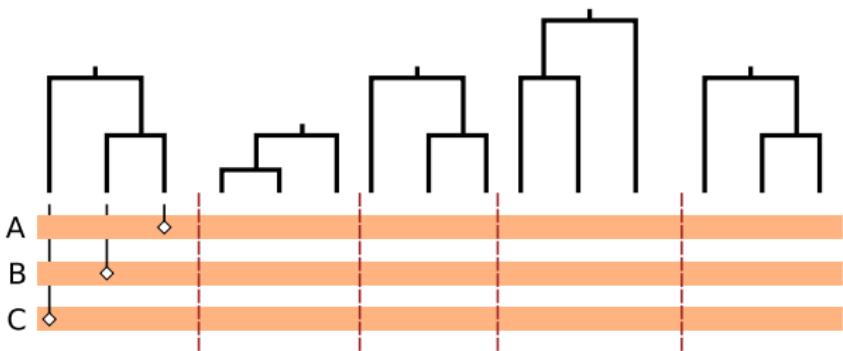
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



- ▶ Different sites correspond to different trees.
- ▶ The further away sites are on the alignment, the more likely they are to possess different ancestry.
- ▶ Single nucleotide polymorphisms (SNPs) are usually widely separated and are thus assumed to be completely unlinked - necessary for the validity of GWAS analyses.
- ▶ Short gene sequences often assumed to be completely linked (one tree for all sites).

## Structured Populations

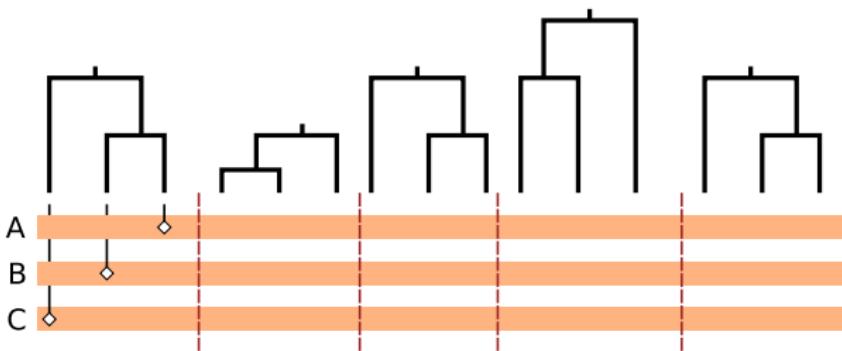
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Effect of recombination on phylogeny



- ▶ Different sites correspond to different trees.
- ▶ The further away sites are on the alignment, the more likely they are to possess different ancestry.
- ▶ Single nucleotide polymorphisms (SNPs) are usually widely separated and are thus assumed to be completely unlinked - necessary for the validity of GWAS analyses.
- ▶ Short gene sequences often assumed to be completely linked (one tree for all sites).
- ▶ Even for asexual entities (viruses, bacteria, etc.) reality is usually somewhere between these extremes.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Consider a Wright-Fisher population with female and male diploid individuals.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Consider a Wright-Fisher population with female and male diploid individuals.
- ▶ Focus on a small segment of a single autosome.
  - ▶ An autosome is a chromosome which is a member of a homologous pair, i.e. not a sex chromosome.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

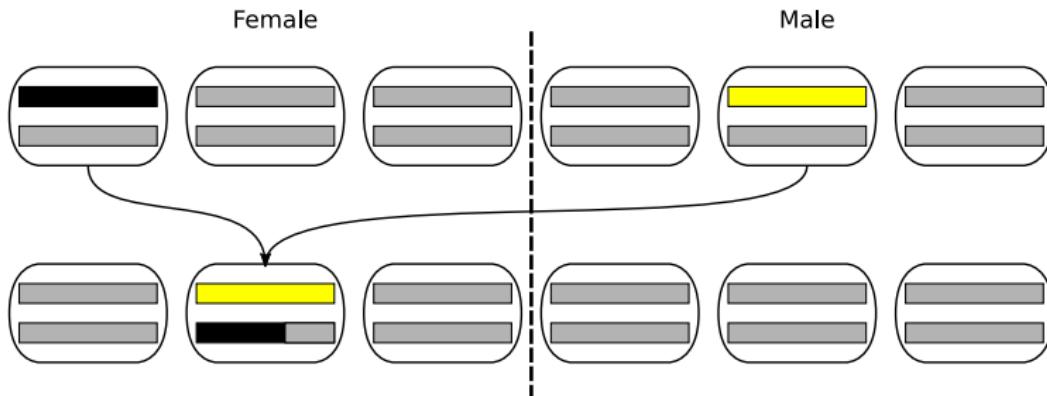
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Consider a Wright-Fisher population with female and male diploid individuals.
- ▶ Focus on a small segment of a single autosome.
  - ▶ An autosome is a chromosome which is a member of a homologous pair, i.e. not a sex chromosome.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

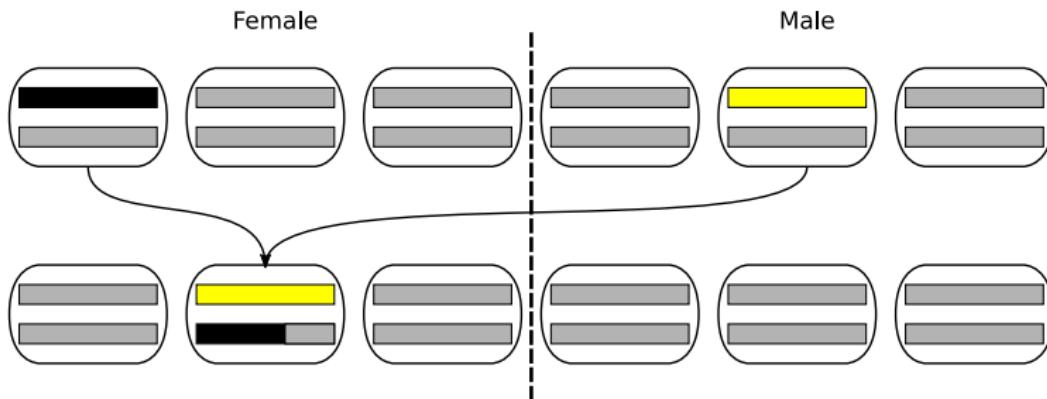
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Consider a Wright-Fisher population with female and male diploid individuals.
- ▶ Focus on a small segment of a single autosome.
  - ▶ An autosome is a chromosome which is a member of a homologous pair, i.e. not a sex chromosome.



- ▶ Each child selects 1 male and 1 female parent randomly from the previous generation.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

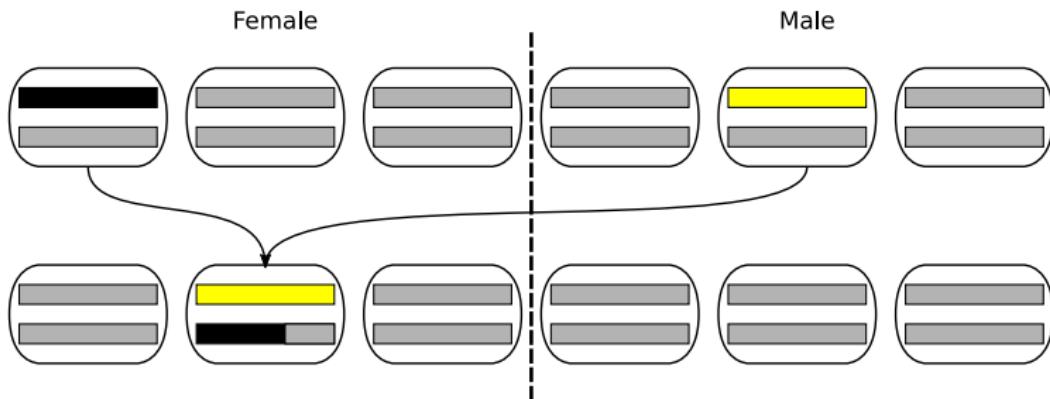
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Consider a Wright-Fisher population with female and male diploid individuals.
- ▶ Focus on a small segment of a single autosome.
  - ▶ An autosome is a chromosome which is a member of a homologous pair, i.e. not a sex chromosome.



- ▶ Each child selects 1 male and 1 female parent randomly from the previous generation.
- ▶ With probability  $r$  (which depends on the segment length) the homologous pair from one of the parents is recombinated.

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

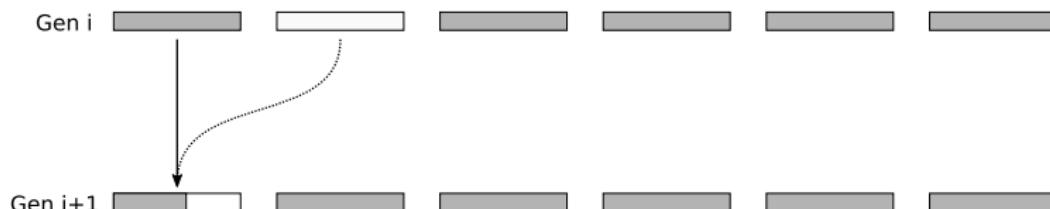
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

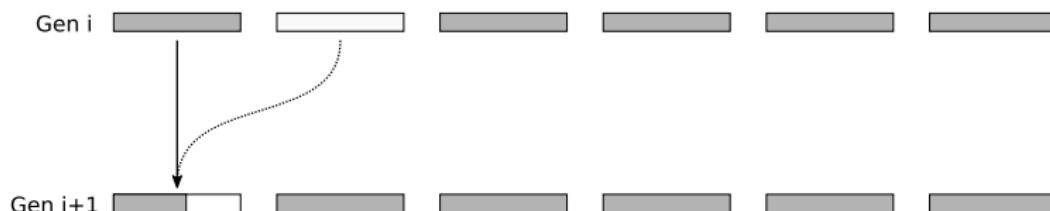
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



- ▶ Each child in  $i + 1$  selects a parent at random from gen.  $i$ .

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



- ▶ Each child in  $i + 1$  selects a parent at random from gen.  $i$ .
- ▶ With probability  $r$  an additional parent is selected.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

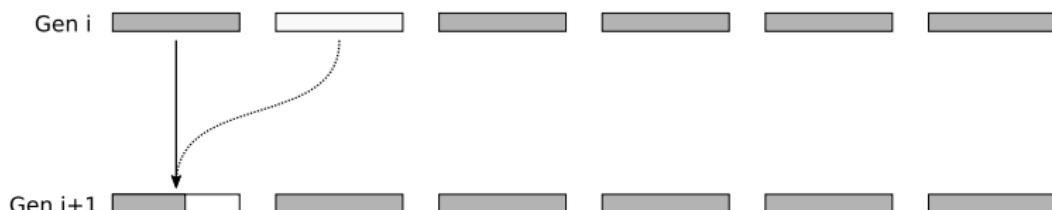
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Wright-Fisher with Recombination

- ▶ Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



- ▶ Each child in  $i + 1$  selects a parent at random from gen.  $i$ .
- ▶ With probability  $r$  an additional parent is selected.
- ▶ In this case, a break-point is chosen randomly on the chromosome, and everything to the right is replaced by the homologous section of the second parent's chromosome.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

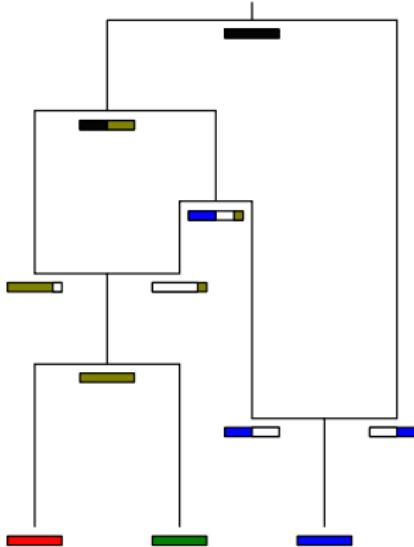
## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

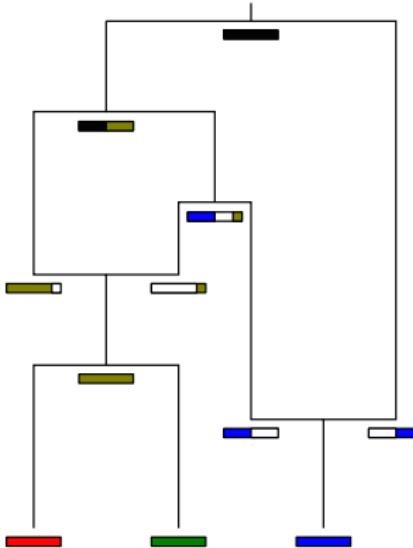
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

► **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

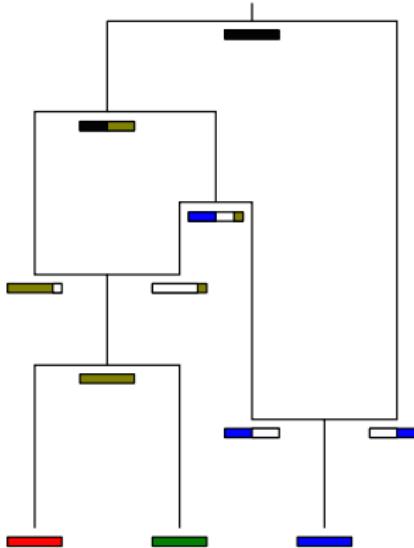
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

- ▶ **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .
- ▶ **Recombination rate:**  $\rho k$ .



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

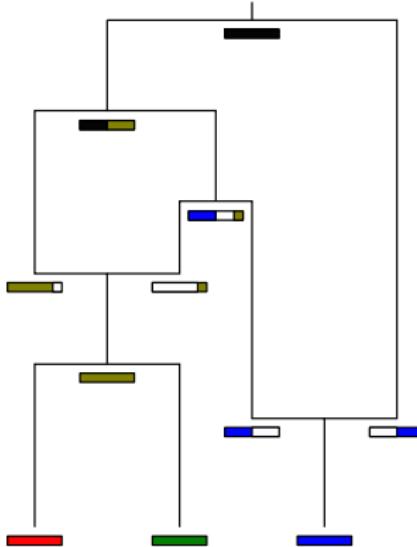
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

- ▶ **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .
- ▶ **Recombination rate:**  $\rho k$ .
- ▶ **Recombination break points:**  
chosen randomly along sequence:  
one parent contributes everything  
to the left, the other everything to  
the right.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

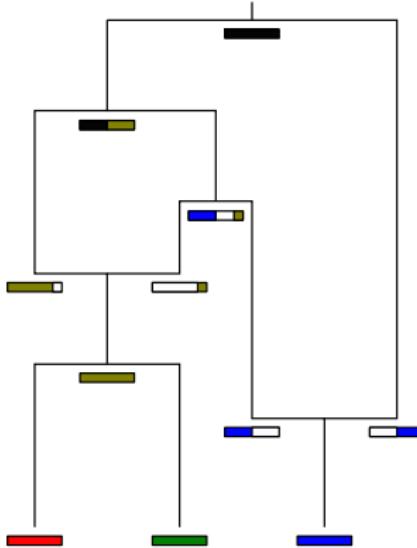
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

- ▶ **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .
- ▶ **Recombination rate:**  $\rho k$ .
- ▶ **Recombination break points:**  
chosen randomly along sequence:  
one parent contributes everything  
to the left, the other everything to  
the right.
- ▶ Each site possesses a local tree.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

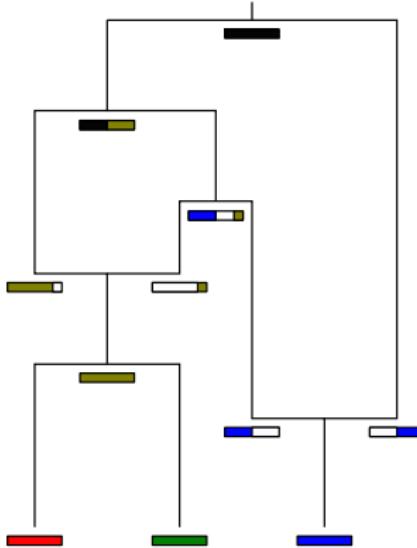
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

- ▶ **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .
- ▶ **Recombination rate:**  $\rho k$ .
- ▶ **Recombination break points:**  
chosen randomly along sequence:  
one parent contributes everything  
to the left, the other everything to  
the right.
- ▶ Each site possesses a local tree.
- ▶ Local trees may find MRCAs  
before **grand (G)MRCA** of the  
process.



## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

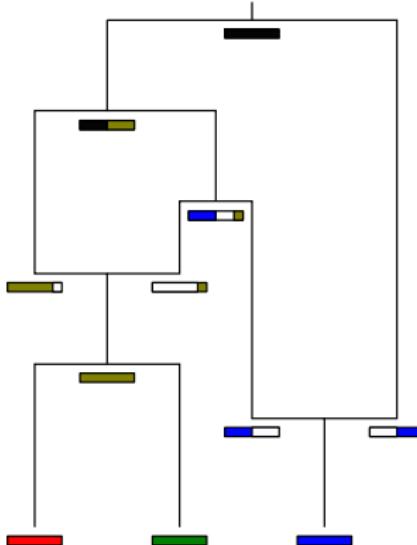
- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Coalescent with Recombination

For fixed recombination rate  $\rho = r/g$  in the limit  $r \ll 1$ ,  $g \ll 1$  and  $N \gg 1$ , the genealogical process is the coalescent with recombination [Hudson, 1983]:

- ▶ **Coalescence rate:**  $\binom{k}{2} \frac{1}{Ng}$ .
- ▶ **Recombination rate:**  $\rho k$ .
- ▶ **Recombination break points:**  
chosen randomly along sequence:  
one parent contributes everything  
to the left, the other everything to  
the right.
- ▶ Each site possesses a local tree.
- ▶ Local trees may find MRCAs  
before **grand (G)MRCA** of the  
process.



The result is the “ancestral recombination graph” or ARG.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

$$P(G, \rho, N, Q | A) = \frac{1}{P(A)} P(A|G, Q) P(G|\rho, N) P(\rho, N, Q)$$

- ▶  $G$  is the recombination graph/network.
- ▶  $Q$  is the substitution rate matrix.
- ▶  $\rho$  is the recombination rate.
- ▶  $N$  is the effective population size.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

$$P(G, \rho, N, Q | A) = \frac{1}{P(A)} P(A|G, Q) P(G|\rho, N) P(\rho, N, Q)$$

- ▶  $G$  is the recombination graph/network.
- ▶  $Q$  is the substitution rate matrix.
- ▶  $\rho$  is the recombination rate.
- ▶  $N$  is the effective population size.

Sampling from this distribution is difficult since:

1. some features of  $G$  do not contribute to the likelihood  
(i.e. these features are “unidentifiable”),

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

$$P(G, \rho, N, Q | A) = \frac{1}{P(A)} P(A|G, Q) P(G|\rho, N) P(\rho, N, Q)$$

- ▶  $G$  is the recombination graph/network.
- ▶  $Q$  is the substitution rate matrix.
- ▶  $\rho$  is the recombination rate.
- ▶  $N$  is the effective population size.

Sampling from this distribution is difficult since:

1. some features of  $G$  do not contribute to the likelihood (i.e. these features are “unidentifiable”),
2. the likelihood surface contains many distinct peaks, and

## Structured Populations

Population structure

Spatial structure

Non-spatial structure

Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

$$P(G, \rho, N, Q | A) = \frac{1}{P(A)} P(A|G, Q) P(G|\rho, N) P(\rho, N, Q)$$

- ▶  $G$  is the recombination graph/network.
- ▶  $Q$  is the substitution rate matrix.
- ▶  $\rho$  is the recombination rate.
- ▶  $N$  is the effective population size.

Sampling from this distribution is difficult since:

1. some features of  $G$  do not contribute to the likelihood (i.e. these features are “unidentifiable”),
2. the likelihood surface contains many distinct peaks, and
3. the volume of the space of phylogenetic networks with significant posterior probability is usually extremely large.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Bayesian phylogenetic network inference

Can easily write down an expression for the posterior distribution for a phylogenetic network given a sequence alignment:

$$P(G, \rho, N, Q | A) = \frac{1}{P(A)} P(A|G, Q) P(G|\rho, N) P(\rho, N, Q)$$

- ▶  $G$  is the recombination graph/network.
- ▶  $Q$  is the substitution rate matrix.
- ▶  $\rho$  is the recombination rate.
- ▶  $N$  is the effective population size.

Sampling from this distribution is difficult since:

1. some features of  $G$  do not contribute to the likelihood (i.e. these features are “unidentifiable”),
2. the likelihood surface contains many distinct peaks, and
3. the volume of the space of phylogenetic networks with significant posterior probability is usually extremely large.

Despite this, many approximate algorithms exist.

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# Algorithms for Bayesian ARG inference

**SMARTIE** : [Bloomquist and Suchard, 2010] (BEAST)  
MCMC sampler under a “non-informative”  
network prior rather than the CwR.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Algorithms for Bayesian ARG inference

**SMARTIE** : [Bloomquist and Suchard, 2010] (BEAST)

MCMC sampler under a “non-informative”  
network prior rather than the CwR.

**ARGweaver** : [Rasmussen et al., 2014]

MCMC sampler under a computationally efficient  
approximation of the CwR.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Algorithms for Bayesian ARG inference

**SMARTIE** : [Bloomquist and Suchard, 2010] (BEAST)

MCMC sampler under a “non-informative” network prior rather than the CwR.

**ARGweaver** : [Rasmussen et al., 2014]

MCMC sampler under a computationally efficient approximation of the CwR.

**ClonalOrigin** : [Didelot et al., 2010]

MCMC sampler for Bacterial ARGs under the coalescent with gene conversion (modification of the CwR to account for homologous gene conversion).

## Structured Populations

Population structure

Spatial structure

Non-spatial structure

Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Algorithms for Bayesian ARG inference

**SMARTIE** : [Bloomquist and Suchard, 2010] (BEAST)

MCMC sampler under a “non-informative” network prior rather than the CwR.

**ARGweaver** : [Rasmussen et al., 2014]

MCMC sampler under a computationally efficient approximation of the CwR.

**ClonalOrigin** : [Didelot et al., 2010]

MCMC sampler for Bacterial ARGs under the coalescent with gene conversion (modification of the CwR to account for homologous gene conversion).

**Bacter** : [Vaughan et al., 2017] (BEAST 2)

Re-implementation of the ClonalOrigin model with fewer restrictions.

## Structured Populations

Population structure

Spatial structure

Non-spatial structure

Structure in phylodynamic analyses

Structured birth-death models

Structured coalescent models

Structural dynamics

Species trees

Species networks

## Phylogenetic Networks

Introduction to networks

Linkage and phylogeny

Wright-Fisher with Recombination

Coalescent with Recombination

Bayesian inference

Population dynamics inference

## References

# Bayesian ARG inference example

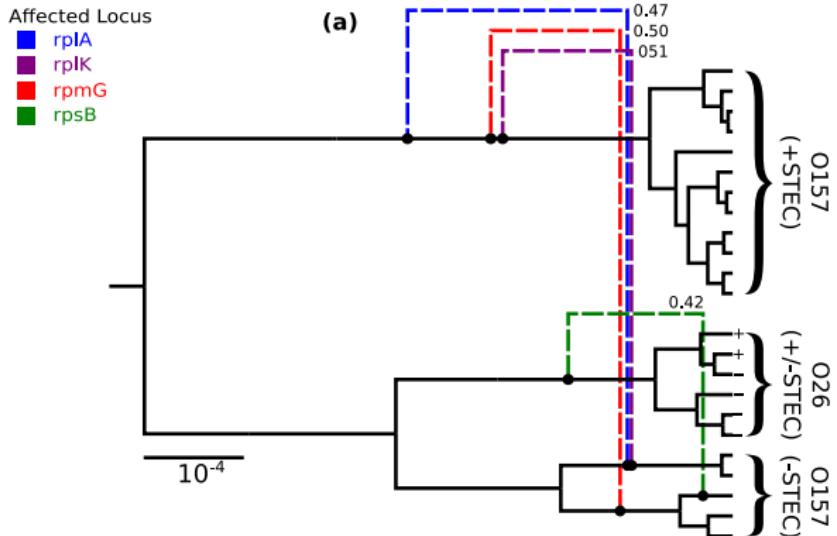


Figure adapted from [Vaughan et al., 2017]

Phylogenetic network describing ancestry of *E. coli* sequences produced using Bayesian inference.

## Structured Populations

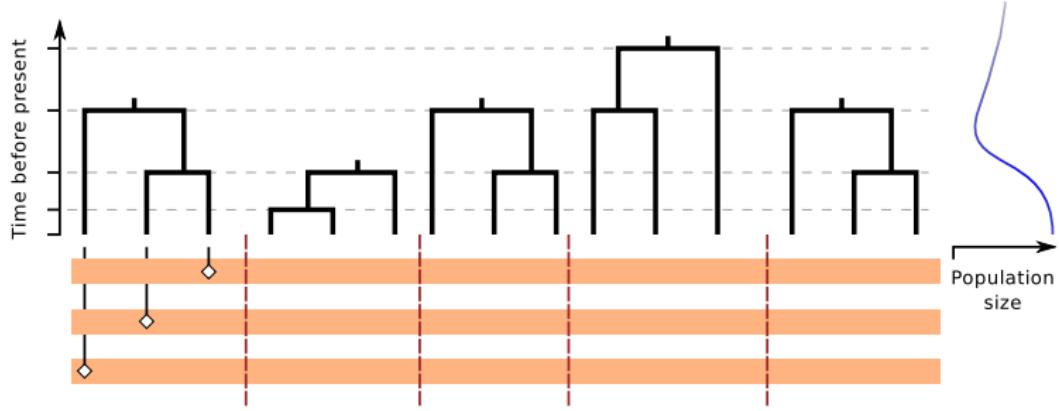
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of population dynamics



## Structured Populations

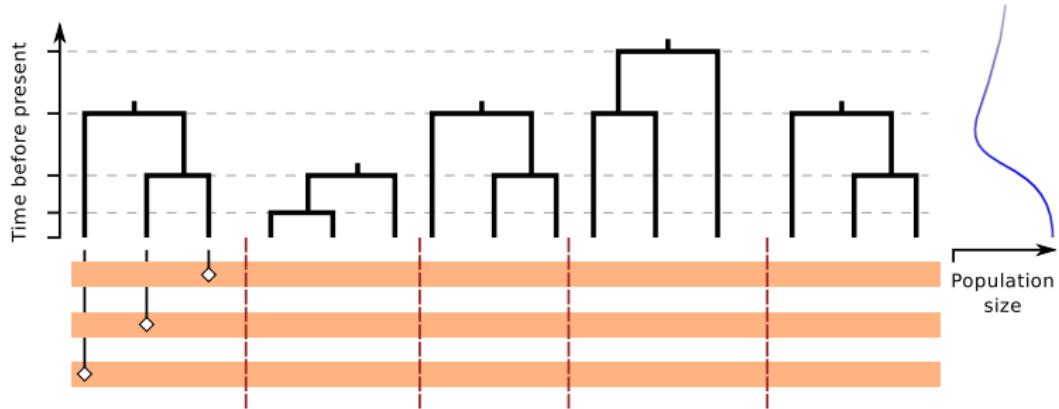
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of population dynamics



- ▶ Each local tree contributes additional information to the inference of population size.

## Structured Populations

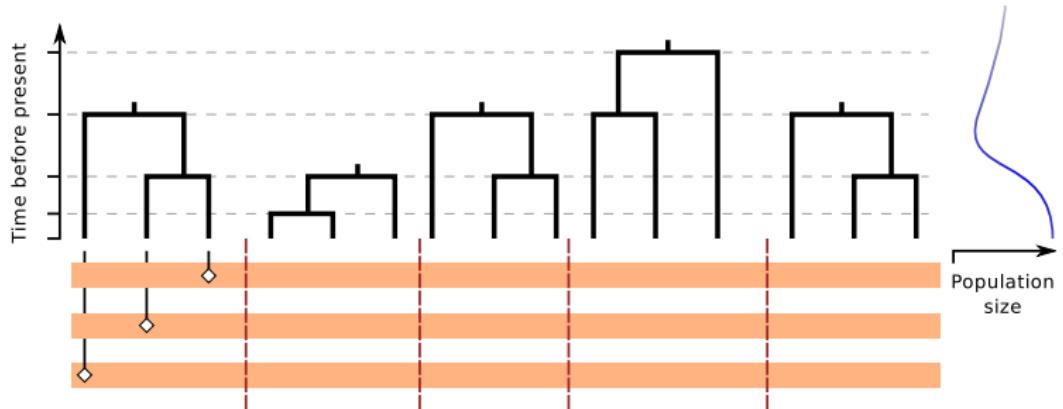
- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of population dynamics



- ▶ Each local tree contributes additional information to the inference of population size.
- ▶ The longer the sequence, assuming the local trees can be accurately inferred, the more powerful the population dynamics inference.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Sequentially Markovian Coalescent

- ▶ [Wiuf and Hein, 1999] reformulated the CwR as a “spatial” process along sequences.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Sequentially Markovian Coalescent

- ▶ [Wiuf and Hein, 1999] reformulated the CwR as a “spatial” process along sequences.
- ▶ [McVean and Cardin, 2005] modified this to produce a process for local trees which is Markovian along sequences.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# The Sequentially Markovian Coalescent

- ▶ [Wiuf and Hein, 1999] reformulated the CwR as a “spatial” process along sequences.
- ▶ [McVean and Cardin, 2005] modified this to produce a process for local trees which is Markovian along sequences.

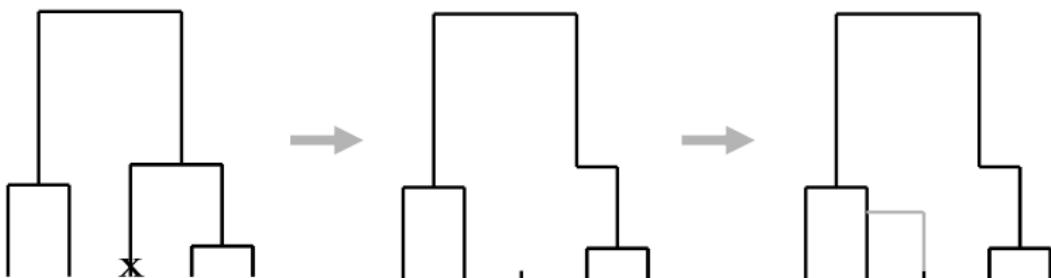


Figure adapted from [McVean and Cardin, 2005]

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# The Sequentially Markovian Coalescent

- ▶ [Wiuf and Hein, 1999] reformulated the CwR as a “spatial” process along sequences.
- ▶ [McVean and Cardin, 2005] modified this to produce a process for local trees which is Markovian along sequences.

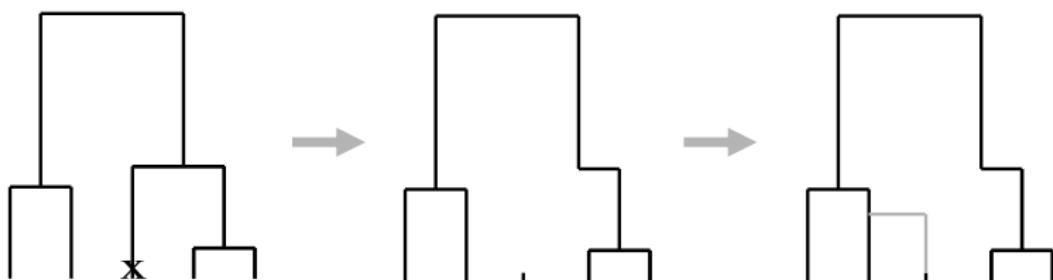


Figure adapted from [McVean and Cardin, 2005]

- ▶ Sites between breakpoints exponentially distributed with rate  $\rho L$  where  $L$  is the total edge length of the current local tree.

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# The Sequentially Markovian Coalescent

- ▶ [Wiuf and Hein, 1999] reformulated the CwR as a “spatial” process along sequences.
- ▶ [McVean and Cardin, 2005] modified this to produce a process for local trees which is Markovian along sequences.

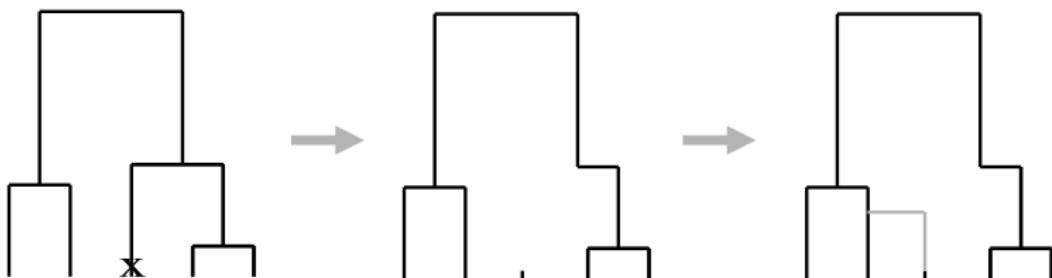


Figure adapted from [McVean and Cardin, 2005]

- ▶ Sites between breakpoints exponentially distributed with rate  $\rho L$  where  $L$  is the total edge length of the current local tree.
- ▶ It neglects some possible recombinations, e.g. those that do not affect the data.

<b>Structured Populations</b>
Population structure
Spatial structure
Non-spatial structure
Structure in phylodynamic analyses
Structured birth-death models
Structured coalescent models
Structural dynamics
Species trees
Species networks
<b>Phylogenetic Networks</b>
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
<b>References</b>

# Inference of human population dynamics

- ▶ [Li and Durbin, 2011] developed an SMC-based hidden Markov model on pairs of alignments.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of human population dynamics

- ▶ [Li and Durbin, 2011] developed an SMC-based hidden Markov model on pairs of alignments.
- ▶ Hidden states of HMM are local tree heights at each site.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of human population dynamics

- ▶ [Li and Durbin, 2011] developed an SMC-based hidden Markov model on pairs of alignments.
- ▶ Hidden states of HMM are local tree heights at each site.
- ▶ Used to jointly infer heights and population size dynamics.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Inference of human population dynamics

- ▶ [Li and Durbin, 2011] developed an SMC-based hidden Markov model on pairs of alignments.
- ▶ Hidden states of HMM are local tree heights at each site.
- ▶ Used to jointly infer heights and population size dynamics.

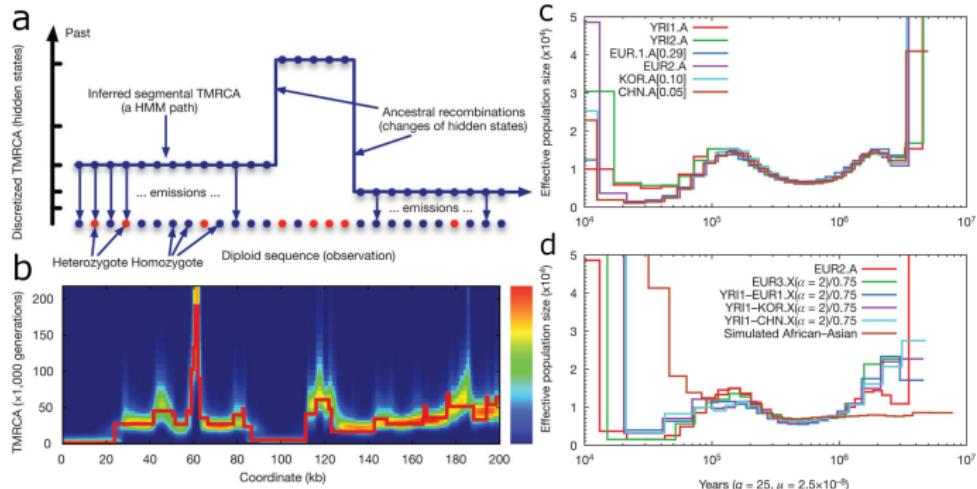


Figure adapted from [Li and Durbin, 2011]

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phydodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# Questions

- ② What is the maximum number of local trees that can correspond to a sequence alignment?
- ② Suppose you have an alignment of extremely long sequences sampled from individuals belonging to a species undergoing frequent recombination. Why might there still be significant uncertainty in the inferred ARG?
- ② We saw how recombination can improve our ability to infer ancestral population dynamics. Would you expect higher recombination rates to always improve this? Why/Why not?

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# References I

- Barido-Sottani, J., Vaughan, T. G., and Stadler, T. (2018). A multi-state birth-death model for Bayesian inference of lineage-specific birth and death rates. *bioRxiv preprint*.
- Beerli, P. (2004). Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. *Molecular Ecology*, 13(4):827–836.
- Bloomquist, E. W. and Suchard, M. A. (2010). Unifying vertical and nonvertical evolution: a stochastic arg-based framework. *Syst Biol*, 59(1):27–41.
- De Maio, N., Wu, C.-H., O'Reilly, K. M., and Wilson, D. (2015). New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genetics*, 11(8):e1005421.
- Didelot, X., Lawson, D., Daarling, A., and Falush, D. (2010). Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*, 186:1435.
- FitzJohn, R. G., Maddison, W. P., and Otto, S. P. (2009). Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, 58(6):595–611.
- Heled, J. and Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Mol Biol Evol*, 27(3):570–80.
- Hey, J. (2009). Isolation with migration models for more than two populations. *Molecular biology and evolution*, 27(4):905–920.
- Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol*, 23(2):183–201.
- Kühnert, D., Stadler, T., Vaughan, T. G., and Drummond, A. J. (2016). Phylodynamics with migration: A computational framework to quantify population structure from genomic data. *Molecular biology and evolution*, page msw064.
- Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496.
- Maddison, W. P., Midford, P. E., and Otto, S. P. (2007). Estimating a binary character's effect on speciation and extinction. *Syst Biol*, 56(5):701–710.
- Marston, A. L. and Amon, A. (2004). Meiosis: cell-cycle controls shuffle and deal. *Nature Reviews Molecular Cell Biology*, 5(12):983–997.
- McVean, G. A. T. and Cardin, N. J. (2005). Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci*, 360(1459):1387–1393.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References

# References II

- Müller, N. F., Rasmussen, D., and Stadler, T. (2018). MASCOT: parameter and state inference under the marginal structured coalescent approximation. *Bioinformatics*.
- Müller, N. F., Rasmussen, D. A., and Stadler, T. (2017). The structured coalescent and its approximations. *Molecular Biology and Evolution*, 34(11):2970–2981.
- Notohara, M. (1990). The coalescent and the genealogical process in geographically structured population. *J Math Biol*, 29(1):59–75.
- Ogilvie, H. A., Heled, J., Xie, D., and Drummond, A. J. (2016). Computational performance and statistical accuracy of \*beast and comparisons with other methods. *Systematic biology*, page syv118.
- Pannell, J. R. (2003). Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, 57(5):949–961.
- Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLoS genetics*, 10(5):e1004342.
- Scire, J., Barido-Sottani, J., Kühnert, D., Vaughan, T. G., and Stadler, T. Improved multi-type birth-death phylodynamic inference in BEAST 2. In preparation.
- Vaughan, T. G., Welch, D., Drummond, A. J., Biggs, P. J., George, T., and French, N. P. (2017). Inferring ancestral recombination graphs from bacterial genomic data. *Genetics*, 205:857–870.
- Wiuf, C. and Hein, J. (1999). Recombination as a point process along sequences. *Theor Popul Biol*, 55(3):248–259.
- Zhang, C., Ogilvie, H. A., Drummond, A. J., and Stadler, T. (2017). Bayesian inference of species networks from multilocus sequence data. *Molecular Biology and Evolution*, page msx307.

## Structured Populations

- Population structure
- Spatial structure
- Non-spatial structure
- Structure in phylodynamic analyses
- Structured birth-death models
- Structured coalescent models
- Structural dynamics
- Species trees
- Species networks

## Phylogenetic Networks

- Introduction to networks
- Linkage and phylogeny
- Wright-Fisher with Recombination
- Coalescent with Recombination
- Bayesian inference
- Population dynamics inference

## References