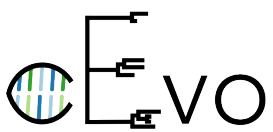
Computational Biology

Lecturers:
Tanja Stadler, Carsten Magnus & Tim Vaughan

Teaching Assistants:
Jūlija Pečerska, Jérémie Sciré,
Sarah Nadeau & Marc Manceau

Computational Evolution
Department of Biosystems Science and Engineering

HS 2019



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Questions

- Under a structured birth-death model, how do the sub-population sizes vary (if at all) through time?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Questions

- Under a structured birth-death model, how do the sub-population sizes vary (if at all) through time?
- 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?

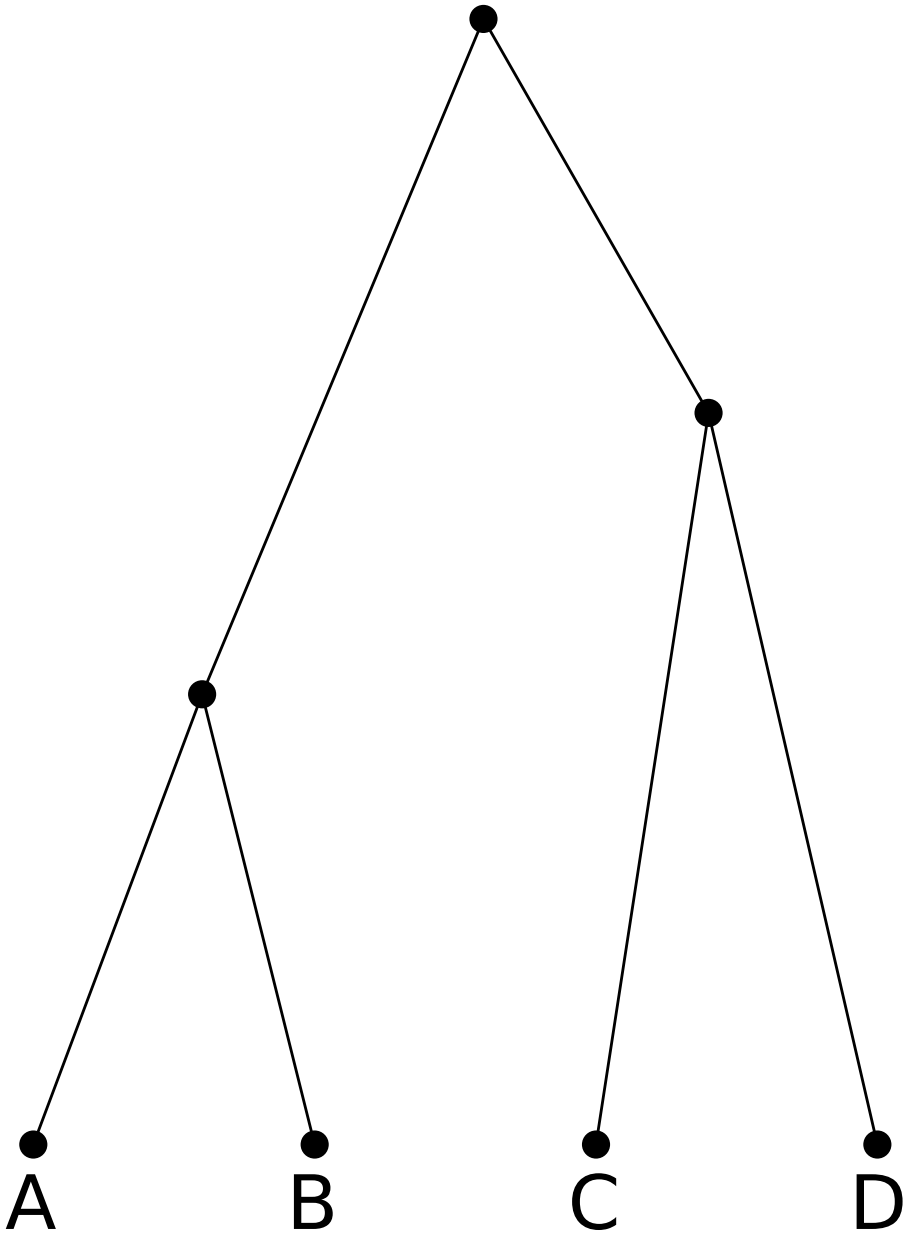
Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Questions

- Under a structured birth-death model, how do the sub-population sizes vary (if at all) through time?
- 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?
- How might the evolution of languages violate the assumptions of a substitution+birth/death phylodynamic model?

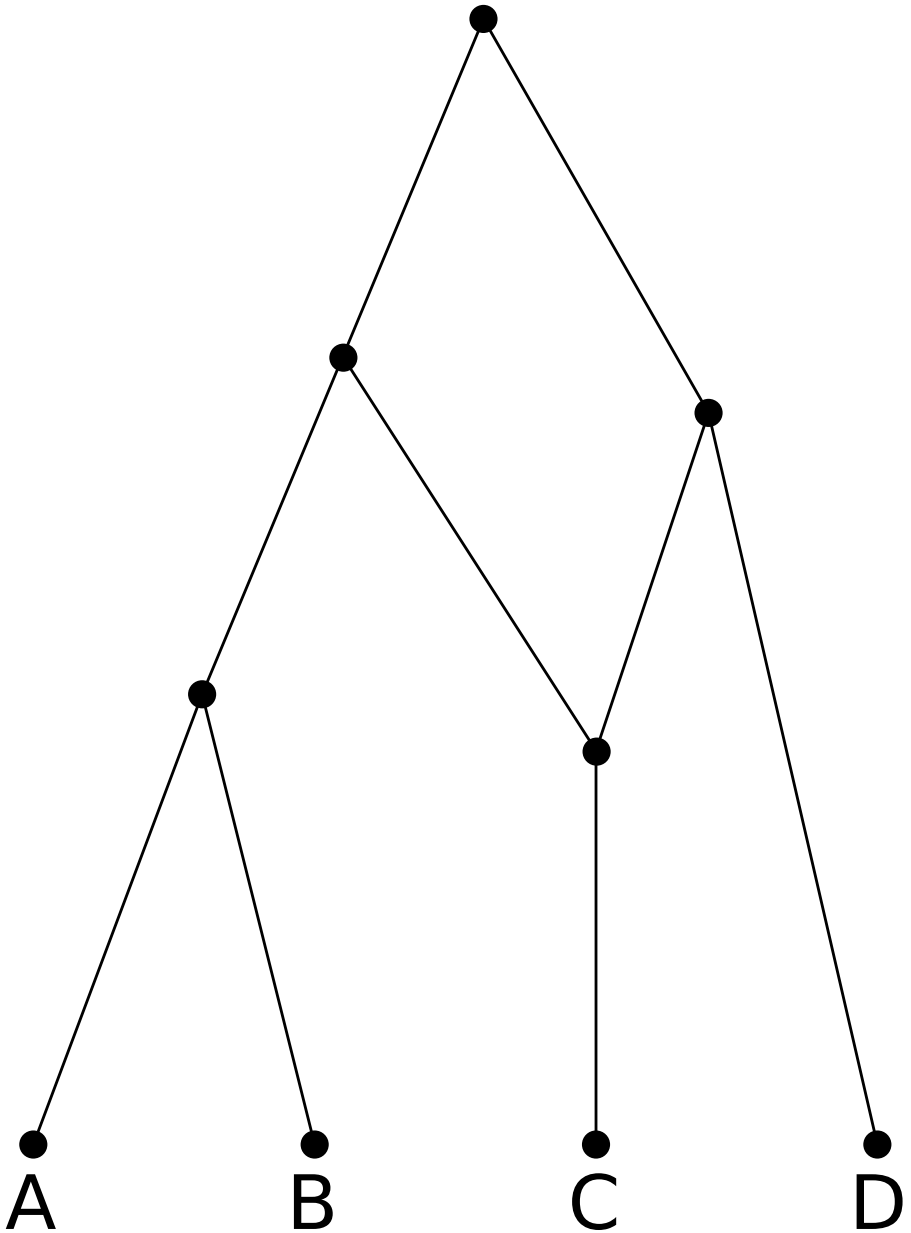
Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Phylogenetic Trees



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

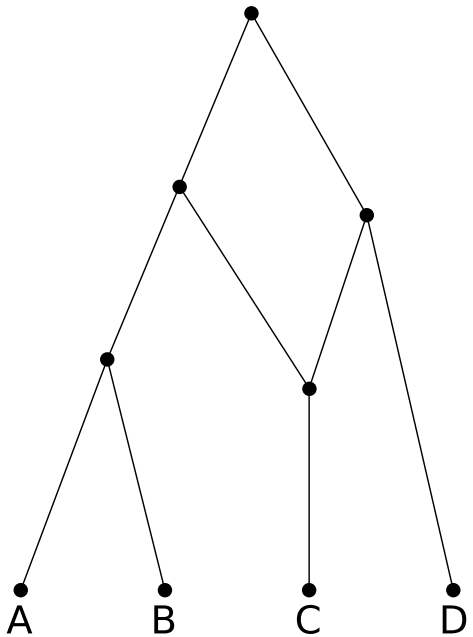
Phylogenetic Networks



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Phylogenetic Networks

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

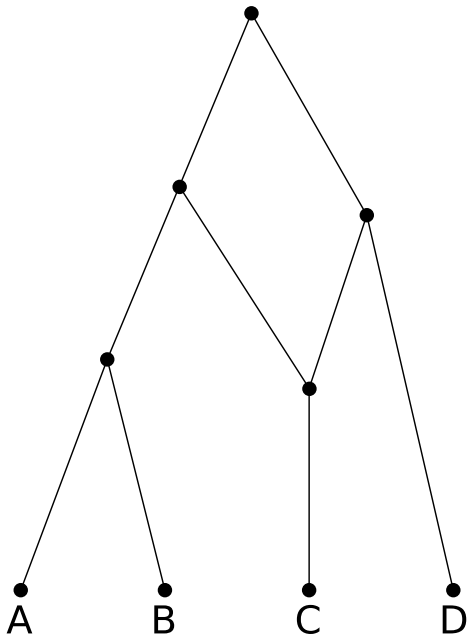


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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.

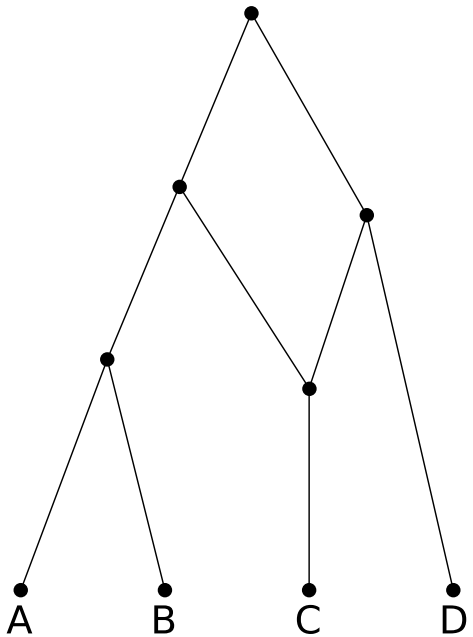


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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.

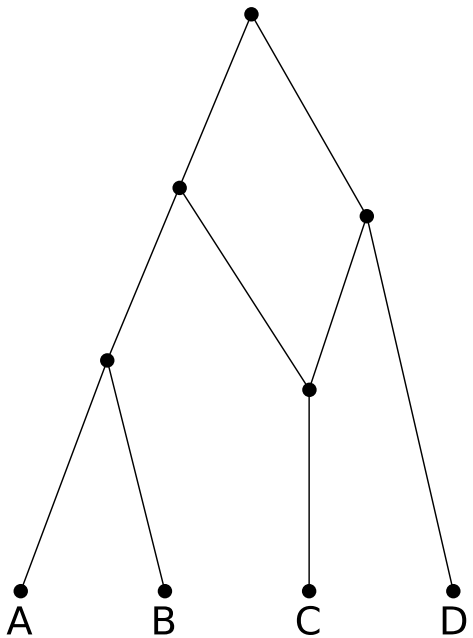


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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.



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

How many networks are there?

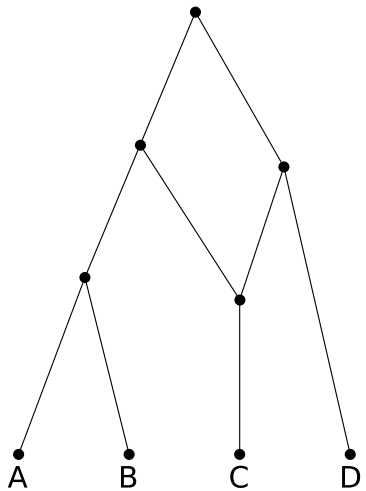
- Consider the ancestry of 4 species. How many distinct network topologies are there?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

How many networks are there?

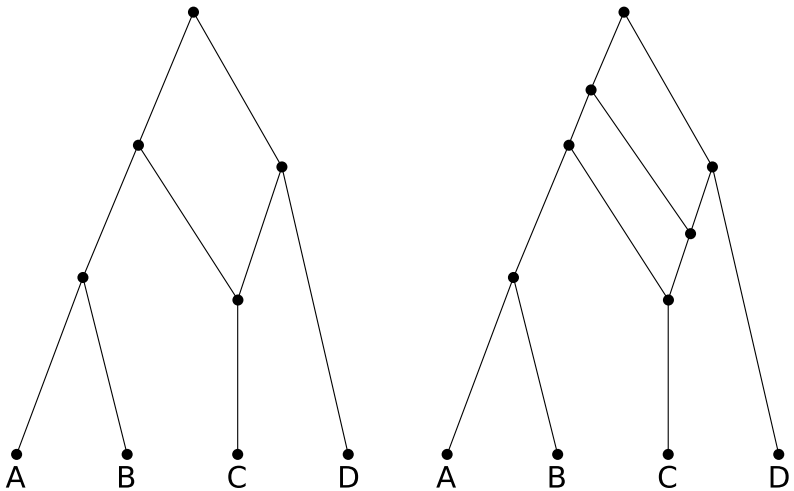
Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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



How many networks are there?

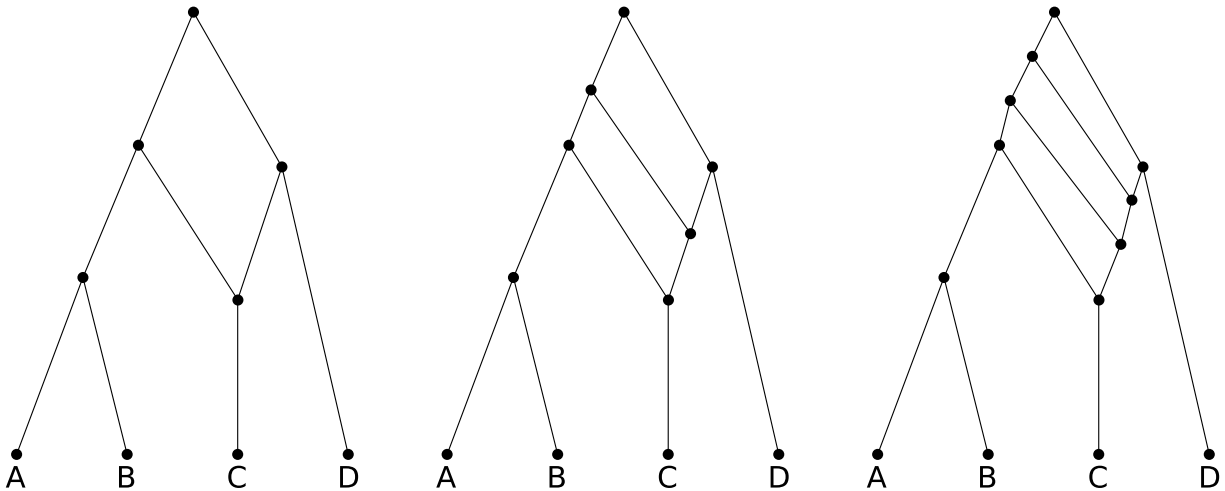
- Consider the ancestry of 4 species. How many distinct network topologies are there?



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

How many networks are there?

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

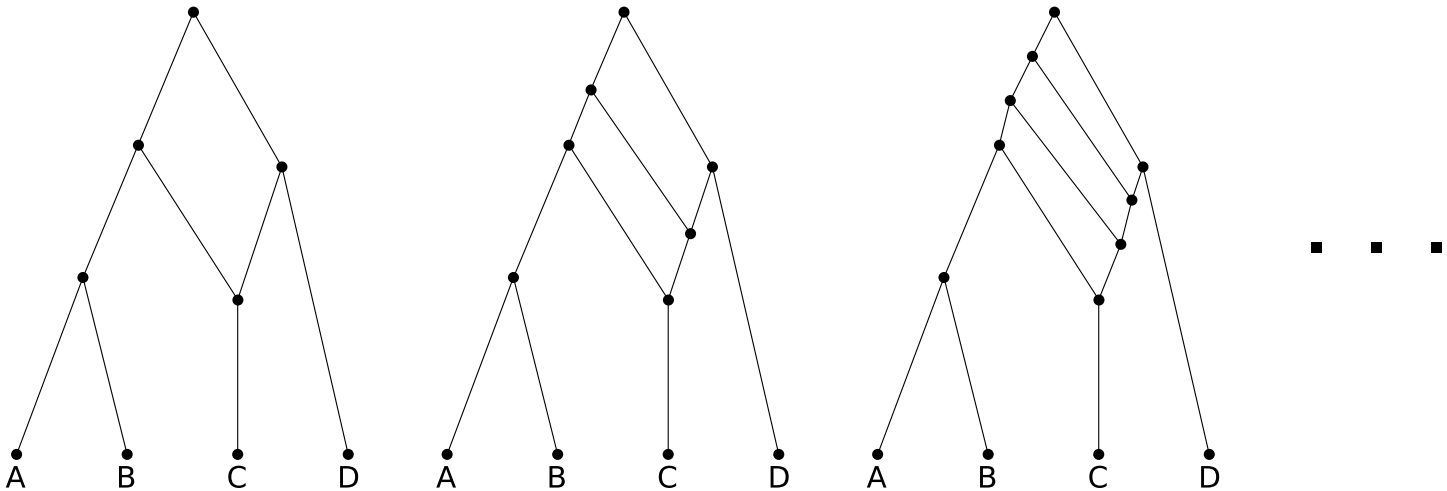


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

How many networks are there?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

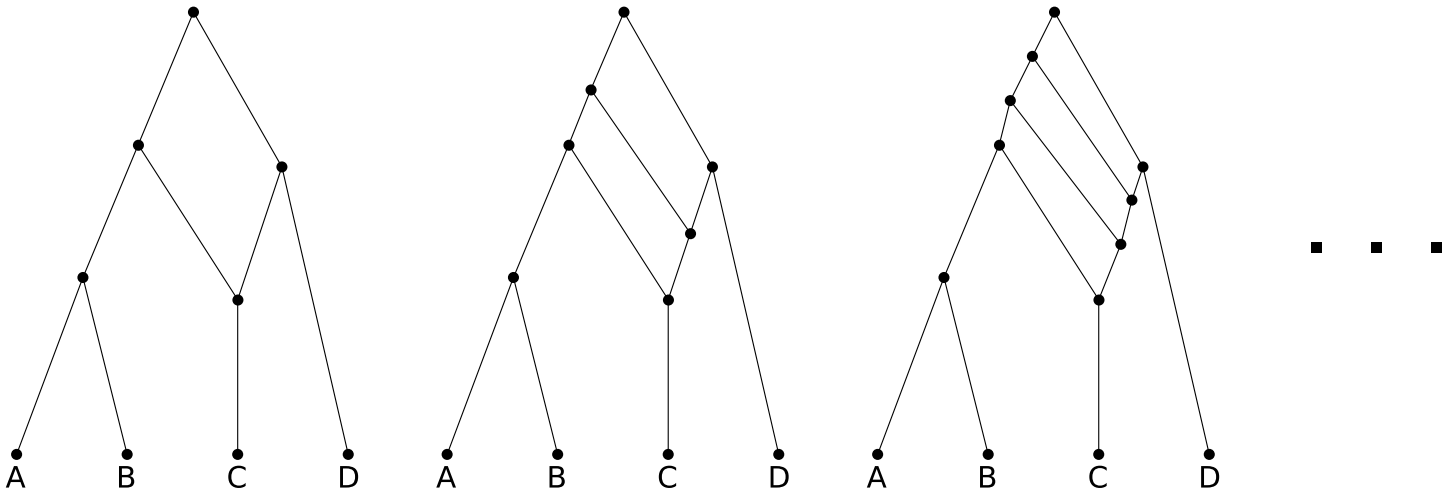
- Consider the ancestry of 4 species. How many distinct network topologies are there?



How many networks are there?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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

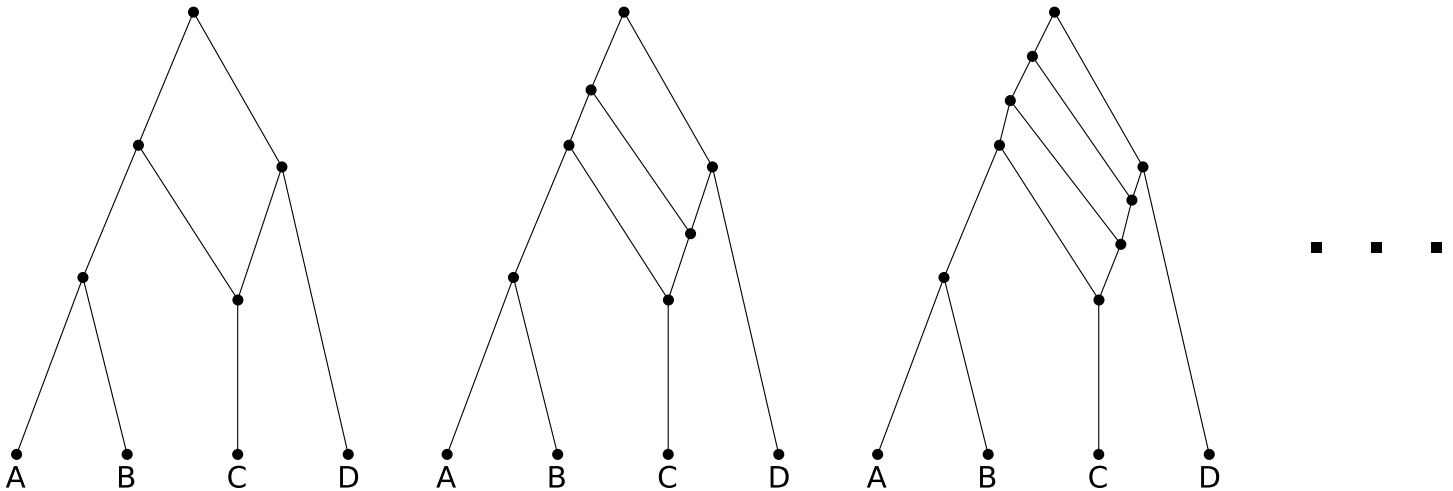


- Recall that there are $(2n-3)!!$ rooted trees with n leaves.

How many networks are there?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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



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

Sexual reproduction and genetic linkage

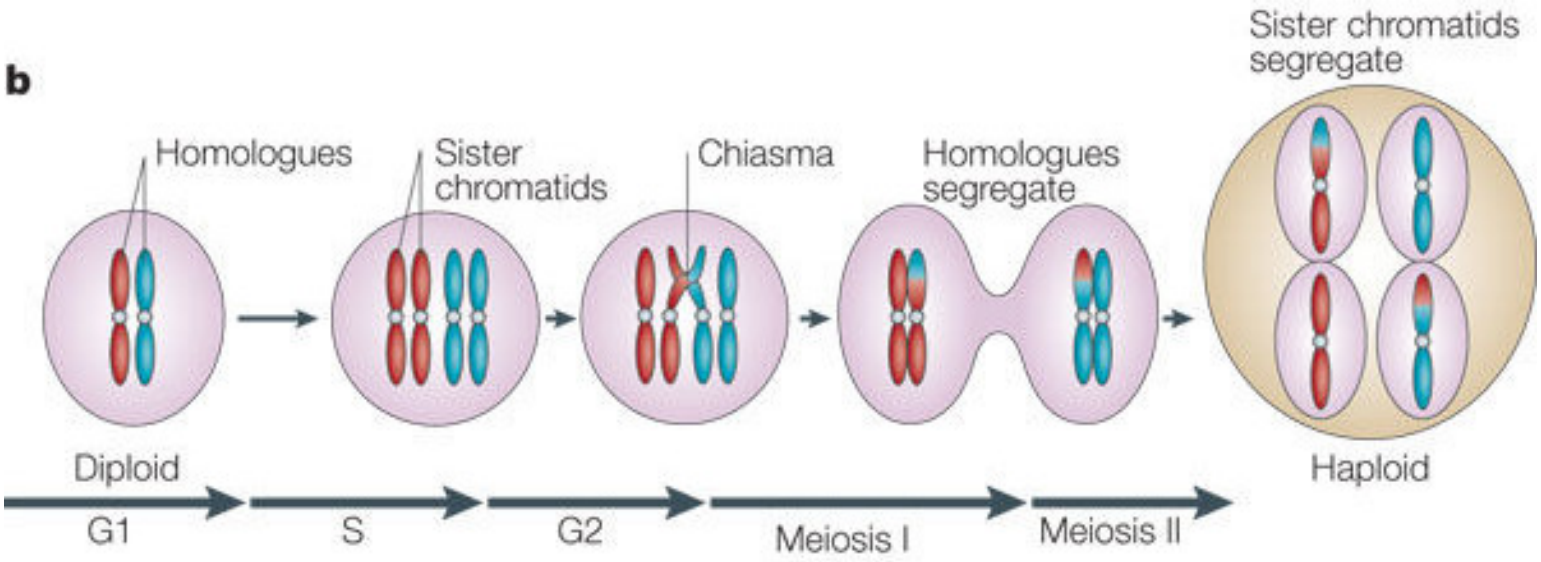


Figure adapted from [Marston and Amon, 2004]

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Sexual reproduction and genetic linkage

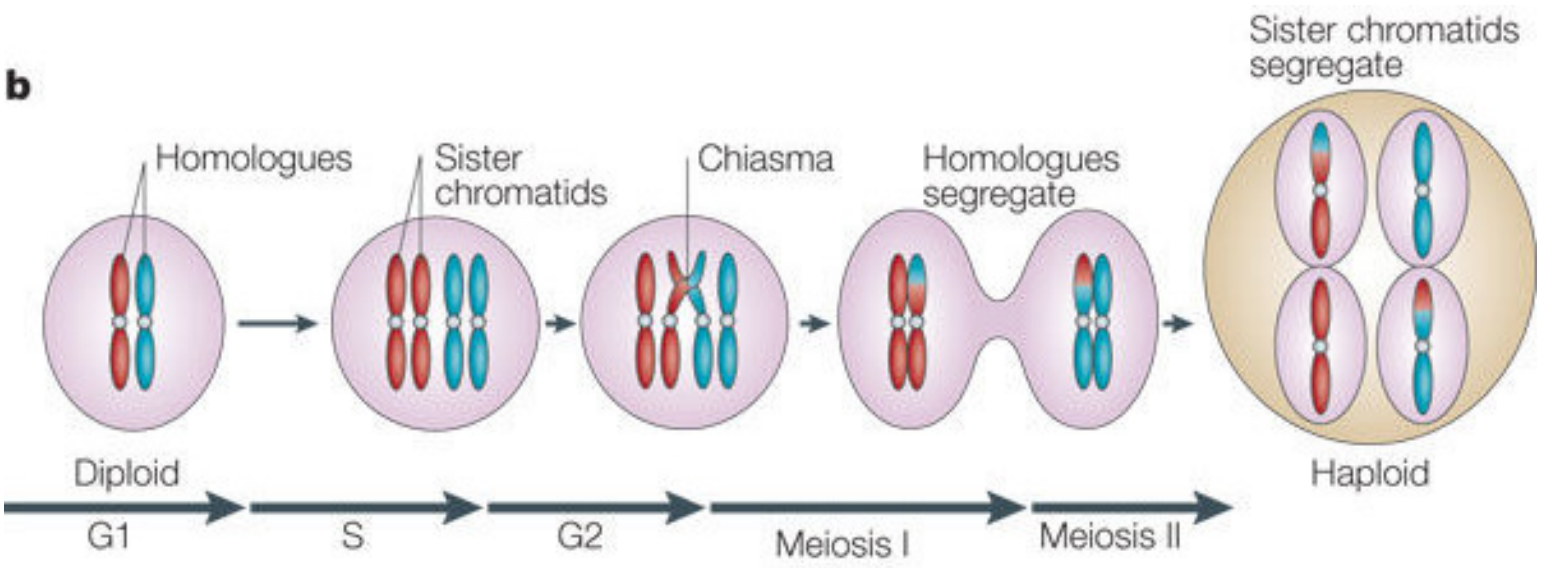


Figure adapted from [Marston and Amon, 2004]

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Sexual reproduction and genetic linkage

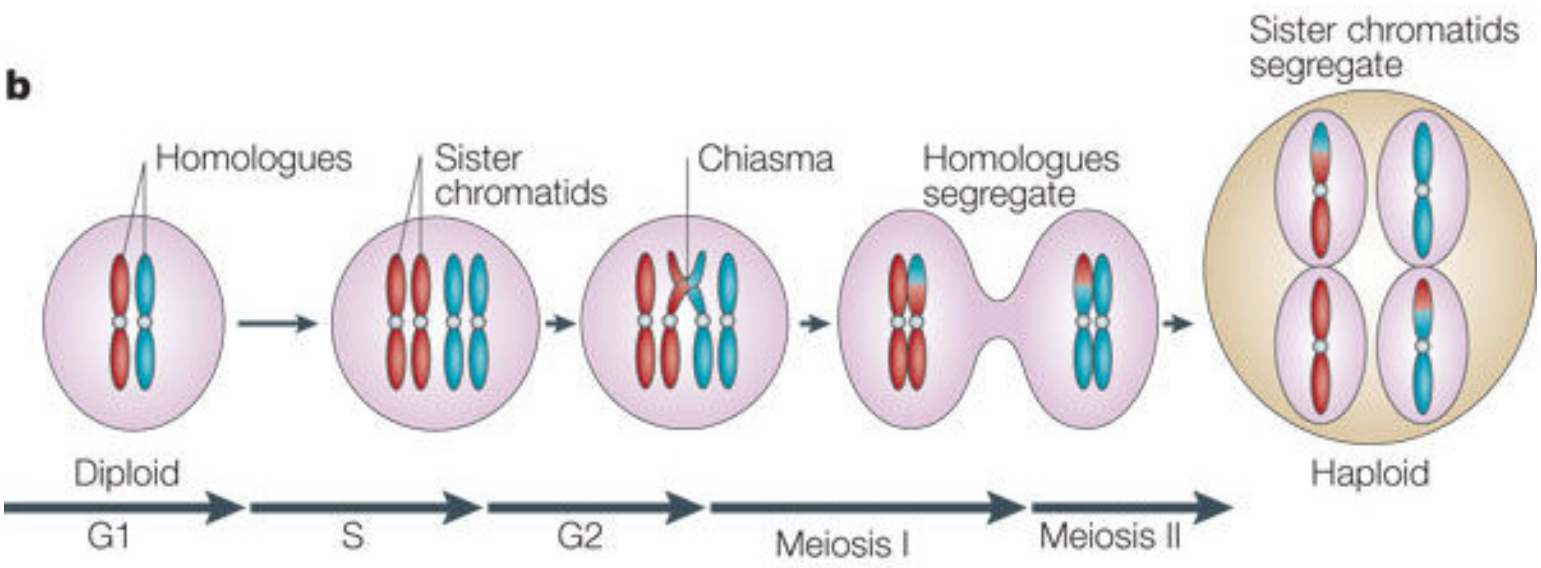


Figure adapted from [Marston and Amon, 2004]

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Sexual reproduction and genetic linkage

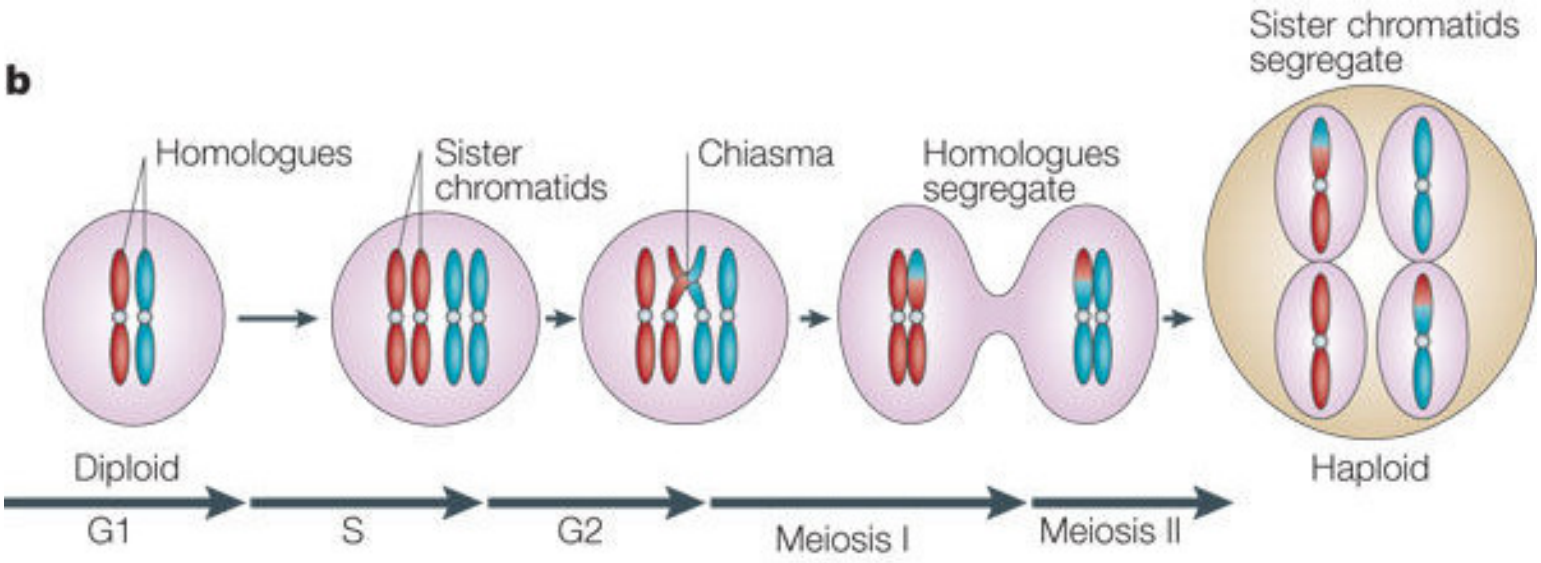
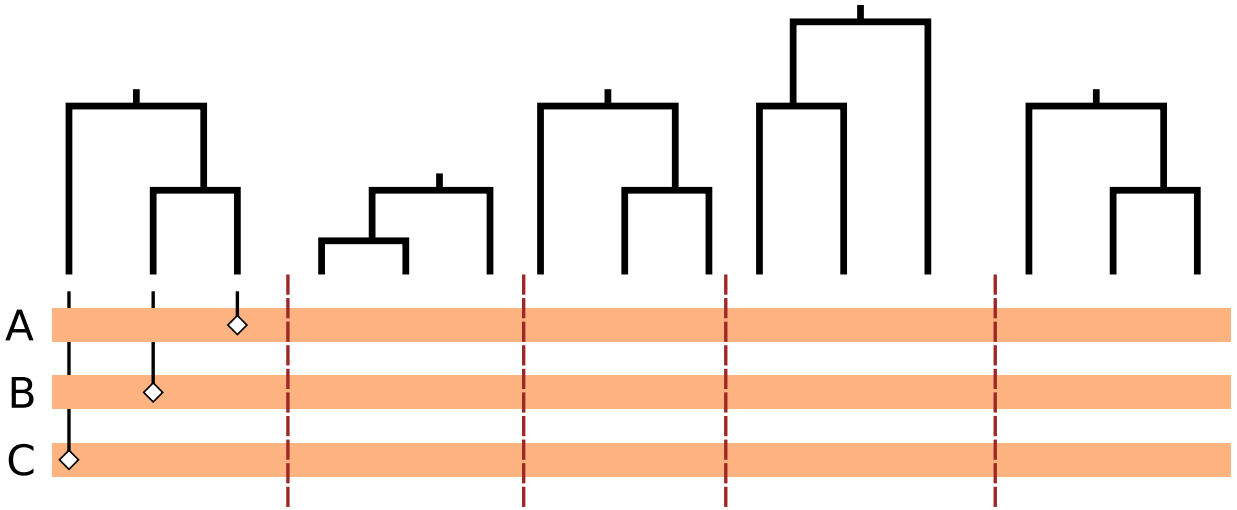


Figure adapted from [Marston and Amon, 2004]

- 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.

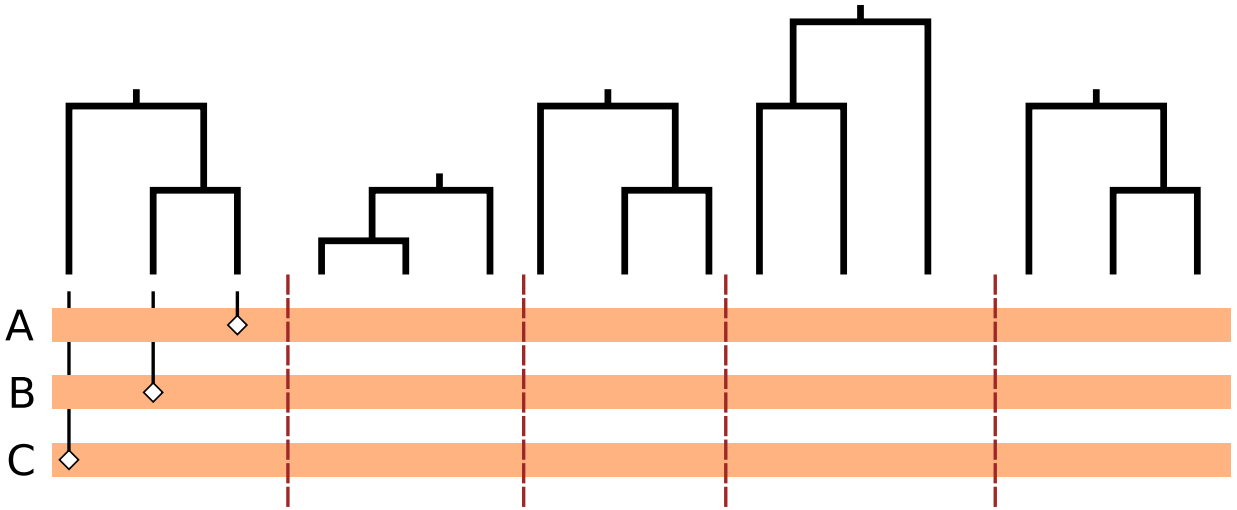
Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Effect of recombination on phylogeny



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

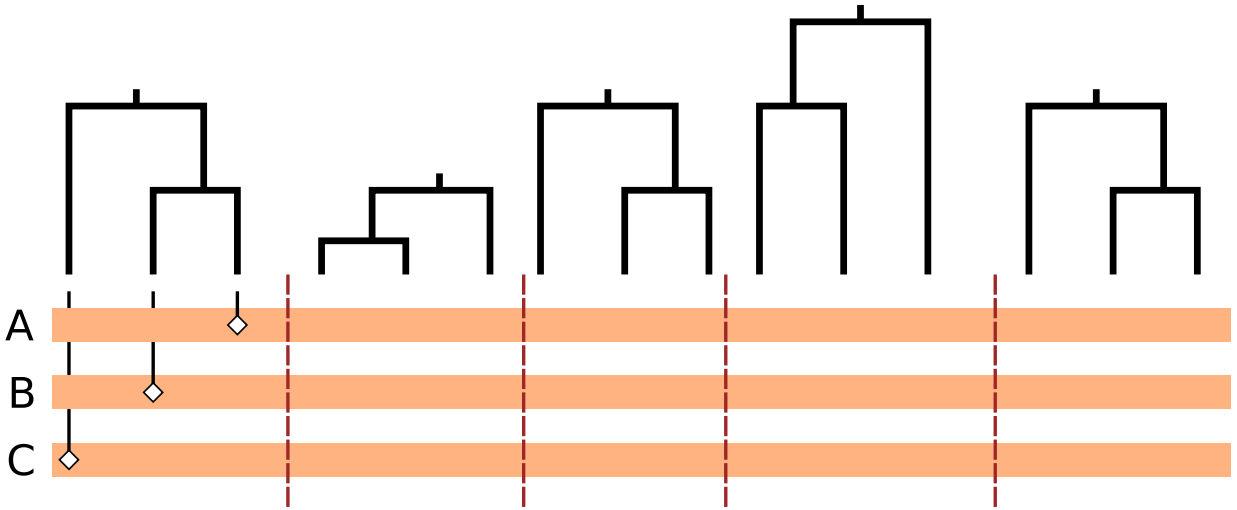
Effect of recombination on phylogeny



- Different sites correspond to different trees.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

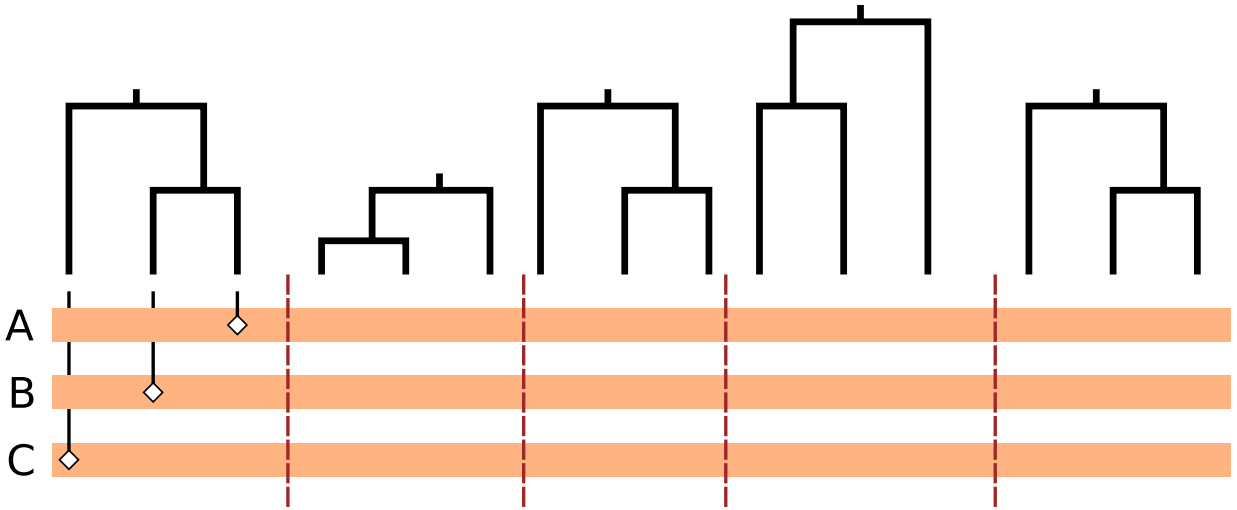
Effect of recombination on phylogeny



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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

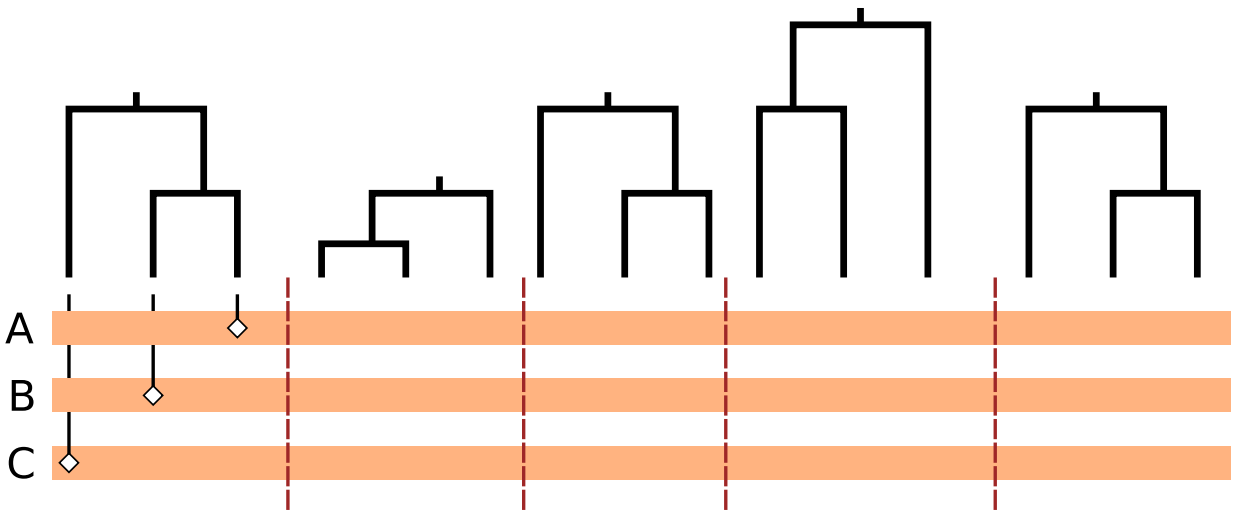
Effect of recombination on phylogeny



- 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

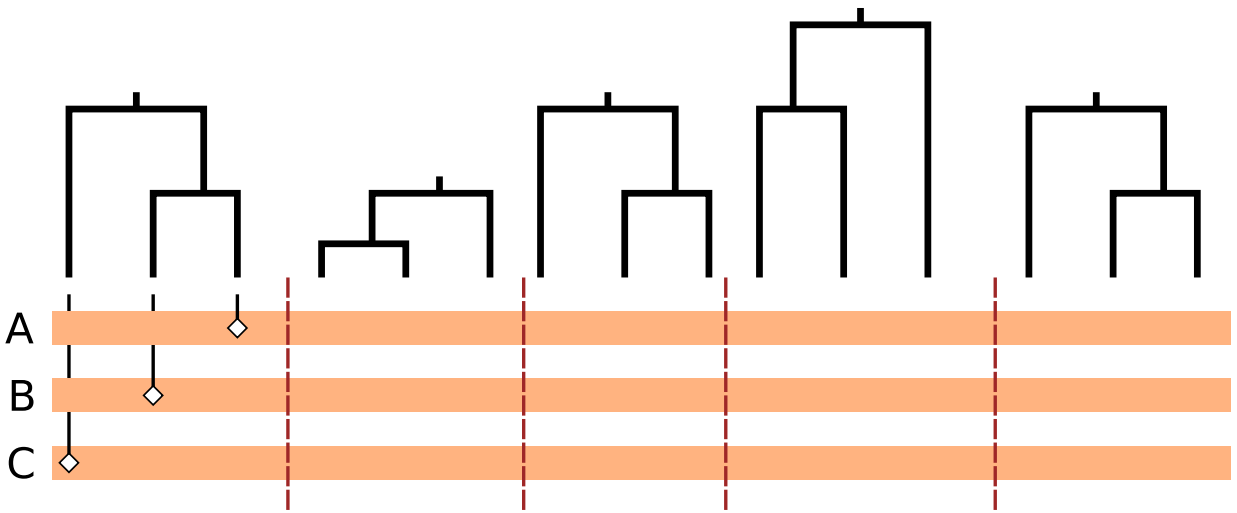
Effect of recombination on phylogeny



- 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).

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Wright-Fisher with Recombination

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

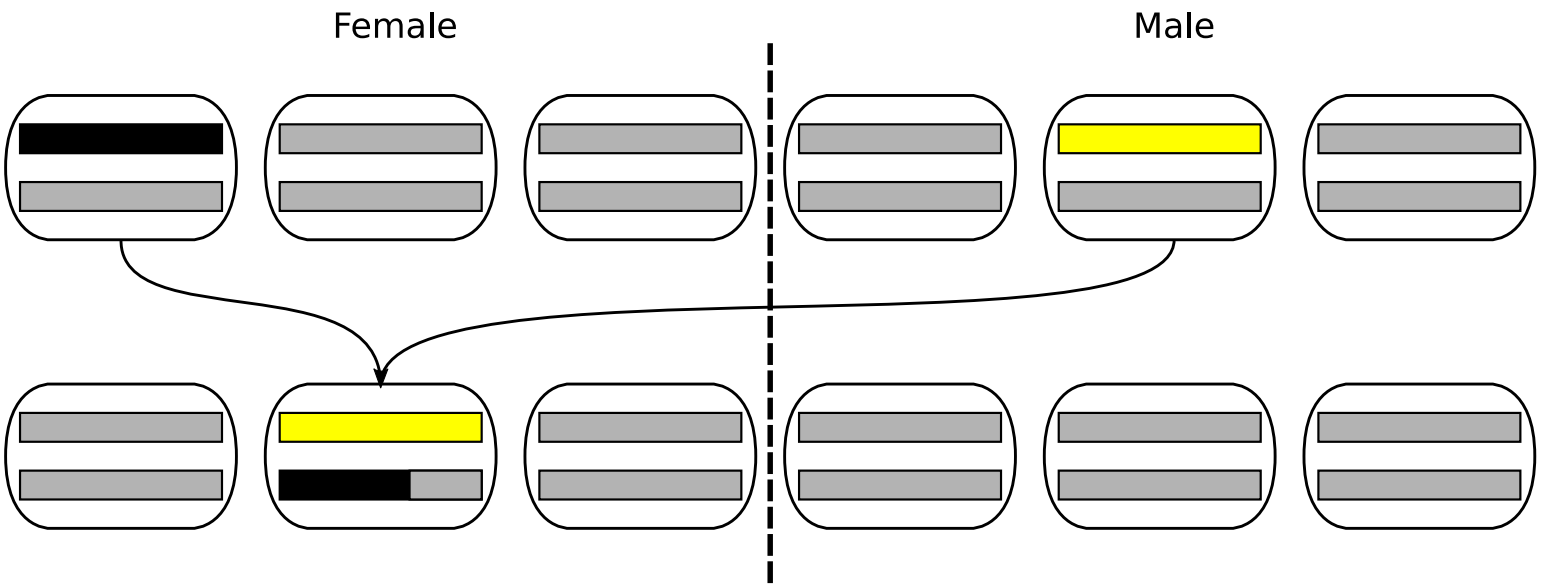
Wright-Fisher with Recombination

- 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Wright-Fisher with Recombination

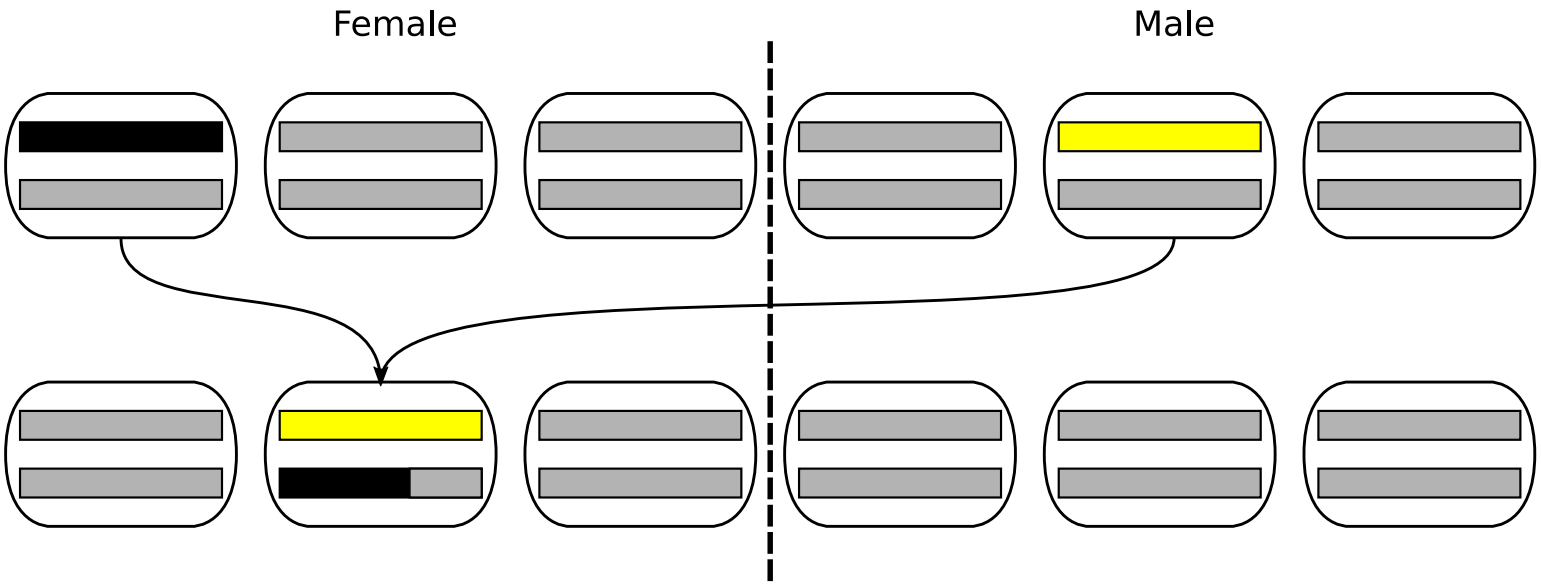
- Consider a Wright-Fisher population with female and male diploid individuals.
- Focus on a small segment of a single autosome.
  - An autosome is a chromosome which is a member of a homologous pair, i.e. not a sex chromosome.



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

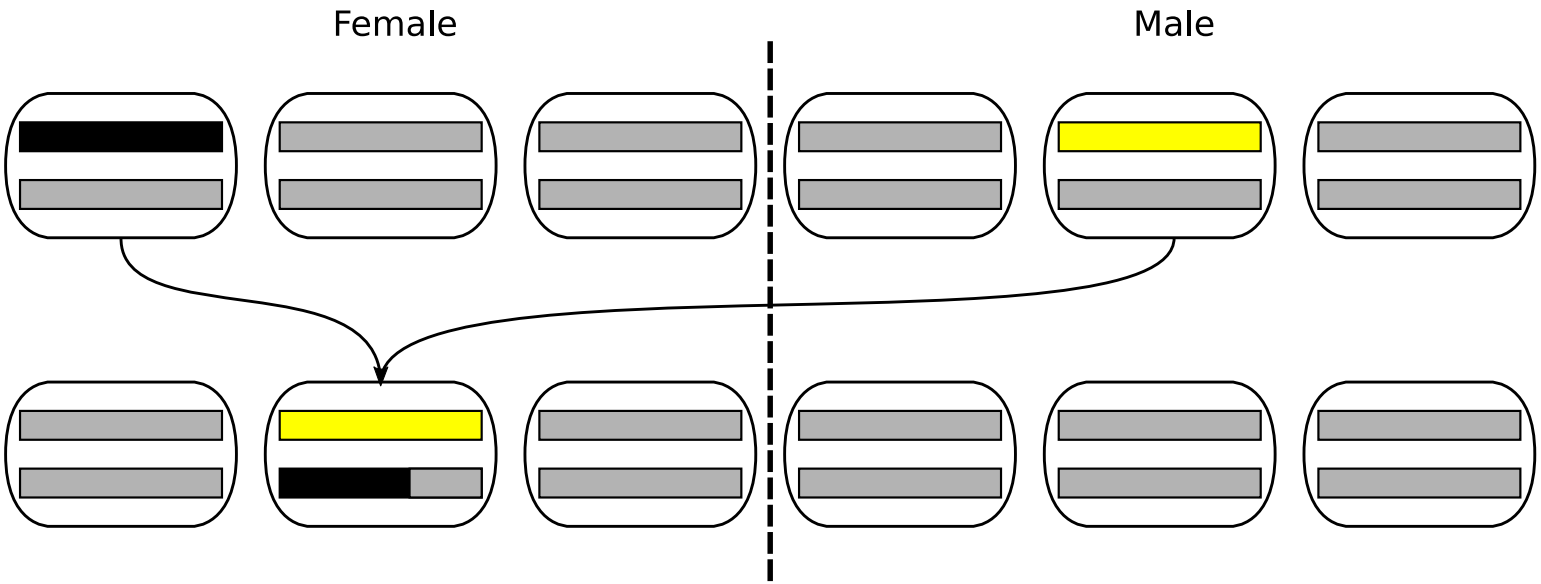


- Each child selects 1 male and 1 female parent randomly from the previous generation.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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 recombined.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Wright-Fisher with Recombination

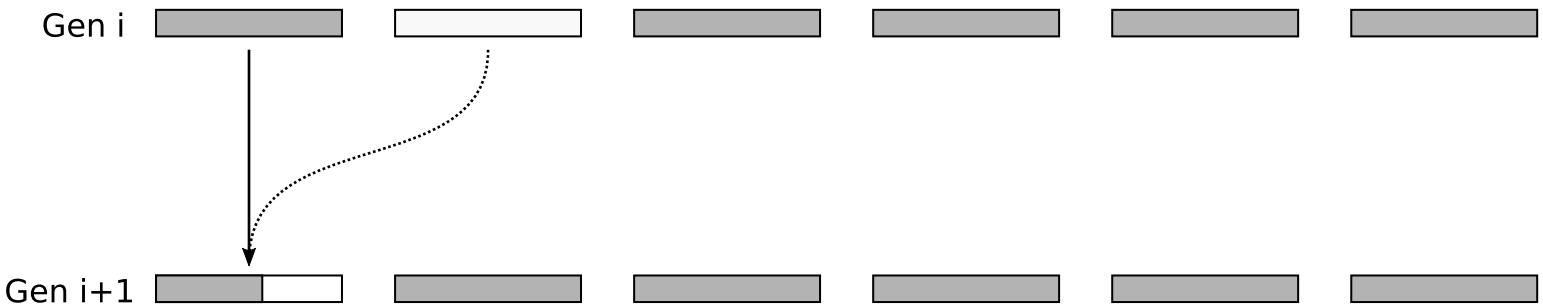
- Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Wright-Fisher with Recombination

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

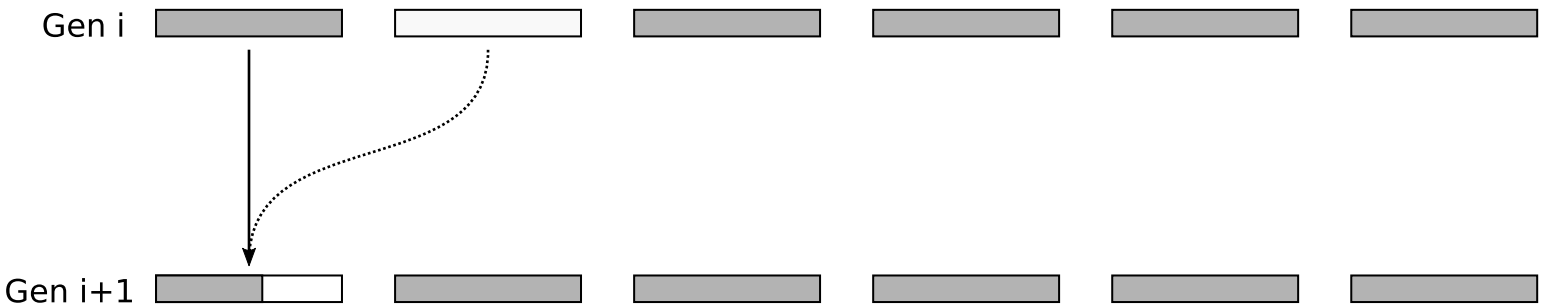
- Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



Wright-Fisher with Recombination

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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

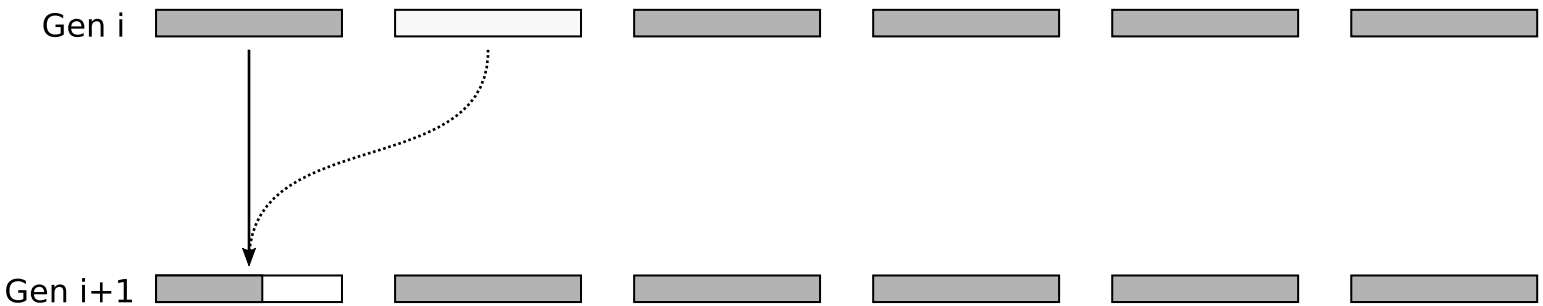


- Each child in $\mathbf{i}+1$ selects a parent at random from gen. $\mathbf{i}$.

Wright-Fisher with Recombination

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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

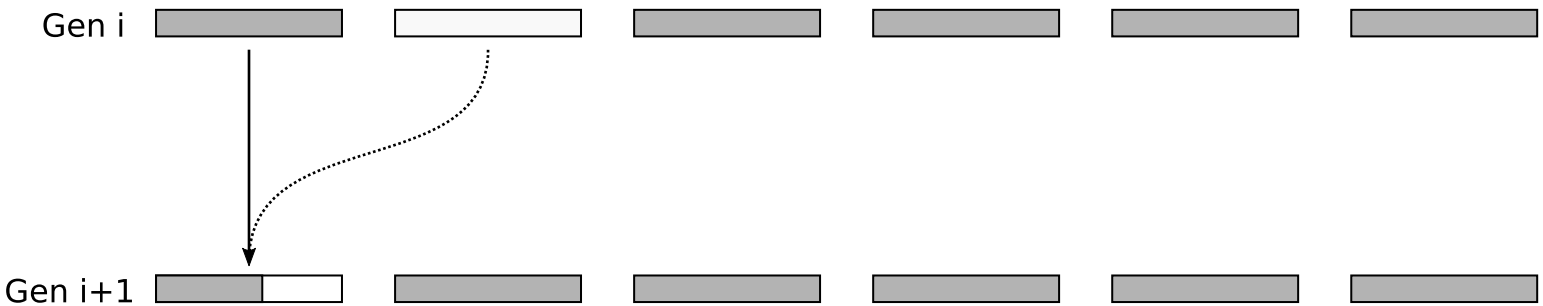


- Each child in $\mathbf{i}+1$ selects a parent at random from gen. $\mathbf{i}$.
- With probability $\mathbf{r}$ an additional parent is selected.

Wright-Fisher with Recombination

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

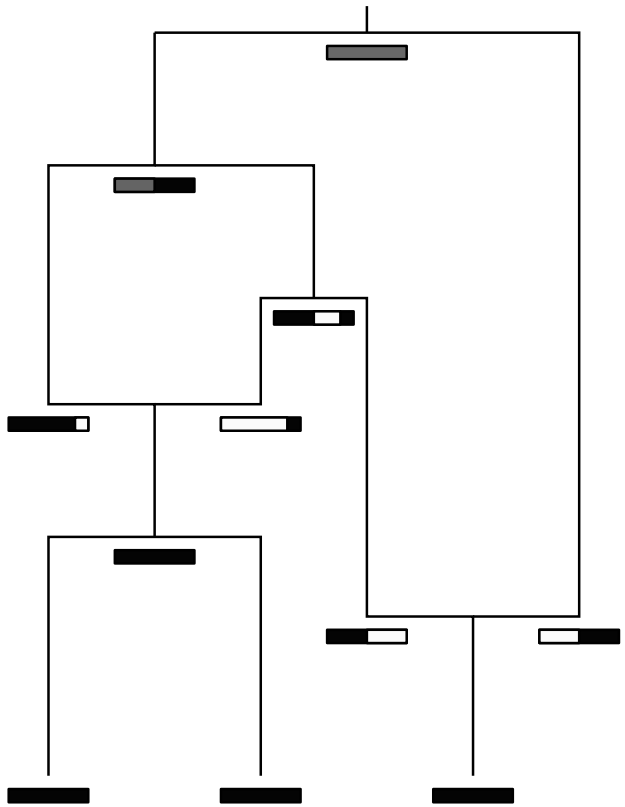
- Since the specific pairing of chromosomes only matters over a single generation, in the long term the haploid approximation is very good:



- Each child in $\mathbf{i}+1$ selects a parent at random from gen. $\mathbf{i}$.
- With probability $\mathbf{r}$ 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.

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]:

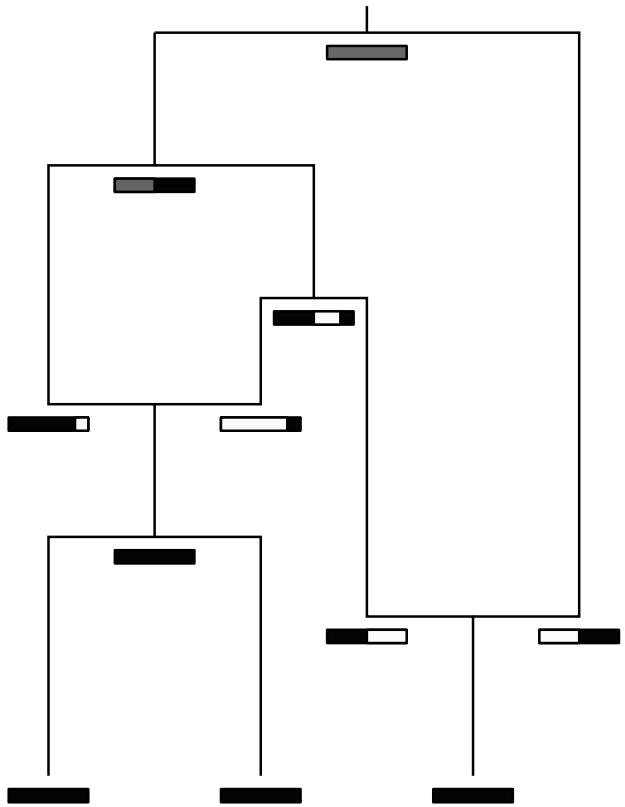


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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

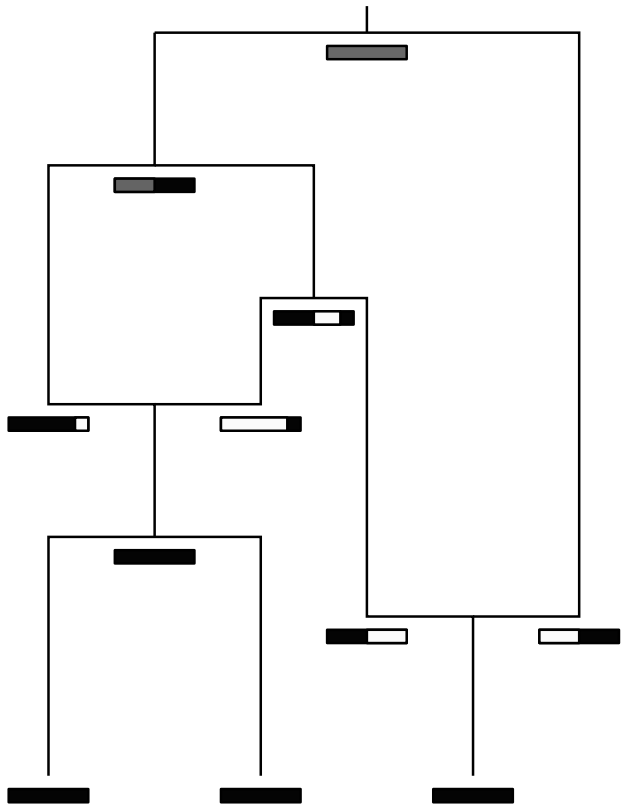


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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:** ρk.

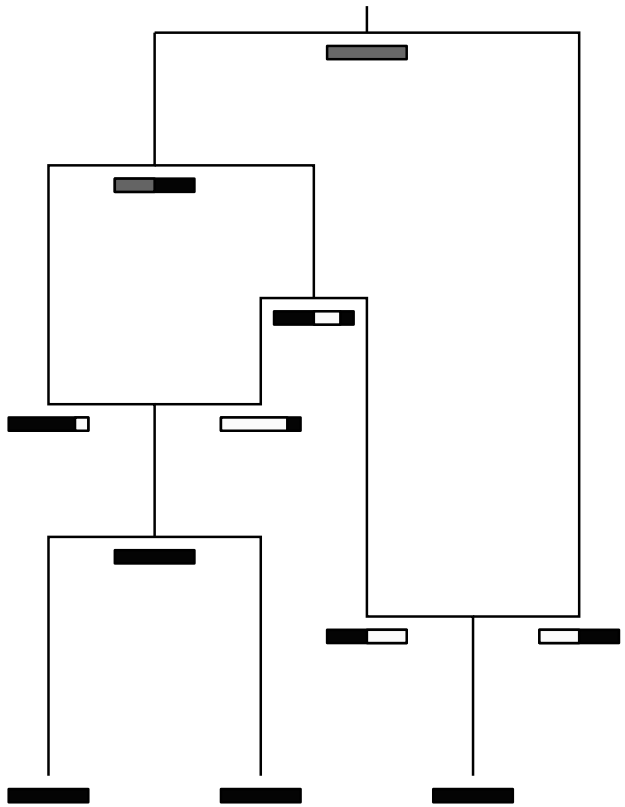


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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:** ρk.
- **Recombination break points:** chosen randomly along sequence: one parent contributes everything to the left, the other everything to the right.

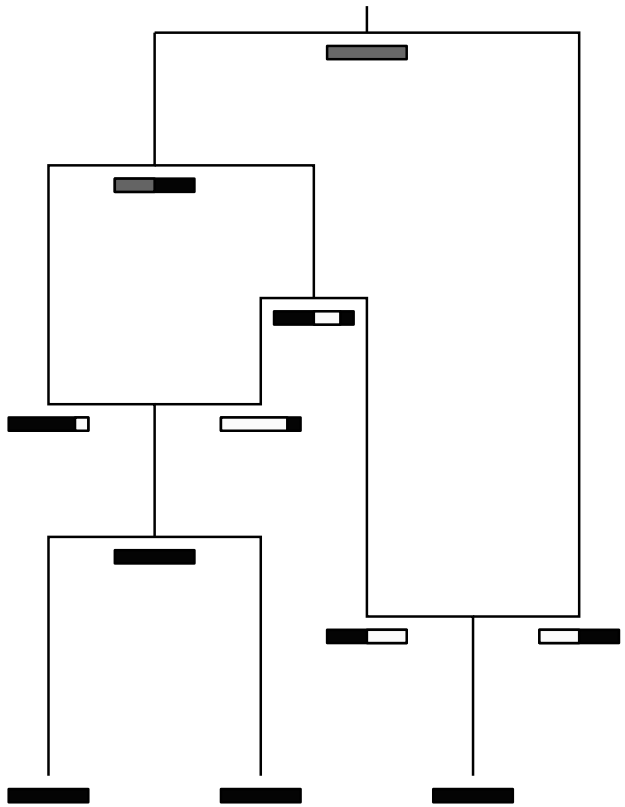


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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:** ρk.
- **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.

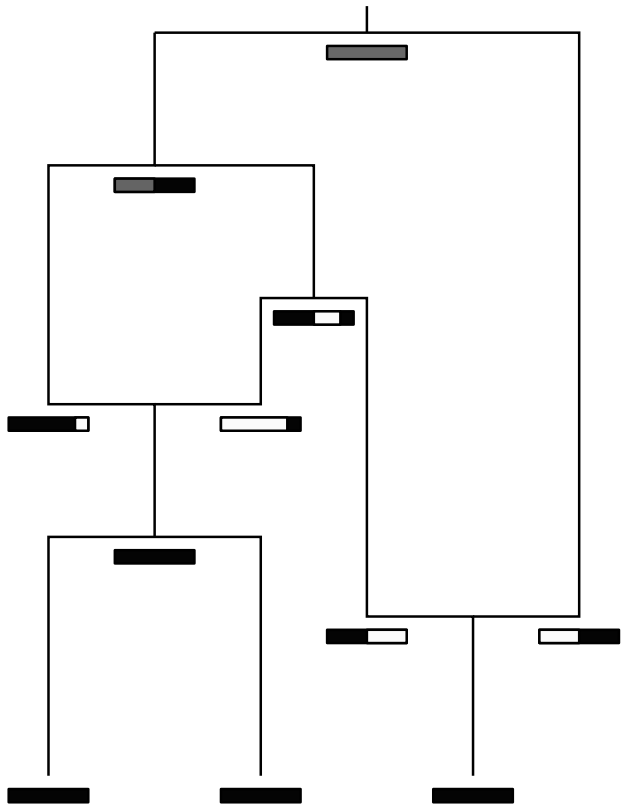


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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:** ρk.
- **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 (grey sites) before **grand (G)MRCA** of the process.

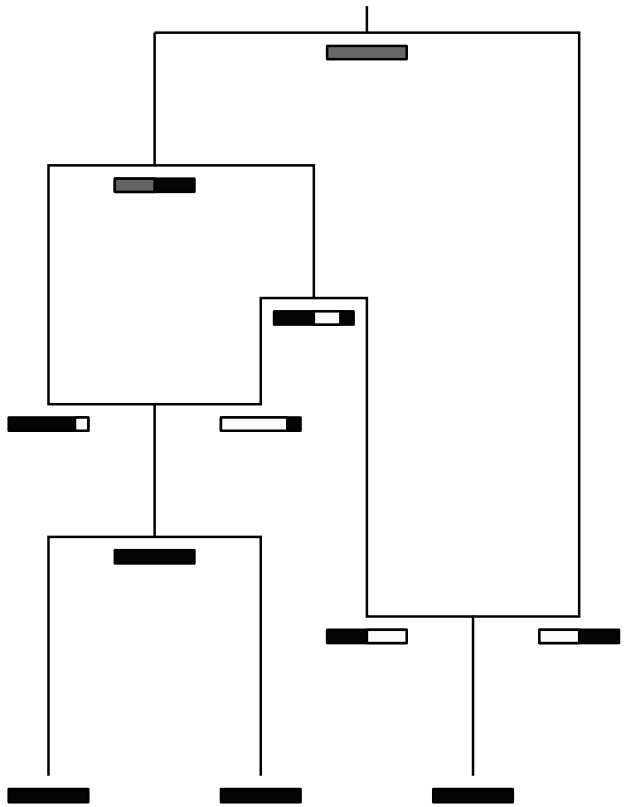


Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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:** ρk.
- **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 (grey sites) before **grand (G)MRCA** of the process.



The result is the "ancestral recombination graph" or ARG.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

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

- G is the recombination graph/network.
- Q is the substitution rate matrix.
- ρ is the recombination rate.
- N is the effective population size.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

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

- G is the recombination graph/network.
- Q is the substitution rate matrix.
- ρ 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"),

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

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

- G is the recombination graph/network.
- Q is the substitution rate matrix.
- ρ 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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

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

- G is the recombination graph/network.
- Q is the substitution rate matrix.
- ρ 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Bayesian phylogenetic network inference

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

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

- G is the recombination graph/network.
- Q is the substitution rate matrix.
- ρ 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Algorithms for Bayesian ARG inference

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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).

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Algorithms for Bayesian ARG inference

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

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.

Bayesian ARG inference example

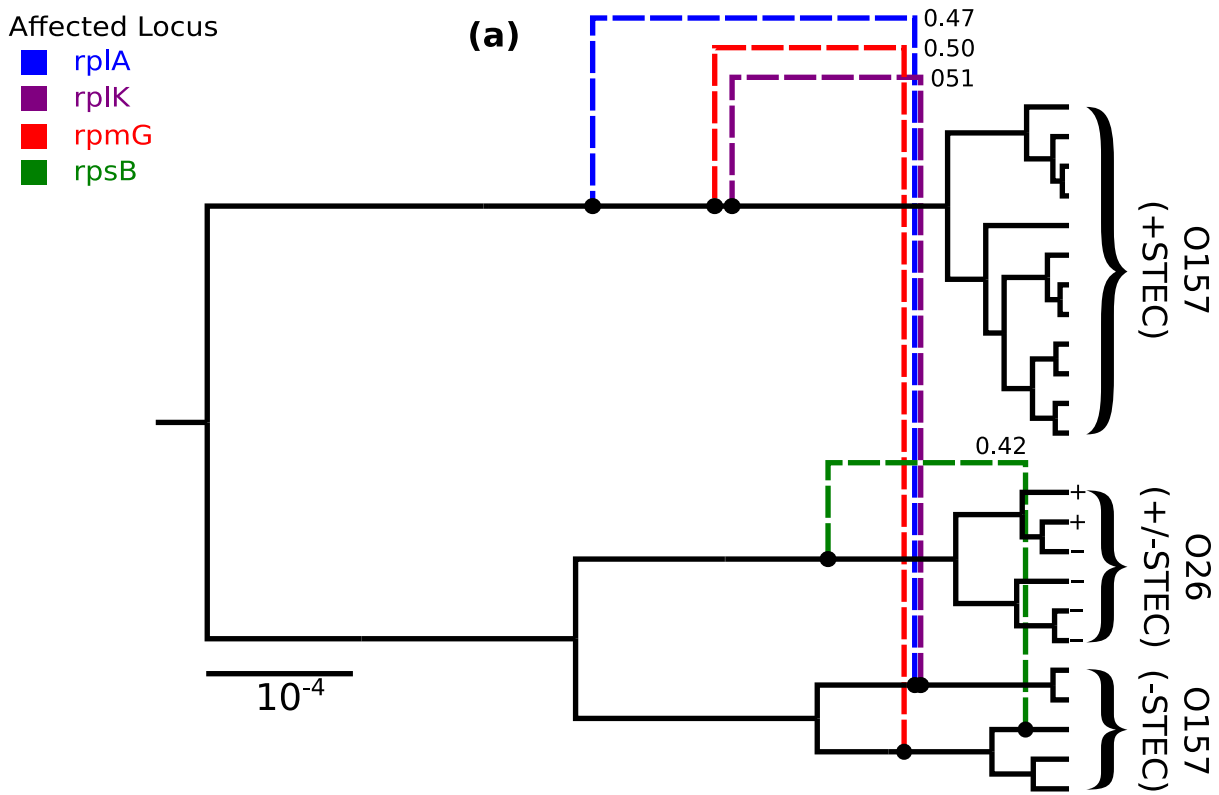
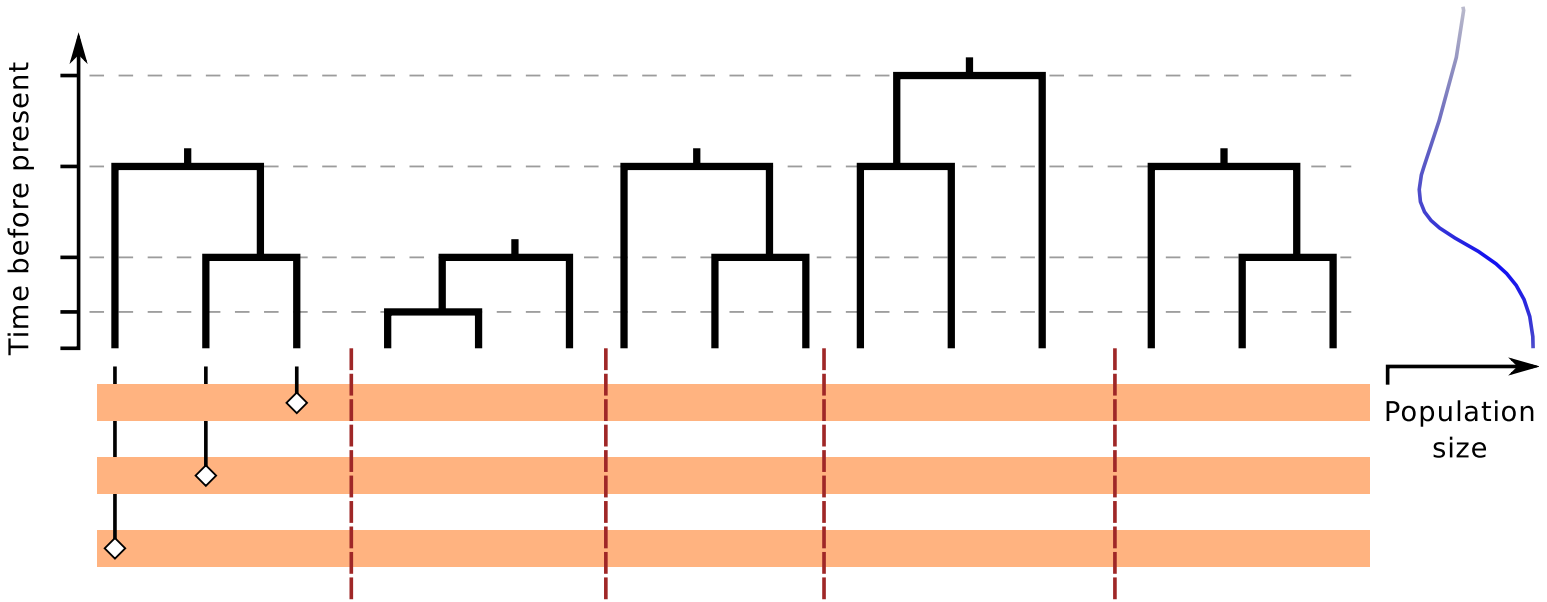


Figure adapted from [Vaughan et al., 2017]

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

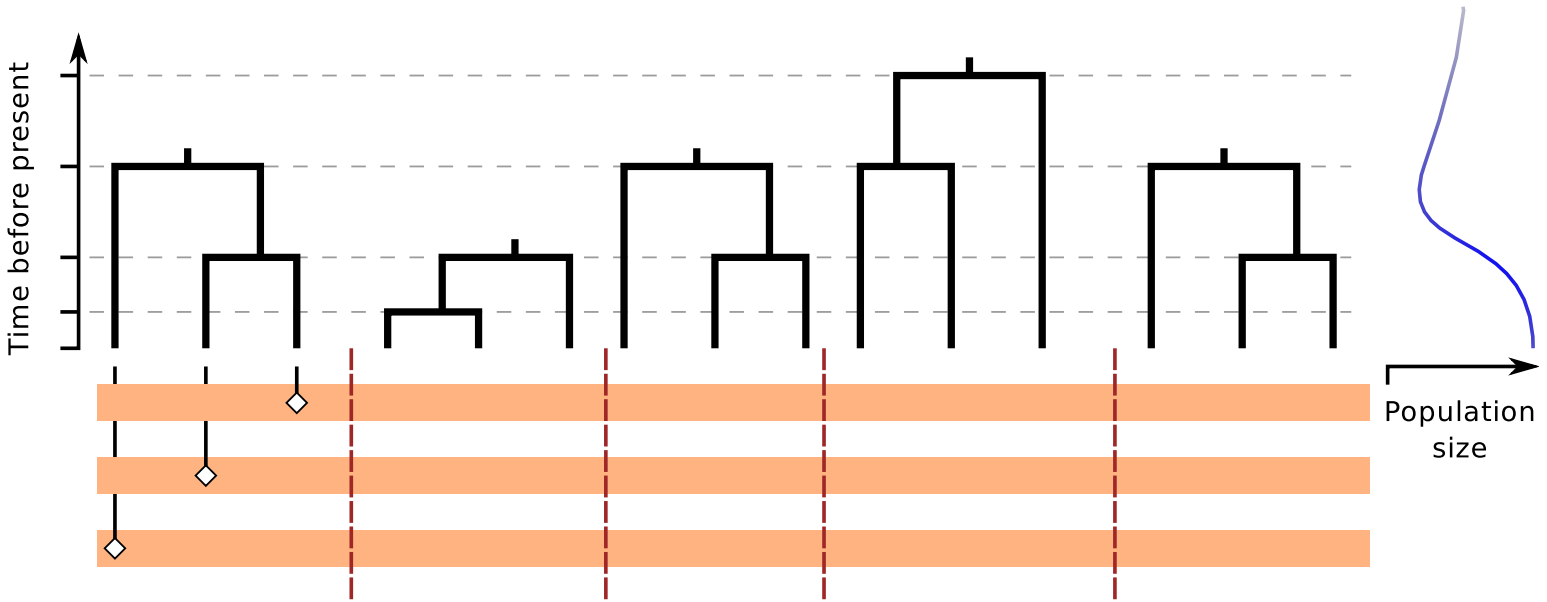
Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Inference of population dynamics



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

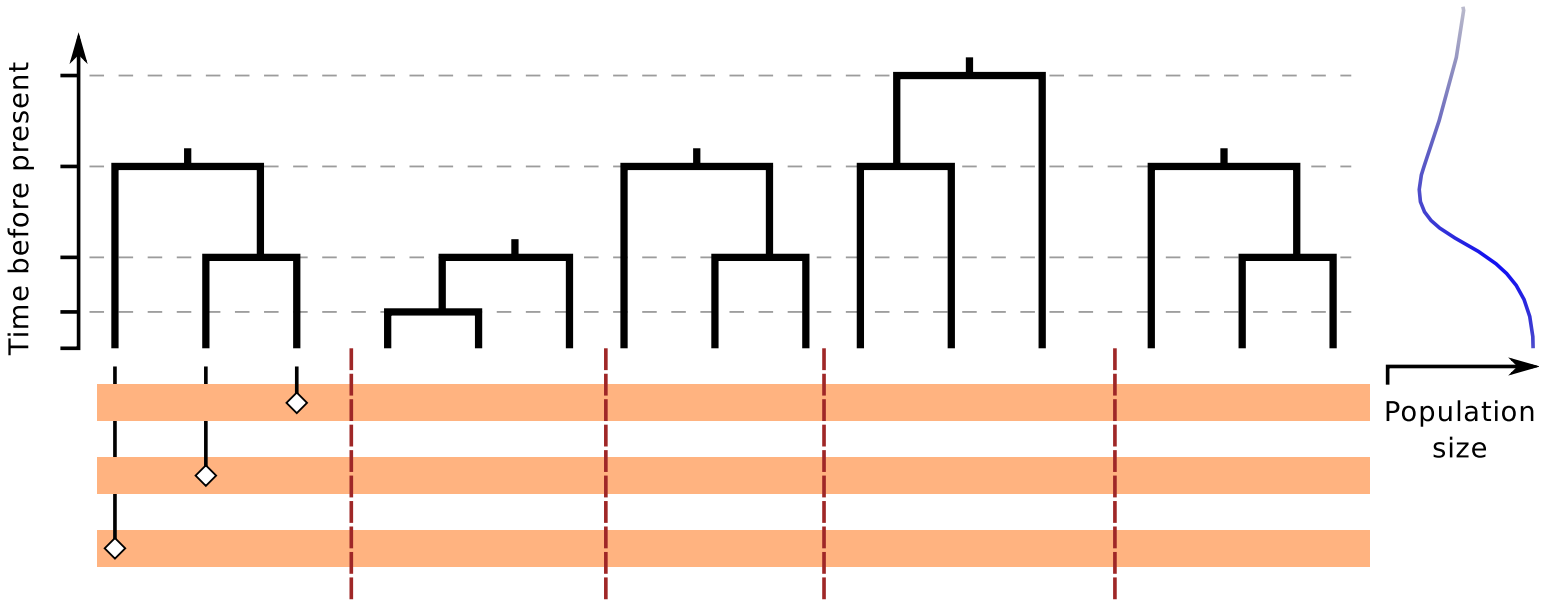
Inference of population dynamics



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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

The Sequentially Markovian Coalescent

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

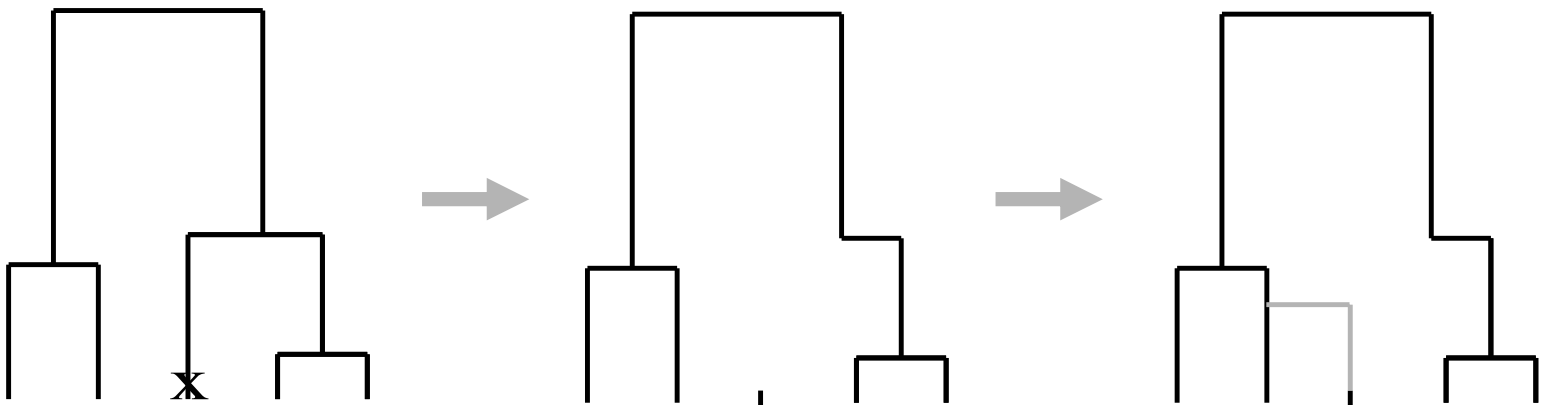


Figure adapted from [McVean and Cardin, 2005]

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

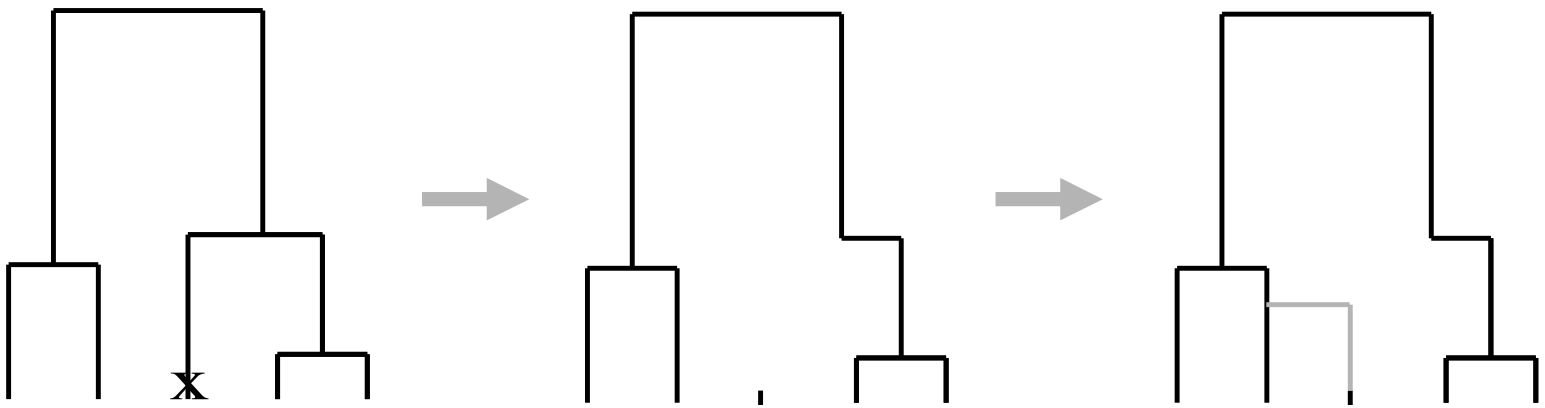


Figure adapted from [McVean and Cardin, 2005]

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

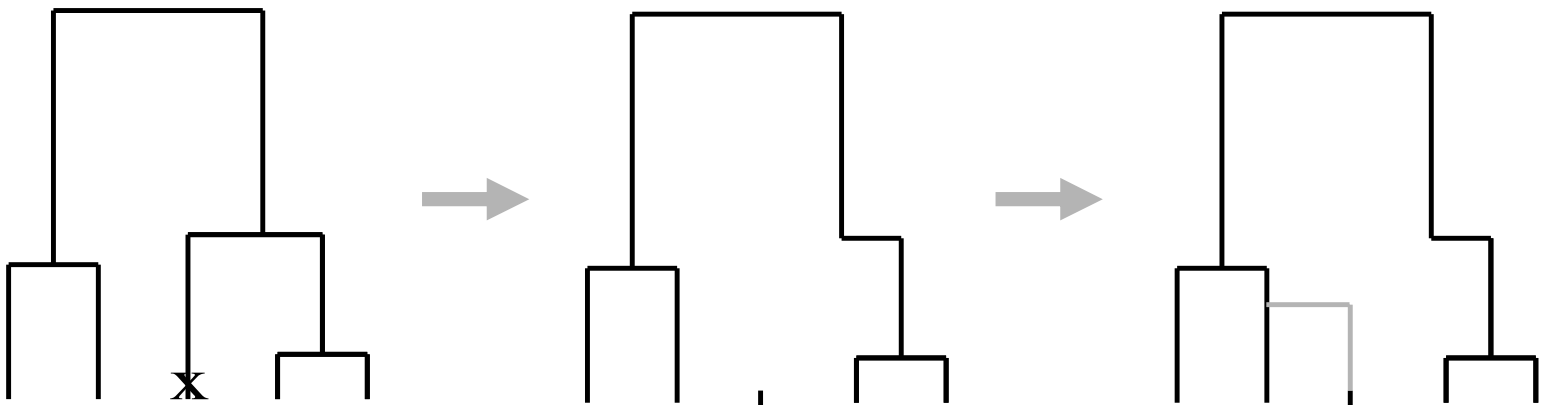


Figure adapted from [McVean and Cardin, 2005]

- Sites between breakpoints exponentially distributed with rate ρL 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Inference of human population dynamics

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
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.

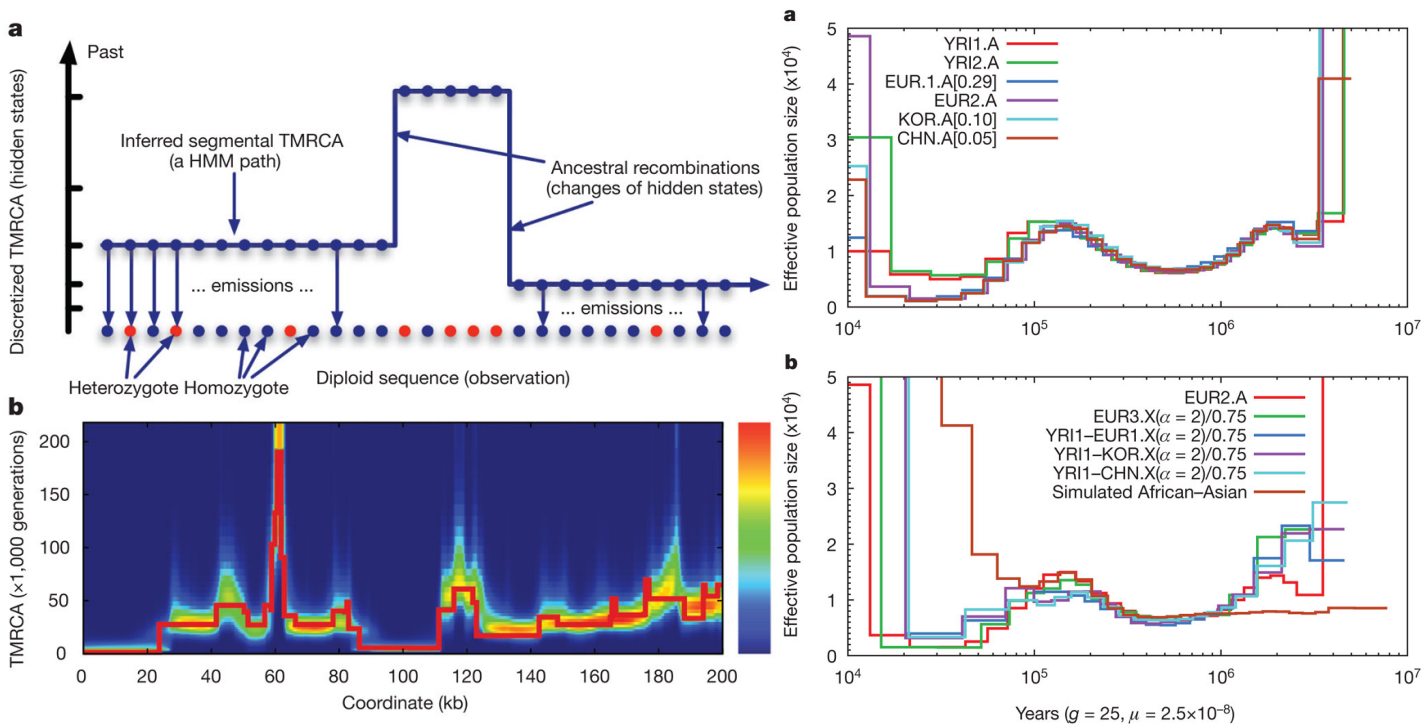


Figure adapted from [Li and Durbin, 2011]

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species tree models: Multi-species coalescent

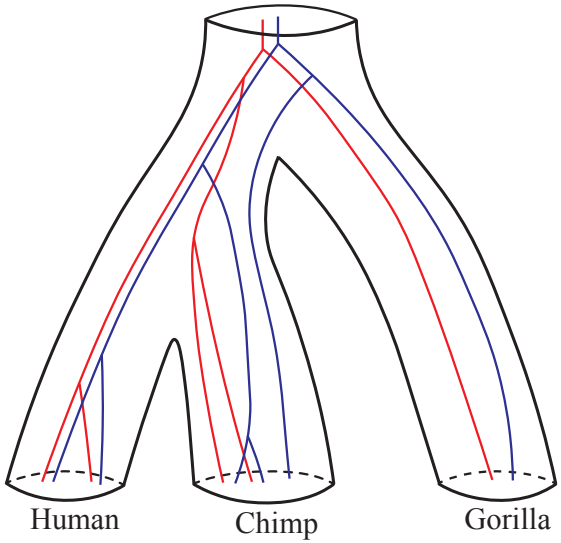


Figure adapted from [Zhang et al., 2017]

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species tree models: Multi-species coalescent

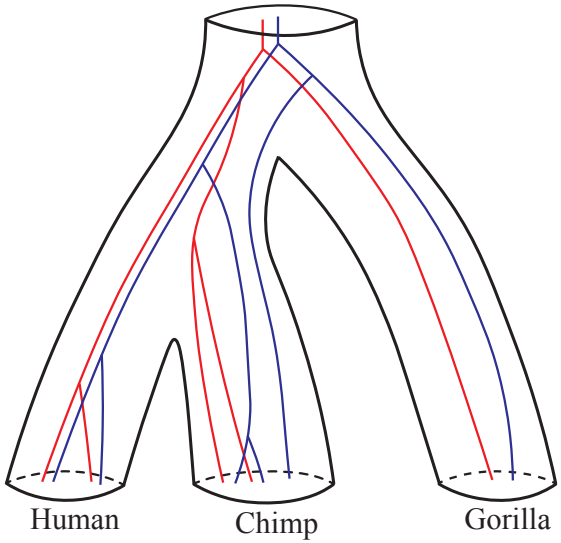


Figure adapted from [Zhang et al., 2017]

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species tree models: Multi-species coalescent

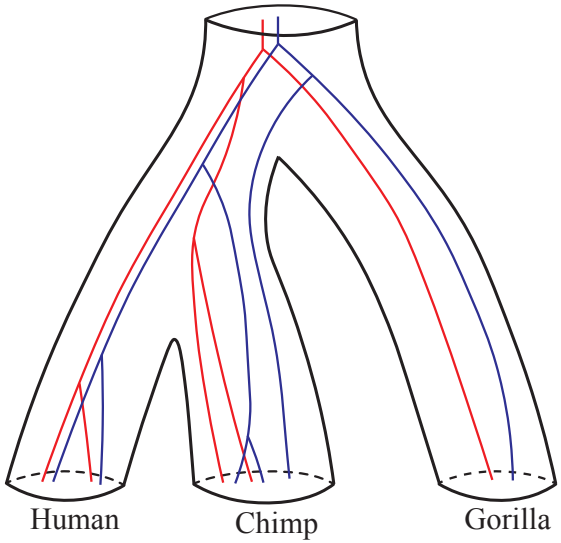


Figure adapted from [Zhang et al., 2017]

- 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species tree models: Multi-species coalescent

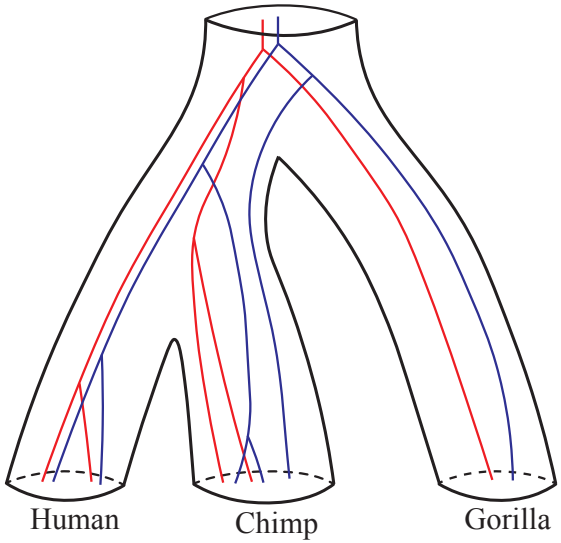


Figure adapted from [Zhang et al., 2017]

- 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**.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species tree models: Multi-species coalescent

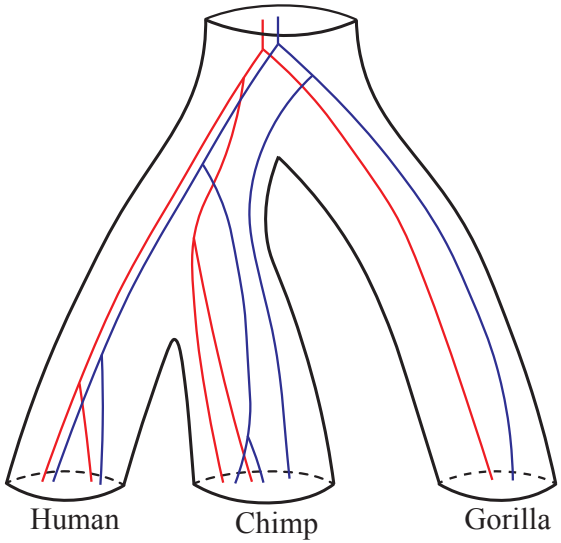


Figure adapted from [Zhang et al., 2017]

- 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].

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species network inference

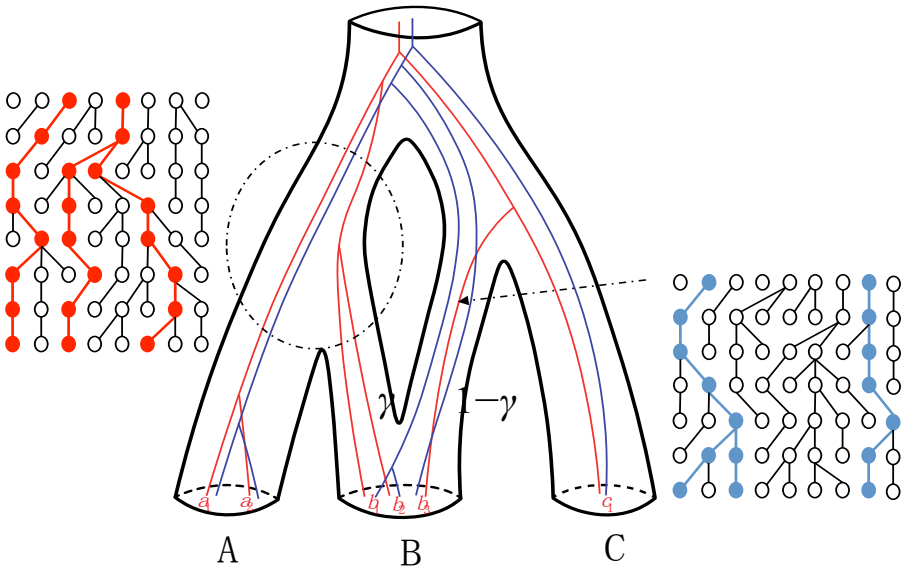


Figure adapted from [Zhang et al., 2017]

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species network inference

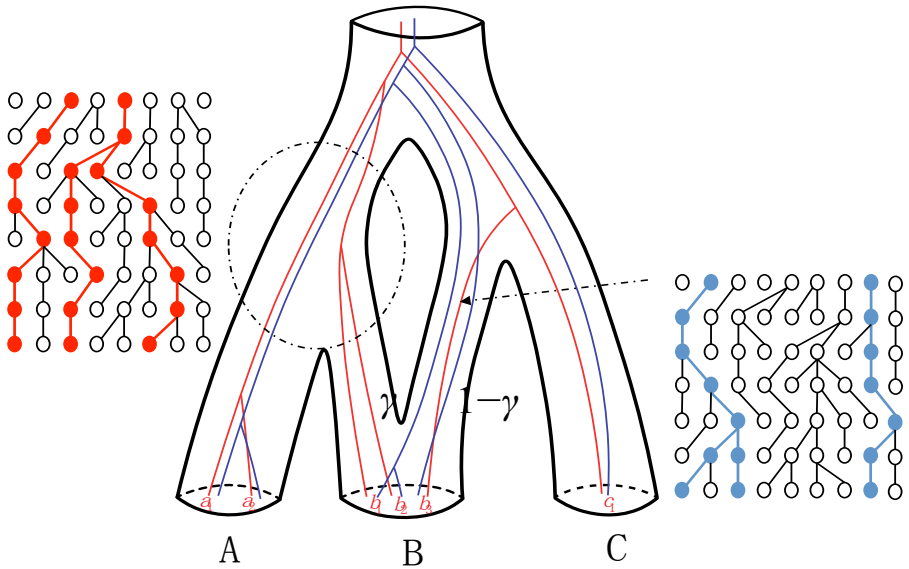


Figure adapted from [Zhang et al., 2017]

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

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Species network inference

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

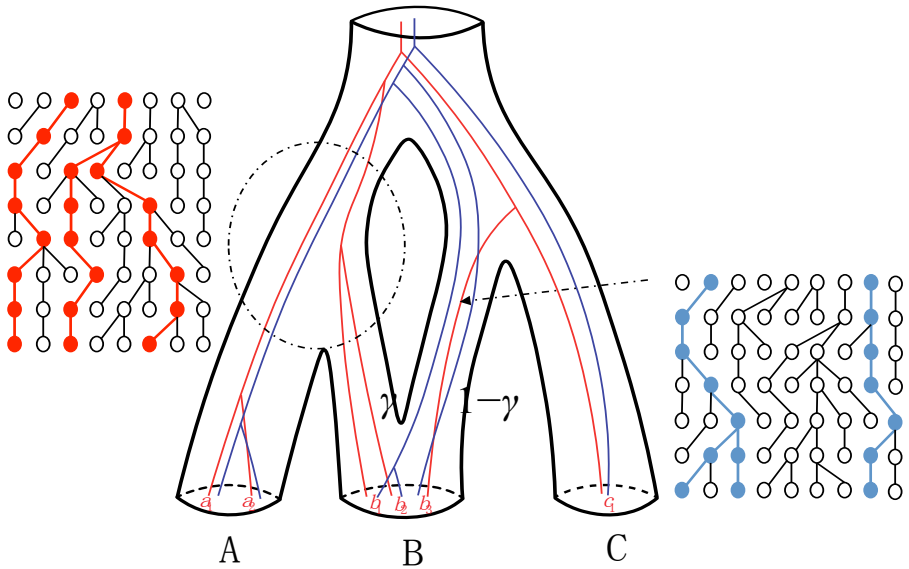


Figure adapted from [Zhang et al., 2017]

- 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.

Species network inference

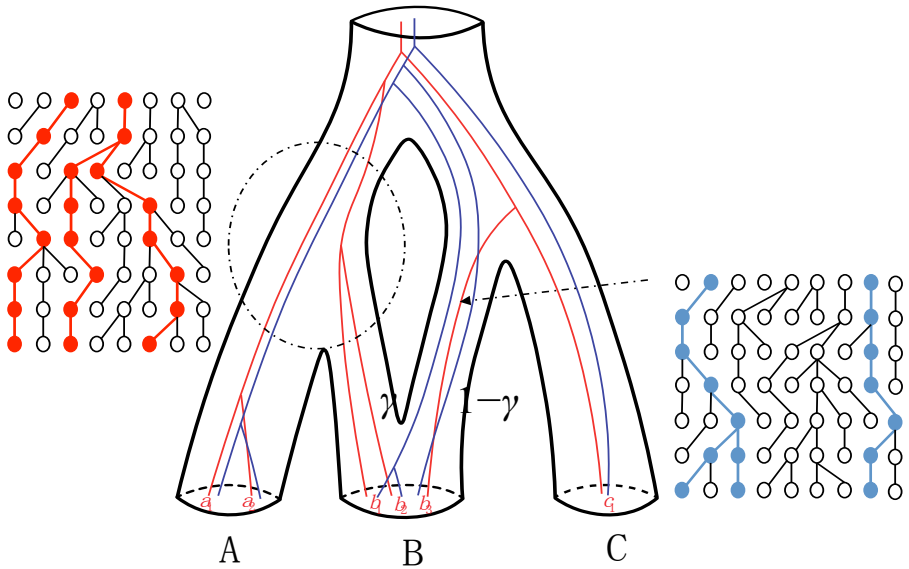
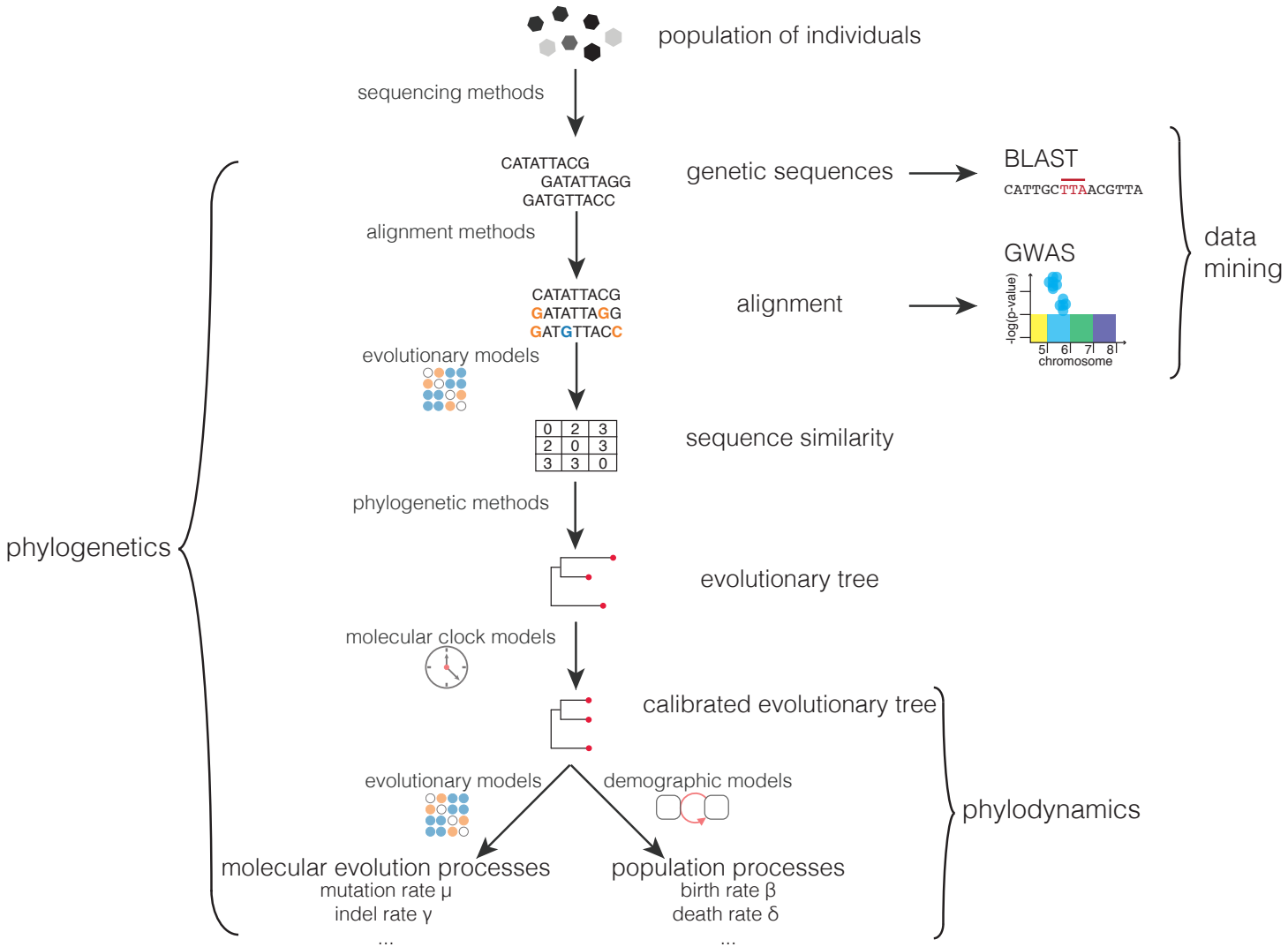


Figure adapted from [Zhang et al., 2017]

- 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

From sequences to population dynamics



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

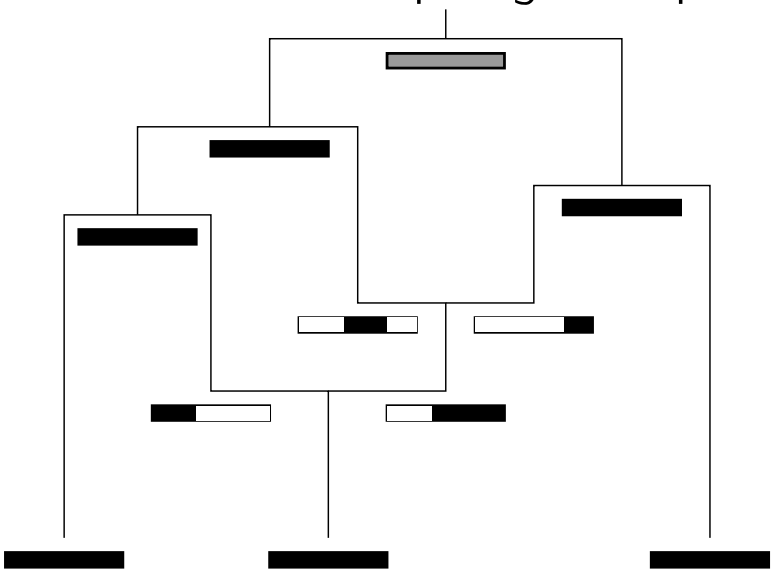
Questions

- What is the maximum number of local trees that can correspond to a sequence alignment?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Questions

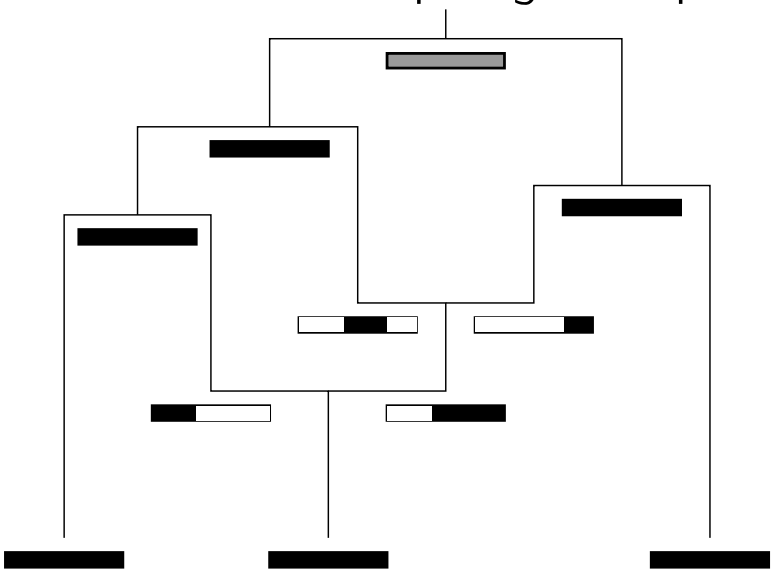
- What is the maximum number of local trees that can correspond to a sequence alignment?
- Draw the local topologies implied by the following ARG:



Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

Questions

- What is the maximum number of local trees that can correspond to a sequence alignment?
- Draw the local topologies implied by the following ARG:



- 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?

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References

References I

- Bloomquist, E. W. and Suchard, M. A. (2010). Unifying vertical and nonvertical evolution: a stochastic arg-based framework. *Syst Biol*, 59(1):27–41.
- Didelot, X., Lawson, D., Daarling, A., and Falush, D. (2010). Inference of homologous recombination in bacteria using whole-genome sequences. *Genetics*, 186:1435.
- Heled, J. and Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Mol Biol Evol*, 27(3):570–80.
- Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol*, 23(2):183–201.
- Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357):493–496.
- 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.
- 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.
- Rasmussen, M. D., Hubisz, M. J., Gronau, I., and Siepel, A. (2014). Genome-wide inference of ancestral recombination graphs. *PLoS genetics*, 10(5):e1004342.
- 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.

Phylogenetic Networks
Introduction to networks
Linkage and phylogeny
Wright-Fisher with Recombination
Coalescent with Recombination
Bayesian inference
Population dynamics inference
Species trees
Species networks
Finale
References