Construction of coalescent trees on partially fixed pedigrees

Renaud Alie, Sorana Froda & Fabrice Larribe

Université du Québec à Montréal

7 August 2018

Genealogies

- Φ : Phenotypes

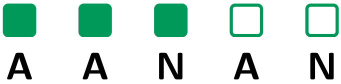
  A : Affected, **N** : Not Affected

- $\mathcal{H}$: Haplotypes

  ■ : Mutant, □ : Non-Mutant

- θ : Unknown parameter

$$\mathcal{L}(\theta) = P\left(\Phi, \mathcal{H} | \theta\right).$$



Genealogies

- Φ : Phenotypes

  A : Affected, **N** : Not Affected

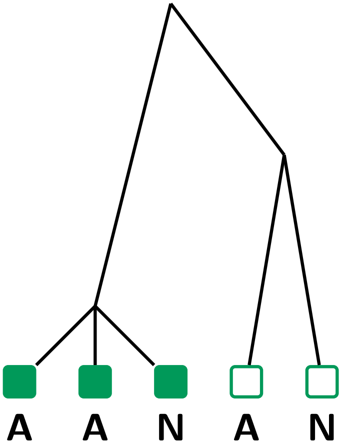
- $\mathcal{H}$: Haplotypes

  ■ : Mutant, □ : Non-Mutant

- θ : Unknown parameter

- $\mathcal{G}$: Genealogy

$$\begin{aligned}\mathcal{L}(\theta) &= P\left(\Phi, \mathcal{H} | \theta\right) \\ &= \int_{\mathcal{G}} P\left(\Phi, \mathcal{H} | \mathcal{G}, \theta\right) P\left(\mathcal{G} | \theta\right) \mathrm{d}\,\mathcal{G}.\end{aligned}$$

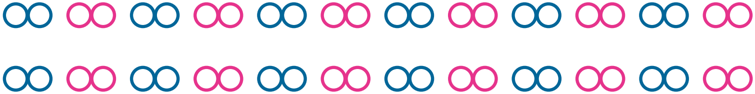


Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males; N females; $4N$ genes.

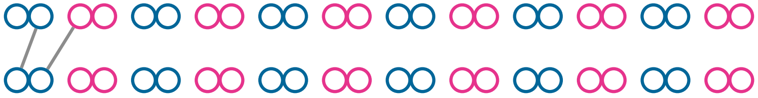
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males; N females; $4N$ genes.



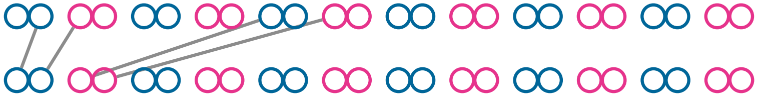
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



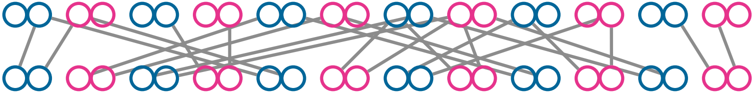
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



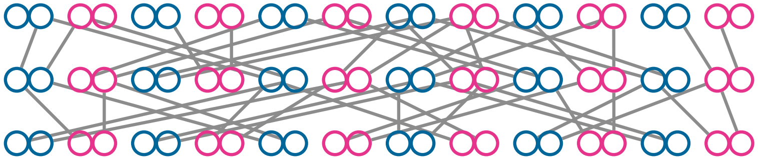
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



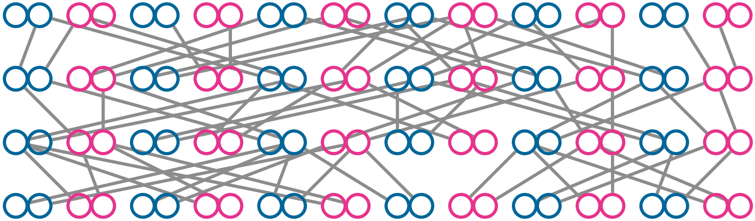
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



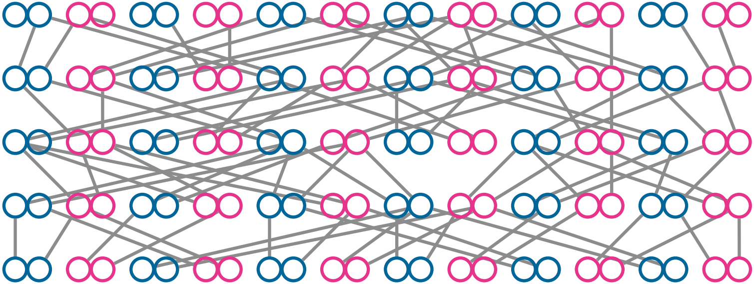
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



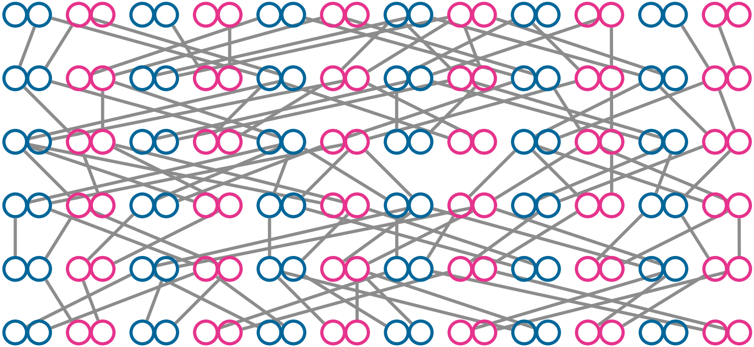
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



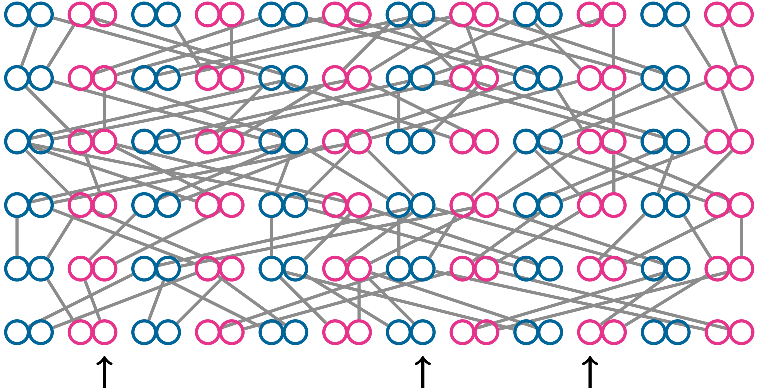
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males; N females; $4N$ genes.



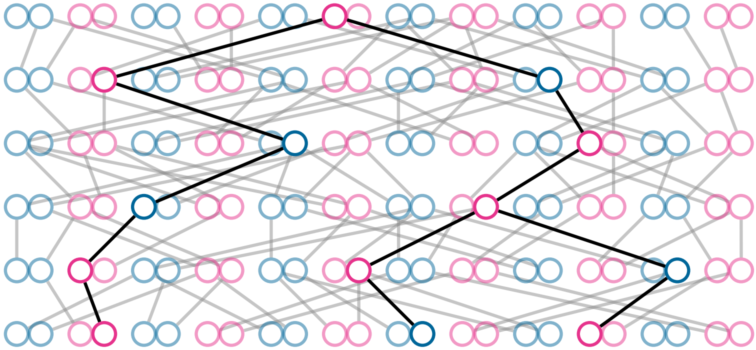
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ◯◯ ; N females ◯◯ ; $4N$ genes.



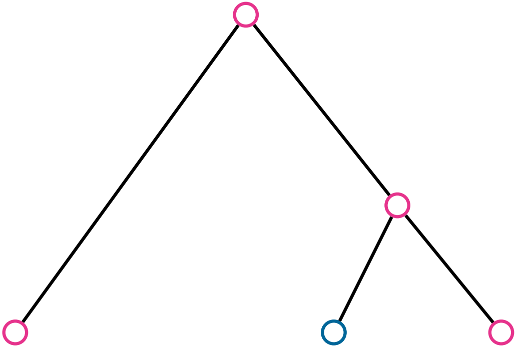
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ; N females ; $4N$ genes.



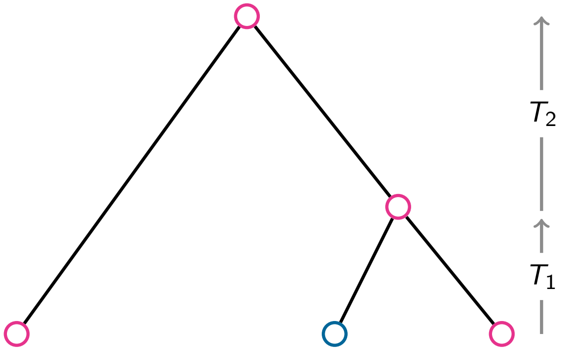
Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ○○ ; N females ○○ ; $4N$ genes.



Diploid, Two-Sex, Wright-Fisher Model

Population size $2N$; N males ○○; N females ○○; $4N$ genes.

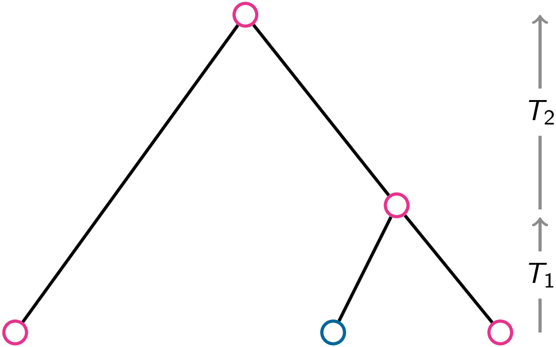


Diploid, Two-Sex, Wright-Fisher Model

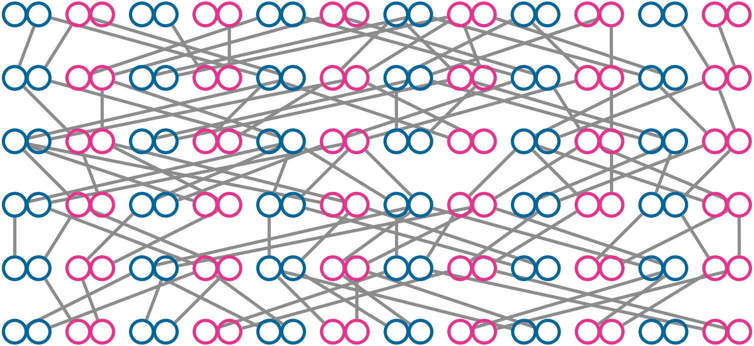
Population size $2N$; N males ○○ ; N females ○○ ; $4N$ genes.

Convergence to the n-coalescent (Möhle [1998])

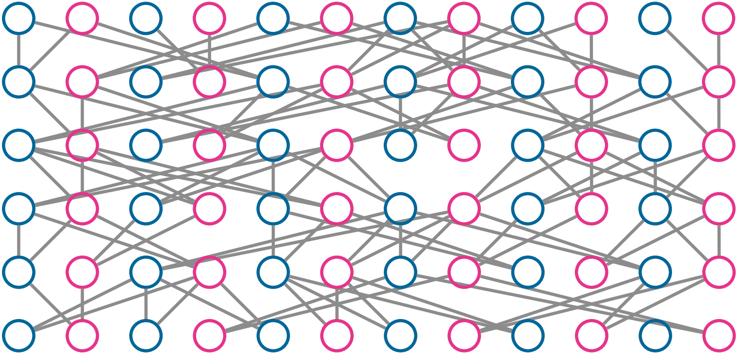
$$T_1/4N \sim \mathcal{E}xp(3),\ T_2/4N \sim \mathcal{E}xp(1); \quad N \text{ large.}$$



Population Pedigree



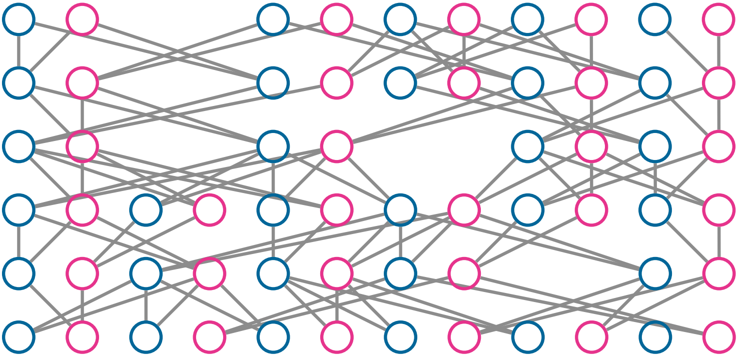
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

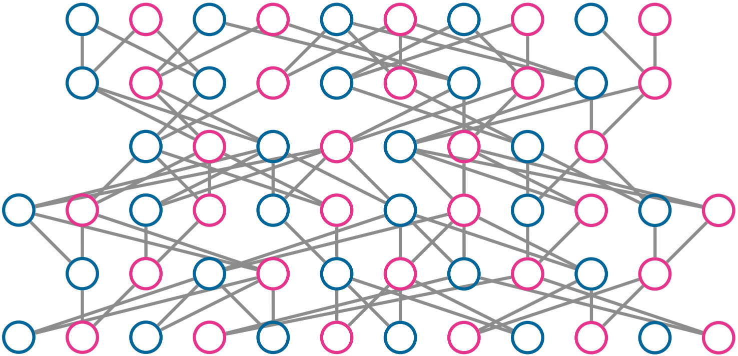
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

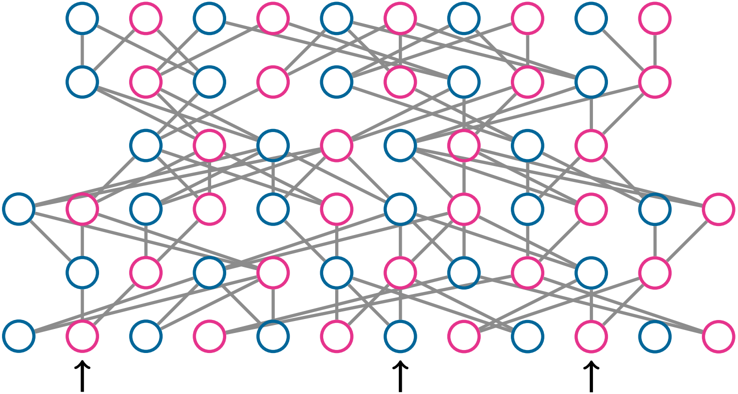
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

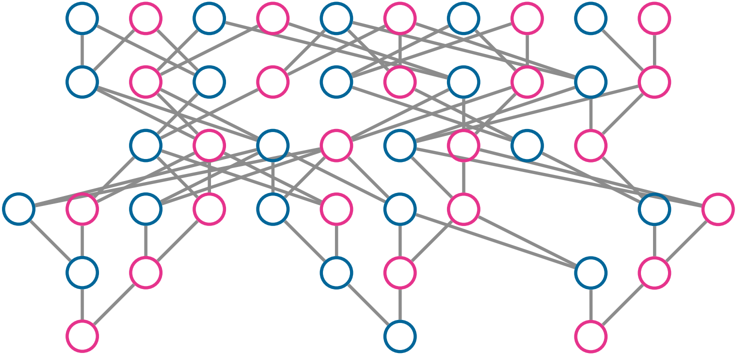
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

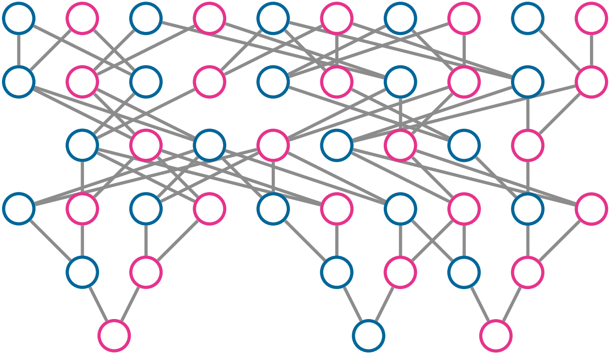
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

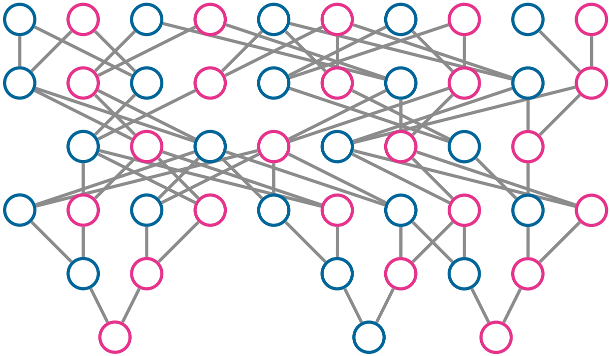
Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

Population Pedigree

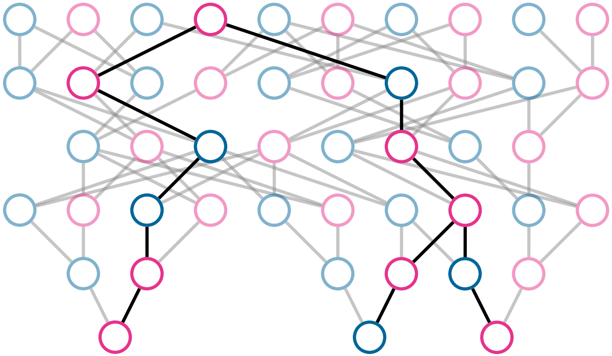


Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

Population Pedigree



Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

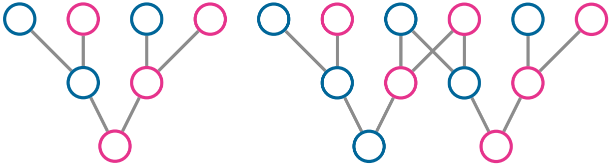
Population Pedigree

Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

Population Pedigree

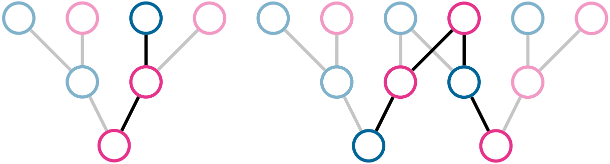


Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

Population Pedigree

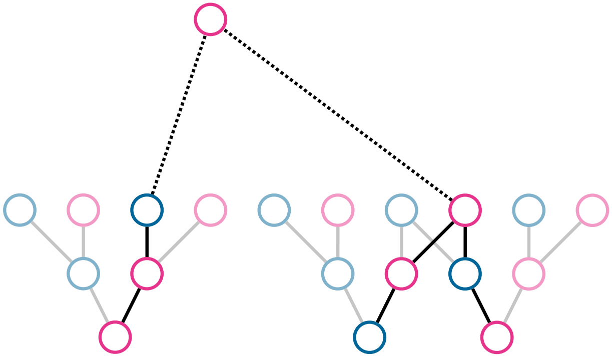


Population Pedigree (Wakeley et al. [2012])

The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

Population Pedigree

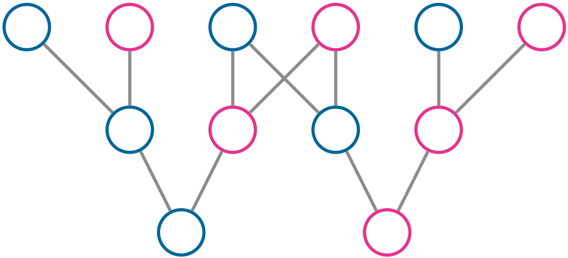


Population Pedigree (Wakeley et al. [2012])

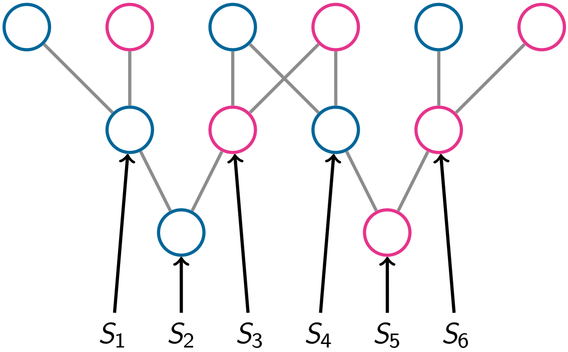
The set of all family relationships among members of the population for every generation.

"There is a fundamental connection between coalescent trees and pedigrees : a pedigree can be thought of as providing a scaffold on which coalescent trees are constructed." Speed and Balding [2015]

Pedigree Genealogy



Pedigree Genealogy

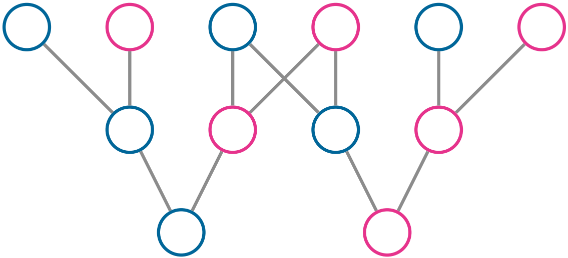


Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,.$$

Mendel's 1st Law

Pedigree Genealogy

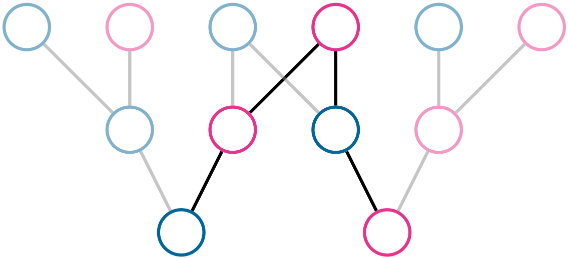


0 1 1 1 0 1

Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\,\mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{st} \text{ Law}$$

Pedigree Genealogy

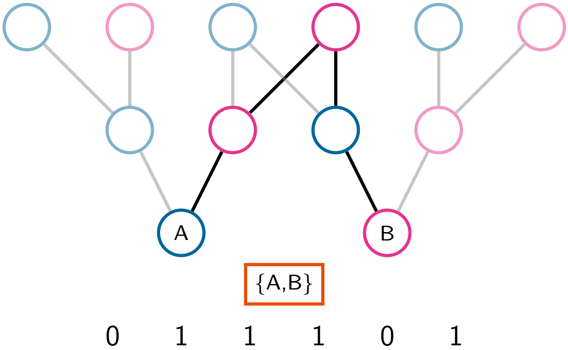


0 1 1 1 0 1

Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d. \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's 1}^{\text{st}} \text{ Law}$$

Pedigree Genealogy



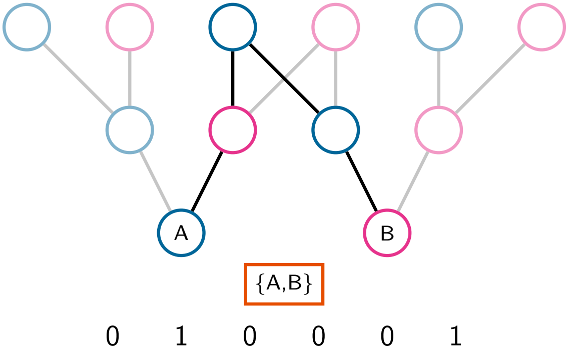
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{\text{st}} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



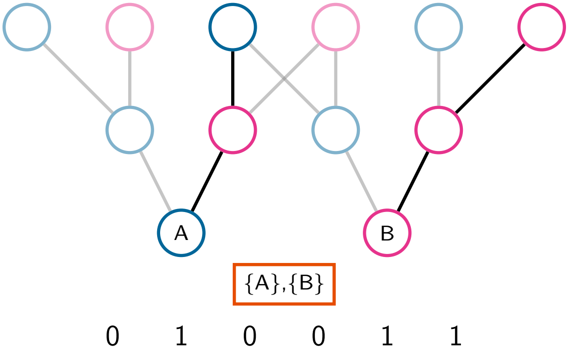
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\,\mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{\text{st}} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



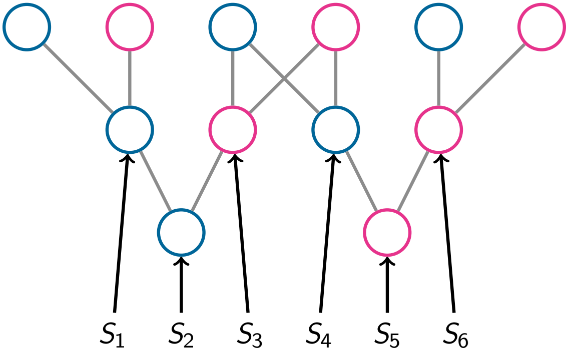
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d. \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{st} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



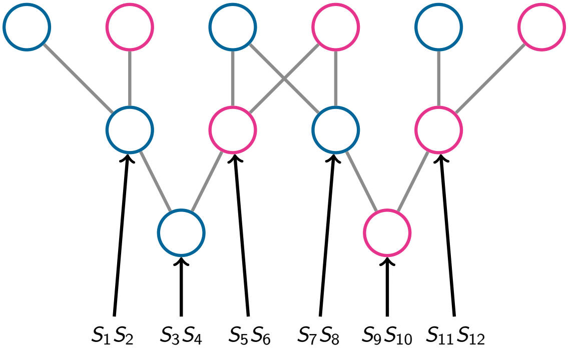
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{\text{st}} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



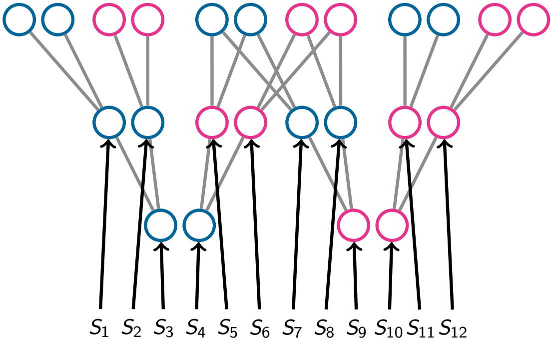
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{\text{st}} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



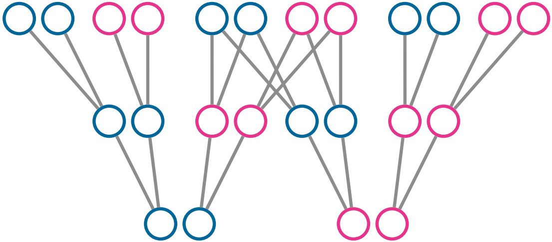
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's 1}^{\text{st}} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



0 0 1 0 1 1 1 0 0 1 0 1

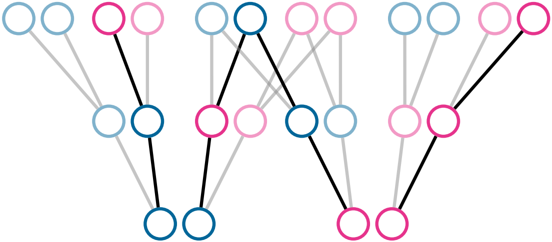
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\,\mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{st} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



0 0 1 0 1 1 1 0 0 1 0 1

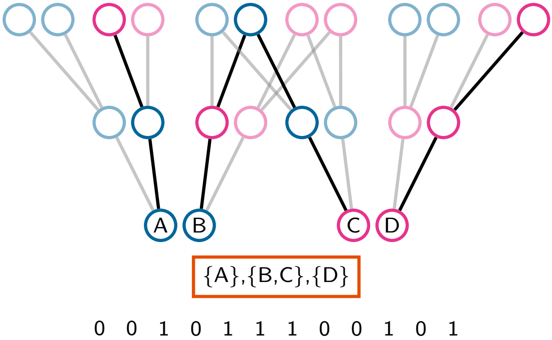
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d.\, \mathcal{B}ernoulli\,(0.5)\,. \qquad \text{Mendel's } 1^{st} \text{ Law}$$

IBD Partition

Partition of genes into subsets that are IBD

Pedigree Genealogy



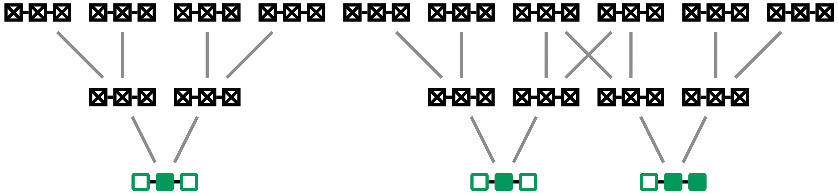
Meiosis Indicator (Thompson [2000])

$$S_i \sim i.i.d. \mathcal{B}ernoulli\,(0.5)\,.$$ Mendel's 1st Law

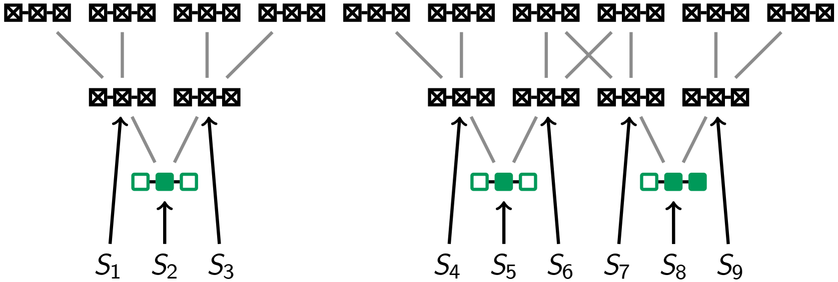
IBD Partition

Partition of genes into subsets that are IBD

SNP Data



SNP Data



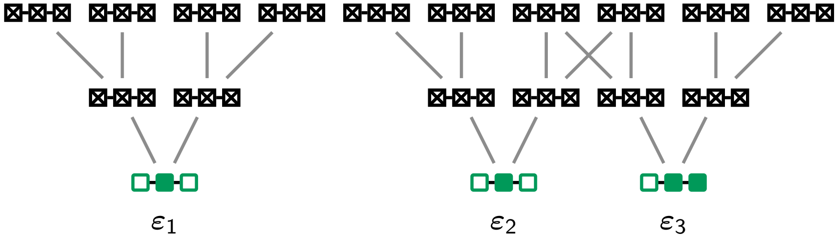
k-locus Meiosis Indicator

$$S_i \in \{0,1\}^k.$$

e.g. No interference (Allen and Darwiche [2008]) :

$$P(S_i = 011) = P(S_i = 100) = 0.5p_{12}(1 - p_{23}).$$

SNP Data

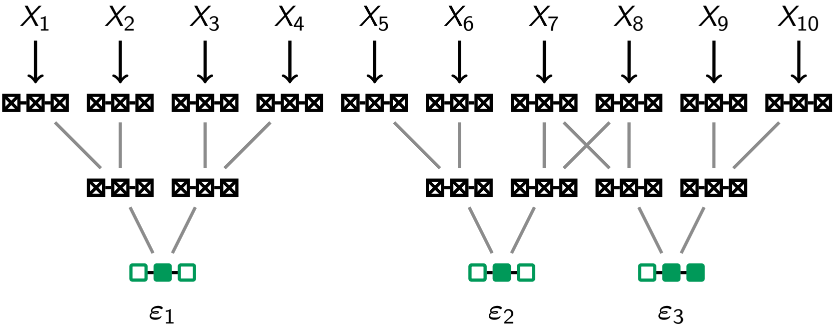


Evidence : ε (Koller and Friedman [2009])

An instantiation of a subset of random variables.

$$P(S_1, S_2, \ldots, S_n|\varepsilon) \neq \prod_{i=1}^{n} P(S_i|\varepsilon).$$

SNP Data

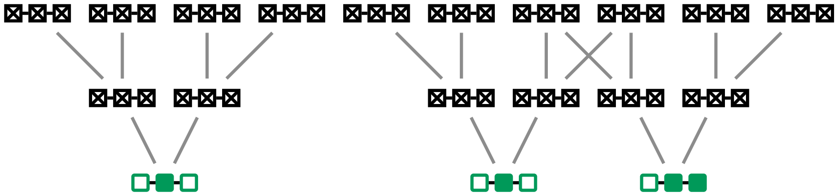


Evidence : ε (Koller and Friedman [2009])

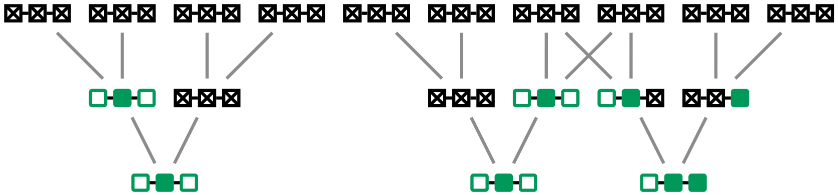
An instantiation of a subset of random variables.

$$P(S_1, S_2, \ldots, S_n|\varepsilon) \neq \prod_{i=1}^{n} P(S_i|\varepsilon).$$

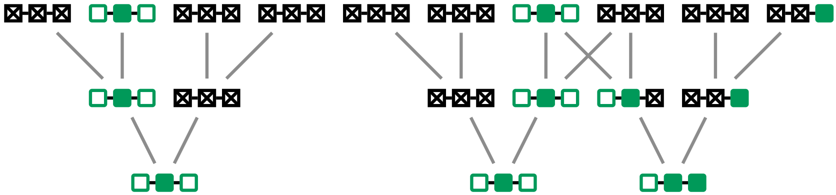
P-Coalescent



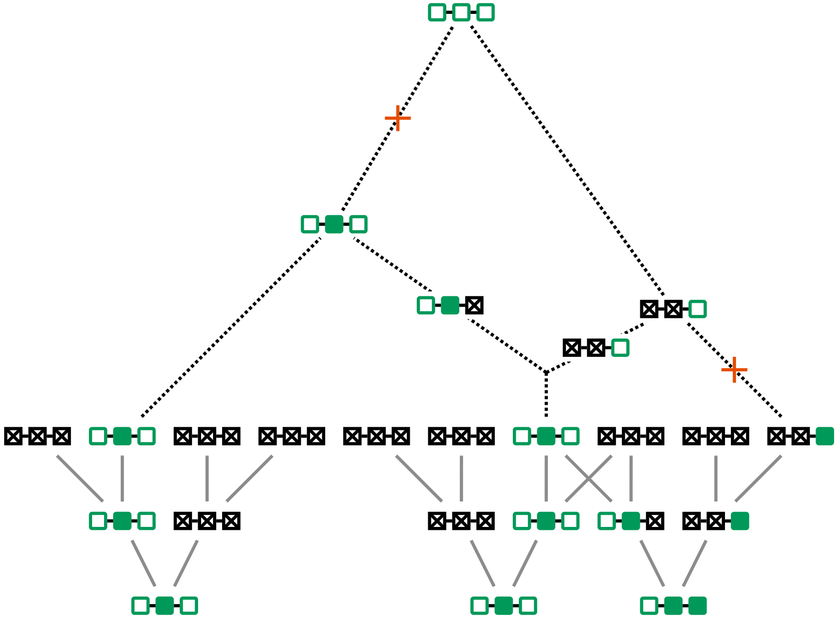
P-Coalescent



P-Coalescent



P-Coalescent



Perspectives

Perspectives

Pedigree Censorship

Big Pedigree $\Leftrightarrow$ High Complexity.

Deviations from the coalescent decreases the more distant the ancestry. (Wakeley et al. [2012])

Perspectives

Pedigree Censorship

Big Pedigree $\Leftrightarrow$ High Complexity.

Deviations from the coalescent decreases the more distant the ancestry. (Wakeley et al. [2012])

Recombination Model

- Independently segregating loci
- Non-interfering recombinations
- Positive interference

Perspectives

Pedigree Censorship

Big Pedigree $\Leftrightarrow$ High Complexity.

Deviations from the coalescent decreases the more distant the ancestry. (Wakeley et al. [2012])

Recombination Model

- Independently segregating loci
- Non-interfering recombinations
- Positive interference

Founder Prior

- Uniform prior
- Linkage equilibrium
- Linkage disequilibrium

David Allen and Adnan Darwiche. Rc_link : Genetic linkage analysis using bayesian networks. *International journal of approximate reasoning*, 48(2) :499–525, 2008.

Alexandre Bureau. *Genetic linkage analysis based on identity by descent using Markov chain Monte Carlo sampling on large pedigrees*. PhD thesis, University of California, Berkeley, 2001.

Robert Fung and Brendan Del Favero. Backward simulation in bayesian networks. In *Uncertainty Proceedings 1994*, pages 227–234. Elsevier, 1994.

Daphne Koller and Nir Friedman. *Probabilistic graphical models : principles and techniques*. MIT press, 2009.

Fabrice Larribe, Sabin Lessard, and Nicholas J Schork. Gene mapping via the ancestral recombination graph. *Theoretical population biology*, 62(2) :215–229, 2002.

References II

M Möhle. Coalescent results for two-sex population models. *Advances in Applied Probability*, 30(2) :513–520, 1998.

Doug Speed and David J Balding. Relatedness in the post-genomic era : is it still useful ? *Nature Reviews Genetics*, 16(1) :33, 2015.

Elizabeth A Thompson. Statistical inference from genetic data on pedigrees. In *NSF-CBMS regional conference series in probability and statistics*, pages i–169. JSTOR, 2000.

Elizabeth A Thompson. Identity by descent : variation in meiosis, across genomes, and in populations. *Genetics*, 194(2) :301–326, 2013.

John Wakeley, Léandra King, Bobbi S Low, and Sohini Ramachandran. Gene genealogies within a fixed pedigree, and the robustness of kingman's coalescent. *Genetics*, pages genetics–111, 2012.