

Computational Biology

Lecturers:

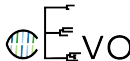
Tanja Stadler, Tim Vaughan & Carsten Magnus

Teaching Assistants:

Antoine Zwaans, Adrian Lison,
James Munday & Marcus Overwater

Computational Evolution
Department of Biosystems Science and Engineering

HS 2023



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

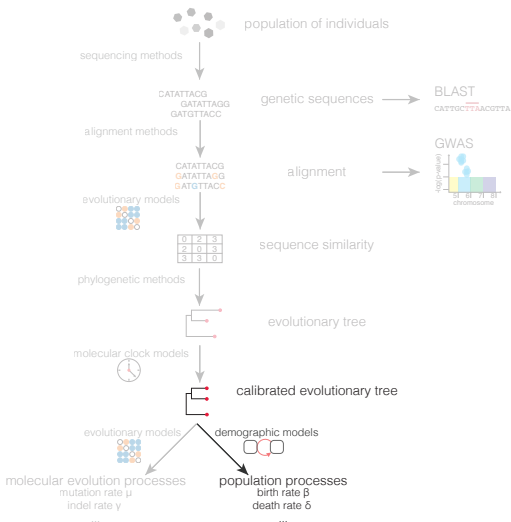
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ❓ How does the approximate number of steps required to calculate the phylogenetic likelihood depend on the number of leaves in a phylogenetic tree? (I.e. what is the time complexity of this calculation?)
- ❓ What kind of population dynamic process could a decrease in slope in the LTT plot reflect?
- ❓ Assume a birth-death process where each individual at present is sampled with probability ρ . How is the derivation of $p(0|t, \rho)$, the probability of sampling no individual at present, different compared to the derivation of $p(0|t)$?



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Alternative tree-generating process: The Coalescent

CB

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- Introduced by [Kingman, 1982] (the Kingman coalescent).

Alternative tree-generating process: The Coalescent

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Introduced by [Kingman, 1982] (the Kingman coalescent).
- ▶ Can be derived as a limiting distribution from several population genetic models.

Alternative tree-generating process: The Coalescent

CB

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Introduced by [Kingman, 1982] (the Kingman coalescent).
- ▶ Can be derived as a limiting distribution from several population genetic models.
- ▶ Common assumption is that the underlying population dynamics are deterministic.

Alternative tree-generating process: The Coalescent

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

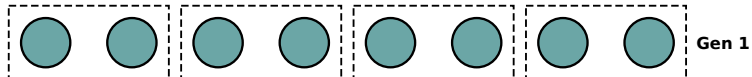
Birth-death vs Coalescent

References

- ▶ Introduced by [Kingman, 1982] (the Kingman coalescent).
- ▶ Can be derived as a limiting distribution from several population genetic models.
- ▶ Common assumption is that the underlying population dynamics are deterministic.
- ▶ Often used as the basis for phylodynamic inference of population size and dynamics.

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

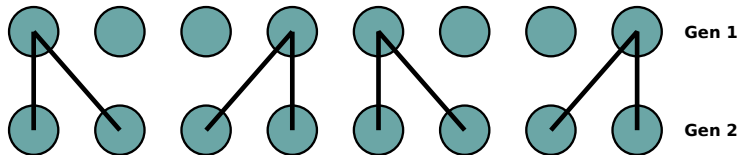
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

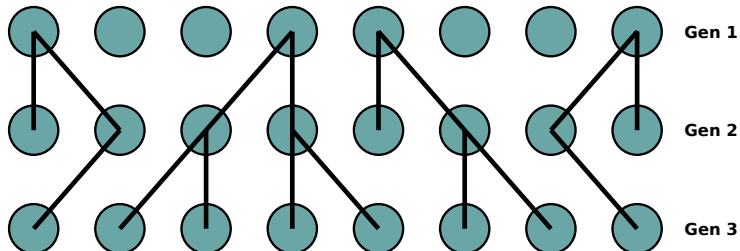
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

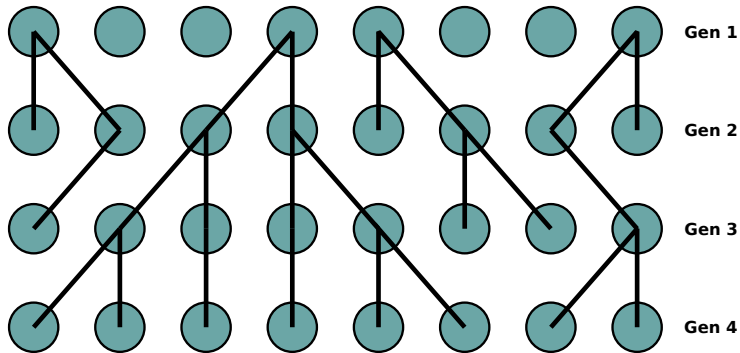
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

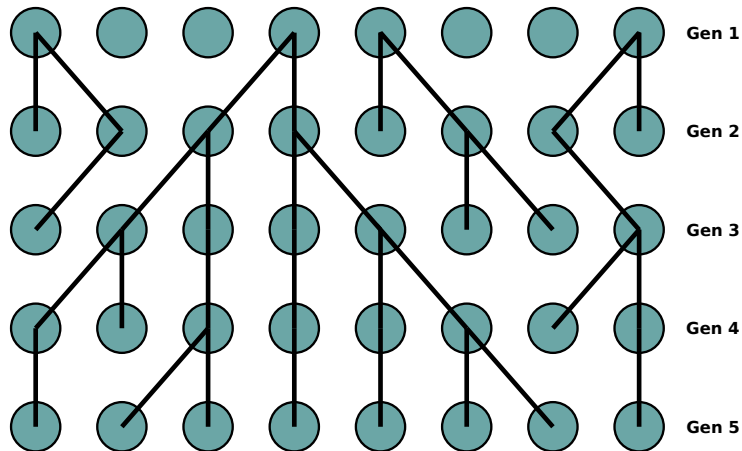
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

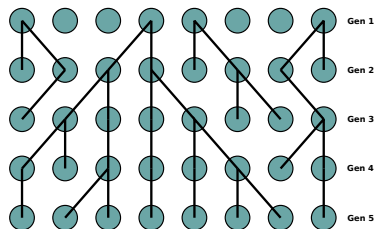
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



- Discrete generations.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

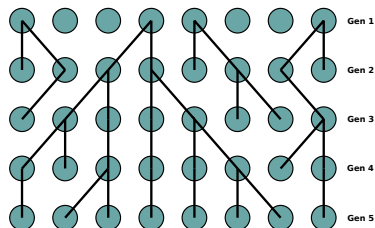
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



- Discrete generations.
- Each generation consists of N individuals.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

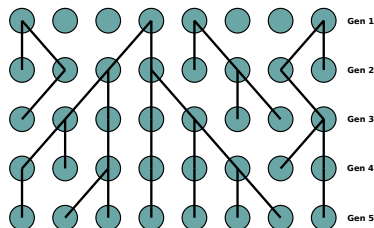
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



- Discrete generations.
- Each generation consists of N individuals.
- Each individual in the offspring population chooses its parent uniformly at random from the N parents.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

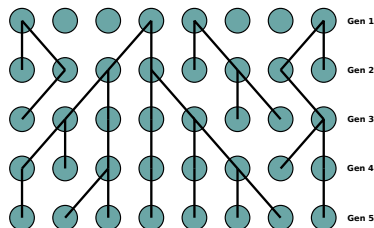
Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process

CB



- ▶ Discrete generations.
- ▶ Each generation consists of N individuals.
- ▶ Each individual in the offspring population chooses its parent uniformly at random from the N parents.
 - ▶ A given parent has a binomially-distributed number of offspring.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

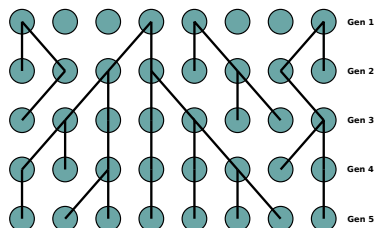
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process



- ▶ Discrete generations.
- ▶ Each generation consists of N individuals.
- ▶ Each individual in the offspring population chooses its parent uniformly at random from the N parents.
 - ▶ A given parent has a binomially-distributed number of offspring.
- ▶ For phylogenies of a particular gene, ploidy can be taken into account by multiplying N by a factor which accounts for the number of copies of a gene present in each individual.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

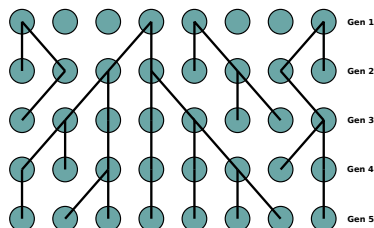
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The Wright-Fisher process



- ▶ Discrete generations.
- ▶ Each generation consists of N individuals.
- ▶ Each individual in the offspring population chooses its parent uniformly at random from the N parents.
 - ▶ A given parent has a binomially-distributed number of offspring.
- ▶ For phylogenies of a particular gene, ploidy can be taken into account by multiplying N by a factor which accounts for the number of copies of a gene present in each individual.
 - ▶ E.g. for a diploid organism, the number of copies of a gene in the population is $2N$.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

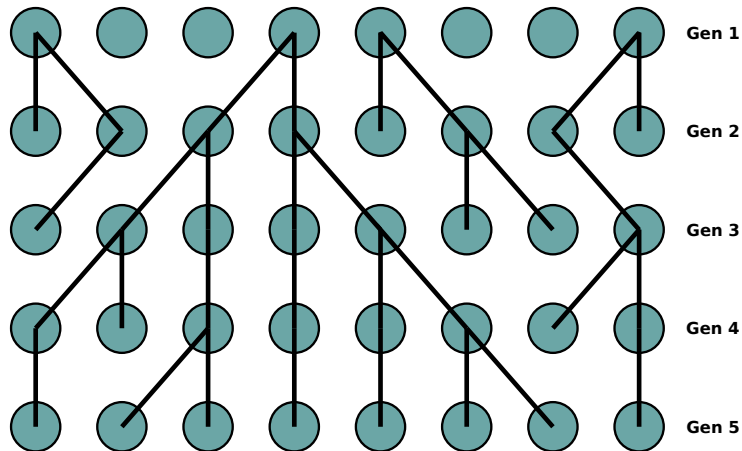
Application to HCV

Birth-death vs Coalescent

References

The Sampled WF Phylogeny

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

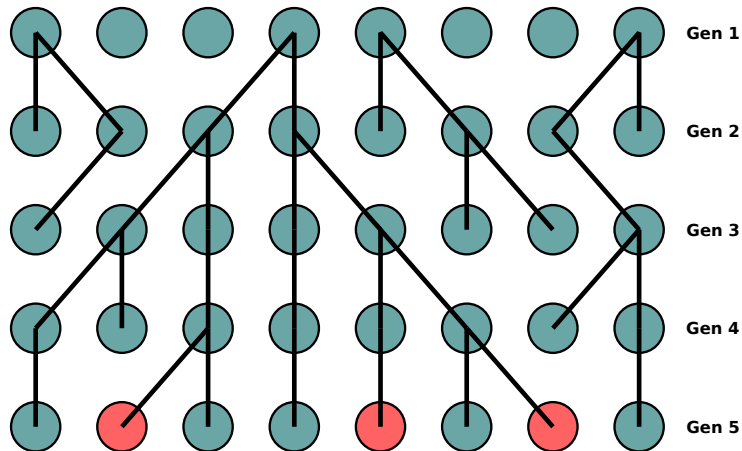
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

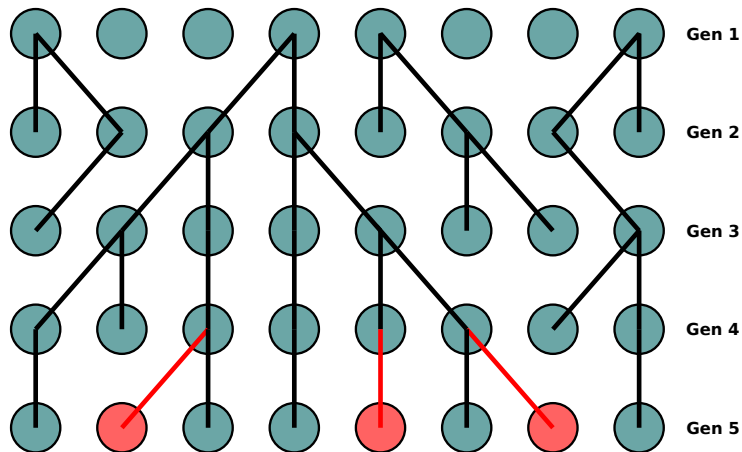
Coalescent trees

Population dynamics

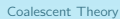
Application to HCV

Birth-death vs Coalescent

References



CB



Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

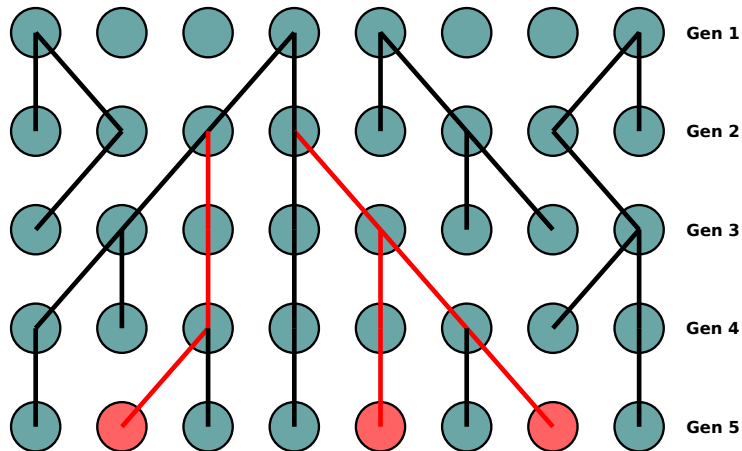
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

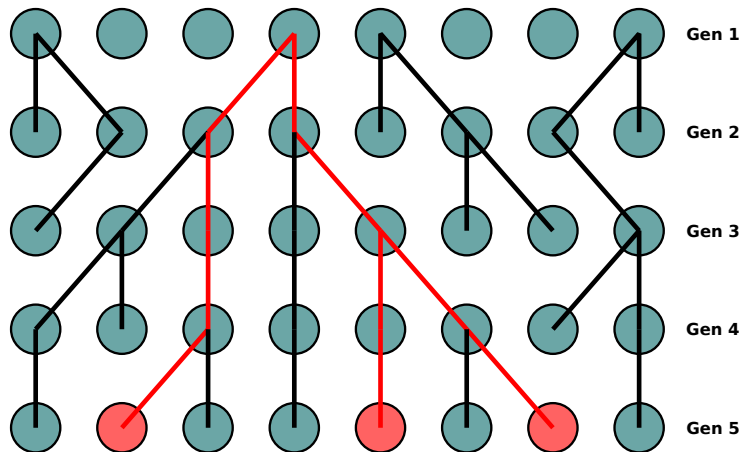
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

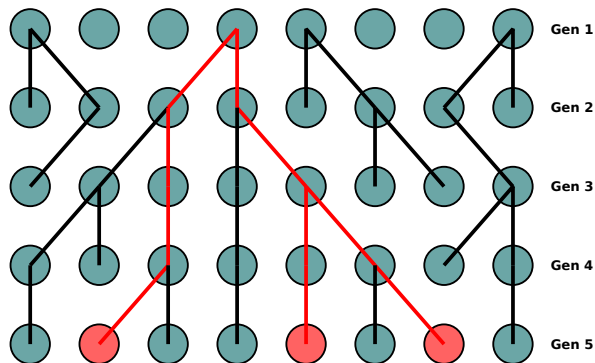
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- Generations between internal nodes related to population size.

The Sampled WF Phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

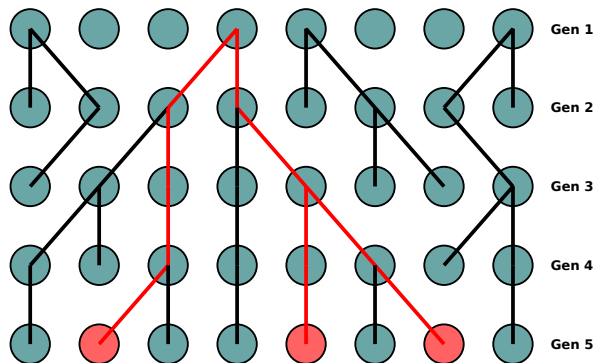
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- Generations between internal nodes related to population size.
- Can we quantify this relationship?

Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

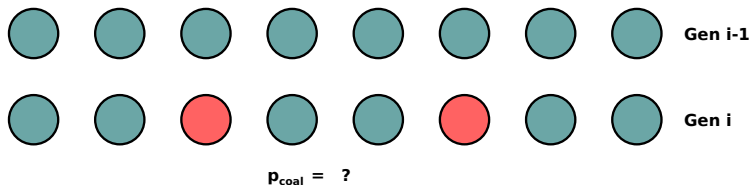
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

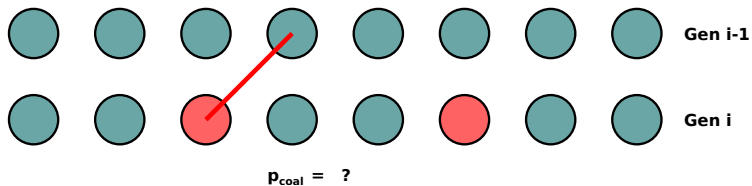
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

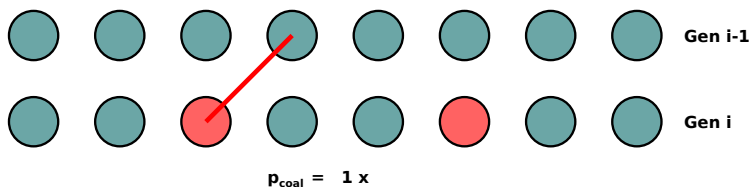
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

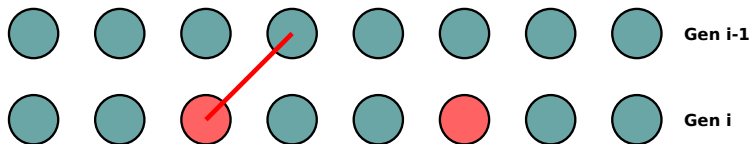
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



$$p_{\text{coal}} = 1 \times 1/N$$

Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

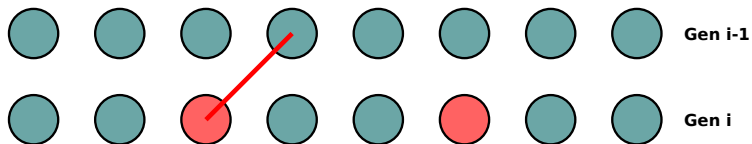
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



$$p_{\text{coal}} = 1 \times 1/N = 1/N$$

Sampled 2-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

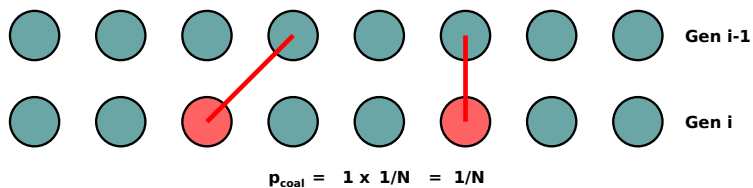
Coalescent trees

Population dynamics

Application to HCV

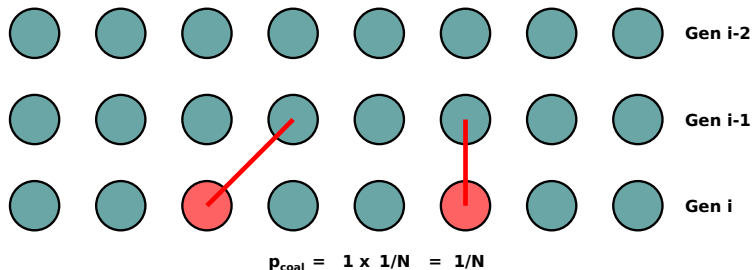
Birth-death vs Coalescent

References



Sampled 2-individual phylogeny

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

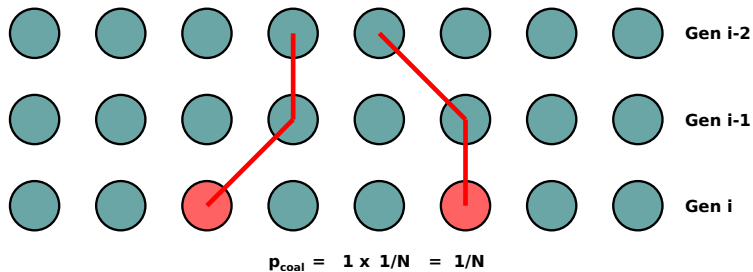
Application to HCV

Birth-death vs Coalescent

References

Sampled 2-individual phylogeny

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

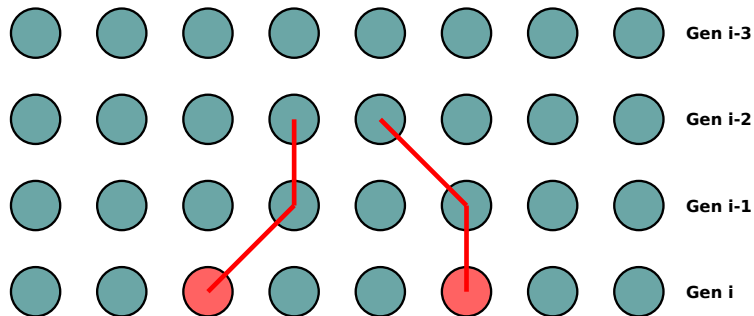
Application to HCV

Birth-death vs Coalescent

References

Sampled 2-individual phylogeny

CB



$$p_{\text{coal}} = 1 \times 1/N = 1/N$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

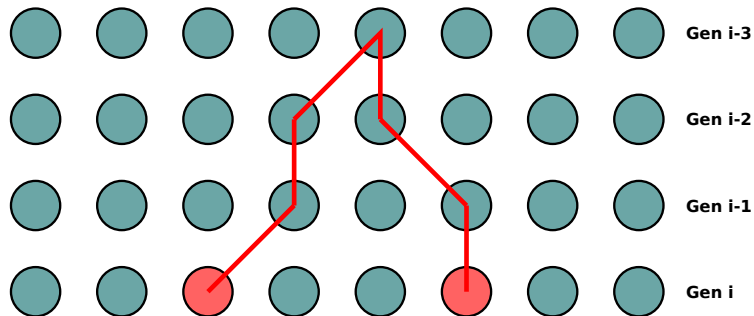
Application to HCV

Birth-death vs Coalescent

References

Sampled 2-individual phylogeny

CB



$$p_{\text{coal}} = 1 \times 1/N = 1/N$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

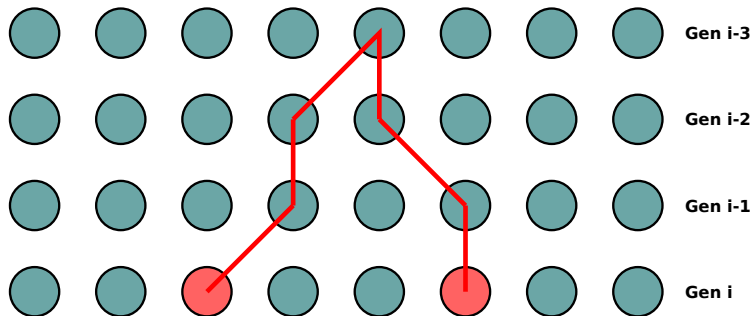
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Sampled 2-individual phylogeny



$$p_{\text{coal}} = 1 \times 1/N = 1/N$$

► Probability of coalescence in generation $i - m$:

$$\begin{aligned} P(m) &= (1 - p_{\text{coal}})^{m-1} p_{\text{coal}} \\ &= \left(1 - \frac{1}{N}\right)^{m-1} \frac{1}{N} \end{aligned}$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

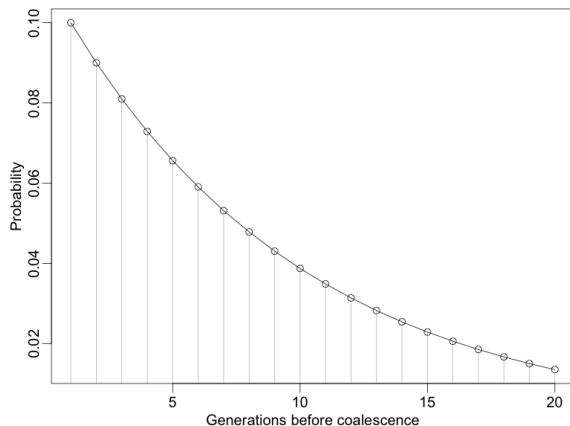
Application to HCV

Birth-death vs Coalescent

References

Coalescence time distribution

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

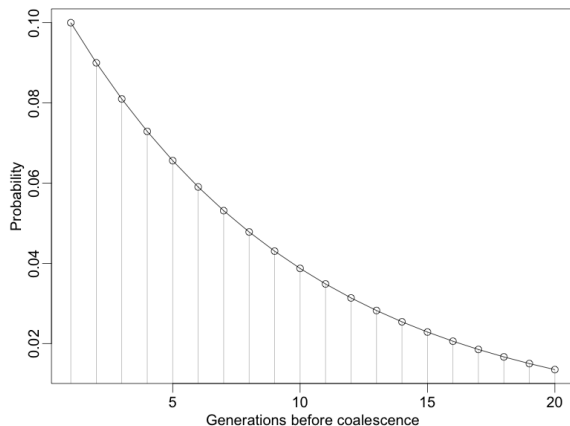
Application to HCV

Birth-death vs Coalescent

References

Coalescence time distribution

CB



► $P(m|N) = \left(1 - \frac{1}{N}\right)^{m-1} \frac{1}{N}$

Coalescent Theory

Introduction

Wright-Fisher process

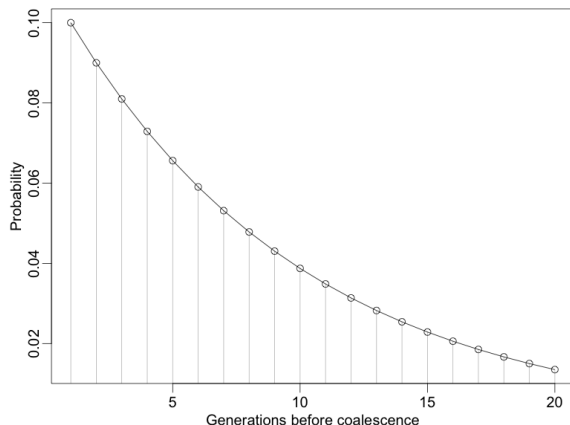
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- $P(m|N) = (1 - \frac{1}{N})^{m-1} \frac{1}{N}$
- For large N , $P(m|N) \rightarrow \exp[-\frac{m-1}{N}] \frac{1}{N}$

Coalescent Theory

Introduction

Wright-Fisher process

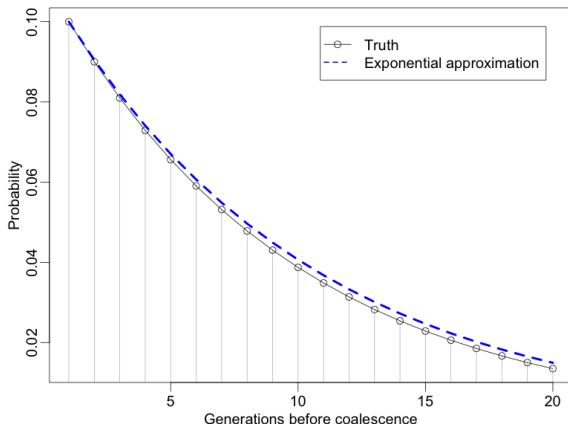
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- $P(m|N) = (1 - \frac{1}{N})^{m-1} \frac{1}{N}$
- For large N , $P(m|N) \rightarrow \exp[-\frac{m-1}{N}] \frac{1}{N}$

Coalescent Theory

Introduction

Wright-Fisher process

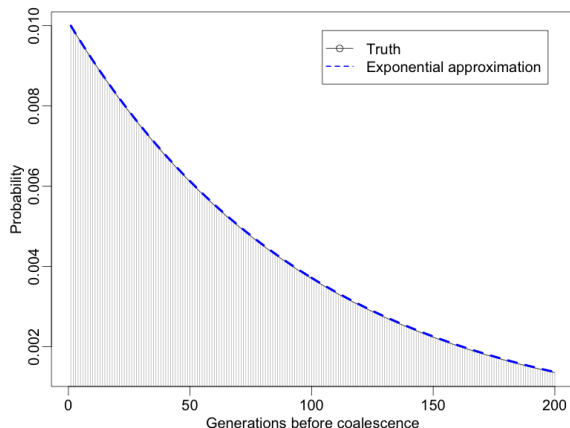
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- $P(m|N) = (1 - \frac{1}{N})^{m-1} \frac{1}{N}$
- For large N , $P(m|N) \rightarrow \exp[-\frac{m-1}{N}] \frac{1}{N}$

Coalescent Theory

Introduction

Wright-Fisher process

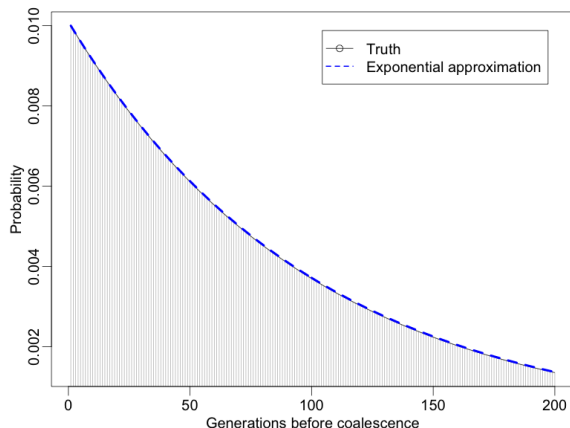
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



- ▶ $P(m|N) = (1 - \frac{1}{N})^{m-1} \frac{1}{N}$
- ▶ For large N , $P(m|N) \rightarrow \exp[-\frac{m-1}{N}] \frac{1}{N}$
- ▶ To see why, note that:
 $\exp[-(m-1)/N] = (\exp[-\frac{1}{N}])^{m-1} = (1 - \frac{1}{N} + O(\frac{1}{N^2}))^{m-1}$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- m is the number of generations. Let g be the calendar time of a generation (e.g. 5 days). Thus $\Delta t = gm$ is the calendar time span of m generations.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ m is the number of generations. Let g be the calendar time of a generation (e.g. 5 days). Thus $\Delta t = gm$ is the calendar time span of m generations.
- ▶ In calendar time the probability *density* function for the coalescence time of two lineages is $\frac{1}{gN} e^{-\frac{\Delta t}{gN}}$.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ m is the number of generations. Let g be the calendar time of a generation (e.g. 5 days). Thus $\Delta t = gm$ is the calendar time span of m generations.
- ▶ In calendar time the probability *density* function for the coalescence time of two lineages is $\frac{1}{gN} e^{-\frac{\Delta t}{gN}}$.
- ▶ In the large N limit, the time to coalescence is exponentially distributed with mean gN .

Sampled k-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled k-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

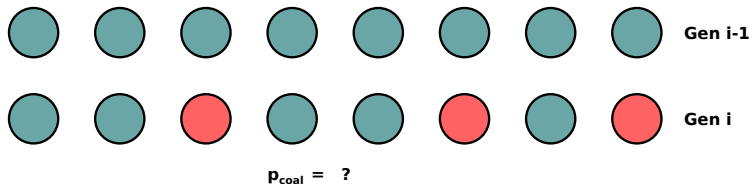
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References



Sampled k-individual phylogeny

CB

Coalescent Theory

Introduction

Wright-Fisher process

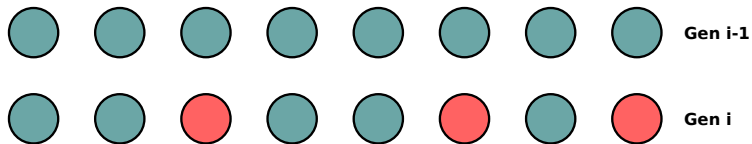
Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

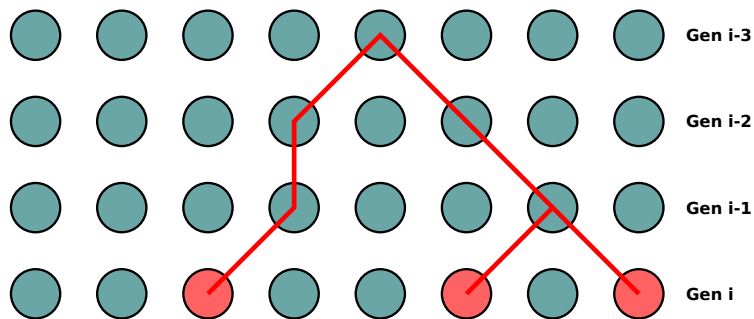
References



$$p_{\text{coal}} = 3 \times 1/N \text{ (approx, assumes } N \gg 3)$$

Sampled k-individual phylogeny

CB



$$p_{\text{coal}} = 3 \times 1/N \text{ (approx, assumes } N \gg 3)$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

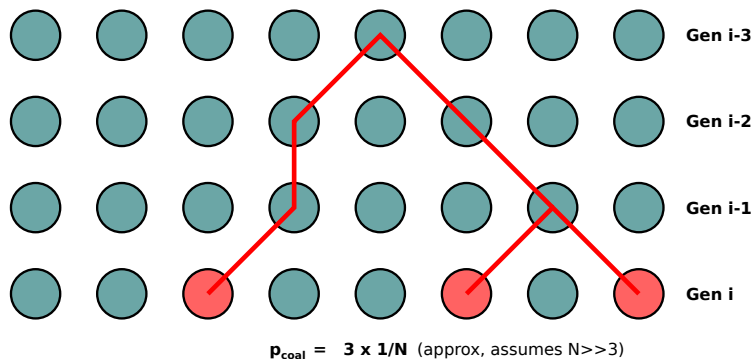
Application to HCV

Birth-death vs Coalescent

References

Sampled k-individual phylogeny

CB



► **Question** How can this be generalized to k samples?

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

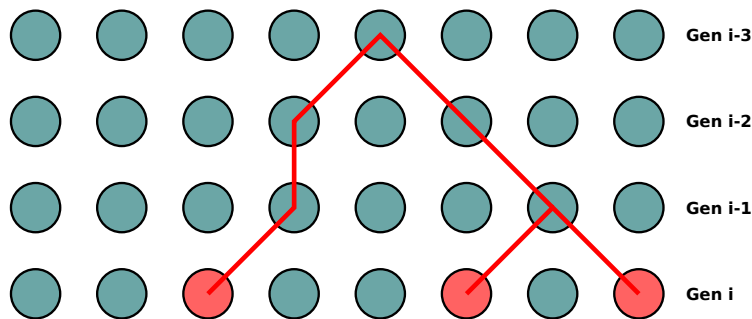
Application to HCV

Birth-death vs Coalescent

References

Sampled k-individual phylogeny

CB



$$p_{\text{coal}} = 3 \times 1/N \text{ (approx, assumes } N \gg 3)$$

► **Question** How can this be generalized to k samples?

► **Answer** $p_{\text{coal}} \simeq \binom{k}{2} \frac{1}{N}$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

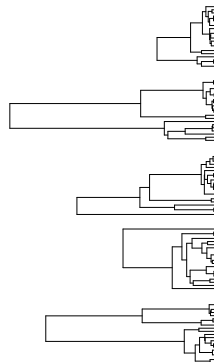
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- Developed by John Kingman in a series of papers in 1982.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

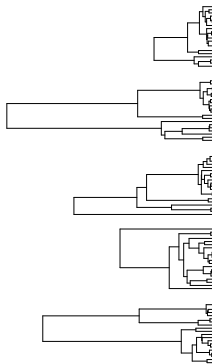
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Developed by John Kingman in a series of papers in 1982.
- ▶ Continuous-time Markov process which produces sampled time trees.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

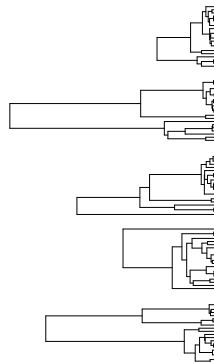
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Developed by John Kingman in a series of papers in 1982.
- ▶ Continuous-time Markov process which produces sampled time trees.
- ▶ Process occurs *backwards in time*.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

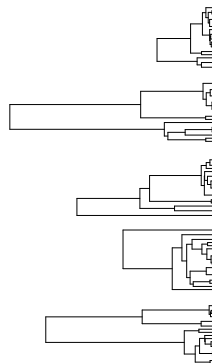
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Developed by John Kingman in a series of papers in 1982.
- ▶ Continuous-time Markov process which produces sampled time trees.
- ▶ Process occurs *backwards in time*.
- ▶ Equivalent to sampled trees produced by WF model when N is much larger than the number of samples.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

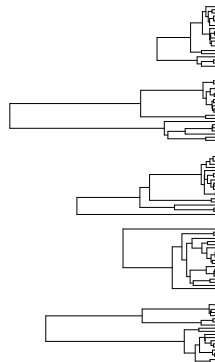
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Developed by John Kingman in a series of papers in 1982.
- ▶ Continuous-time Markov process which produces sampled time trees.
- ▶ Process occurs *backwards in time*.
- ▶ Equivalent to sampled trees produced by WF model when N is much larger than the number of samples.
- ▶ Times between coalescence events are drawn from exponential distributions with rate parameters $\binom{k}{2} \frac{1}{Ng}$, i.e.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

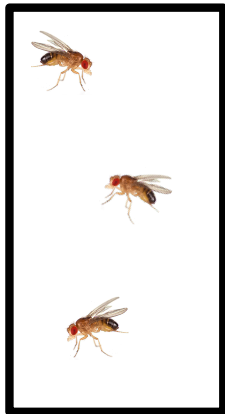
Birth-death vs Coalescent

References

$$P(\Delta t | N, g, k) = \exp \left[-\Delta t \binom{k}{2} \frac{1}{Ng} \right] \binom{k}{2} \frac{1}{Ng}.$$

“Bugs in a Box” analogy

Consider a box full of “hyperactive, indiscriminate, voracious, and insatiable bugs.” [Felsenstein, 2004]



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

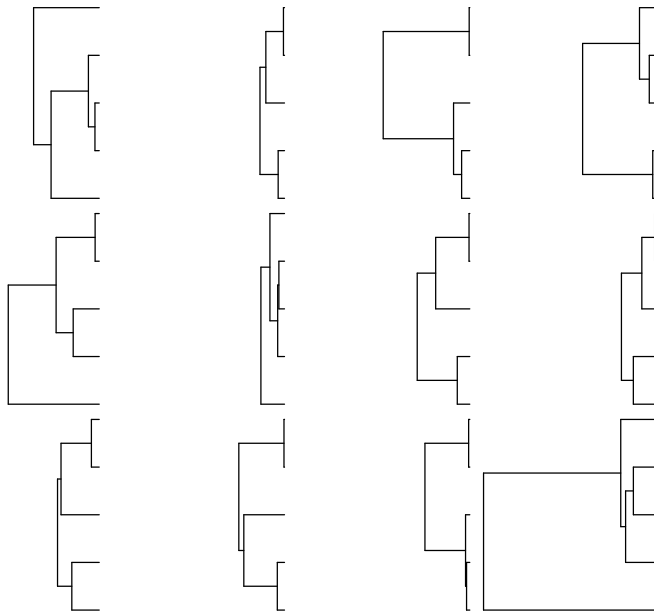
Application to HCV

Birth-death vs Coalescent

References

Simulated coalescent trees (5 leaves, $g = 1$)

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

CB

Under the coalescent model, the average time required for n lineages to coalesce into one is

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

► Can write $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=2}^n \frac{2}{k(k-1)}.$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

- ▶ Can write $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=2}^n \frac{2}{k(k-1)}.$
- ▶ Can expand $\frac{2}{k(k-1)} = \frac{2}{k-1} - \frac{2}{k}.$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

- ▶ Can write $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=2}^n \frac{2}{k(k-1)}$.
- ▶ Can expand $\frac{2}{k(k-1)} = \frac{2}{k-1} - \frac{2}{k}$.
- ▶ Then $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=1}^{n-1} \frac{2}{k} - \sum_{k=2}^n \frac{2}{k} = 2(1 - \frac{1}{n})$.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

- ▶ Can write $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=2}^n \frac{2}{k(k-1)}$.
- ▶ Can expand $\frac{2}{k(k-1)} = \frac{2}{k-1} - \frac{2}{k}$.
- ▶ Then $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=1}^{n-1} \frac{2}{k} - \sum_{k=2}^n \frac{2}{k} = 2(1 - \frac{1}{n})$.
- ▶ We therefore find that $E[t_{\text{root}}] \rightarrow 2Ng$ as the number of lineages n (i.e. number of leaves in the coalescent tree) becomes large.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

The age of a coalescent tree

Under the coalescent model, the average time required for n lineages to coalesce into one is

$$E[t_{\text{root}}] = \sum_{k=2}^n \frac{Ng}{\binom{k}{2}} = Ng \sum_{k=2}^n \frac{1}{\binom{k}{2}}.$$

- Can write $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=2}^n \frac{2}{k(k-1)}$.
- Can expand $\frac{2}{k(k-1)} = \frac{2}{k-1} - \frac{2}{k}$.
- Then $\sum_{k=2}^n \frac{1}{\binom{k}{2}} = \sum_{k=1}^{n-1} \frac{2}{k} - \sum_{k=2}^n \frac{2}{k} = 2(1 - \frac{1}{n})$.
- We therefore find that $E[t_{\text{root}}] \rightarrow 2Ng$ as the number of lineages n (i.e. number of leaves in the coalescent tree) becomes large.

This is an upper bound on the *expectation*: individual coalescent trees can be older than this.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

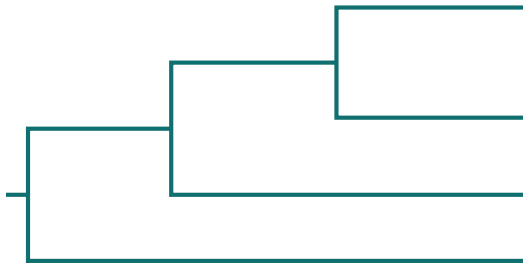
Application to HCV

Birth-death vs Coalescent

References

The probability of a coalescent tree

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

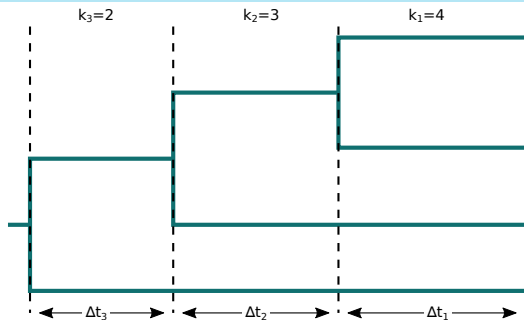
Application to HCV

Birth-death vs Coalescent

References

The probability of a coalescent tree

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

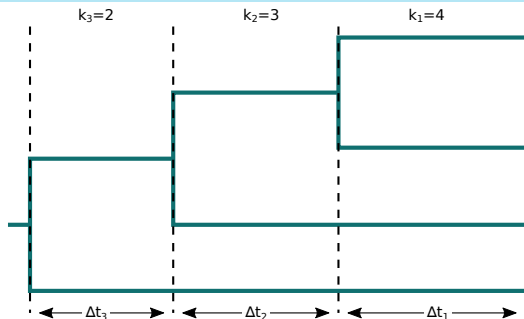
Application to HCV

Birth-death vs Coalescent

References

The probability of a coalescent tree

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

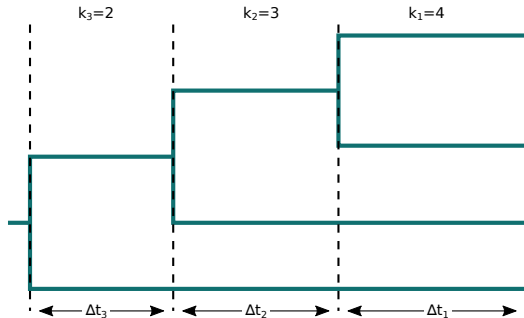
Application to HCV

Birth-death vs Coalescent

References

$$P(\mathcal{T}|Ng) = e^{-\Delta t_1 \binom{4}{2} \frac{1}{Ng}} \times \frac{1}{Ng} \times e^{-\Delta t_2 \binom{3}{2} \frac{1}{Ng}} \times \frac{1}{Ng} \times e^{-\Delta t_3 \binom{2}{2} \frac{1}{Ng}} \times \frac{1}{Ng}$$
$$= \prod_{i=1}^{n-1} \left(\exp \left[-\Delta t_i \binom{k_i}{2} \frac{1}{Ng} \right] \frac{1}{Ng} \right)$$

The probability of a coalescent tree



$$\begin{aligned}
 P(\mathcal{T}|Ng) &= e^{-\Delta t_1 \binom{4}{2} \frac{1}{Ng}} \times \frac{1}{Ng} \times e^{-\Delta t_2 \binom{3}{2} \frac{1}{Ng}} \times \frac{1}{Ng} \times e^{-\Delta t_3 \binom{2}{2} \frac{1}{Ng}} \times \frac{1}{Ng} \\
 &= \prod_{i=1}^{n-1} \left(\exp \left[-\Delta t_i \binom{k_i}{2} \frac{1}{Ng} \right] \frac{1}{Ng} \right)
 \end{aligned}$$

The exponentials give the probability of nothing happening in interval Δt_i , and the $1/gN$ factors are the probability densities of the particular coalescent events. (Note the units!)

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

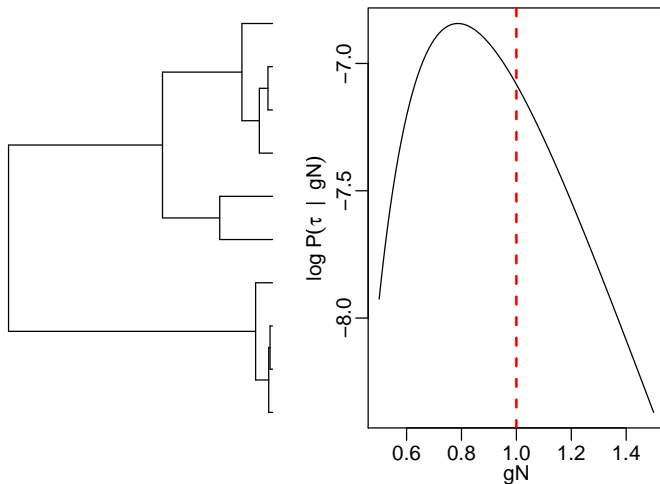
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- For a given coalescent tree \mathcal{T} , $L(gN; \mathcal{T}) = P(\mathcal{T}|gN)$.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- If we take a real tree (e.g. one inferred from genetic data sampled from a real biological population) and infer the population size using the coalescent distribution, we can expect our result to be biased since the real dynamics differ from the WF dynamics.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ If we take a real tree (e.g. one inferred from genetic data sampled from a real biological population) and infer the population size using the coalescent distribution, we can expect our result to be biased since the real dynamics differ from the WF dynamics.
- ▶ One of the important ways that reality differs is that real populations are *structured*, where the WF population is assumed to be completely homogeneous.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ If we take a real tree (e.g. one inferred from genetic data sampled from a real biological population) and infer the population size using the coalescent distribution, we can expect our result to be biased since the real dynamics differ from the WF dynamics.
- ▶ One of the important ways that reality differs is that real populations are *structured*, where the WF population is assumed to be completely homogeneous.
- ▶ In any case, the inferred population size is referred to as the *effective* population size (sometimes written N_e).

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ If we take a real tree (e.g. one inferred from genetic data sampled from a real biological population) and infer the population size using the coalescent distribution, we can expect our result to be biased since the real dynamics differ from the WF dynamics.
- ▶ One of the important ways that reality differs is that real populations are *structured*, where the WF population is assumed to be completely homogeneous.
- ▶ In any case, the inferred population size is referred to as the *effective* population size (sometimes written N_e).
- ▶ This is the size of a WF population which shares some statistical similarity with the real population.

[Coalescent Theory](#)[Introduction](#)[Wright-Fisher process](#)[Coalescent trees](#)[Population dynamics](#)[Application to HCV](#)[Birth-death vs Coalescent](#)[References](#)

- ▶ If we take a real tree (e.g. one inferred from genetic data sampled from a real biological population) and infer the population size using the coalescent distribution, we can expect our result to be biased since the real dynamics differ from the WF dynamics.
- ▶ One of the important ways that reality differs is that real populations are *structured*, where the WF population is assumed to be completely homogeneous.
- ▶ In any case, the inferred population size is referred to as the *effective* population size (sometimes written N_e).
- ▶ This is the size of a WF population which shares some statistical similarity with the real population.
- ▶ Care must be taken when drawing conclusions from effective population sizes.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example
 - ▶ The Cannings model (generalization of the Wright-Fisher model)

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example
 - ▶ The Cannings model (generalization of the Wright-Fisher model)
 - ▶ The Moran model (overlapping generations, fixed population size)

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example
 - ▶ The Cannings model (generalization of the Wright-Fisher model)
 - ▶ The Moran model (overlapping generations, fixed population size)
 - ▶ some stochastic logistic models (continuous time, population fluctuations)

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example
 - ▶ The Cannings model (generalization of the Wright-Fisher model)
 - ▶ The Moran model (overlapping generations, fixed population size)
 - ▶ some stochastic logistic models (continuous time, population fluctuations)
 - ▶ ...

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ The coalescent distribution/process is often derived as a limit of the Wright-Fisher process, as we have done here.
- ▶ It also appears as the limit of many other population processes, for example
 - ▶ The Cannings model (generalization of the Wright-Fisher model)
 - ▶ The Moran model (overlapping generations, fixed population size)
 - ▶ some stochastic logistic models (continuous time, population fluctuations)
 - ▶ ...
- ▶ The fact that the coalescent distribution persists in the face of many departures from the WF model is sometimes termed the “robustness of the coalescent”.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

1. Samples are members of a population that is at demographic equilibrium.
 - ▶ Justifies use of fixed or slowly varying population size.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

1. Samples are members of a population that is at demographic equilibrium.
 - ▶ Justifies use of fixed or slowly varying population size.
2. Number of samples is small compared to the total population size.
 - ▶ Justifies neglect of > 2 lineages coalescing in the same generation.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

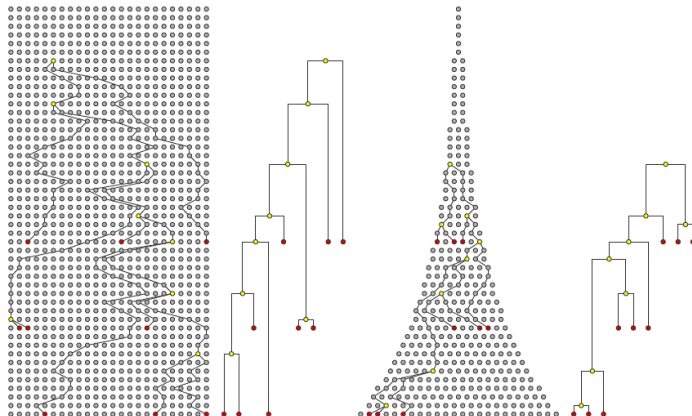
Birth-death vs Coalescent

References

1. Samples are members of a population that is at demographic equilibrium.
 - ▶ Justifies use of fixed or slowly varying population size.
2. Number of samples is small compared to the total population size.
 - ▶ Justifies neglect of > 2 lineages coalescing in the same generation.
3. Population is “well mixed”, samples are drawn uniformly at random.
 - ▶ Justifies the coalescent rate between any pair of sampled lineages being equal.
 - ▶ Population structure violates this assumption.

Extension: population size changes

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

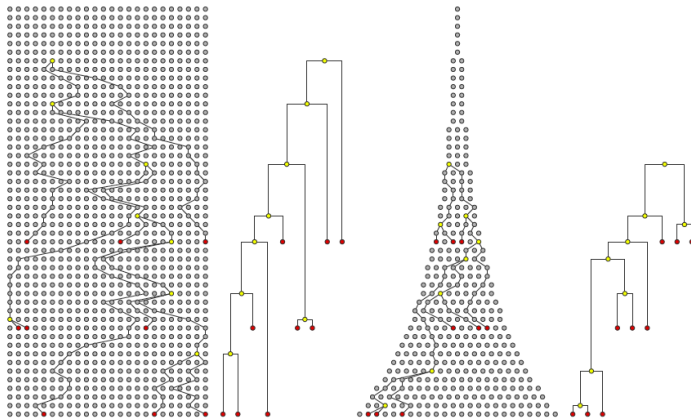
Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Extension: population size changes



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- $P(\mathcal{T}|N(t))$ is calculated via the rate of coalescence $\frac{1}{N(t)}$ (where $N(t)$ is the population size as a function of time). Where $N(t)$ is large we have a slower coalescence rate and thus longer branches.

- Under a Wright-Fisher model with a deterministically varying population size $N(t)$, the probability of a sampled tree becomes:

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- Under a Wright-Fisher model with a deterministically varying population size $N(t)$, the probability of a sampled tree becomes:

$$P(\mathcal{T}|N(t)) = \prod_{i=1}^{n-1} \left(\exp \left[- \int_{t_i}^{t_{i+1}} \binom{k_i}{2} \frac{dt}{N(t)g} \right] \frac{1}{N(t_{i+1})g} \right)$$

where t_i is the time at the beginning of interval i .

- Under a Wright-Fisher model with a deterministically varying population size $N(t)$, the probability of a sampled tree becomes:

$$P(\mathcal{T}|N(t)) = \prod_{i=1}^{n-1} \left(\exp \left[- \int_{t_i}^{t_{i+1}} \binom{k_i}{2} \frac{dt}{N(t)g} \right] \frac{1}{N(t_{i+1})g} \right)$$

where t_i is the time at the beginning of interval i .

- For a given parametric form, for example $N(t) = N_0 e^{-\gamma t}$, yields $P(\mathcal{T}|N_0, \gamma) = L(N_0, \gamma; \mathcal{T})$ i.e. the likelihood for the demographic model parameters.

- Under a Wright-Fisher model with a deterministically varying population size $N(t)$, the probability of a sampled tree becomes:

$$P(\mathcal{T}|N(t)) = \prod_{i=1}^{n-1} \left(\exp \left[- \int_{t_i}^{t_{i+1}} \binom{k_i}{2} \frac{dt}{N(t)g} \right] \frac{1}{N(t_{i+1})g} \right)$$

where t_i is the time at the beginning of interval i .

- For a given parametric form, for example $N(t) = N_0 e^{-\gamma t}$, yields $P(\mathcal{T}|N_0, \gamma) = L(N_0, \gamma; \mathcal{T})$ i.e. the likelihood for the demographic model parameters.
- Thus we can directly compare and test different demographic scenarios for a given tree.

The Epidemic Behavior of the Hepatitis C Virus

Oliver G. Pybus,* Michael A. Charleston, Sunetra Gupta,
Andrew Rambaut, Edward C. Holmes, Paul H. Harvey

Hepatitis C virus (HCV) is a leading worldwide cause of liver disease. Here, we use a new model of HCV spread to investigate the epidemic behavior of the virus and to estimate its basic reproductive number from gene sequence data. We find significant differences in epidemic behavior among HCV subtypes and suggest that these differences are largely the result of subtype-specific transmission patterns. Our model builds a bridge between the disciplines of population genetics and mathematical epidemiology by using pathogen gene sequences to infer the population dynamic history of an infectious disease.

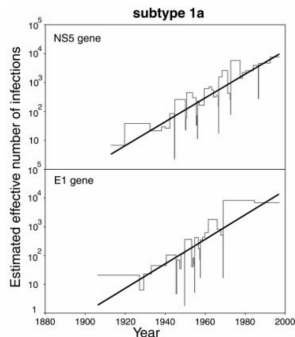


Table 1. Maximum likelihood parameter estimates for each HCV type or subtype. Seq., number of sequences.

HCV	Gene	Seq.	ML parameter estimates (approximate 95% CIs)		
			$N(0)$	r	c
1a	E1	34	13726 (6143, 32615)	0.098 (0.081, 0.114)	∞ (0.28, ∞)
	NS5	59	9858 (5488, 18430)	0.095 (0.079, 0.109)	∞ (0.24, ∞)
1b	E1	76	46484 (11408, 114430)	0.079 (0.068, 0.092)	3.3 (0.12, ∞)
	NS5	53	11800 (3254, 62084)	0.088 (0.068, 0.117)	0.44 (0.015, ∞)
4	E1	22	1817 (1012, 8797)	0.026 (0.008, 0.045)	0.015 (3×10^{-5} , ∞)
	NS5	18	2498 (1328, 5820)	0.043 (0.022, 0.071)	6×10^{-5} (4×10^{-9} , 0.037)
6	E1	23	1579 (948, 3433)	0.012 (0.003, 0.029)	0.028 (2×10^{-6} , ∞)
	NS5	40	2500 (1680, 4532)	0.008 (0.002, 0.018)	0.066 (8×10^{-5} , ∞)

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

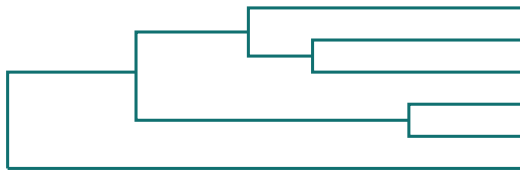
Application to HCV

Birth-death vs Coalescent

References

Non-parametric population dynamics inference

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

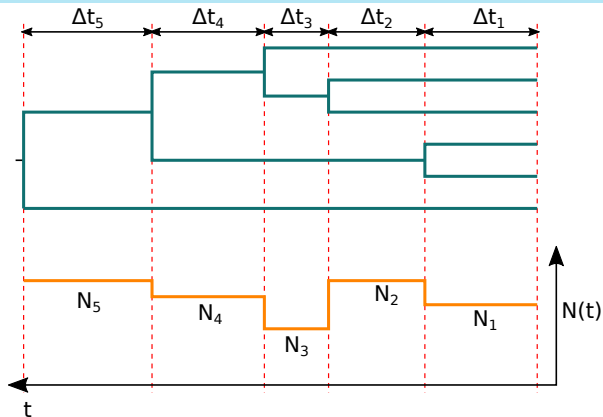
Application to HCV

Birth-death vs Coalescent

References

Non-parametric population dynamics inference

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

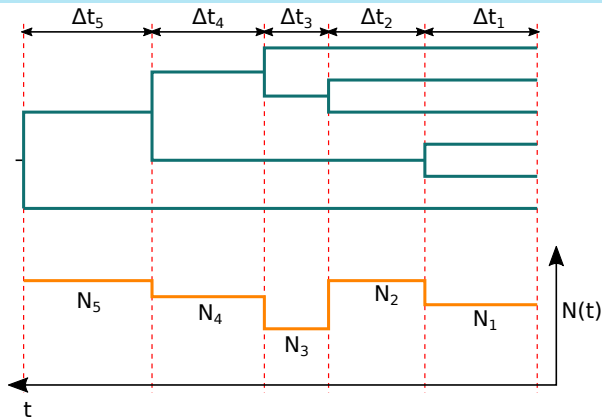
Application to HCV

Birth-death vs Coalescent

References

Non-parametric population dynamics inference

CB



- Assume population has distinct constant values in each interval between coalescent events.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

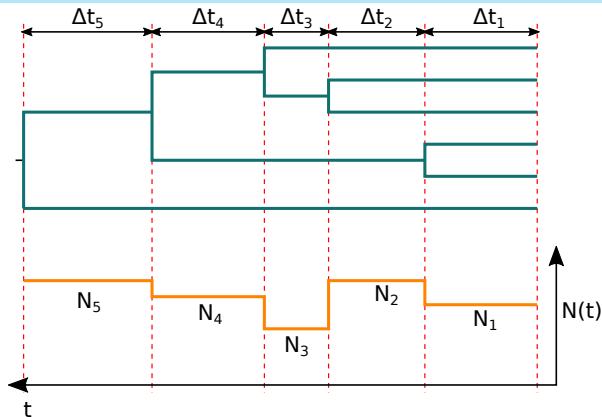
Application to HCV

Birth-death vs Coalescent

References

Non-parametric population dynamics inference

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

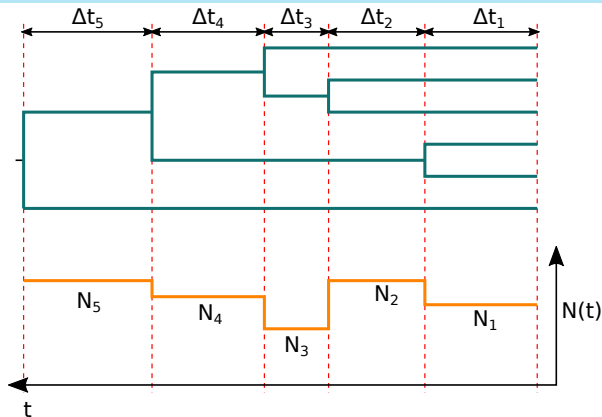
Birth-death vs Coalescent

References

- Assume population has distinct constant values in each interval between coalescent events.
- Can obtain a separate ML estimate for each population size.

Non-parametric population dynamics inference

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

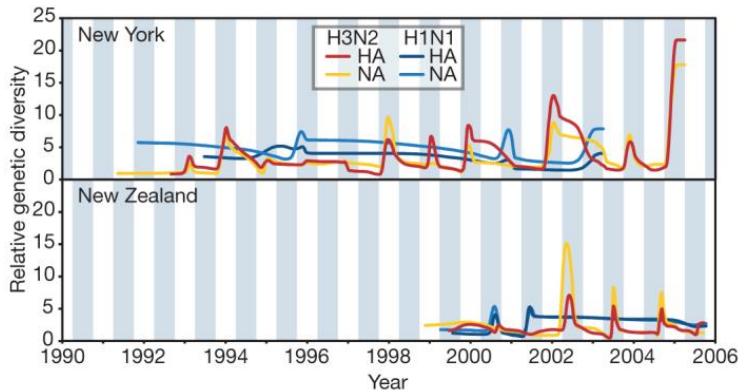
Birth-death vs Coalescent

References

- Assume population has distinct constant values in each interval between coalescent events.
- Can obtain a separate ML estimate for each population size.
- Resulting population function estimate is the “skyline plot” [Pybus et al., 2000].

Example: N_e dynamics for influenza (H3N2)

CB



[Rambaut et al., 2008]

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

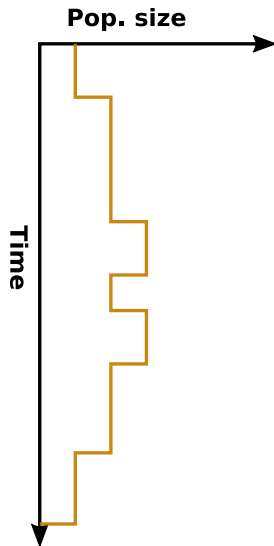
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Sampled Tree

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

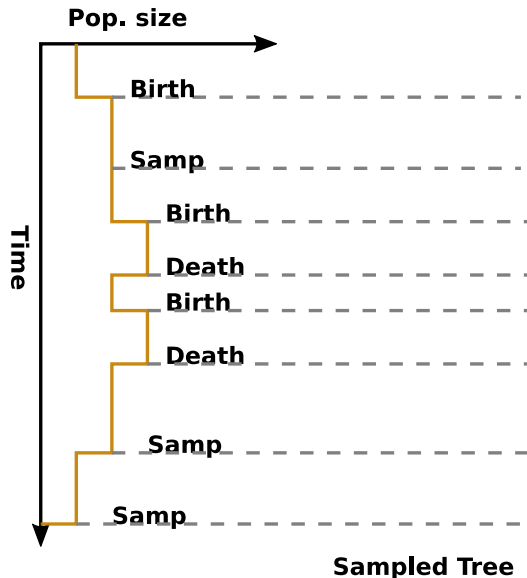
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

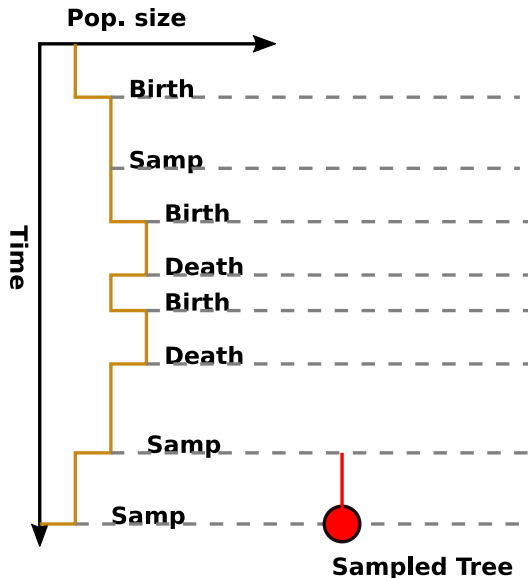
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

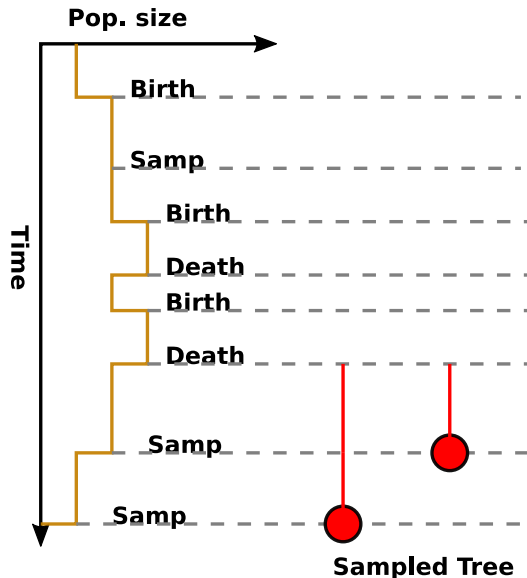
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

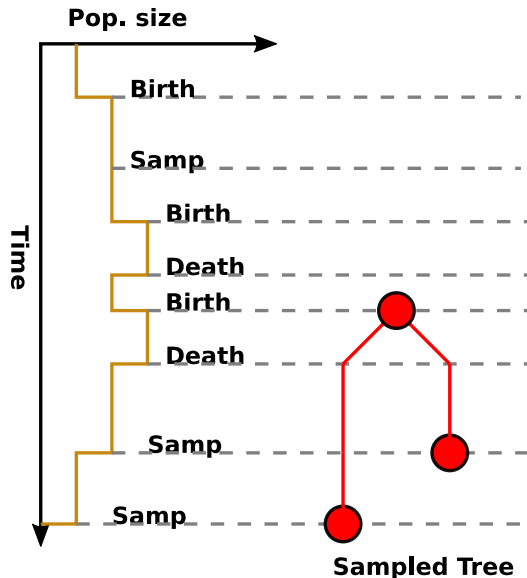
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

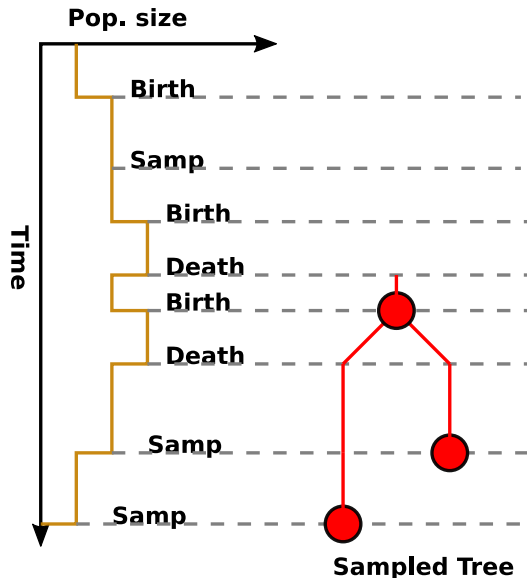
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

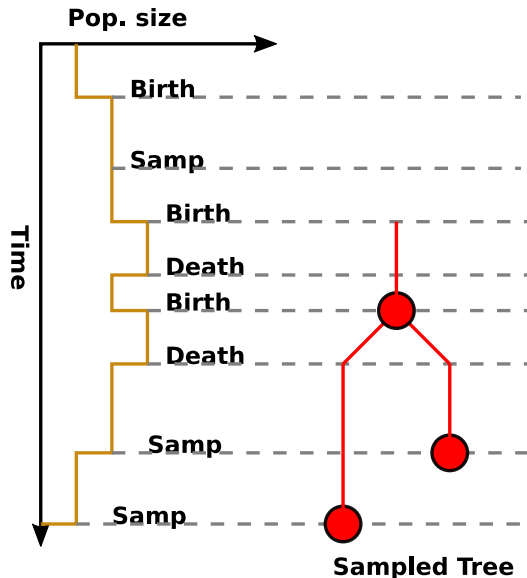
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

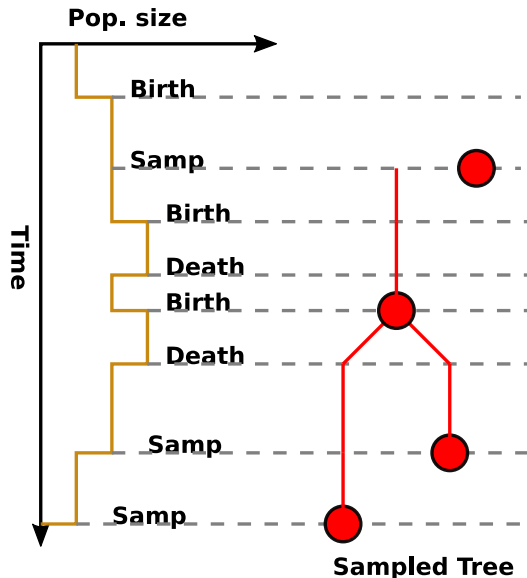
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

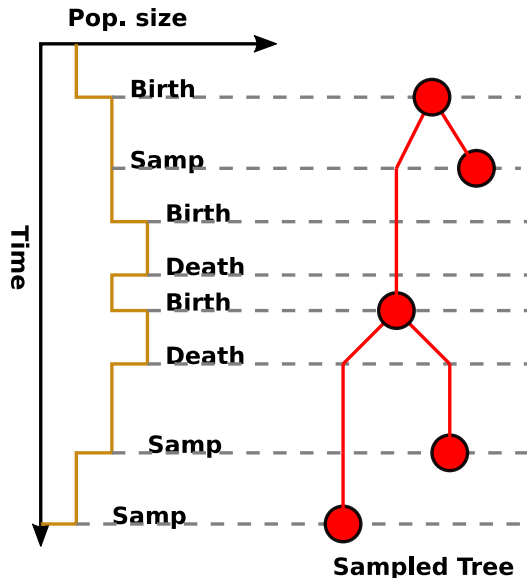
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

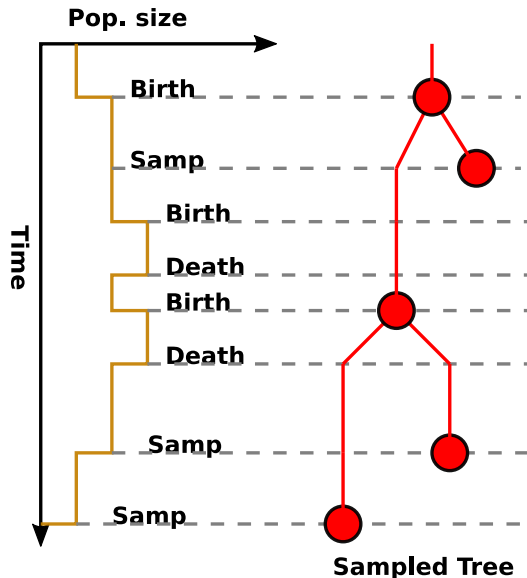
Application to HCV

Birth-death vs Coalescent

References

A connection to birth-death models

CB



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.
- ▶ Birth events occur at time t with the overall rate of $\beta I(t)$.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.
- ▶ Birth events occur at time t with the overall rate of $\beta I(t)$.
- ▶ Every birth is a potential coalescence between sampled lineages.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.
- ▶ Birth events occur at time t with the overall rate of $\beta I(t)$.
- ▶ Every birth is a potential coalescence between sampled lineages.
- ▶ Probability of choosing a sampled lineage pair is $\binom{k}{2} / \binom{I(t)}{2}$.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.
- ▶ Birth events occur at time t with the overall rate of $\beta I(t)$.
- ▶ Every birth is a potential coalescence between sampled lineages.
- ▶ Probability of choosing a sampled lineage pair is $\binom{k}{2} / \binom{I(t)}{2}$.
- ▶ Approximate coalescence rate is

$$\beta I(t) \frac{k(k-1)}{I(t)(I(t)-1)} \simeq \binom{k}{2} \frac{2\beta}{I(t)}$$

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ It is possible to develop coalescent distributions that approximate the probability density of sampled phylogenies generated by birth-death processes. [Volz et al., 2009]
- ▶ Assume the approximate ODE solution $I(t) = I(T)e^{(\beta-\delta)(T-t)}$ for the linear birth-death process (last lecture) is correct.
- ▶ Birth events occur at time t with the overall rate of $\beta I(t)$.
- ▶ Every birth is a potential coalescence between sampled lineages.
- ▶ Probability of choosing a sampled lineage pair is $\binom{k}{2} / \binom{I(t)}{2}$.
- ▶ Approximate coalescence rate is

$$\beta I(t) \frac{k(k-1)}{I(t)(I(t)-1)} \simeq \binom{k}{2} \frac{2\beta}{I(t)}$$

- ▶ One can use this coalescence rate to compute an approximation to $P(\mathcal{T}|\beta, \delta, T)$.

Coalescent approximation of birth-death models

- Quality of approximation depends heavily on how well birth-death population dynamics are approximated by deterministic ODE solution.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

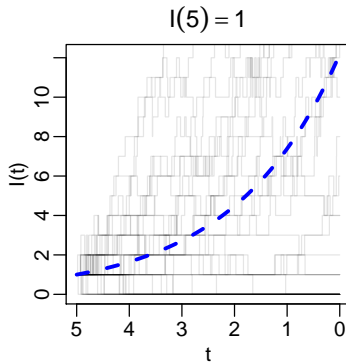
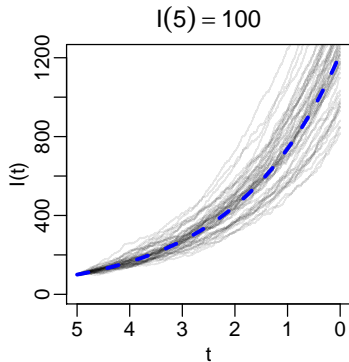
Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- Quality of approximation depends heavily on how well birth-death population dynamics are approximated by deterministic ODE solution.



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

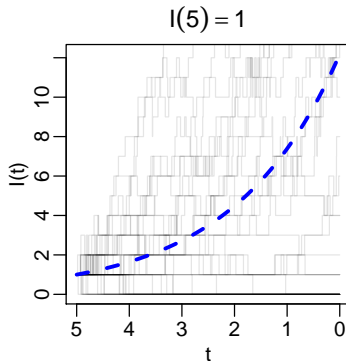
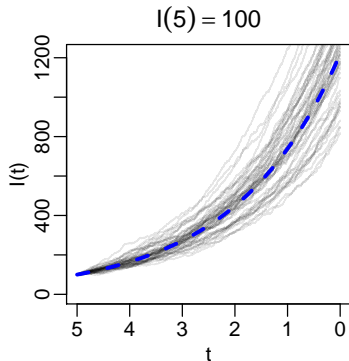
Application to HCV

Birth-death vs Coalescent

References

Coalescent approximation of birth-death models

- Quality of approximation depends heavily on how well birth-death population dynamics are approximated by deterministic ODE solution.



- This approximation can perform very poorly when population size is small, *as it always is at the start of an epidemic*.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

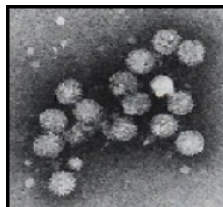
Application to HCV

Birth-death vs Coalescent

References

Hepatitis C (HCV)

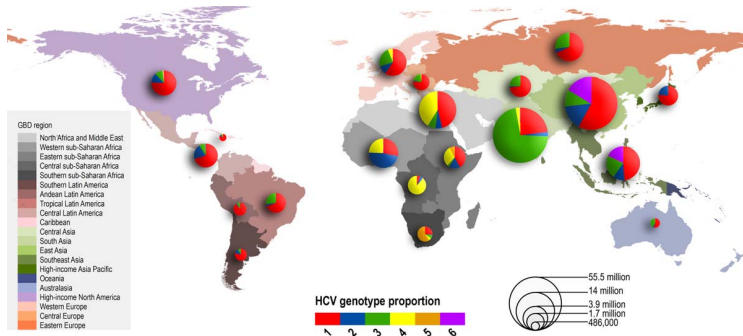
- ▶ Identified in 1989
- ▶ 9.6kb single-stranded RNA genome
- ▶ Tissue culture system only recently developed



Prevalence of HCV world wide

CB

> 185 million people infected worldwide



- ▶ ~80% infections are chronic
- ▶ Liver cirrhosis and cancer risk
- ▶ No protective immunity?

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

By percutaneous exposure to infected blood

- ▶ Blood transfusion / blood products
- ▶ Sexual transmission
- ▶ Vertical transmission (mother - child)
- ▶ Injections with contaminated needles



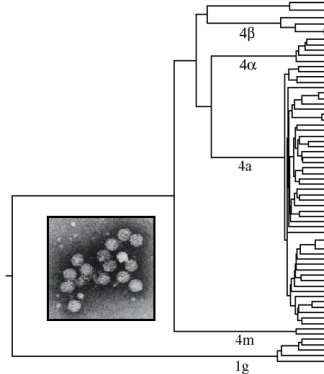
Inference of HCV spread in Egypt

CB

Egyptian HCV gene sequences

n=61

E1 gene, 411bp



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Inference of HCV spread in Egypt

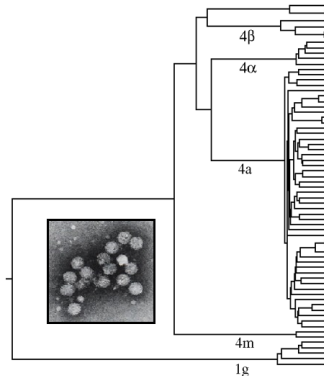
CB

Egyptian HCV gene sequences

n=61

E1 gene, 411bp

- All sequences are contemporaneous, i.e. sampled at the same time



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Inference of HCV spread in Egypt

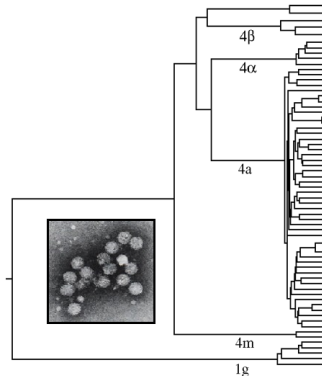
CB

Egyptian HCV gene sequences

n=61

E1 gene, 411bp

- ▶ All sequences are contemporaneous, i.e. sampled at the same time
- ▶ Egypt has the highest prevalence of HCV worldwide (10-20%)



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Inference of HCV spread in Egypt

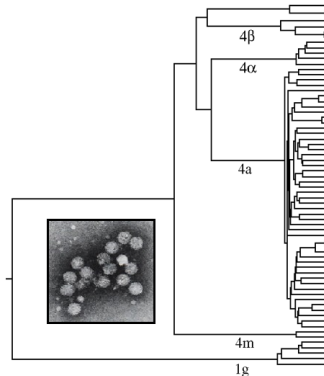
CB

Egyptian HCV gene sequences

n=61

E1 gene, 411bp

- ▶ All sequences are contemporaneous, i.e. sampled at the same time
- ▶ Egypt has the highest prevalence of HCV worldwide (10-20%)
- ▶ But low prevalence in neighbouring states



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Inference of HCV spread in Egypt

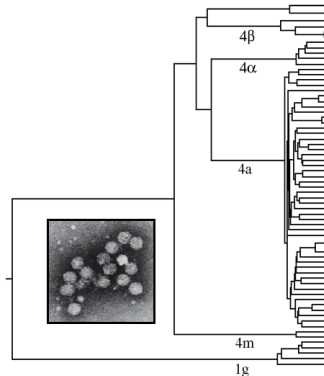
CB

Egyptian HCV gene sequences

n=61

E1 gene, 411bp

- ▶ All sequences are contemporaneous, i.e. sampled at the same time
- ▶ Egypt has the highest prevalence of HCV worldwide (10-20%)
- ▶ But low prevalence in neighbouring states
- ▶ Why is Egypt so seriously affected?



Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Infected population size $N(t)$ of HCV in Egypt

CB

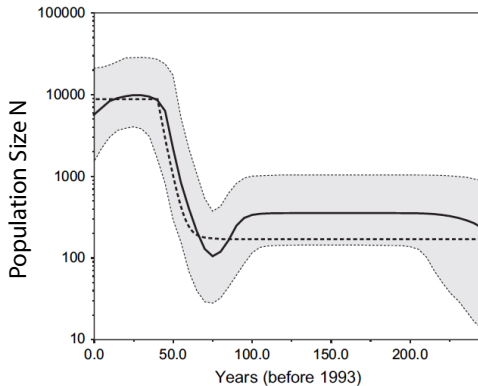


Figure adapted from [Drummond et al., 2005]

Peak in epidemic spread in first half of 20th century based on the coalescent.

[The estimates are Bayesian estimates (rather than Maximum Likelihood), which I will explain next week!]

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

$R(t) = \beta(t)/\delta(t)$ of HCV in Egypt

CB

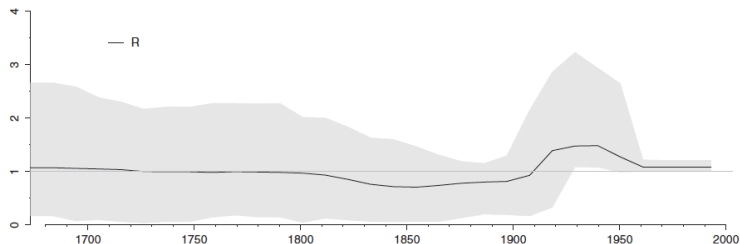


Figure adapted from [Stadler et al., 2013]

Peak in epidemic spread in first half of 20th century based on the birth-death model.

[The estimates are Bayesian estimates (rather than Maximum Likelihood), which I will explain next week!]

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

What caused this peak in infections around 70 years ago?

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ In first half of the 20th century schistosomal (worm) infections were treated with injection therapy.
- ▶ Needles were frequently reused and thus possibly contaminated.
- ▶ In the 1970s, antischistosomal injection therapy was gradually replaced by oral therapy.
- ▶ The use of antischistosomal injections with potentially contaminated needles coincides with the time period of HCV spread, and is thought to explain the high prevalence of HCV in Egypt.

How do I choose?

Birth-death

Coalescent

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

How do I choose?

Birth-death

Coalescent

- Parameters: transmission rates, removal rates, sampling rates/proportions.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

How do I choose?

Birth-death

- Parameters: transmission rates, removal rates, sampling rates/proportions.

Coalescent

- Parameters: effective population size (NOT actual population size).

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

How do I choose?

Birth-death

- ▶ Parameters: transmission rates, removal rates, sampling rates/proportions.
- ▶ Models sampling process (sample times/locations are data).

Coalescent

- ▶ Parameters: effective population size (NOT actual population size).

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

How do I choose?

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Birth-death

- Parameters: transmission rates, removal rates, sampling rates/proportions.
- Models sampling process (sample times/locations are data).

Coalescent

- Parameters: effective population size (NOT actual population size).
- Assumes number of sampled lineages are small ($k \ll N$).

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- Accounts for stochastic variability in population dynamics.

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Accounts for stochastic variability in population dynamics.
- ▶ Generally easier interpretation of parameters.

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Accounts for stochastic variability in population dynamics.
- ▶ Generally easier interpretation of parameters.
- ▶ Uses information about sampling.

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Accounts for stochastic variability in population dynamics.
- ▶ Generally easier interpretation of parameters.
- ▶ Uses information about sampling.

Disadvantages

- ▶ Sensitive to unmodeled changes in sampling fractions.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Accounts for stochastic variability in population dynamics.
- ▶ Generally easier interpretation of parameters.
- ▶ Uses information about sampling.

Disadvantages

- ▶ Sensitive to unmodeled changes in sampling fractions.
- ▶ Difficult (currently!) to extend to complex population models.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- Generally fast likelihood calculations.

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Generally fast likelihood calculations.
- ▶ Easy to extend to complex population dynamics (see [Volz, 2012]).

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Generally fast likelihood calculations.
- ▶ Easy to extend to complex population dynamics (see [Volz, 2012]).
- ▶ Naturally account for incomplete sampling.

Disadvantages

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Generally fast likelihood calculations.
- ▶ Easy to extend to complex population dynamics (see [Volz, 2012]).
- ▶ Naturally account for incomplete sampling.

Disadvantages

- ▶ Sensitive to uncertainty in population dynamics at high sampling ([Boskova et al., 2014], Masters thesis in our group).

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Advantages

- ▶ Generally fast likelihood calculations.
- ▶ Easy to extend to complex population dynamics (see [Volz, 2012]).
- ▶ Naturally account for incomplete sampling.

Disadvantages

- ▶ Sensitive to uncertainty in population dynamics at high sampling ([Boskova et al., 2014], Masters thesis in our group).
- ▶ Sensitive to hidden population structure and nonrandom sampling.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- Both birth-death and coalescent models probabilistically relate a population's demography to its phylogenetic history.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Both birth-death and coalescent models probabilistically relate a population's demography to its phylogenetic history.
- ▶ Both allow for model-based inference of demographic and epidemiological parameters, but may differ in their parameterization.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ▶ Both birth-death and coalescent models probabilistically relate a population's demography to its phylogenetic history.
- ▶ Both allow for model-based inference of demographic and epidemiological parameters, but may differ in their parameterization.
- ▶ Which model to use depends on which assumptions you are willing to make and what you want to infer.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References

- ❓ Under the Wright-Fisher model, how many generations do we have to go back before we find the common ancestor of a pair of genes sampled from a haploid population of size N ?
- ❓ Suppose you had a tree inferred using present-day samples from a population that experienced a severe bottleneck in its recent past. How and why would this bottleneck likely affect our ability to infer ancestral population dynamics?
- ❓ Imagine spreading a Wright-Fisher population across islands in an archipelago, so that movement between the islands is restricted but within each island the population is “well-mixed”. Qualitatively, how would you expect this population structure to influence estimates of the effective population size?

- Boskova, V., Bonhoeffer, S., and Stadler, T. (2014). Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLOS Comput Biol*, 10(11):e1003913.
- Drummond, A. J., Rambaut, A., Shapiro, B., and Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular biology and evolution*, 22(5):1185–1192.
- Felsenstein, J. (2004). *Inferring phylogenies*. Sinauer associates Sunderland.
- Kingman, J. F. (1982). On the genealogy of large populations. *Journal of Applied Probability*, pages 27–43.
- Pybus, O. G., Rambaut, A., and Harvey, P. H. (2000). An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics*, 155(3):1429–1437.
- Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., and Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza a virus. *Nature*, 453:615–619.
- Stadler, T., Kühnert, D., Bonhoeffer, S., and Drummond, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in hiv and hepatitis c virus (hcv). *Proceedings of the National Academy of Sciences*, 110(1):228–233.
- Volz, E. M. (2012). Complex population dynamics and the coalescent under neutrality. *Genetics*, 190(1):187–201.
- Volz, E. M., Kosakovsky Pond, S. L., Ward, M. J., Leigh Brown, A. J., and Frost, S. D. W. (2009). Phylodynamics of infectious disease epidemics. *Genetics*, 183(4):1421–1430.

Coalescent Theory

Introduction

Wright-Fisher process

Coalescent trees

Population dynamics

Application to HCV

Birth-death vs Coalescent

References