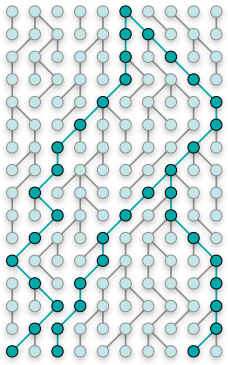


Viral epidemiology and the Coalescent



Philippe Lemey and
Marc A. Suchard

Department of Microbiology and Immunology
K.U. Leuven, and

Departments of Biomathematics and Human Genetics
David Geffen School of Medicine at UCLA
Department of Biostatistics
UCLA School of Public Health

Population genetics



- ✓ population size
- ✓ change in population size
- ✓ migration and selection



Genealogy-based population genetics

Phylogeny



Genealogy



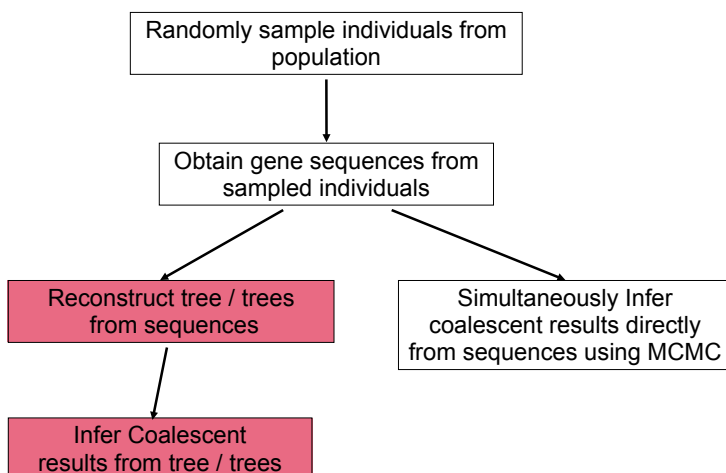
Genealogy-based population genetics

- the Coalescent is a **model of the ancestral relationships** of a random sample of individuals taken from a large background population.
- the Coalescent describes a **probability distribution/density** on ancestral genealogies (trees) given a population history.
therefore the Coalescent can convert information from ancestral genealogies into information about population history and vice versa.
- the Coalescent is a model of ancestral genealogies, not sequences, and its simplest form **assumes neutral evolution**.

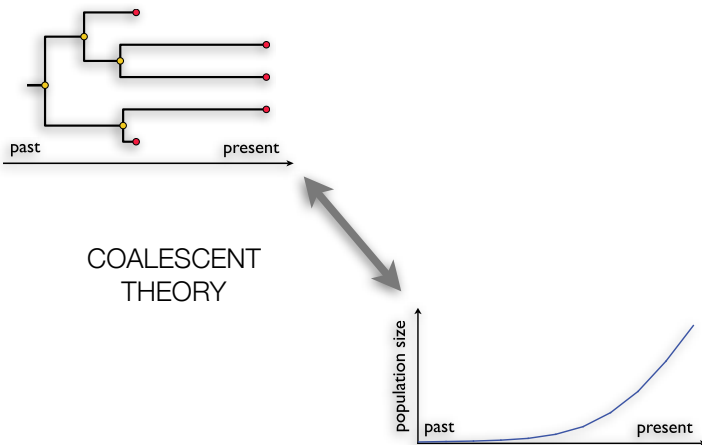
Demographic inference

- change in population size through time
- applications include
 - reconstructing infectious disease epidemics
 - investigating viral dynamics within hosts
 - using viral sequences as genetic markers for their wild hosts and the host demographics
 - population bottlenecks caused by
 - › change in climate/environment? Aridification, ices ages, et cetera
 - › competition with other species? Humans?
 - › transmission bottlenecks in viruses

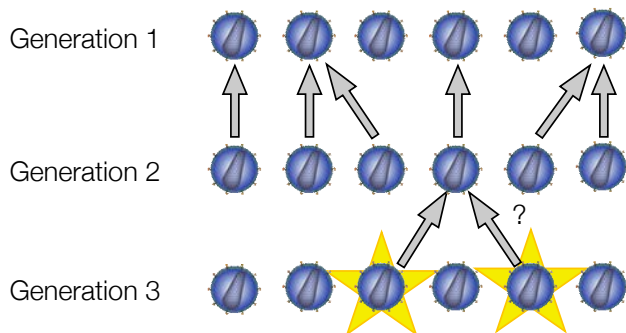
Information pipe-line



Coalescent inference

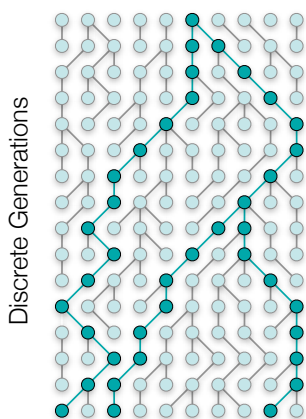


A model of virus reproduction

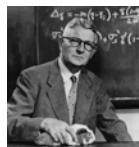


- For a randomly chosen pair of individuals, the probability that they share a common ancestor in the previous generation is $1/N$

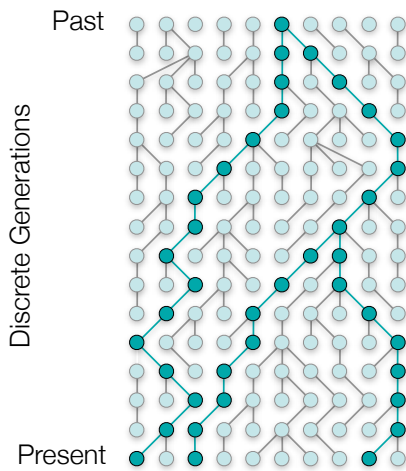
Wright-Fisher reproduction model



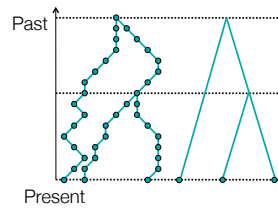
- A constant population size of N individuals (usually $2N$)
- Each new (non-overlapping) generation “chooses” its parent from the previous generation at random with replacement
- No geographic/social structure, no recombination, no selection



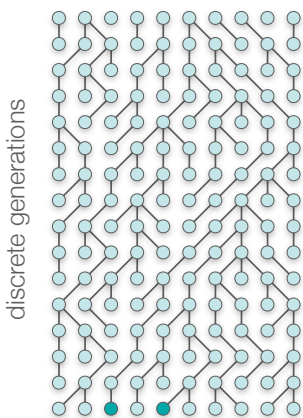
A sample genealogy from an idealized Wright-Fisher population



A sample genealogy of 3 sequences from a population ($N = 10$).



Kingman (discrete-time) coalescent



- 2 individuals coalesce in 1 generation w.p. $\frac{1}{N}$

- 2 individuals coalesce in j generations w.p.

$$\frac{1}{N} \left(1 - \frac{1}{N}\right)^{j-1}$$

- k individuals coalesce in j generations w.p.

$$\binom{k}{2} \frac{1}{N} \left[1 - \binom{k}{2} \frac{1}{N}\right]^{j-1}$$



Kingman (continuous-time) coalescent



- Let $t = j / N$ define a rescaled time in past, and
- Assume a sample of n individuals with $n \ll N$
- Then, the waiting time for k individuals to have $k - 1$ ancestors

$$P(T_k \leq t) = 1 - e^{-\binom{k}{2}t}$$

- Exponential (memoryless), defines a continuous-time Markov chain

$$E(T_k) = \frac{2N}{k(k-1)}$$

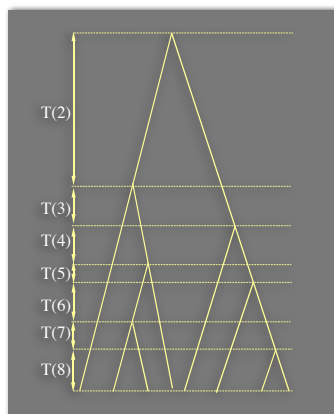
Kingman (1982) *J Appl Prob* 19A, 27-42
 Kingman (1982) *Stoch Proc Appl* 13, 235-48

Kingman coalescent: CTMC

- the number of sampled lineages decreases by one at each coalescence
- the process continues until the most recent common ancestor (MRCA) is reached.
- What is the expected time to MRCA?

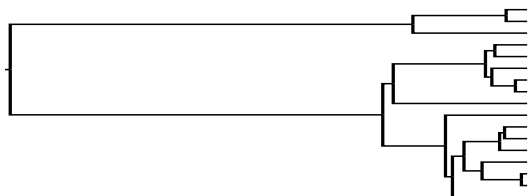
$$E\left(\sum_{k=2}^n T_k\right) = \sum_{k=2}^n E(T_k) \\ = \sum_{k=2}^n \frac{2N}{k(k-1)} = 2N \left(1 - \frac{1}{n}\right)$$

- $t_{\text{MRCA}} < 2 \times E(T_2)$ (??)



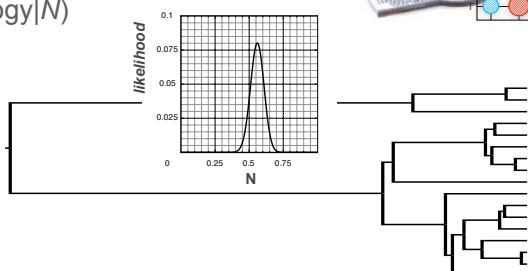
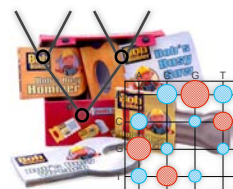
Kingman coalescent: its use here

- If we obtain a genealogy for a sample of individuals from a population
- We can calculate the probability $P(\text{genealogy}|N)$



Kingman coalescent: its use here

- If we **reconstruct** a genealogy for a sample of **gene sequences** from a population
- We can calculate the probability $P(\text{genealogy}|N)$



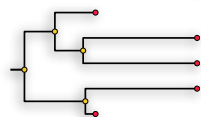
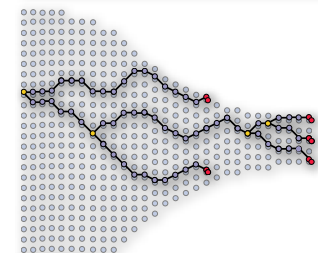
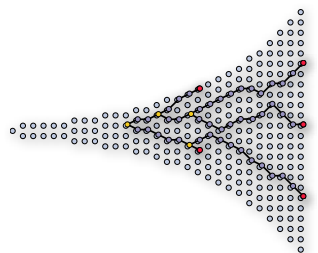
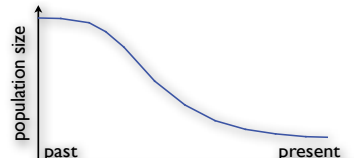
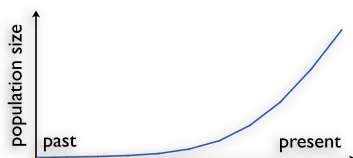
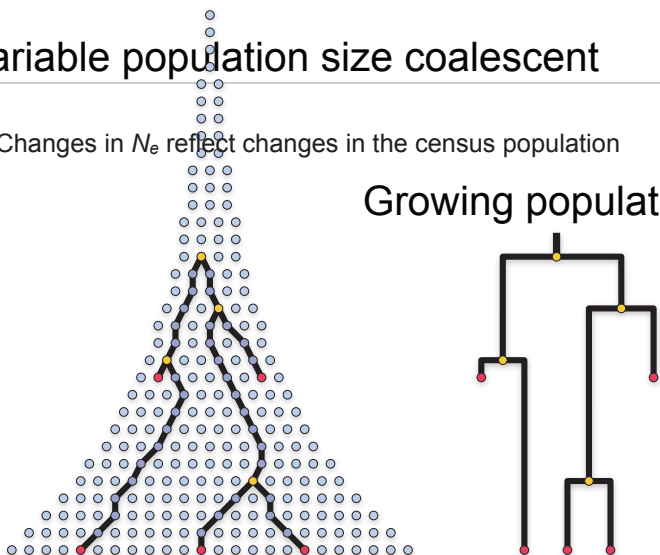
Solution: *effective* population size

- We consider an abstract parameter, the effective population size (N_e),
- The N_e of a real biological population is the size of an idealized Fisher-Wright population that loses or gains genetic diversity at exactly the same rate
- N_e is generally smaller than the census population
- The coalescent N_e provides the time-to-ancestry distribution for a sample genealogy from a real population

Variable population size coalescent

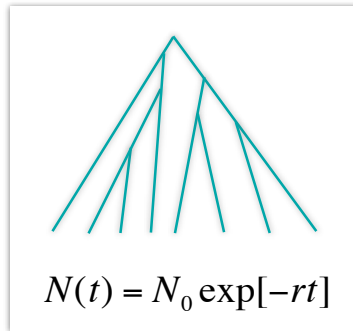
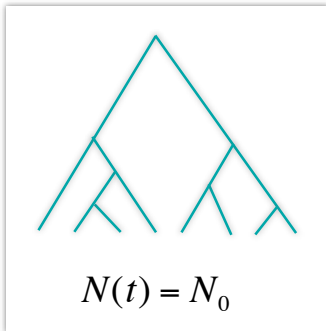
- Changes in N_e reflect changes in the census population

Growing population



Demographic models and tree shape

- The standard coalescent can be extended to accommodate various scenarios of demographic change through time



Demographic models and tree shape

- The standard coalescent can be extended to accommodate various scenarios of demographic change through time

