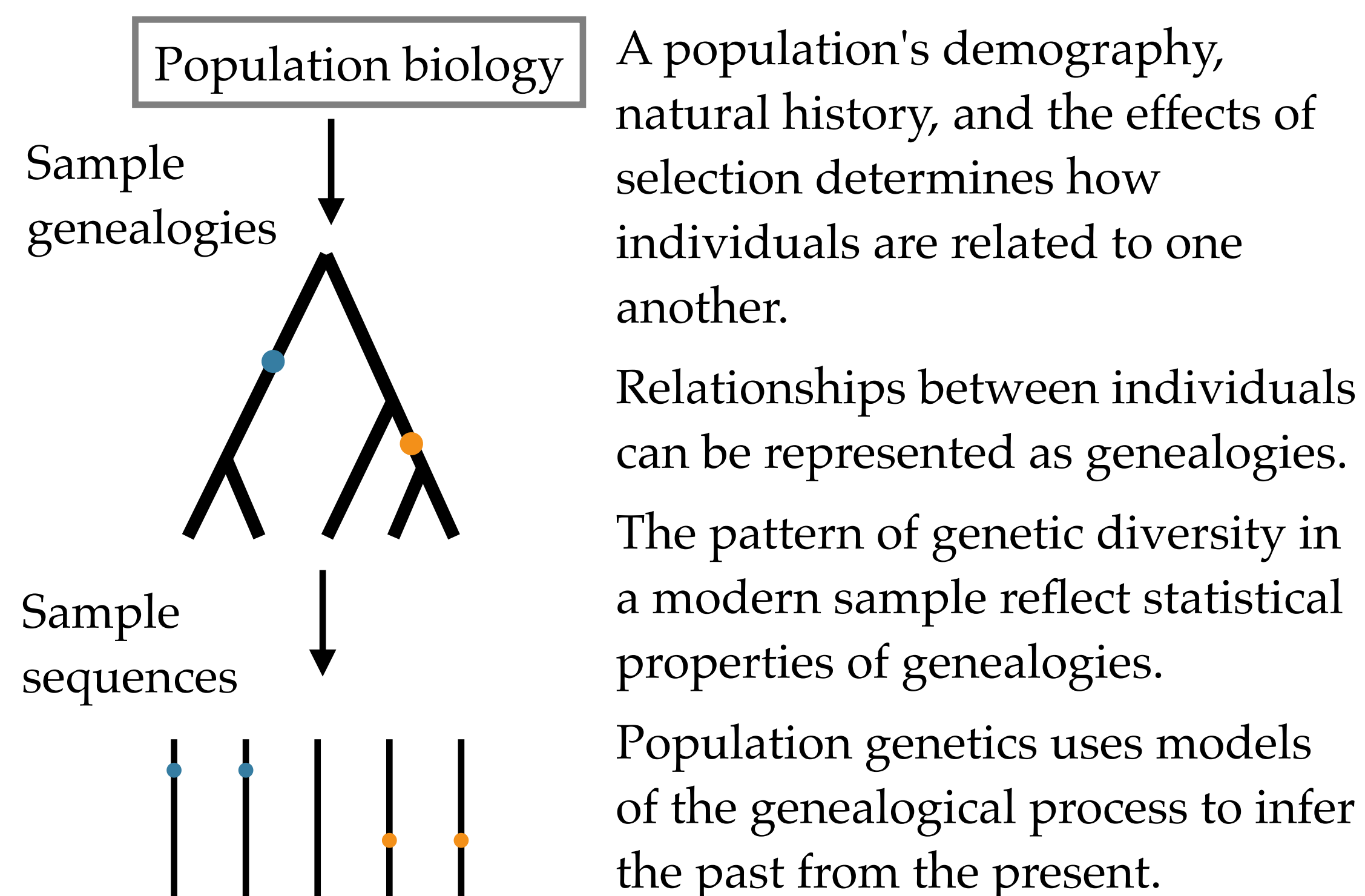


Distinguishing among coalescent models using two-site joint allele frequency spectra

Daniel P. Rice^{1*}, John Novembre¹, Michael M. Desai²

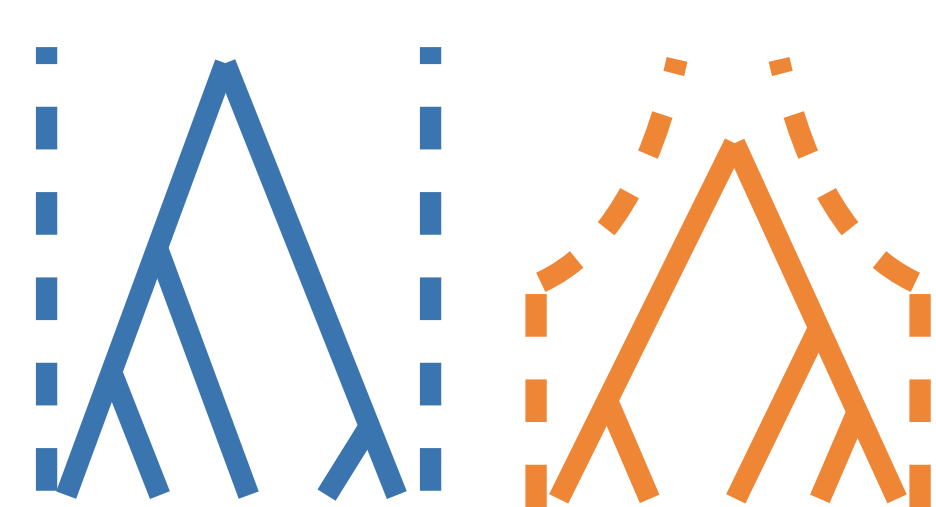
¹University of Chicago, ²Harvard University, *dpr@uchicago.edu

Genetic diversity depends on history through a genealogical process.



Signals of demographic and evolutionary history are confounded.

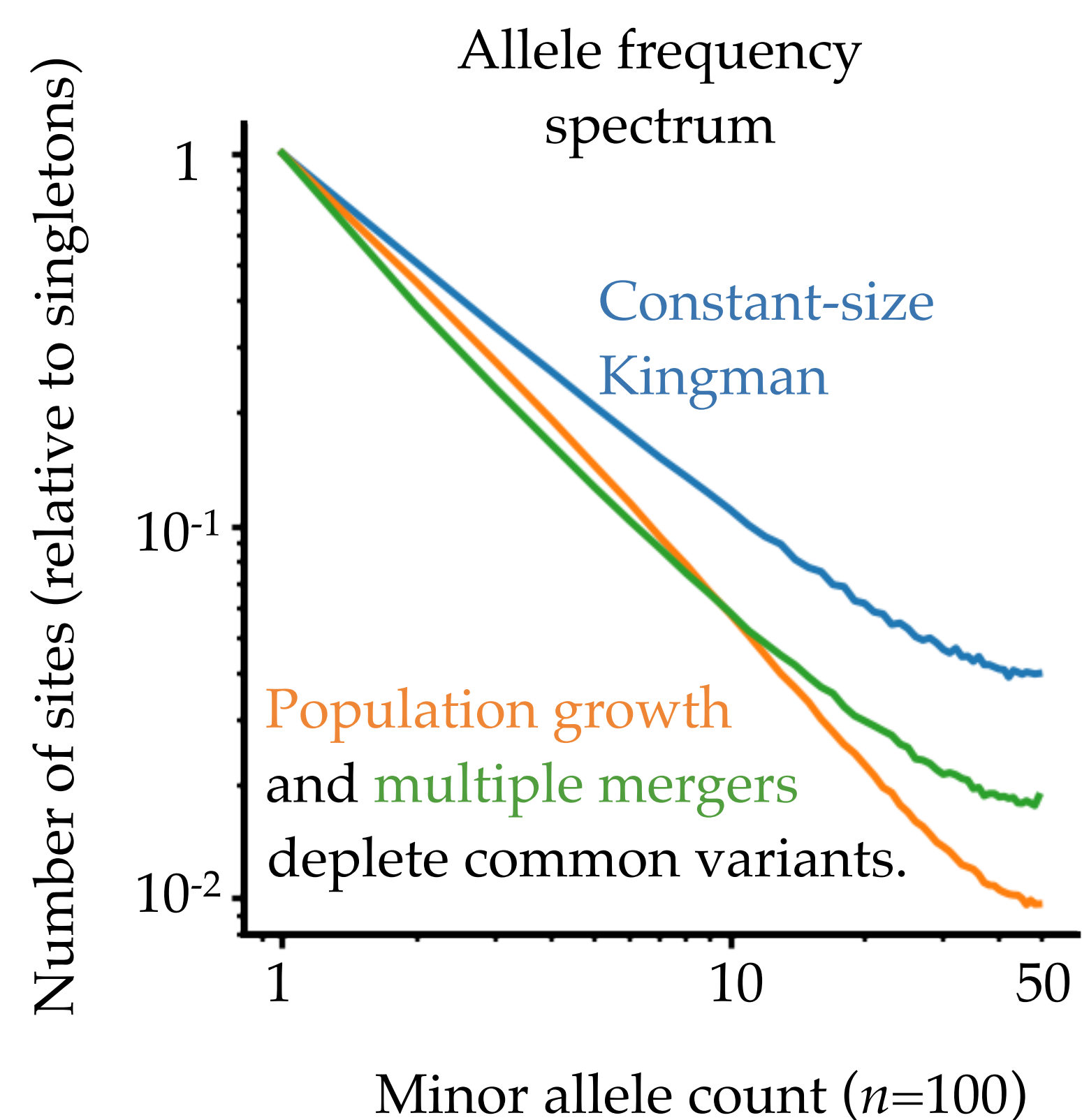
Kingman coalescent



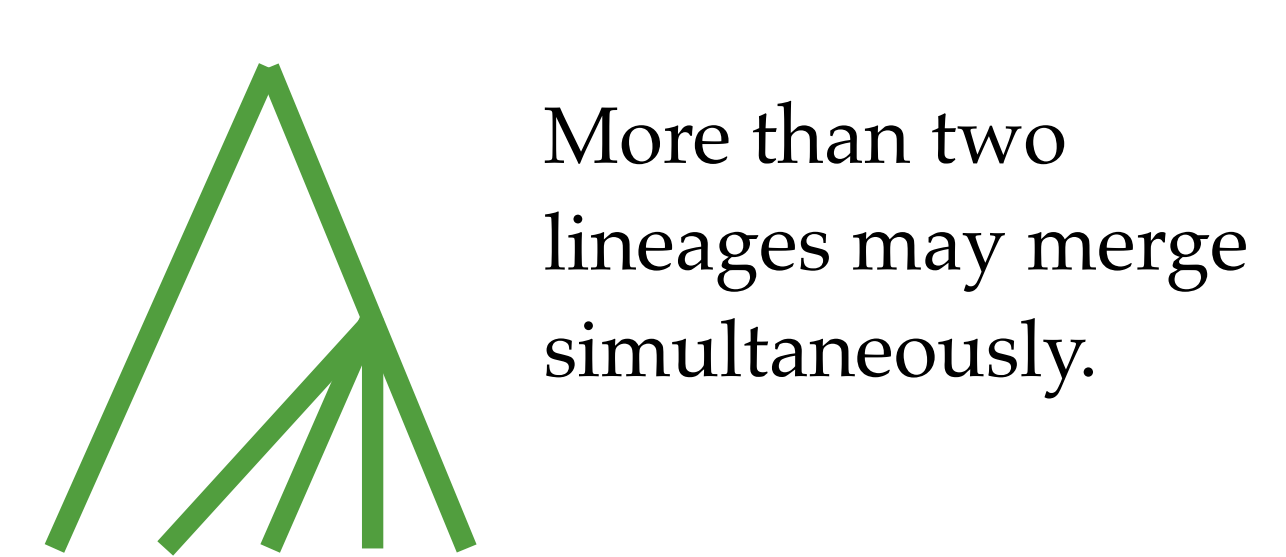
Only two lineages may merge at the same time.

Branch lengths depend on slow variations in the population size.

Demographic inference is possible from genetic diversity data.



Multiple-merger coalescent

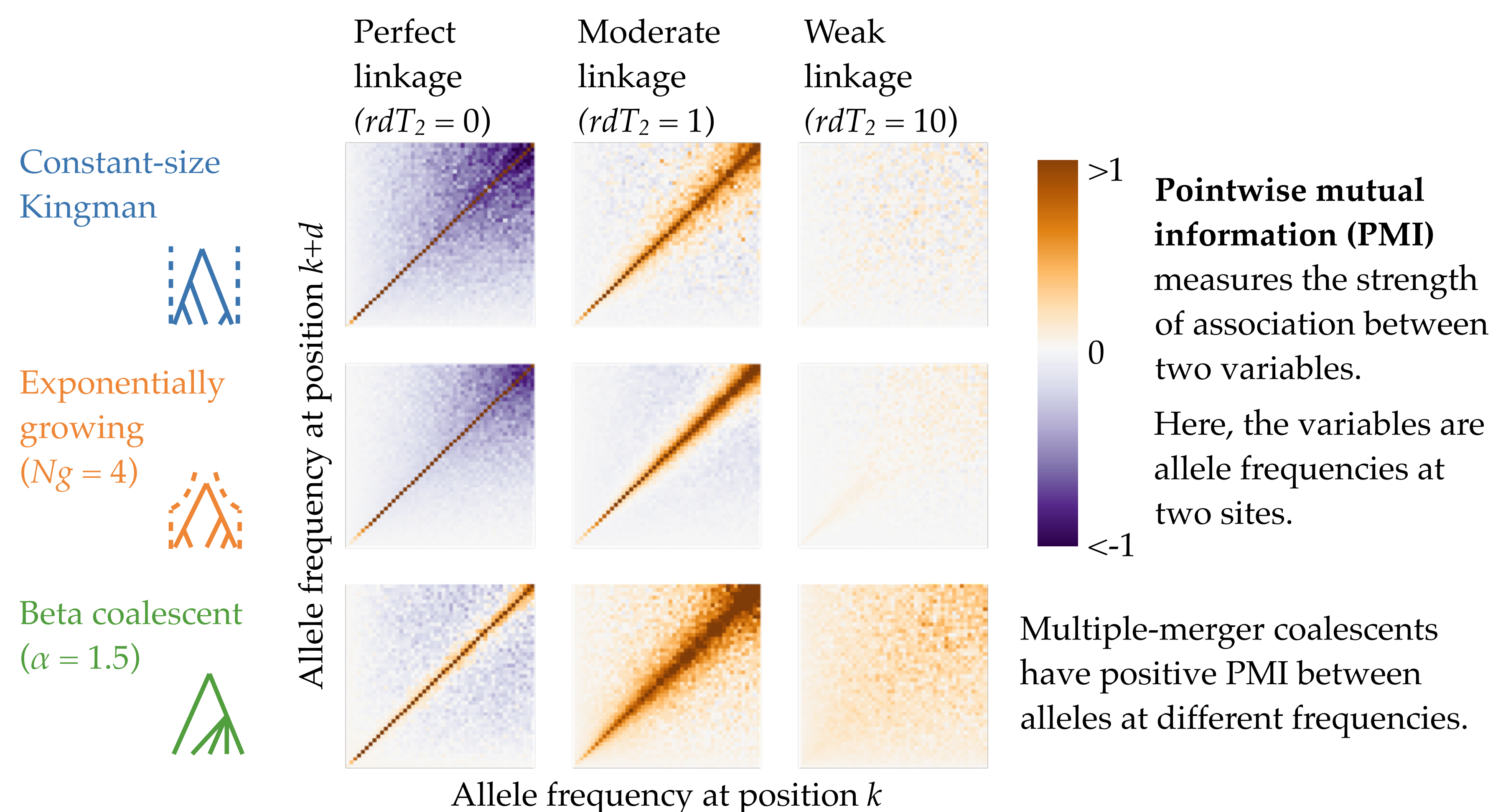


More than two lineages may merge simultaneously.

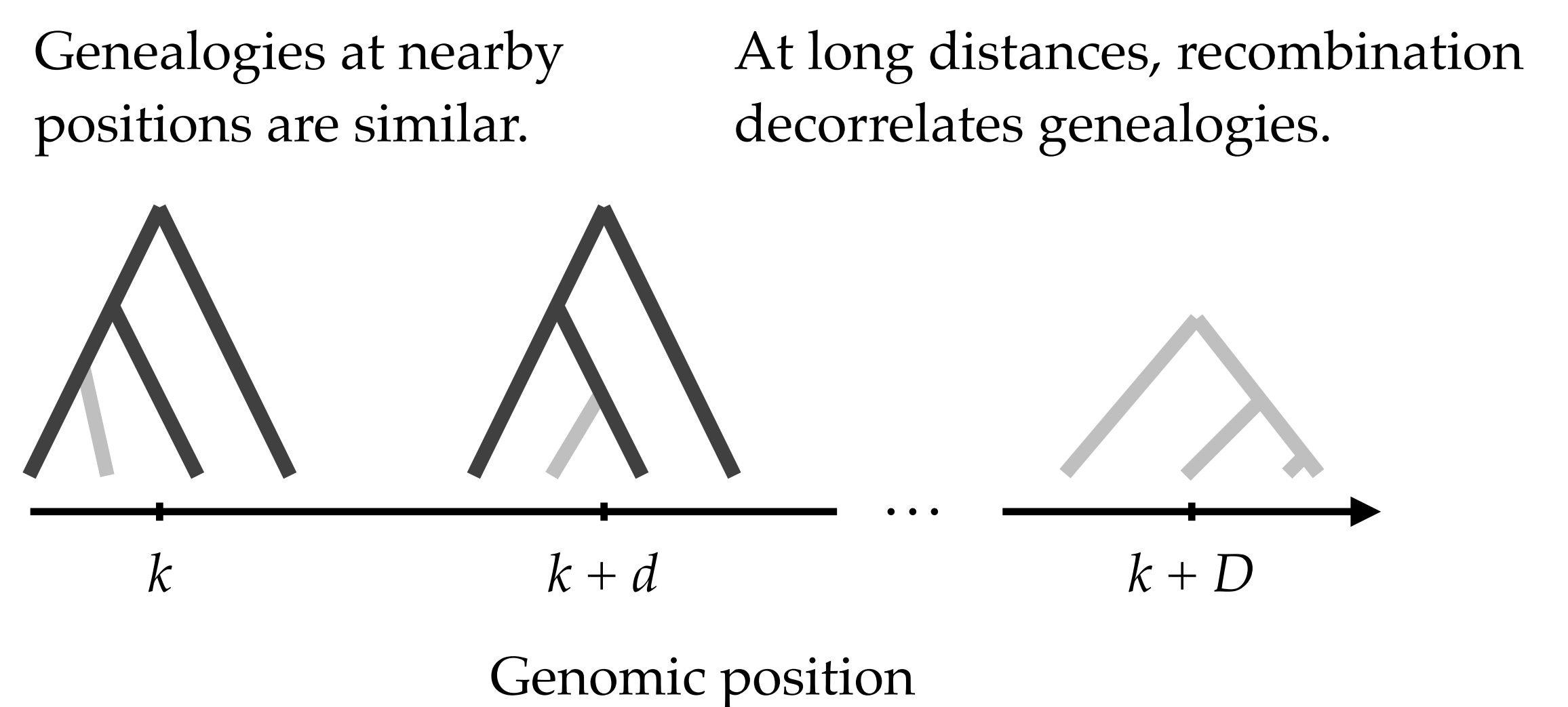
Branch lengths may depend on:

- rate of selective sweeps
- population variation in fitness
- rate of strong bottlenecks
- offspring number tail statistics

The joint allele frequency spectrum at pairs of linked sites distinguishes Kingman from multiple-merger coalescent models.



The joint allele frequency spectrum depends on the distance between two sites.



Multiple merger coalescents are generated by processes with multiple timescales.

The average pairwise diversity (π) depends on the mutation rate through:

$$\pi \propto \mu T_{\text{diversity}}.$$

The length scale of allele frequency correlations depends on the recombination rate through:

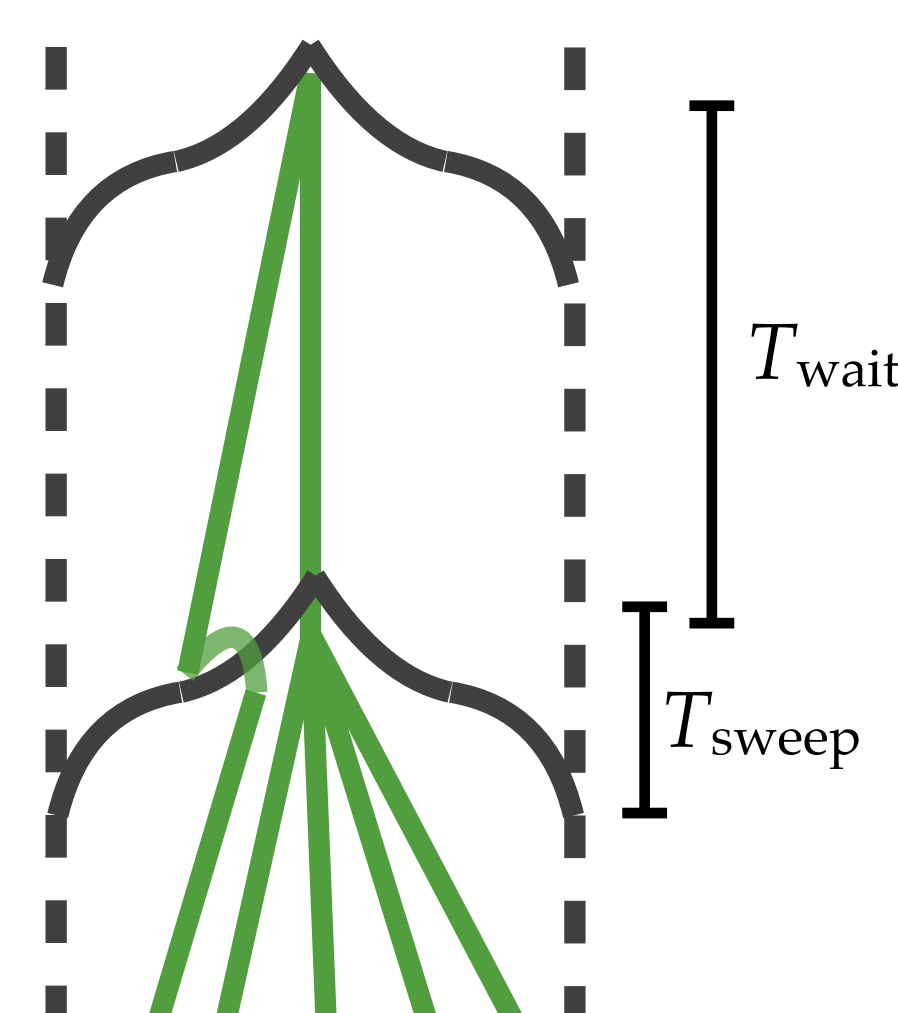
$$d \propto (r T_{\text{correlation}})^{-1}.$$



The Kingman coalescent is characterized by a single timescale:

$$T_{\text{diversity}} = T_{\text{correlation}} = T_2.$$

$$\text{So, } d \sim (r T_{\text{diversity}})^{-1}.$$



Multiple mergers generated by selective sweeps at linked sites has a separation of timescales between:

$$T_{\text{diversity}} \propto T_{\text{wait}}$$

and

$$T_{\text{correlation}} \propto T_{\text{sweep}}.$$

$$\text{So, } d \gg (r T_{\text{diversity}})^{-1}.$$

Mutual information between allele frequencies persists at long distances in multiple merger coalescents.

