

Gene Trees vs. Species Trees

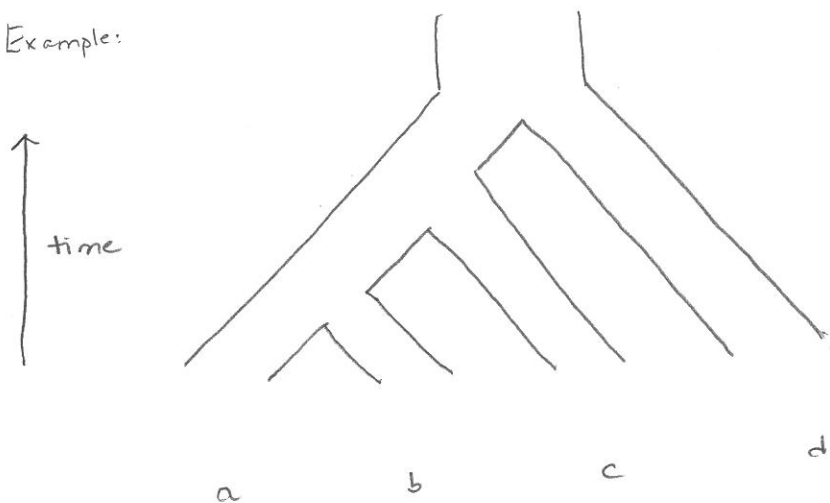
1

In the first lecture, it was mentioned that gene trees and species trees are NOT the same thing. (Yeast dataset, why? 13 weeks on gene trees, ...)

There are many possible explanations for this (LGT, hybridization, etc.), and INCOMPLETE LINEAGE SORTING (ILS). Many data analyses incorporate ILS now since it is modelled by the MULTISPECIES COALESCENT MODEL.

To start, it is important to note that species are made up of populations of individuals so that when gene lineages are sampled from individuals in populations, the gene history depicted as a genealogy or gene tree does not have to agree with the species tree.

Example:



Species tree parameter $\sigma =$
rooted, metric leaf-labelled tree

Within each species or
population are individuals
(or individual genes really)

Eg. Various gene trees can arise on σ due to ILS. (show)

since gene lineage coalescent events predate species divergence events.

\equiv ILS

It should be plausible, even at this stage in our development, that

any possible gene tree topology can arise on σ (and most metric gene trees, but there are some restrictions due to the branching

pattern on σ .)

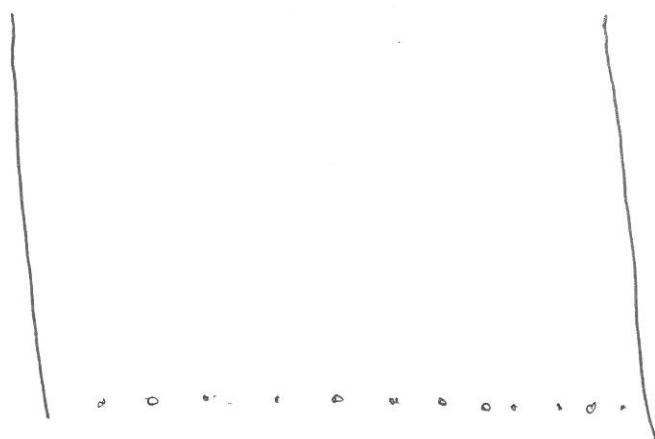
Explains some cause for gene tree discordance.

The model:

limit of Wright-Fisher model

(I) Kingman Coalescent

ONE POPULATION



- Single population
- time measured backwards from the present in number of generations
- N_t = population at time t

First example, $N_t = N_c = \text{constant}$. Assume $K=4$ individuals (gene lineages) A, B, C, D are sampled in the present. Then uniformly at random each lineage chooses its "parent" in the previous generation. This continues until all lineages have coalesced to a common ancestor.

Some assumptions:

The population is panmictic \equiv random mating

Each pair of lineages is equally like to coalesce (consequence really)

Prob(lineages A and B coalesce in previous generation) = _____
 Prob(" " " " \checkmark do not " " " ") = _____