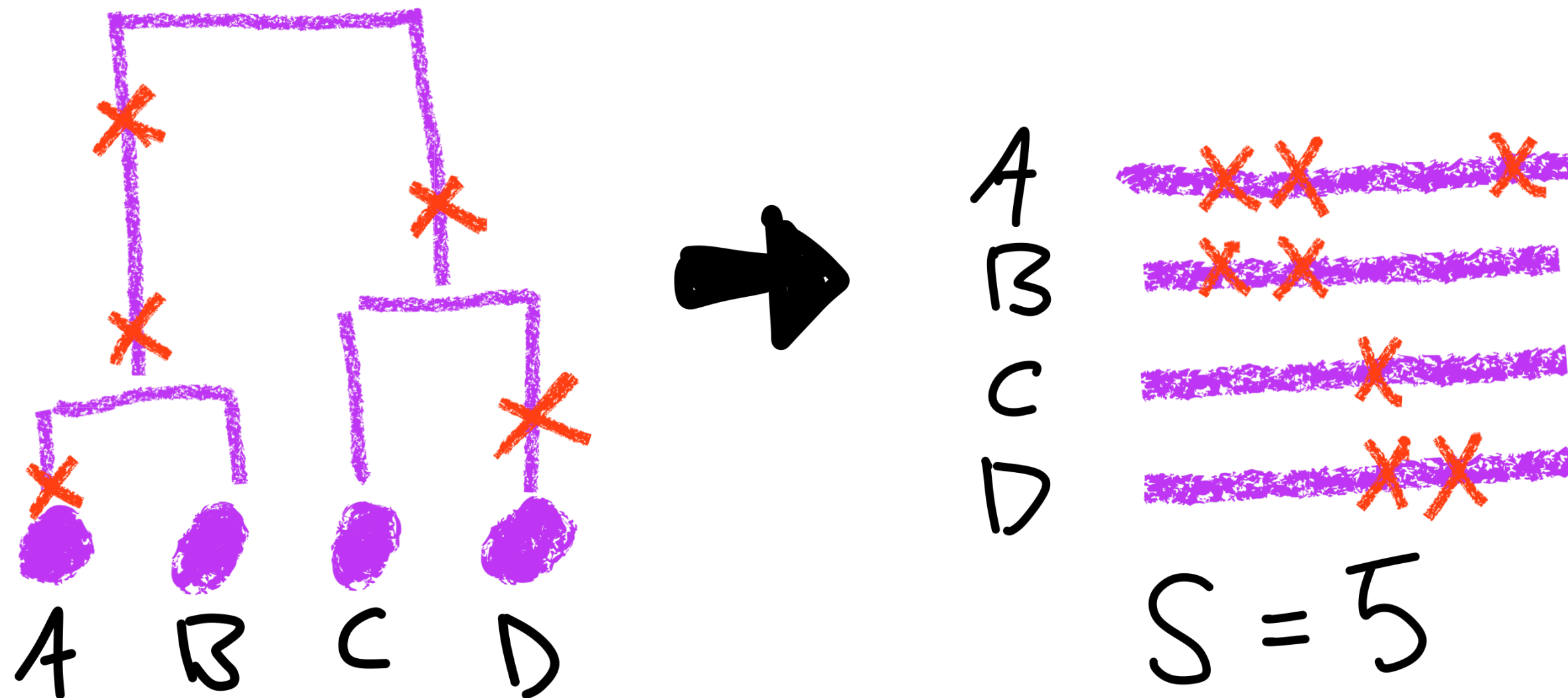


# Coalescent theory

## Genetic diversity

# segregating sites,  $S$ , equals # mutations in the sample's history (infinite sites approximation)



Constant N case :

$$\begin{aligned} E[S] &= \mu E[T_{\text{total}}] \\ &= \mu \sum_{i=2}^n i E[T_i] \\ &= \mu \sum_{i=2}^n i \frac{2N}{\binom{n}{2}} \\ &= 4\mu N \sum_{i=1}^{n-1} \frac{1}{i} \end{aligned}$$

*total branch length*

# Recap

Good to know for homework

Intercoalescent times are indep. exponential rvs

intercoalescent interval  $i$

diploid pop size

# sampled haplotypes

$$T_i \sim \exp\left(\frac{2N}{\binom{i}{2}}\right), \text{ for } i = n, n-1, \dots, 2$$

# mutations on branches are indep. Poisson rvs

$$k \sim \text{Pois}(ut)$$

# mutations

mutation rate

branch length