Coalescent theory
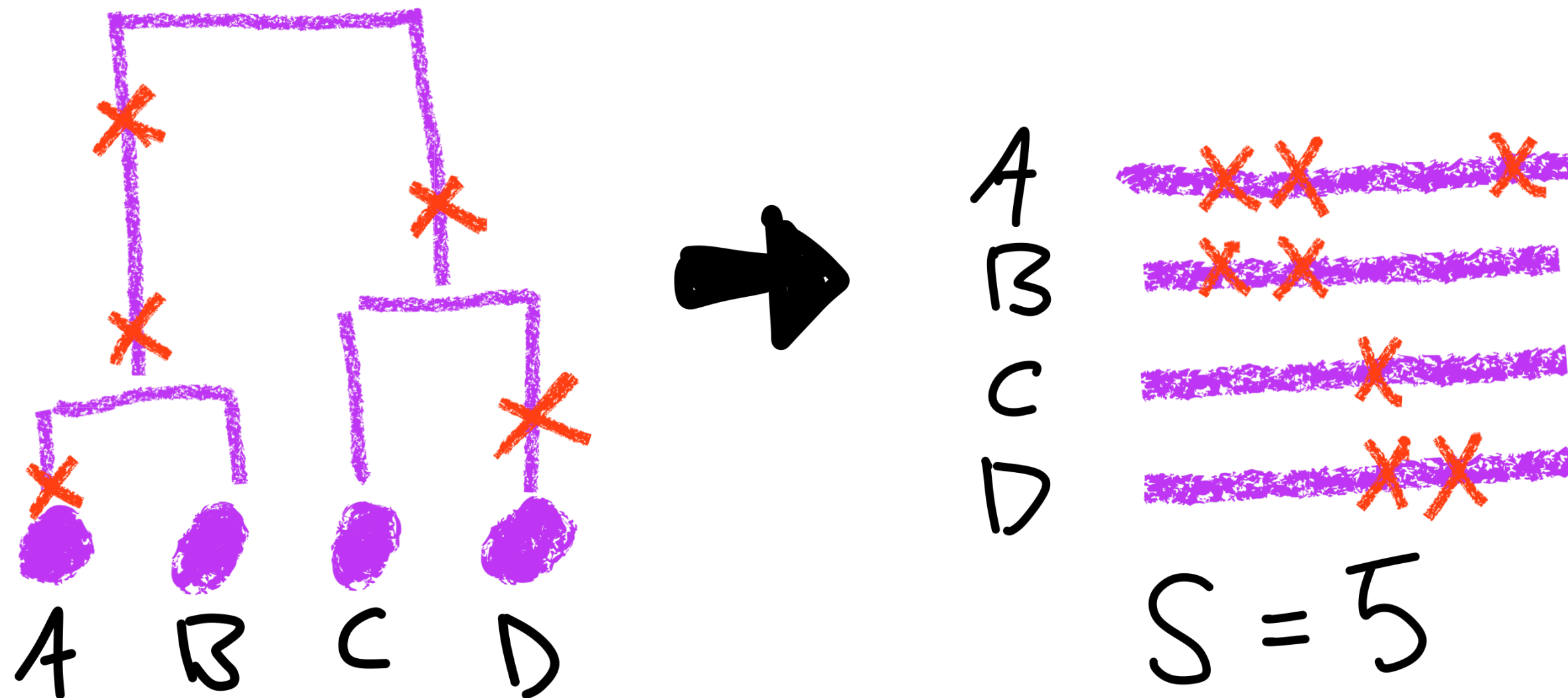**Genetic diversity**

segregating sites, S, equals # mutations in the samples history (infinite sites approximation)

$S = 5$

Constant N case:

$$\mathbb{E}[S] = \mu \, \mathbb{E}[T_{total}]$$ ← total branch length

$$= \mu \sum_{i=2}^{n} i \, \mathbb{E}[T_i]$$

$$= \mu \sum_{i=2}^{n} i \, \frac{2N}{\binom{i}{2}}$$

$$= 4 \mu N \sum_{i=1}^{n-1} \frac{1}{i}$$

Recap
Good to know for homework

Intercoalescent times are indep. exponential rvs

Intercoalescent interval i

diploid pop size

#sampled haplotypes

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

#mutations on branches are indep. Poisson rvs

$$k \sim \text{Pois}\left(\mu t\right)$$

#mutations

mutation rate

branch length