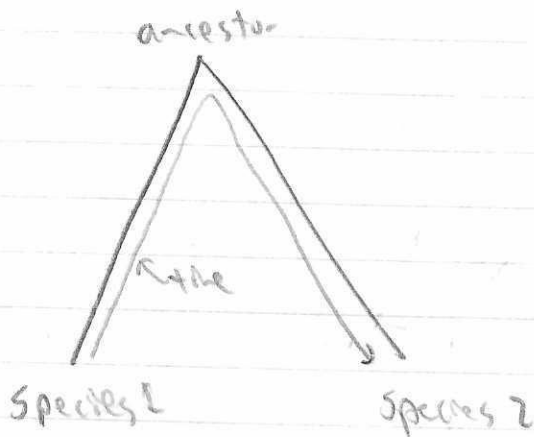


## §2.4: Molecular Evolution

Let  $\rho$  = rate of substitution

$$= \frac{\# \text{ mutations}}{t + ne}$$



Eg. 36 fixed differences between melanogaster and erecta  
23,000,000 b/t common ancestor

$$\rho = \frac{36}{2 \cdot 23,000,000} = 7.83 \cdot 10^{-7} = \frac{\text{substitutions}}{\text{year}} \quad \text{from 768 sites at locus}$$

substitution per site

$$\rho = \frac{7.83 \cdot 10^{-7}}{768} = 1 \cdot 10^{-9} \frac{\text{substitutions}}{\text{year site}}$$

Usual models of evolution in terms of generations, not years

— substitution includes drift that eliminated alleles

Gillespie gives  
different units

But this is  
in units of  
substitutions  
per generation

Surprising result:  $\rho = \mu$  ← does not depend on drift

Multiply average # mutations entering population  
w/ fraction of the mutations that fix

$$\rho = \frac{\text{mutation}}{\text{generation}} \cdot P_i(\text{fix})$$

recall allele freq

$$= 2Nu \cdot \frac{1}{2N}$$

$$\rho = 2Nu \cdot \frac{1}{2N} = \mu$$

Note:  $\rho = \mu$  just means  
substitutions per generation = mutations per generation  
Not years

Above was derivation in infinite allele model

New Model: In finite sites

Choose allele (single piece of DNA)

Look back in time

Each generation, probability of nucleotide change =  $\mu$

$t$  generations  $\Rightarrow$  about  $t\mu$  mutations on average

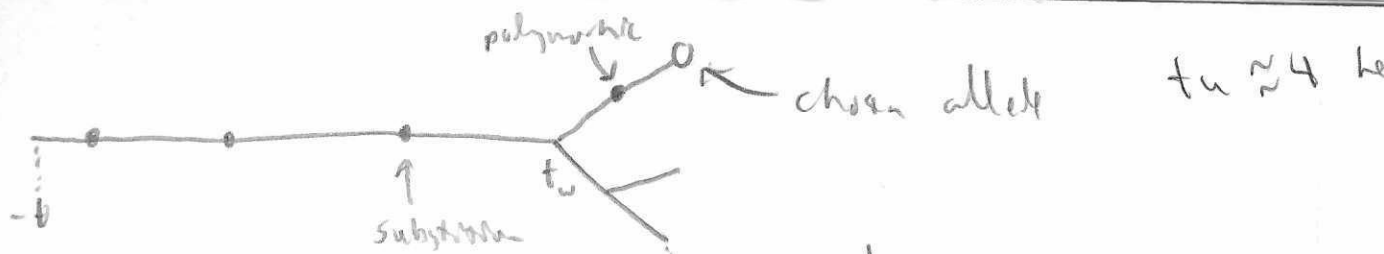
$$\rho = \frac{t\mu}{t} = \mu$$

Not obvious that

6CT

6TT

Still write A/T/G/C  
but each mutation  
brings about a new  
allele that eventually  
results in a sub

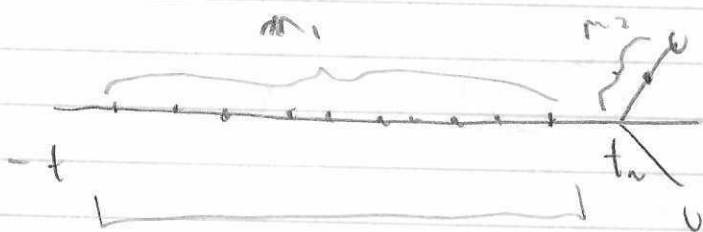


Suppose  $t_w$  generations ago there was a common ancestor to all alleles

All mutations before  $t_w$  are shared by whole population (substitutions)

Mutations after  $t_w$  are polymorphic (not fixed)

If  $t \gg t_w$ , then substitution rate  $\approx \frac{\# \text{ substitutions}}{\text{time}}$



substitution rate is this over  $t$

But we include the lost as well, but that's ok

True substitution rate:  $\frac{m_1}{t}$

Estimated substitution rate  $\frac{t_w}{t} \approx \frac{m_1 + m_2}{t}$

But if  $t \gg t_w$ , then  $m_1 \gg m_2$

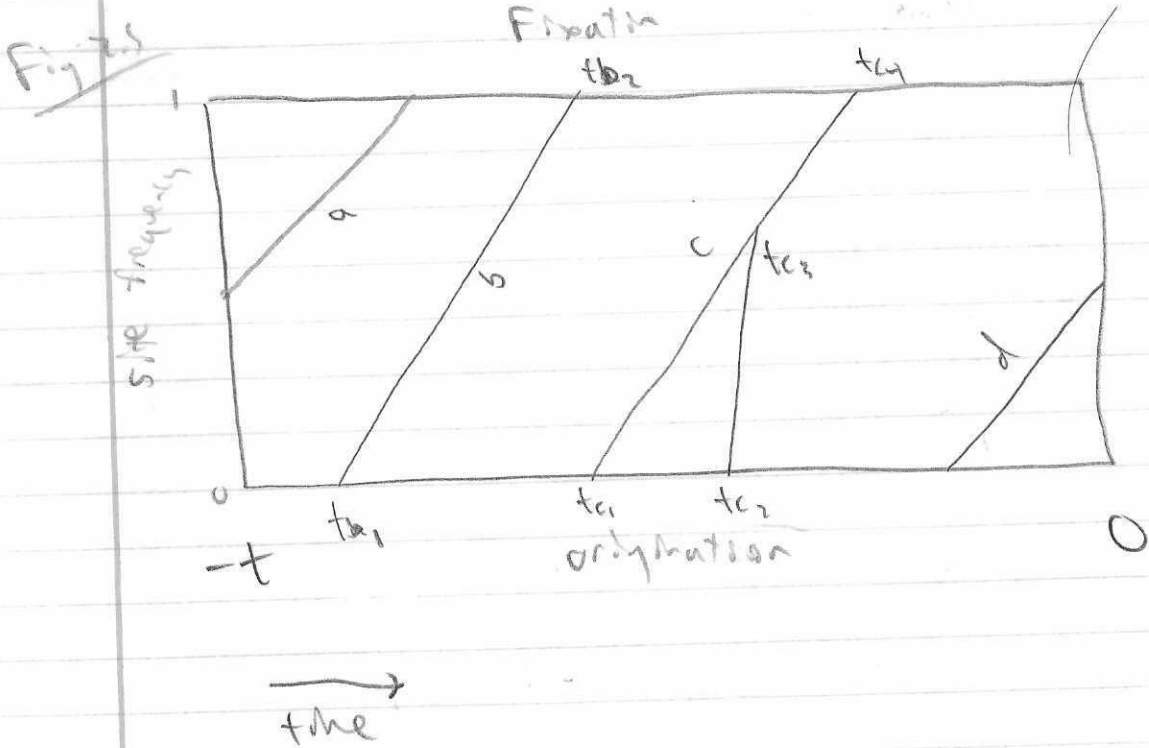
• Usually  $t \gg t_w$  except in closely related species

• Let  $X = \#$  mutations on lineage,

then  $X \sim \text{Poi}(tu)$

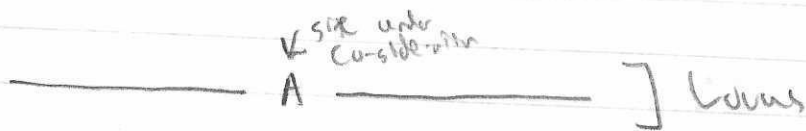
- Poisson if
- ① many opportunities for event to occur
  - ②  $P_r(\text{event})$  is small
  - ③ Events are independent

•  $P_r(i \text{ mutations on lineage}) = \frac{e^{-tu} (tu)^i}{i!}$  } Poisson prob.



↑ only mutations that fix are plotted

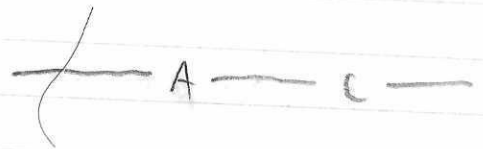
Let's talk about b



@  $t_{b1}$ , all individuals have A at that locus but one individual mutates to c

@  $t_{b2}$ , all individuals have c

Let's talk about c



@  $t_{c1}$ , all individuals have A-c

But one individual gets G-c

@ time  $t_{c2}$ , one individual gets G-T

@ time  $t_{c3}$  only A-c and G-T individuals are left

@ time  $t_{c4}$ , only G-T individuals are left

Exercise: Is it possible to plot?

