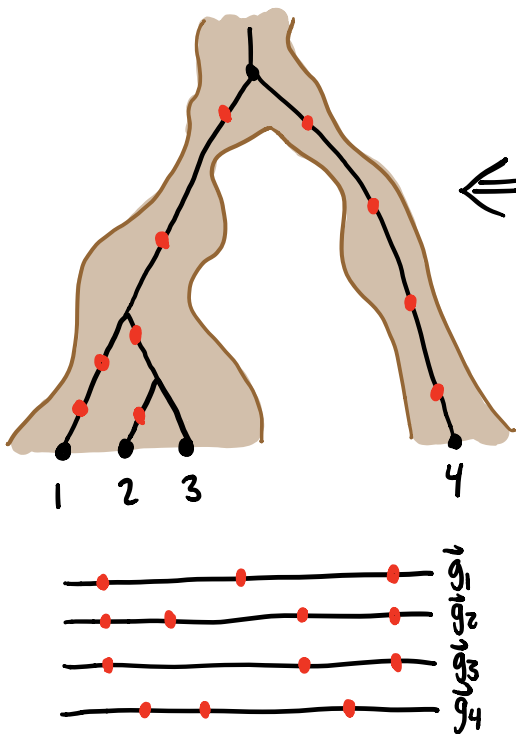


Announcements: PSET 4 DUE **TUES**; Office Hrs **12-1pm Monday**

Last time: Coalescent theory for **neutral** + **asexual** genomes



$$\frac{d\bar{s}(\bar{j})}{dt} = \underbrace{[\bar{x}(\bar{j}) - \bar{x}(t)] f(\bar{j})}_{\text{selection (non-linear)}} + \underbrace{\sum_{\bar{j}} M(\bar{j} \rightarrow \bar{j}) f(\bar{j}) - M(\bar{j} \rightarrow \bar{j}) f(\bar{j})}_{\text{mutation (linear, "local")}}$$

$$+ \underbrace{\rho \sum_{\bar{j}} \gamma(\bar{j} \rightarrow \bar{j}) f(\bar{j}) - \rho f(\bar{j})}_{\text{recombination (non-linear, non-local)}}$$

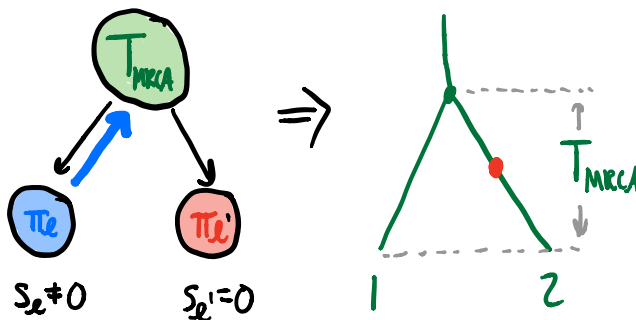
$$+ \underbrace{\sqrt{\frac{f(\bar{j})}{N}} \eta(\bar{j}) - f(\bar{j}) \sum_{\bar{j}} \sqrt{\frac{f(\bar{j})}{N}} \eta(\bar{j})}_{\text{genetic drift (stochastic)}}$$

2 simple rules:

- (i) genealogy: $P(\delta\delta) = 1/N(\tau)$
- (ii) mutations: Poisson(μt)

Today: How can we get **selection** + **recombination** back in picture?

⇒ Selection is hard
(alters causation)
diagram



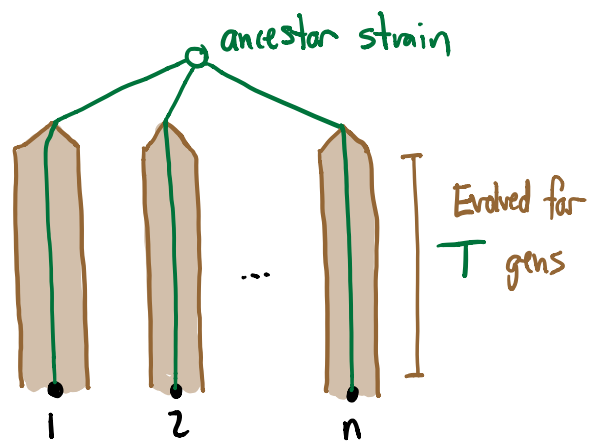
⇒ in some cases, coalescent picture can be salvaged if

① mainly care about predicting **neutral sites** (e.g. synonymous)

② can find some other way to predict **genealogy**

e.g. evolution experiment
in HW 3, Problem 2:

⇒ picked 1 individual
from each population



⇒ know exactly what genealogy looks like!

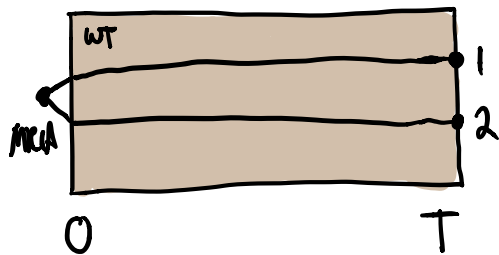
⇒ # synonymous mutations $\sim \text{Poisson}(L_{\text{syn}} \mu T)$

⇒ why can't this work for larger samples?

⇒ why can't this work for larger samples?

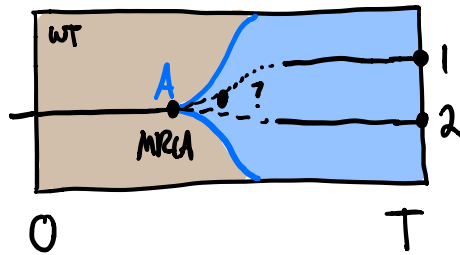
⇒ consider 2 scenarios:

(a) Truly Neutral
($N \gg T$)



$$\rightarrow p_i = \frac{1}{N}$$
$$\text{prob} \sim \frac{T}{N} \ll 1$$

(b) Selective Sweep

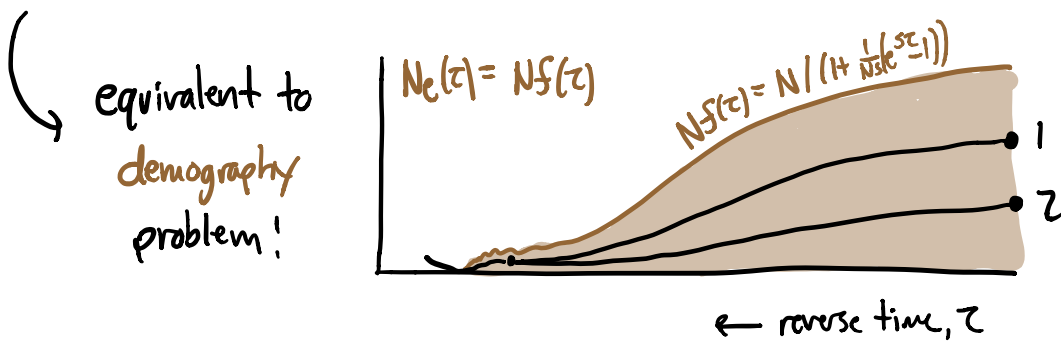
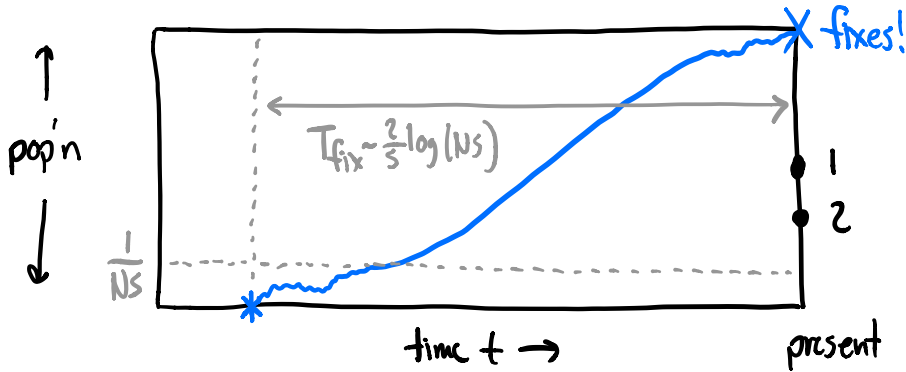


⇒ genealogies for $n \geq 2$ can be very different!

⇒ when **selected mut'n** is from **successive mut'n's regime**

⇒ can make some quantitative progress

in this case, know entire trajectory of **selected mut'n**:

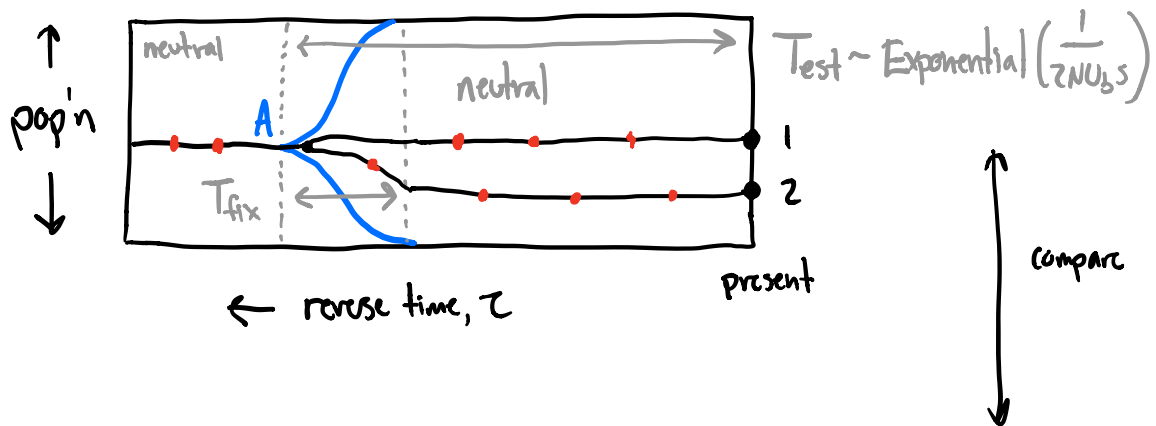


$$\Rightarrow \Pr[T_{MRCA} \geq \tau] = \exp\left[-\int_0^\tau \frac{dz'}{N_e(z')}\right] = e^{-\int_0^\tau \frac{(1-\frac{1}{Ns}) + \frac{1}{Ns} e^{sz'}}{N} dz'}$$

$$= \exp\left[-\frac{(1-\frac{1}{Ns})\tau}{N} - \frac{1}{Ns} \frac{e^{s\tau} - 1}{s}\right] \approx \exp\left[-e^{-s(T_{fix}-\tau)}\right]$$

\Rightarrow no coalescence until $\tau \sim T_{fix} \pm \mathcal{O}(\frac{1}{s})!$ $[f(\tau) \approx \frac{1}{Ns}]$

what if mutation had fixed before time of sampling?



Two characteristic regimes:

$$T_{MRCR} \sim \text{Exp}(N)$$

\uparrow

① if $N \ll T_{\text{test}} \Rightarrow$ coalescence before sweep \Rightarrow neutral!

② if $T_{\text{test}} \ll N \Rightarrow T_{MRCR} \approx T_{\text{test}} = \text{Exponential}\left(\frac{1}{2N u_b s}\right)$

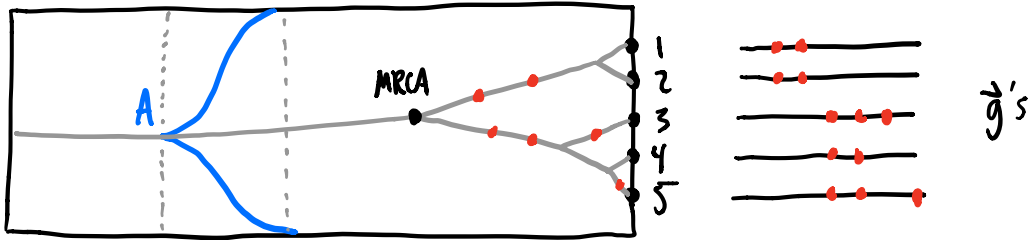
$$\Downarrow \pi_{\text{syn}} = 2\mu \langle T_{MRCR} \rangle = \left(\frac{N}{u_b}\right) \frac{1}{Ns}$$

\Rightarrow anticorrelated w/ N !

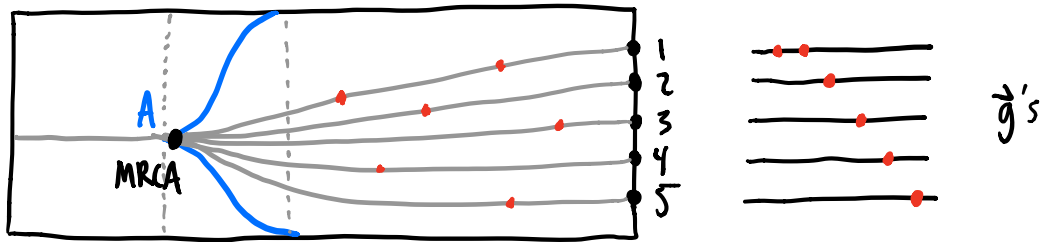
\Downarrow " N_e " $\propto 1/N$

can extend to larger sample sizes:

① $T_{fix} \ll N \ll T_{est} \Rightarrow$ effectively neutral

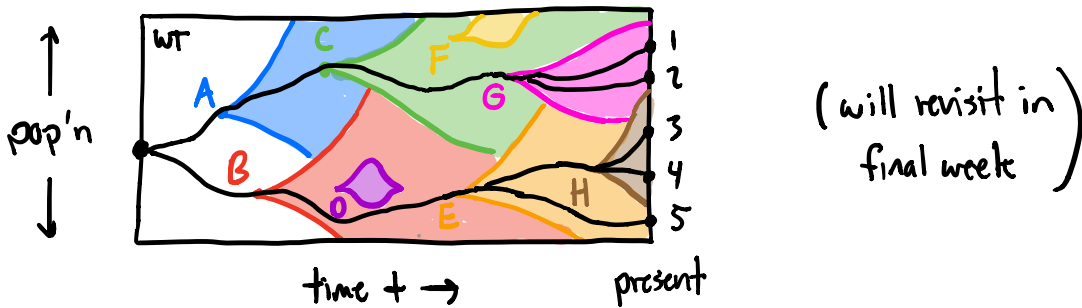


② $T_{fix} \ll T_{est} \ll N \Rightarrow$ "star-like genealogy"



\Rightarrow not just a difference in scale \Rightarrow difference in shape!

② $T_{est} \ll T_{fix} \ll N \Rightarrow$ "clonal interference"



Next: How can we account for **recombination**?

⇒ start w/ neutral case

$$\frac{ds(g)}{dt} = \cancel{\sim \frac{1}{N}} + \sim L \times \mu + \sim e + \sim \frac{1}{N}$$

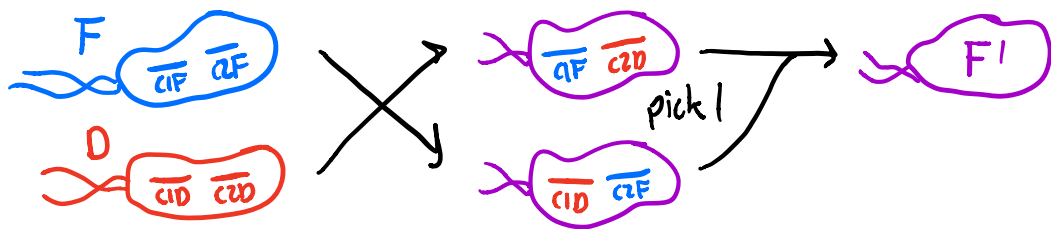
⇒ start w/ **reassortment** model of recombination

w/ 2 chromosomes of length L

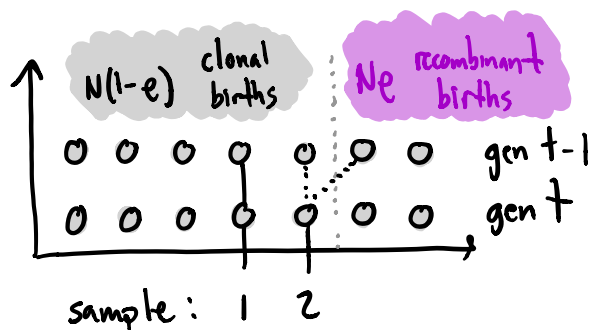


(e.g. HA + NA genes in influenza)

⇒ Recall: @ per capita rate e :



Backwards in time:



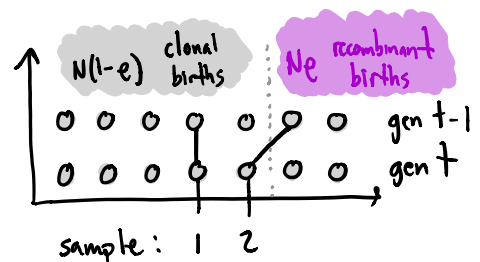
⇒ probability that individual was **recombinant** = $\frac{Ne}{N} = e$

⇒ coalescence probability = $\frac{1}{N}$ (same as before)

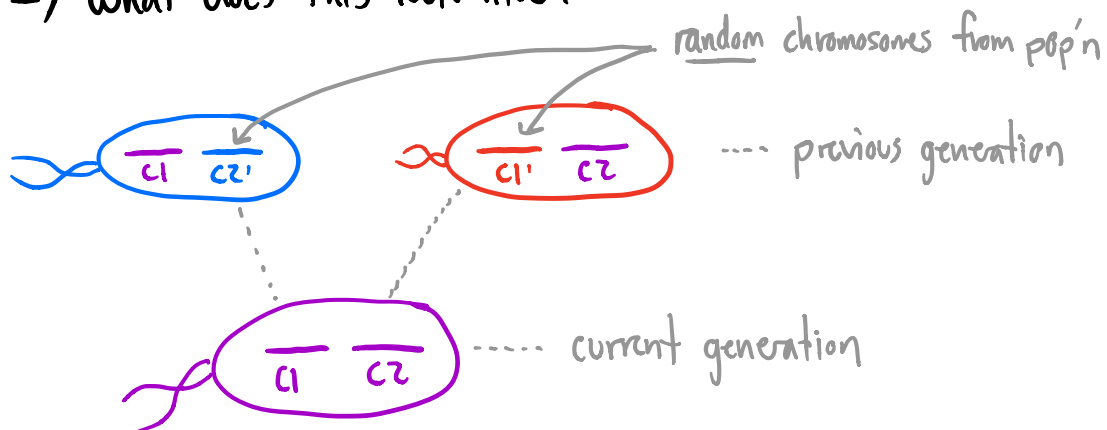
$$\Rightarrow \Pr[\text{coalesce before recombine}] = \frac{\frac{1}{N}}{\frac{1}{N} + 2e} = \frac{1}{1 + 2Ne}$$

⇒ if $Ne \ll 1 \rightarrow$ effectively asexual!

⇒ if $Ne \gg 1$, good chance that some ancestral individuals were result of **recombination event...**



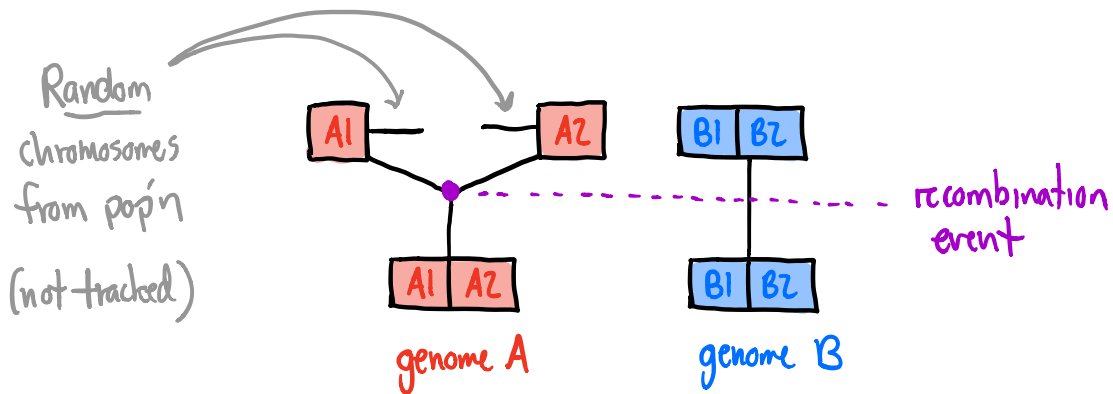
⇒ what does this look like?



⇒ ancestors of 2 chromosomes are different!

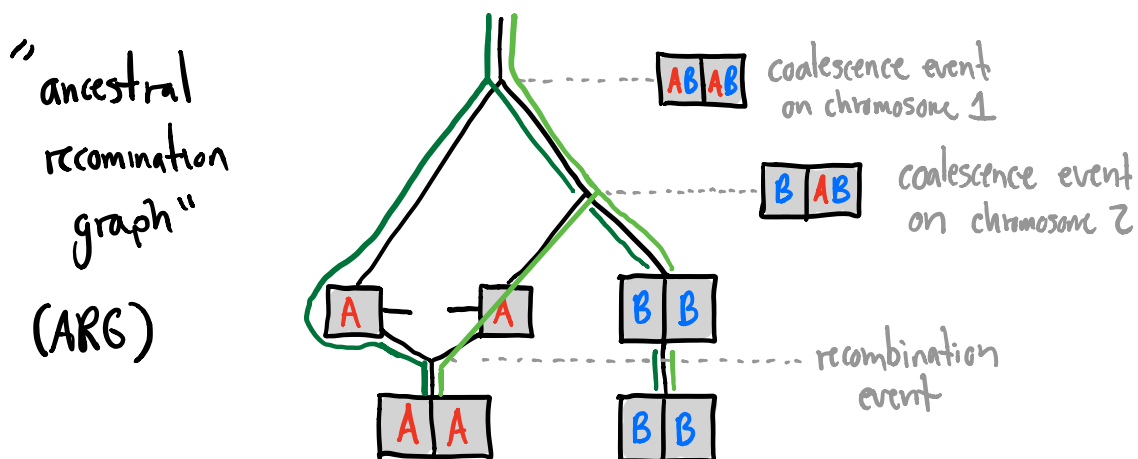
⇒ i.e. genealogies must separate!

⇒ can represent this in coalescent picture as:

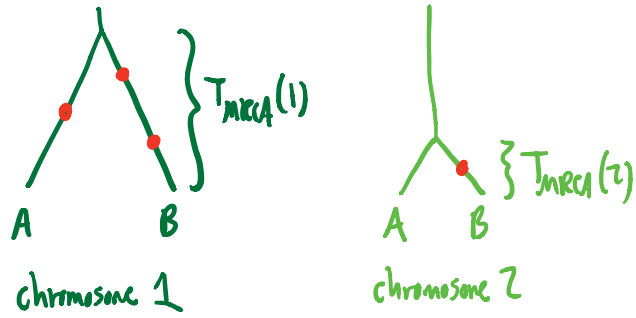


⇒ now coalescent continues w/ larger sample ($n=3$)

⇒ e.g. if no more recomb events, could have:



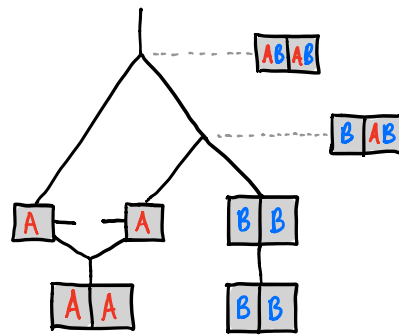
⇒ can extract genealogies
for each chromosome:



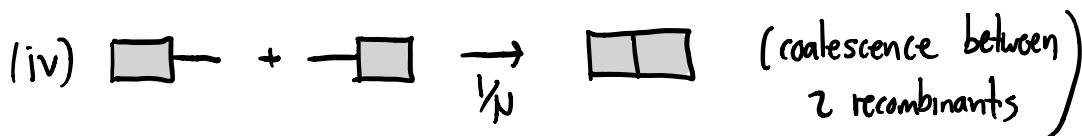
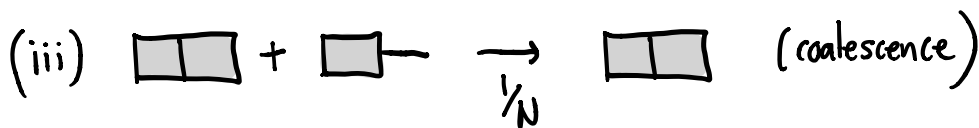
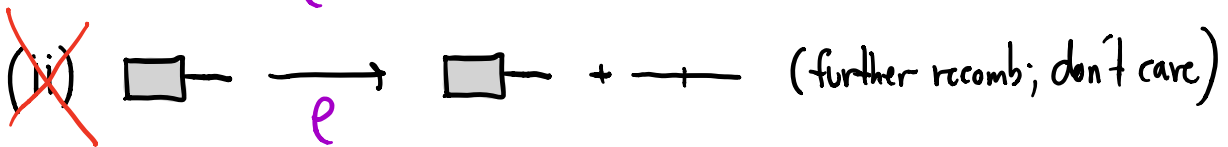
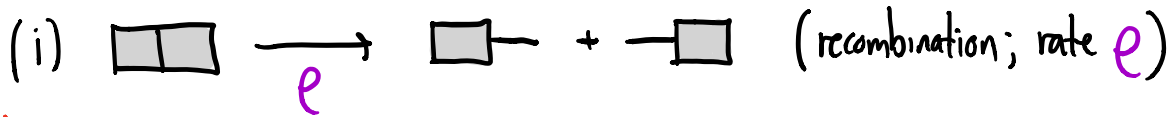
⇒ i.e. recombination allows genealogies to differ
@ different locations along genome

[compare to asexual case where $T_{MRCIA(1)} = T_{MRCIA(2)}$]

⇒ this was just one
possible ARG...



⇒ more generally, @ each step will have 4 types of events:



just as likely per pair!

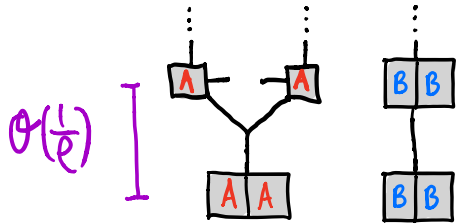
⇒ can we simulate this process in our heads when $N_e \gg 1$?

Start with sample:  

① Total coalescence rate = $1/N$ (1 pair)

② Total recombination rate = $2e$

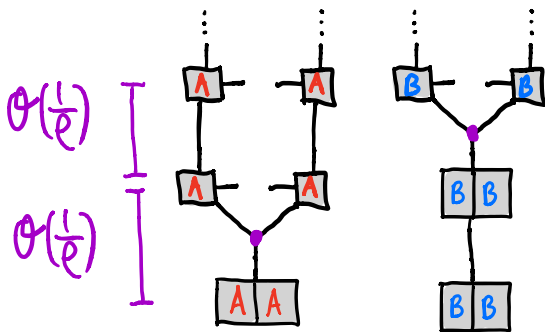
Step 2:



① coalescence = $\frac{1}{N} \binom{3}{2} = \frac{3}{N}$

② recomb = ρ

Step 3:

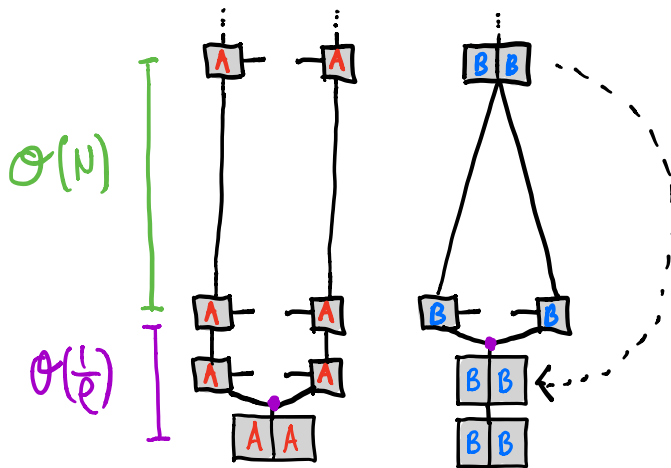


① coalescence = $\frac{1}{N} \binom{4}{2} = \frac{6}{N}$

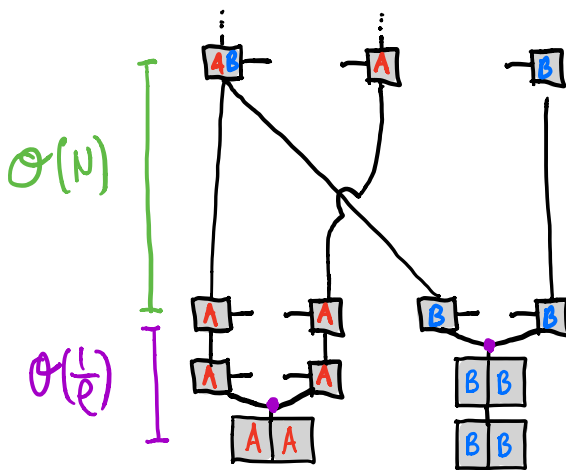
② recomb = 0

2 different types of coalescent events:

(i) coalescence of recombinant chromosomes ($\square + -$)



(ii) coalescence involving sampled genetic material ($\square + \square$)



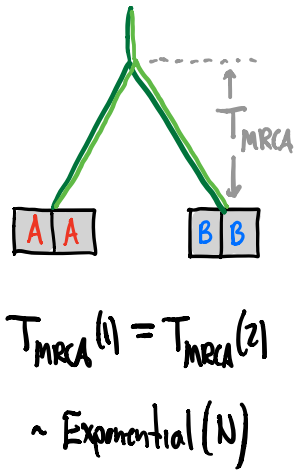
\Rightarrow each coalescence happens \approx independently w/ rate $1/N$.

$\Rightarrow T_{MRCU}(1) \sim \text{Exponential}(N)$

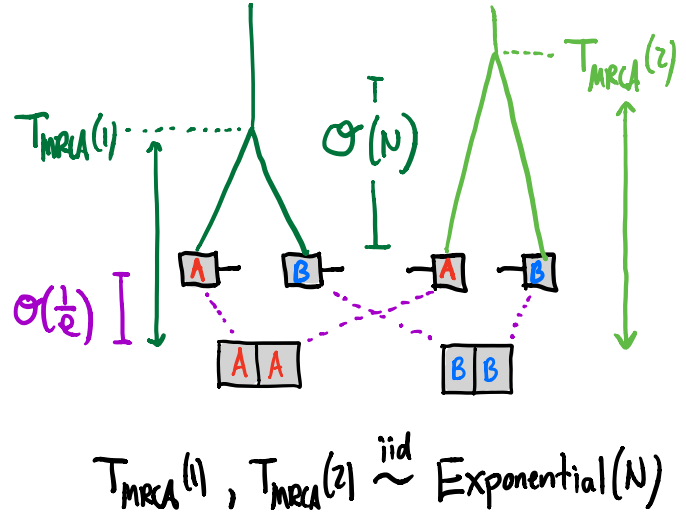
$T_{MRCU}(2) \sim \text{Exponential}(N)$

Putting everything together:

$N_e \ll 1$ (effectively asexual)



$N_e \gg 1$ (effectively independent)

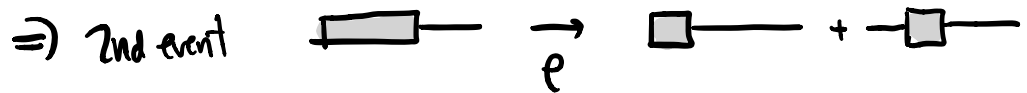
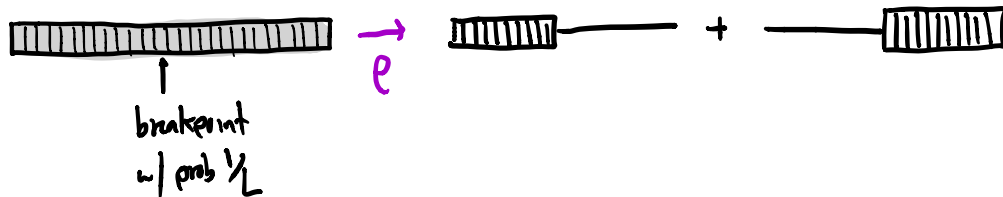


\Rightarrow same idea works for > 2 chromosomes:

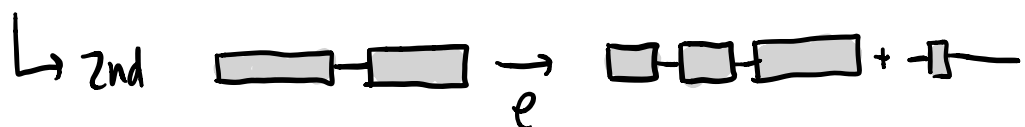


⇒ also works for other forms of recombination:

e.g. crossover:



e.g. HGT / gene conversion:



⇒ easy to simulate, but hard to calculate (even for $n=2$!)

⇒ effective sample size $\sim 2^{\text{effective \# chromosomes}}$

⇒ again, hard to add

selection back to picture...

$$\frac{ds(\vec{q})}{dt} = \cancel{\sim \vec{L} \cdot \vec{\mu}} + \sim L \cdot \mu + \sim e + \sim \frac{E}{\sqrt{u}}$$

⇒ Next: back to forward-time
approach to see if
we can make some progress...