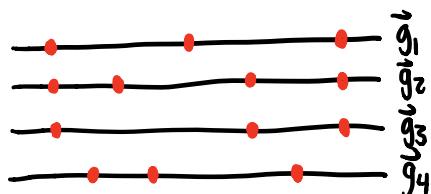
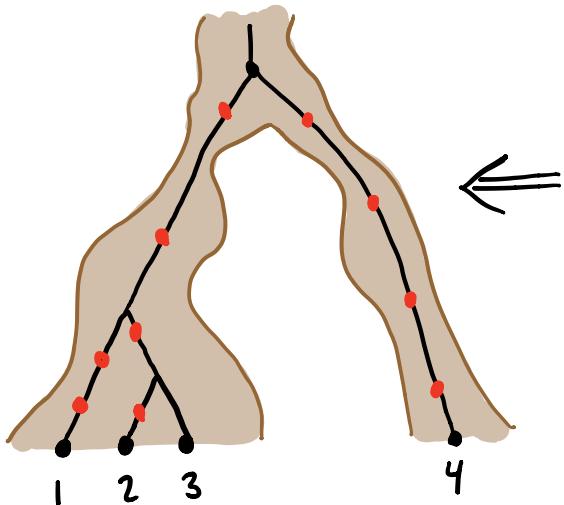


Announcements: PSET 4 DUE **TUES**; Office Hrs 12-1pm **Monday**

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



$$\frac{\partial f(\vec{g})}{\partial t} = \cancel{\left[X(\vec{g}) - \bar{X}(t) \right] f(\vec{g})} + \sum_{\vec{g}'} M(\vec{g} \rightarrow \vec{g}') f(\vec{g}) - M(\vec{g}' \rightarrow \vec{g}) f(\vec{g}') \quad \text{mutation (linear, "local")}$$

$$+ \cancel{\rho \sum_{\vec{g}_1, \vec{g}_2} T(\vec{g}_1, \vec{g}_2)} f(\vec{g}_1) f(\vec{g}_2) - \rho f(\vec{g}) \quad \text{recombination (nonlinear, non-local)}$$

$$+ \sqrt{\frac{f(\vec{g})}{N}} \eta(\vec{g}) - f(\vec{g}) \sum_{\vec{g}'} \frac{f(\vec{g}')}{N} \eta(\vec{g}') \quad \text{genetic drift (stochastic)}$$

2 simple rules:

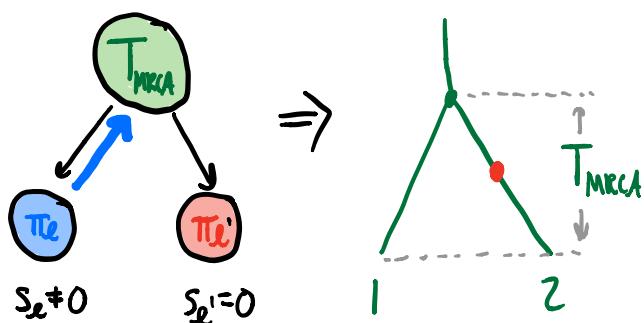
(i) genealogy: $p(\delta\delta) = 1/N(t)$

(ii) mutations: Poisson(μt)

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

\Rightarrow Selection is hard

(alters causation
diagram)



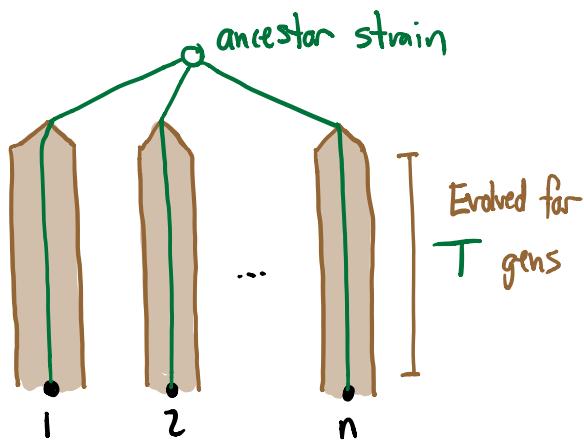
\Rightarrow in some cases, coalescent picture can be salvaged if

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:

\Rightarrow picked 1 individual
from each population



\Rightarrow know exactly what genealogy looks like!

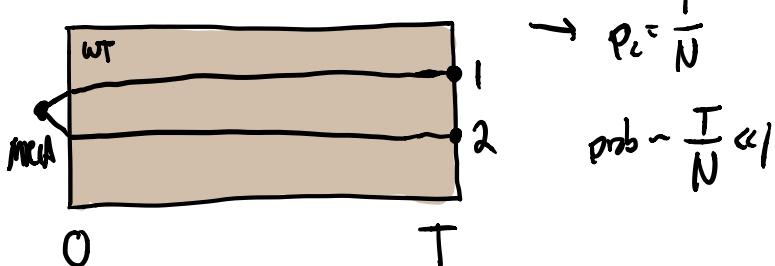
\Rightarrow # Synonymous mutations \sim Poisson($L_{\text{syn}} \nu T$)

\Rightarrow why can't this work for larger samples?

\Rightarrow why can't this work for larger samples?

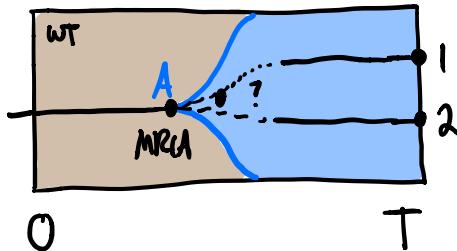
\Rightarrow consider 2 scenarios:

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



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

(b) Selective Sweep

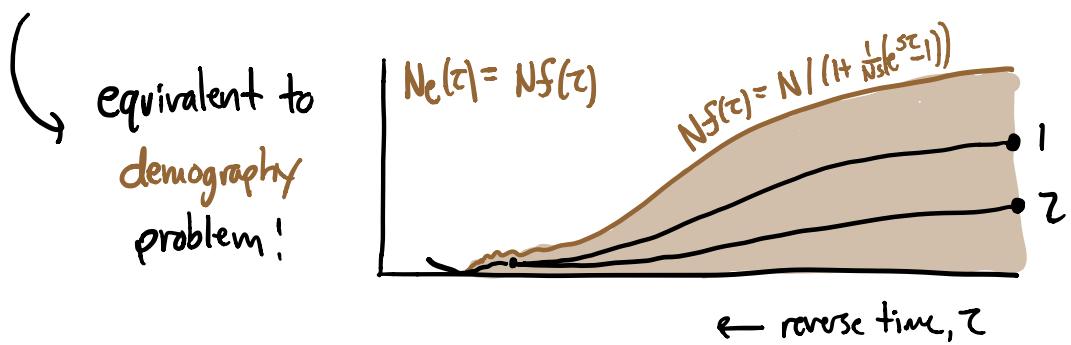
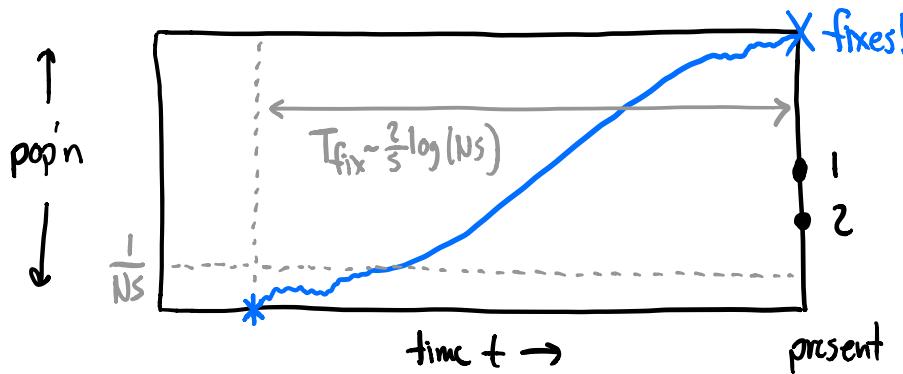


\Rightarrow genealogies for $n \geq 2$ can be very different!

\Rightarrow when selected mutant is from successive mut's regime

\Rightarrow can make some quantitative progress

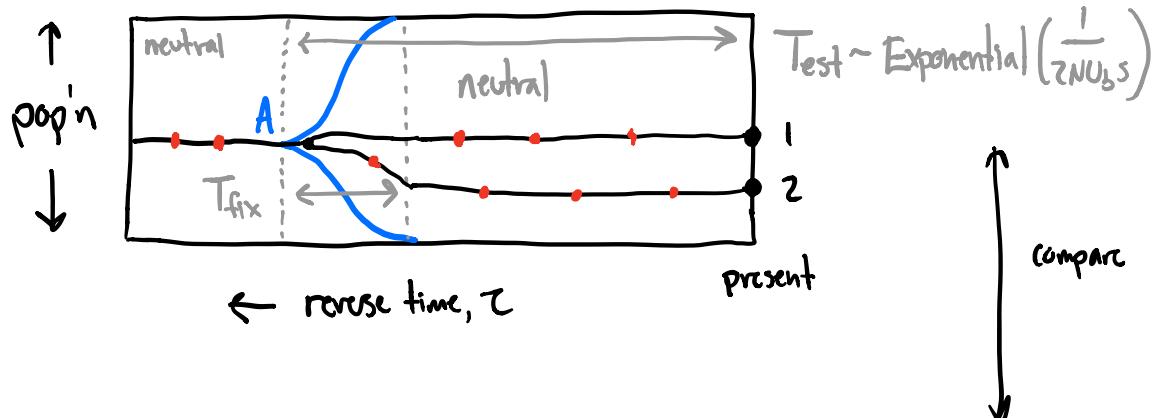
in this case, know entire trajectory of selected mut'n:



$$\begin{aligned} \Rightarrow \Pr[T_{\text{MRCA}} \geq \tau] &= \exp \left[- \int_0^\tau \frac{dc}{Nf(c)} \right] = e^{- \int_0^\tau \frac{(2c - 1)(1 - c)}{Nf(c)} dc} \\ &= \exp \left[- \frac{(2c - 1)^2}{N} - \left(\frac{1}{Ns} \right)^2 \left(e^{\tau} - 1 \right) \right] \approx \exp \left[- e^{-s(T_{\text{fix}} - \tau)} \right] \end{aligned}$$

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

what if mutation had fixed before time of sampling?



Two characteristic regimes:

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

¶

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

② if $T_{\text{test}} \ll N$ \Rightarrow $T_{\text{MRCA}} \approx T_{\text{test}} = \text{Exponential}\left(\frac{1}{2NU_b\tau}\right)$

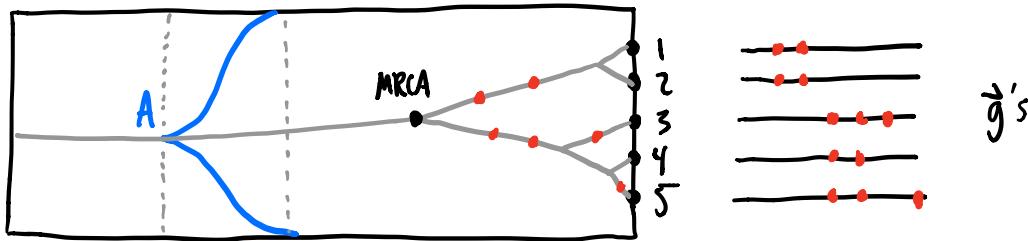
$$\hookrightarrow \pi_{\text{syn}} = 2\mu \langle T_{\text{MRCA}} \rangle = \left(\frac{N}{U_b}\right) \frac{1}{N\tau}$$

\Rightarrow anticorrelated w/ N !

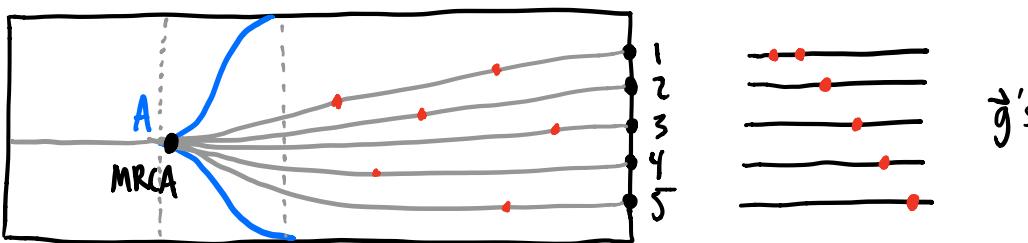
$$\hookrightarrow "N_e" \propto 1/N$$

can extend to larger sample sizes:

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

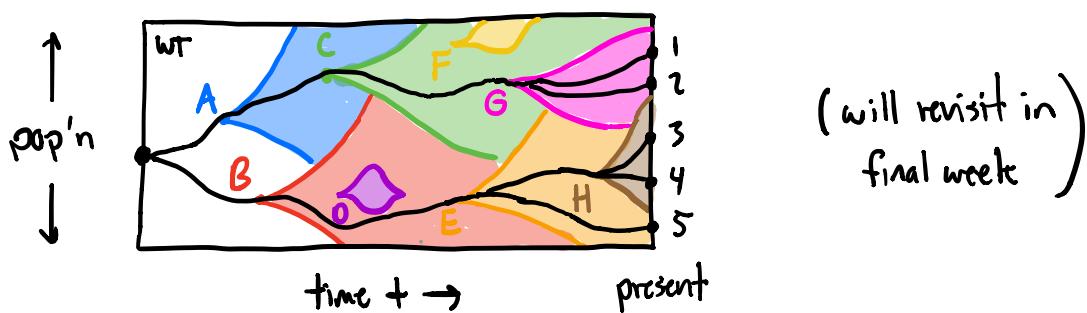


- ② $T_{fix} \ll T_{est} \ll N$ \Rightarrow "star-like genealogy"



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

- ③ $T_{est} \ll T_{fix} \ll N$ \Rightarrow "clonal interference"



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

⇒ Start w/ neutral case

$$\frac{\partial f(\vec{g})}{\partial t} = \cancel{\sim (\vec{v})} + \sim L^* \mu + \sim e + \sim \frac{z}{JN}$$

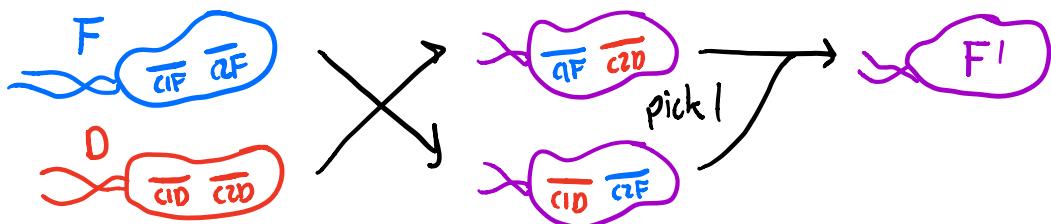
⇒ Start w/ reassortment model of recombination

w/ 2 chromosomes of length L

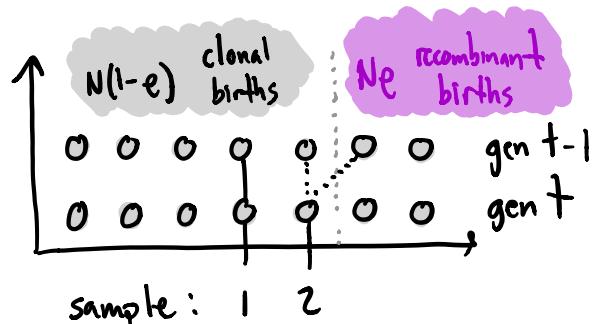


(e.g. HA + NA genes in influenza)

⇒ Recall: @ per capita rate ρ :



Backwards in time:



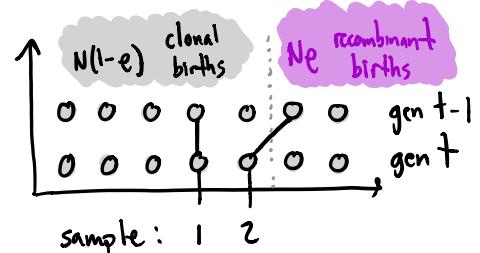
\Rightarrow probability that individual was recombinant = $\frac{Ne}{N} = \rho$

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

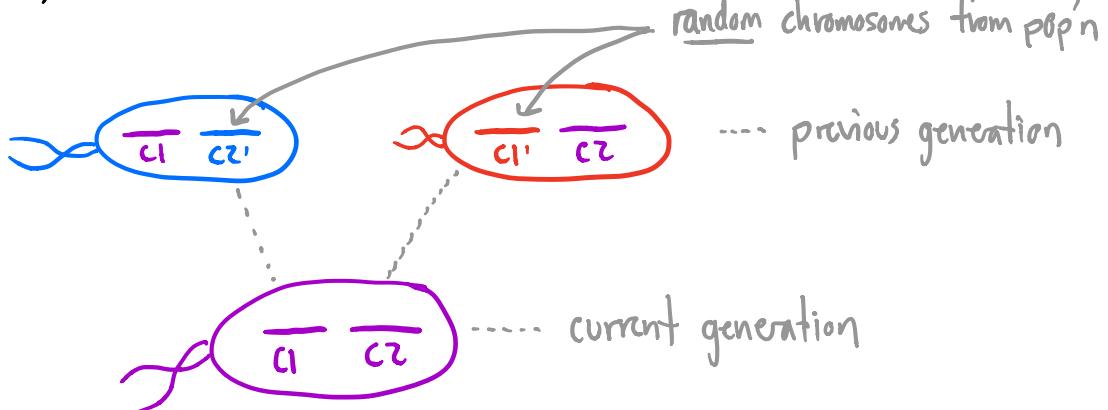
$$\Rightarrow \Pr[\text{coalesce before recombine}] = \frac{\frac{1}{N}}{\frac{1}{N} + 2\rho} = \frac{1}{1 + 2Ne\rho}$$

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

\Rightarrow if $Ne \gg 1$, good chance that some ancestral individuals were result of recombinant event...



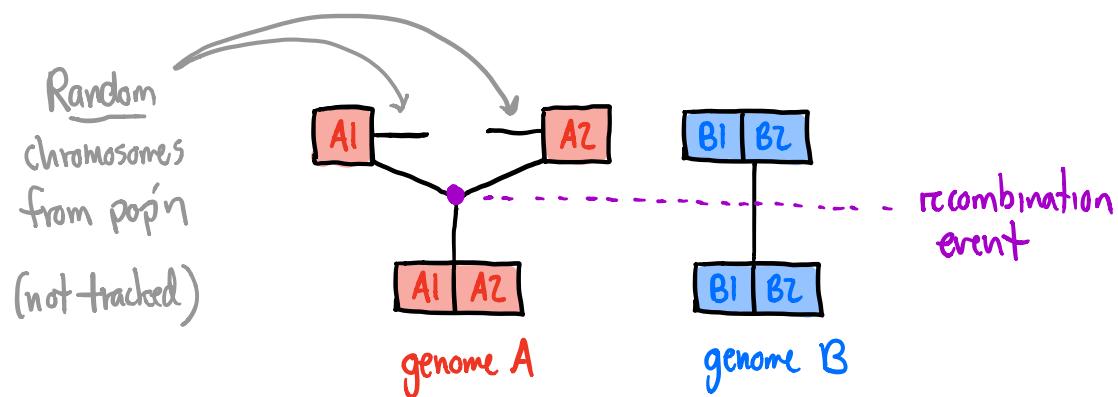
\Rightarrow what does this look like?



\Rightarrow ancestors of 2 chromosomes are different!

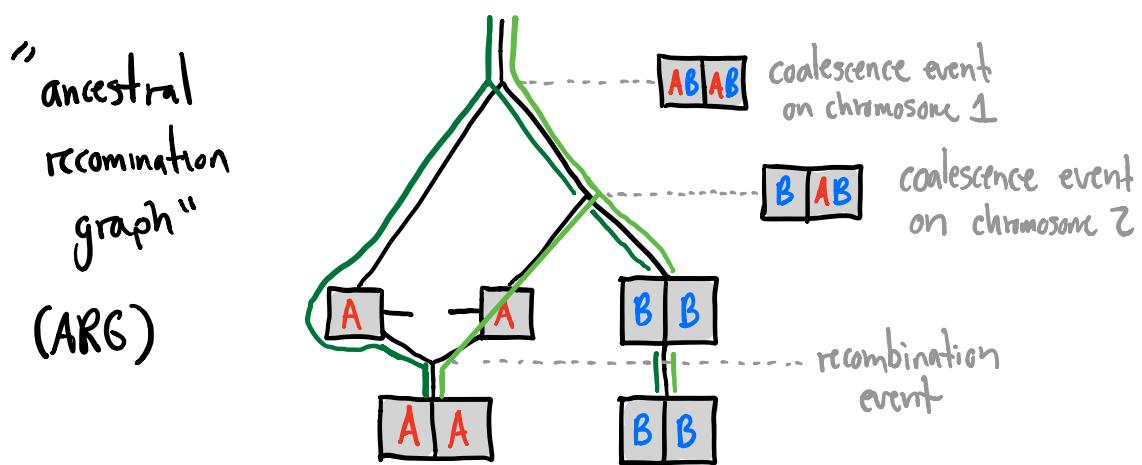
\Rightarrow i.e. genealogies must separate!

\Rightarrow can represent this in coalescent picture as:

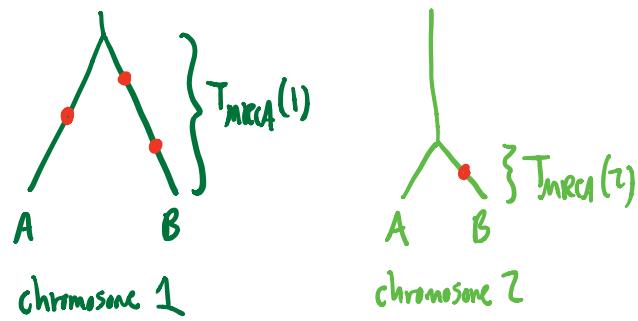


\Rightarrow now coalescent continues w/ larger sample ($n=3$)

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



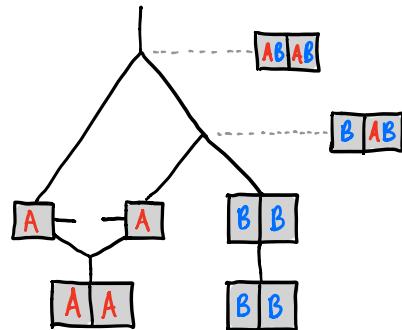
\Rightarrow can extract genealogies
for each chromosome:



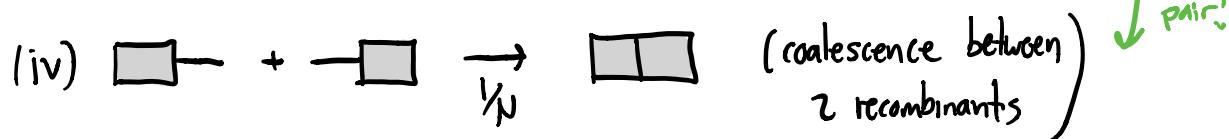
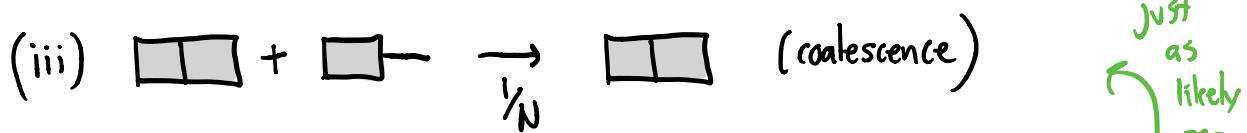
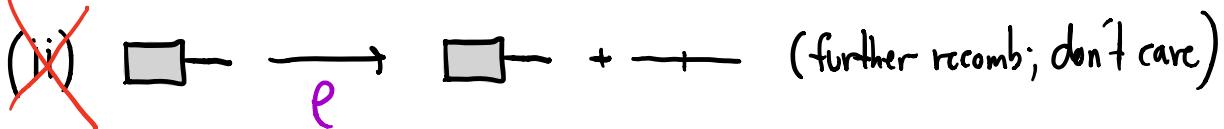
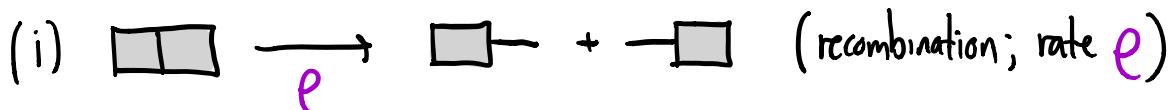
\Rightarrow i.e. recombination allows genealogies to differ
@ different locations along genome

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

\Rightarrow this was just one
possible ARG...



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



just
as
likely
per
pair!

\Rightarrow can we simulate this process in our heads when $N \gg 1$?

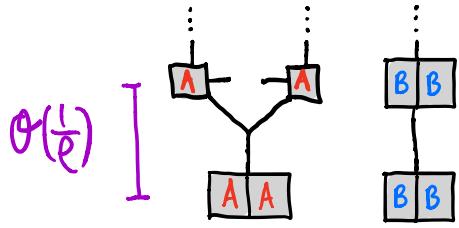
Start with sample:



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

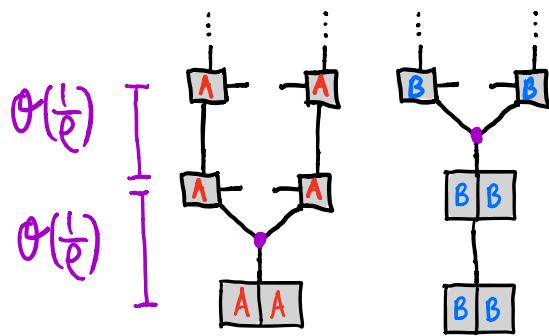
② Total recombination rate = 2ρ

Step 2:



- ① coalescence = $\frac{1}{N} \cdot \binom{3}{2} = \frac{3}{N}$
- ② recomb = ρ

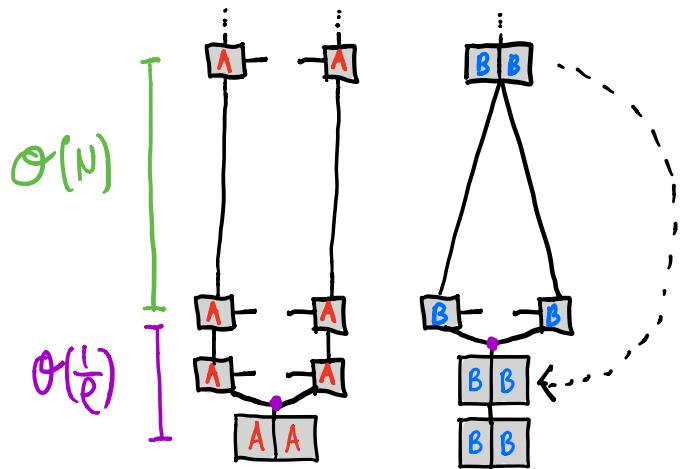
Step 3:



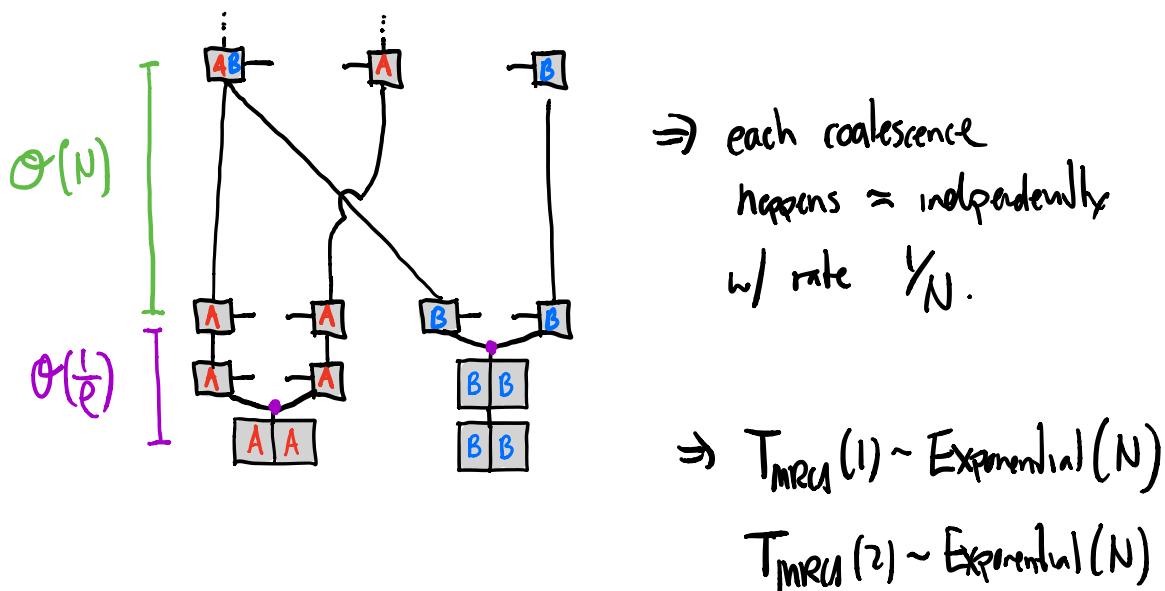
- ① coalescence = $\frac{1}{N} \binom{4}{2} = \frac{6}{N}$
- ② recomb = ρ

2 different types of coalescent events:

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

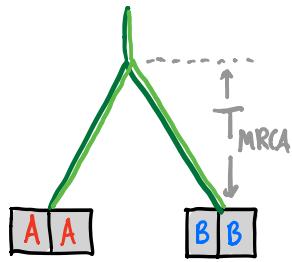


(ii) coalescence involving sampled genetic material ($\square + \square$)



Putting everything together:

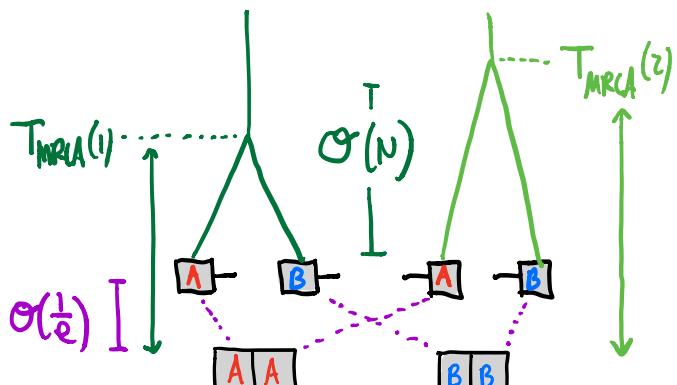
$N\varrho \ll 1$ (effectively asexual)



$$T_{\text{MRCA}}(1) = T_{\text{MRCA}}(2)$$

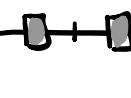
$\sim \text{Exponential}(N)$

$N\varrho \gg 1$ (effectively independent)



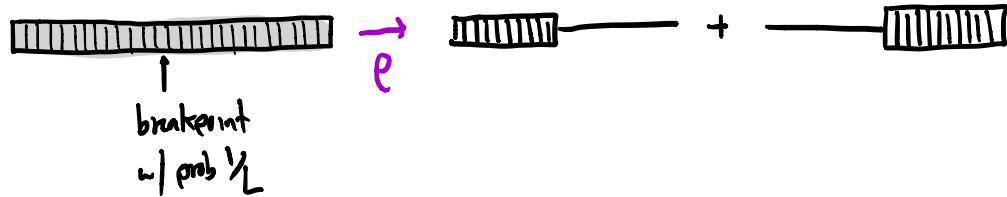
$$T_{\text{MRCA}}(1), T_{\text{MRCA}}(2) \stackrel{\text{iid}}{\sim} \text{Exponential}(N)$$

\Rightarrow same idea works for > 2 chromosomes:

e.g.  $\xrightarrow{\varrho}$  + 

\Rightarrow also works for other forms of recombination:

e.g. crossover:



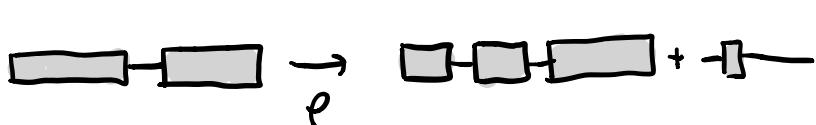
\Rightarrow 2nd event



e.g. HGT / gene conversion:



\hookrightarrow 2nd



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

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

\Rightarrow again, hard to add

selection back to picture ...

$$\frac{\partial c(j)}{\partial t} = \cancel{\sim -\bar{x})} + \sim -L^* \mu \\ + \sim e + \sim \frac{e}{JN}$$

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