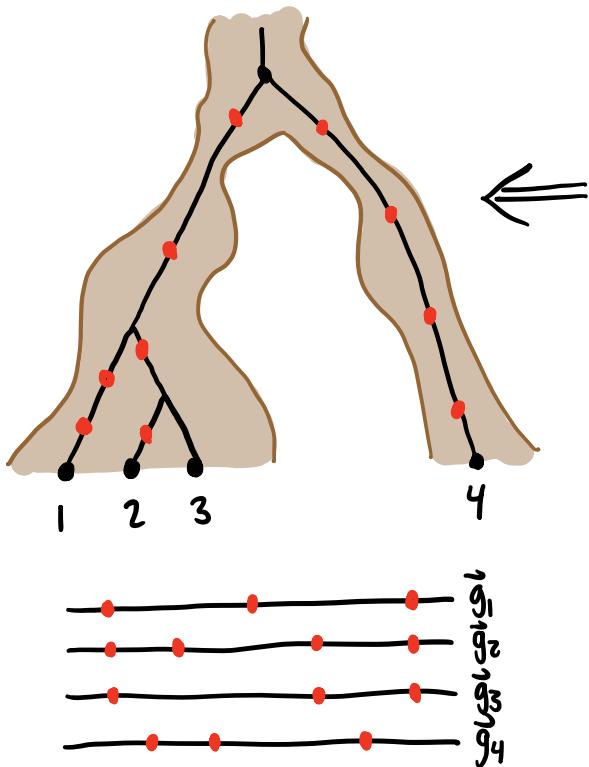


Chapter 12

Genealogies with selection and recombination

Last time: Coalescent theory for neutral + asexual genomes



$$\frac{df(j)}{dt} = \left[X(j) - X(t) \right] f(j) + \sum_{j'} M(j \rightarrow j') f(j') - M(j' \rightarrow j) f(j') \\ + e^{\sum_j f(j)} - p f(j) \quad \text{recombination (nonlinear, non-local)} \\ + \sqrt{\frac{f(j)}{N}} \eta(j) - f(j) \sum_{j'} \sqrt{\frac{f(j')}{N}} \eta(j') \quad \text{genetic drift (stochastic)}$$

~~selection (unlinear)~~

~~mutation (linear, "local")~~

2 simple rules:

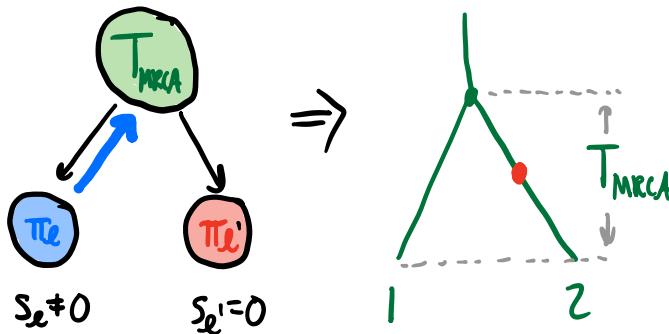
- (i) genealogy: $p(\delta\delta) = 1/N(t)$
- (ii) mutations: Poisson (μt)

⇒ E.g. pairwise diversity:

$$\langle \pi \rangle = 2\mu \cdot \langle T_{MRCA} \rangle = 2\mu \int_0^\infty e^{-\int_0^t \frac{de'}{N(t')}} dt'$$

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

⇒ Selection is hard
(alters causation diagram)



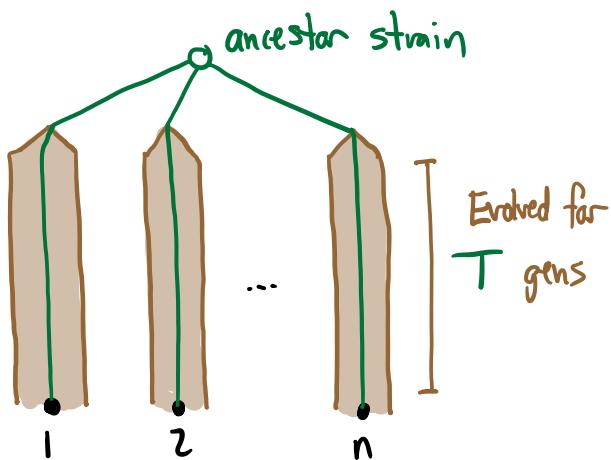
\Rightarrow in some cases, coalescent picture can be salvaged if

- ① mainly care about predicting **neutral sites** (e.g. synonymous mut's)
- ② can find some other way to predict **genealogy**

Simple example:

evolution experiment
in HW 3, Problem 2:

\Rightarrow picked 1 individual
from each population



\Rightarrow know exactly what genealogy looks like!

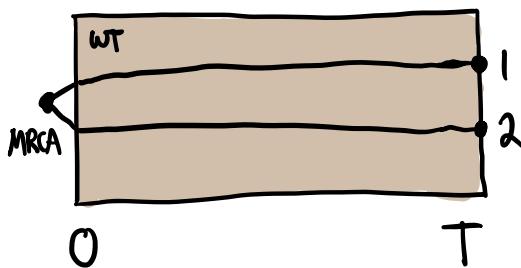
\Rightarrow # synonymous mut's / clone \sim Poisson($l_{syn} \mu T$)

\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
 $(N \gg T)$

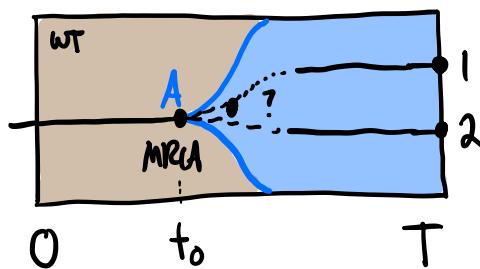


$$\Rightarrow p(\delta\delta) = 1/N \text{ (per gen)} \Rightarrow \Pr(T_{\text{MRC}} < T \ll N) \approx T/N \ll 1$$

i.e., \approx no coalescence during experiment!

"drift is weak"

(b) Selective Sweep



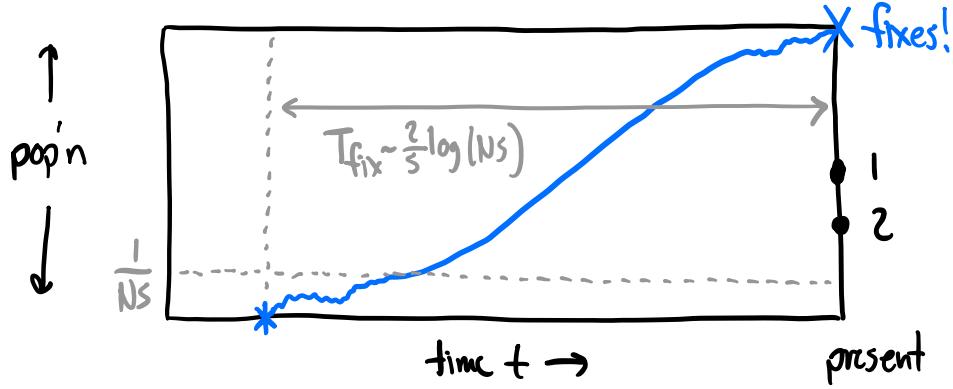
lineages must
coalesce by t_0 !

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

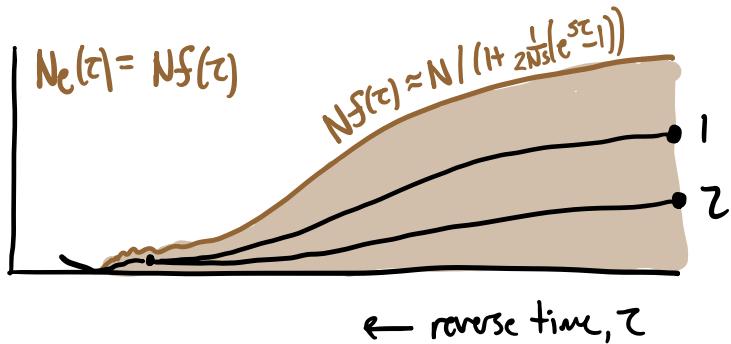
\Rightarrow when selected mutin is from successive mut'n's regime

\Rightarrow can make some quantitative progress

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



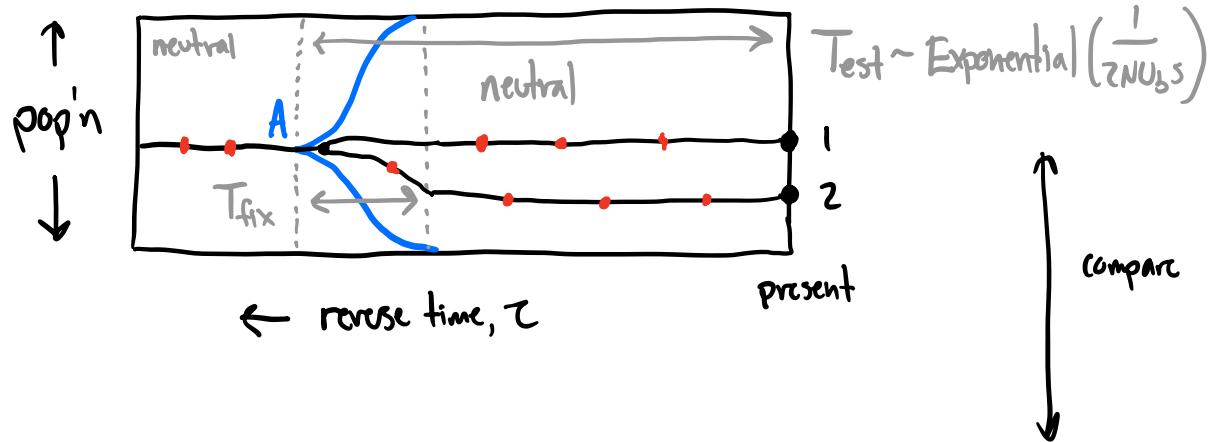
equivalent to
demography
problem!



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

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

what if mutation had fixed before time of sampling?



Two characteristic regimes:

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

π

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

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

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

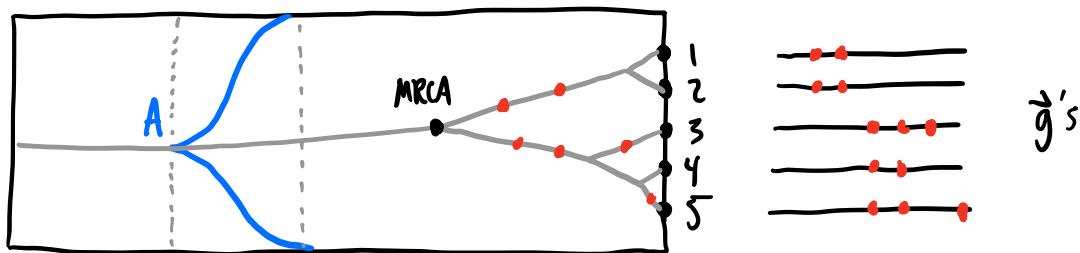
\Rightarrow anticorrelated w/ N !

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

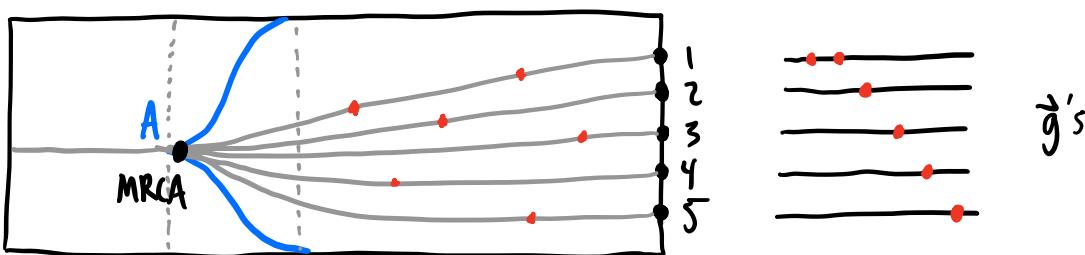
(doesn't make much sense
to think about it as an " N_e ")

can extend to larger sample sizes:

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

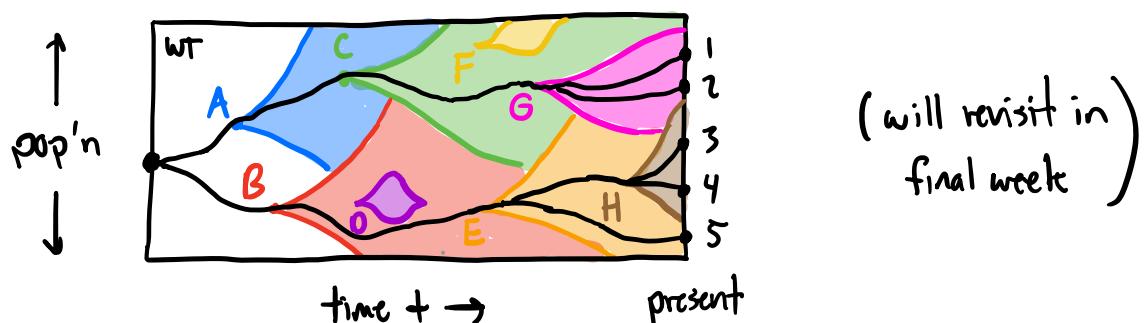


- ② $T_{fix} \ll T_{est} \ll N \Rightarrow$ "star-like genealogy"



\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

$$\frac{dS(\vec{g})}{dt} = \cancel{\sim -\bar{x}} + \sim L \times \mu + \sim e + \sim \frac{z}{JN}$$

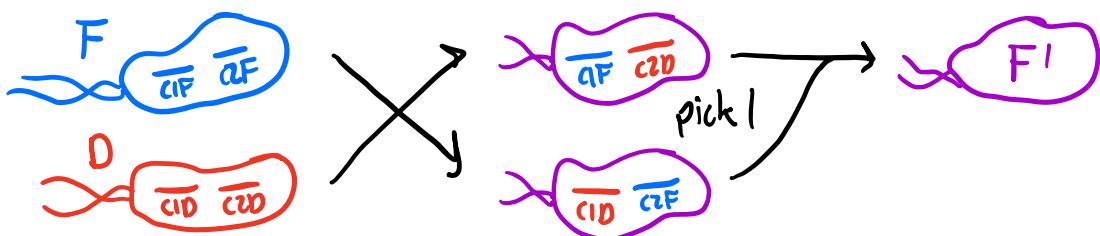
⇒ start w/ reassortment model of recombination

w/ 2 chromosomes of length L

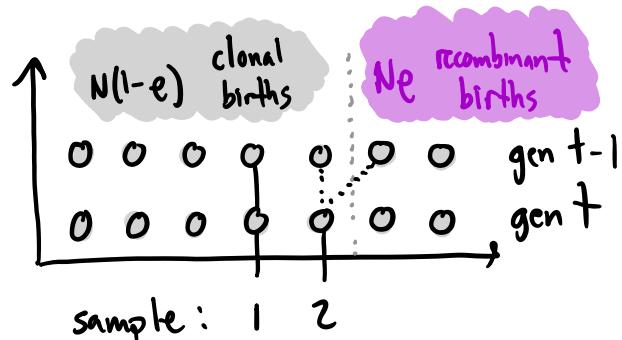


(e.g. HA + NA genes in influenza)

⇒ Recall: @ per capita rate ρ :



Backwards in time:



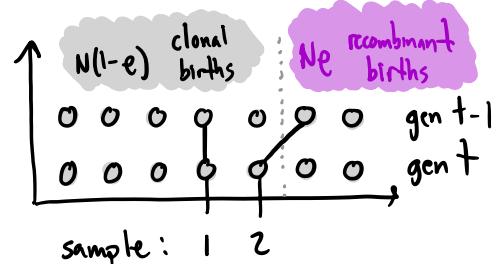
\Rightarrow probability that individual was recombinant = $\frac{Ne}{N} = e$

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

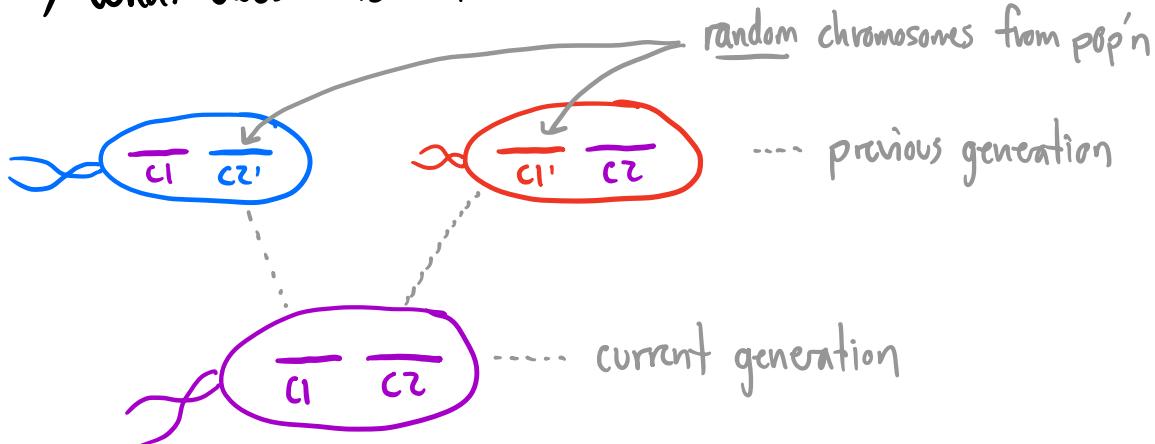
$$\Rightarrow \Pr[\text{coalesce before recombine}] = \frac{\frac{1}{N}}{\frac{1}{N} + 2e} = \frac{1}{1 + 2Ne}$$

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

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



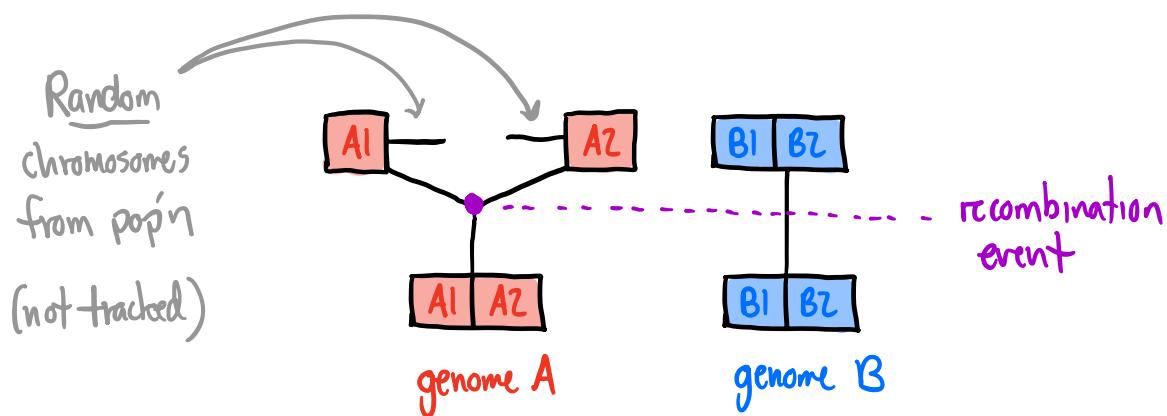
\Rightarrow what does this look like?



\Rightarrow ancestors of 2 chromosomes are different!

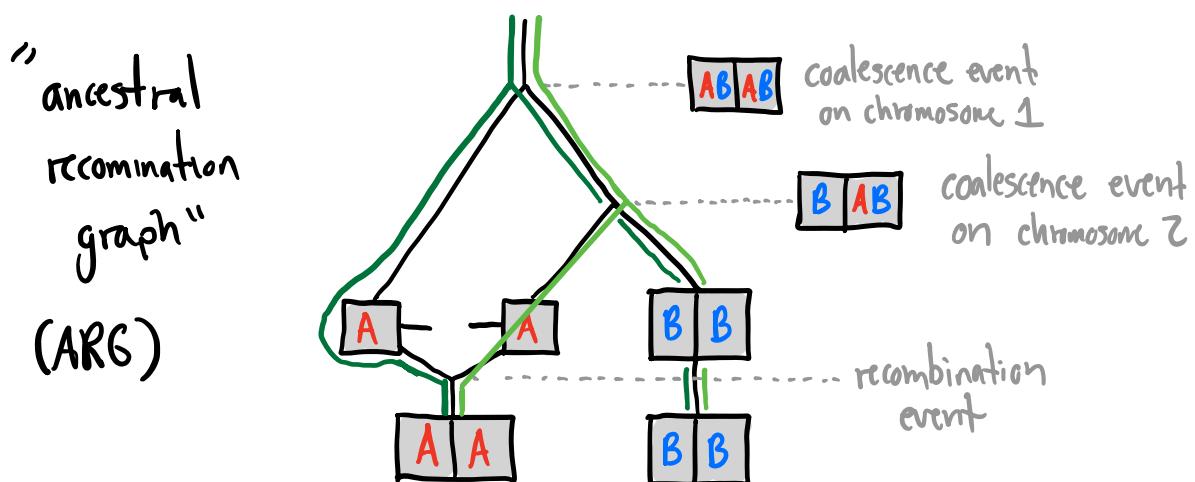
\Rightarrow i.e. genealogies must separate!

\Rightarrow can represent this in coalescent picture as:

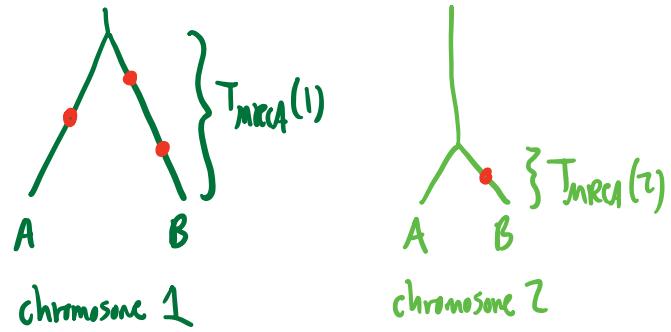


\Rightarrow now coalescent continues w/ larger sample ($n=3$)

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



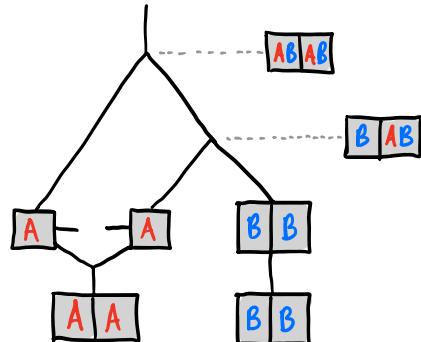
\Rightarrow can extract genealogies
for each chromosome:



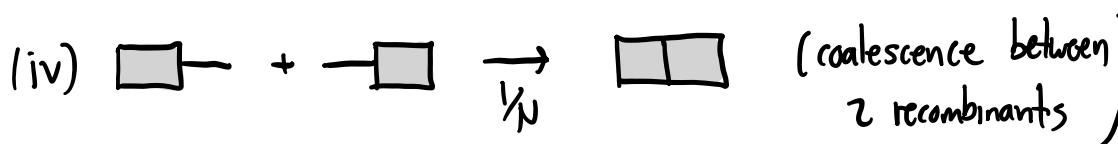
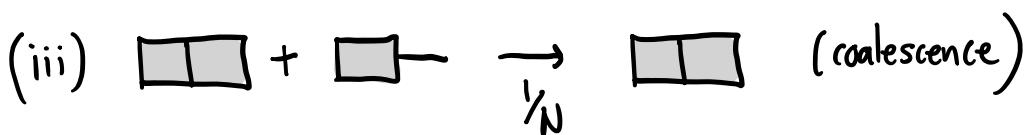
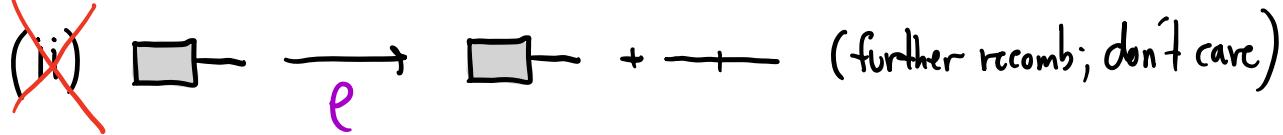
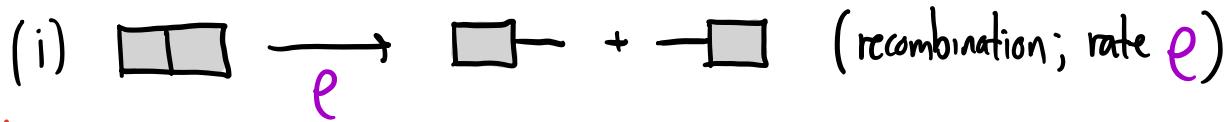
\Rightarrow i.e. recombination allows genealogies to differ
@ different locations along genome

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

\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 $N\rho \gg 1$?

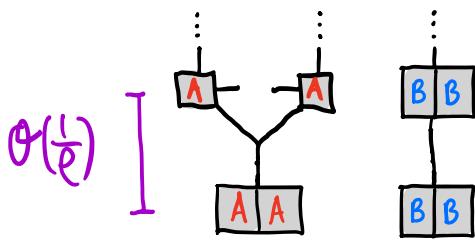
Start with sample:



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

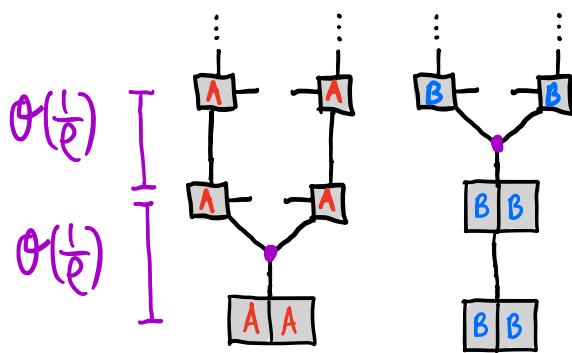
② Total recombination rate = 2ρ

Step 2:



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

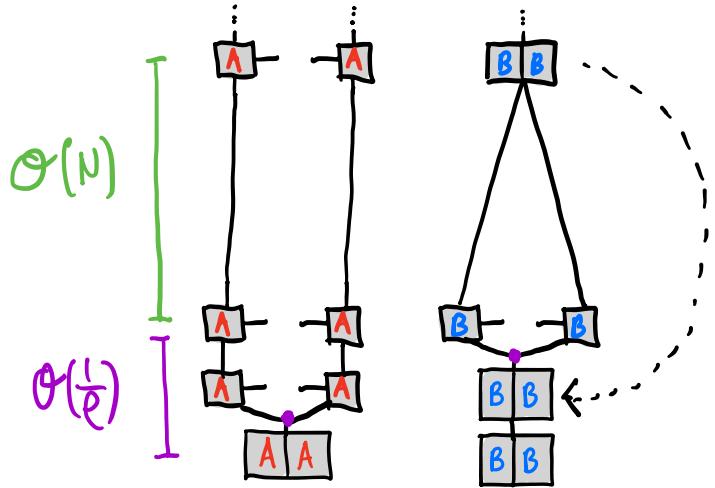
Step 3:



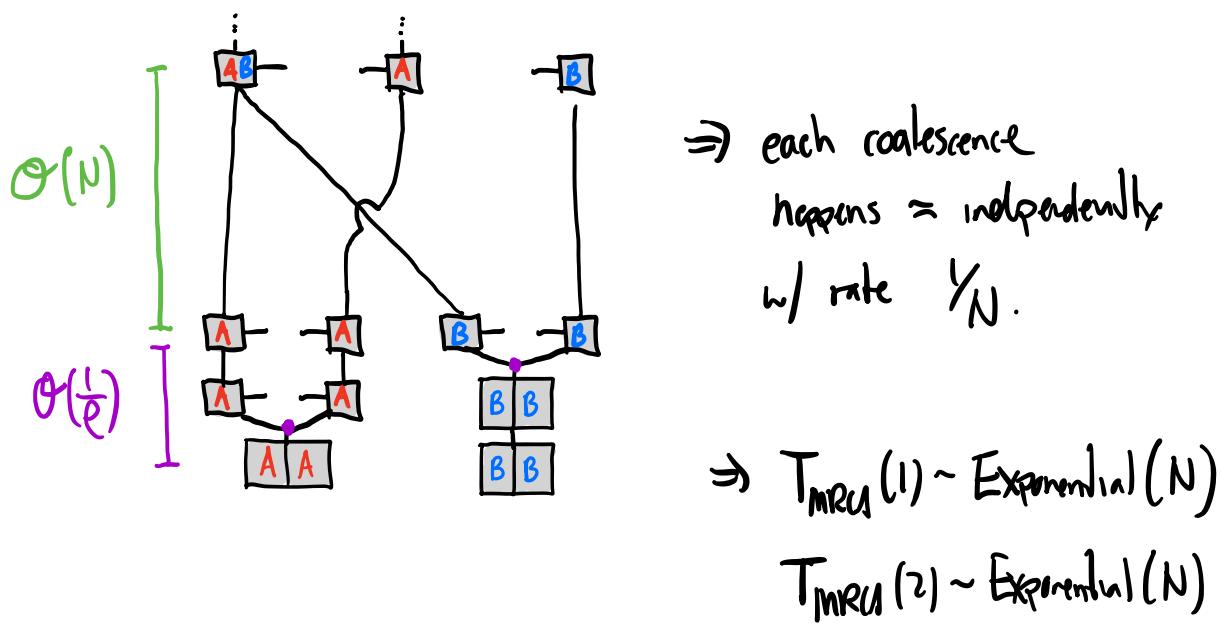
- ① coalescence = $\frac{1}{N} \binom{4}{2} = \frac{6}{N}$
- ② recomb = 0

2 different types of coalescent events:

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

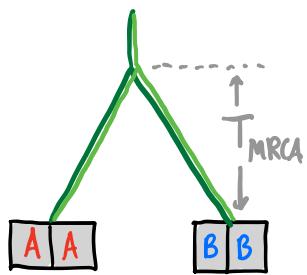


(ii) coalescence involving sampled genetic material ($\square + \square$)



Putting everything together:

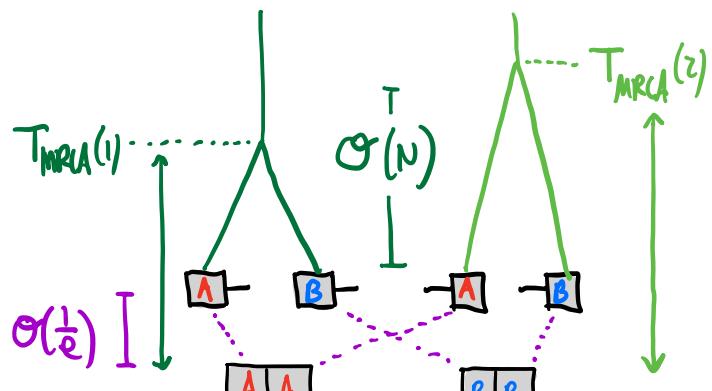
$N_e \ll 1$ (effectively asexual)



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

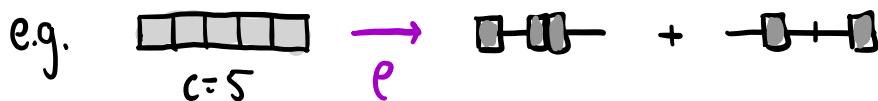
$\sim \text{Exponential}(N)$

$N_e \gg 1$ (effectively independent)



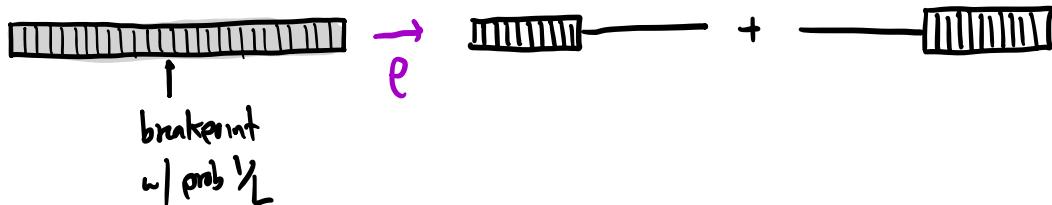
$$T_{\text{MRCA}}^{(1)}, T_{\text{MRCA}}^{(2)} \stackrel{\text{iid}}{\sim} \text{Exponential}(N)$$

\Rightarrow same idea works for > 2 chromosomes:



\Rightarrow also works for other forms of recombination:

e.g. crossover:



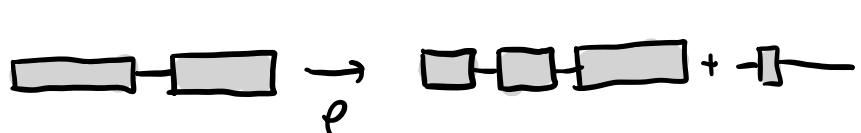
\Rightarrow 2nd event



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 ...

$$\frac{ds(\vec{s})}{dt} = \cancel{\sim (\vec{x})} + \sim L \times \mu$$
$$+ \sim e + \sim \frac{z}{J_N}$$

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