

Lecture 28. Coalescents, part 1 (various cases)

Joe Felsenstein

Department of Genome Sciences and Department of Biology

The Wright-Fisher model

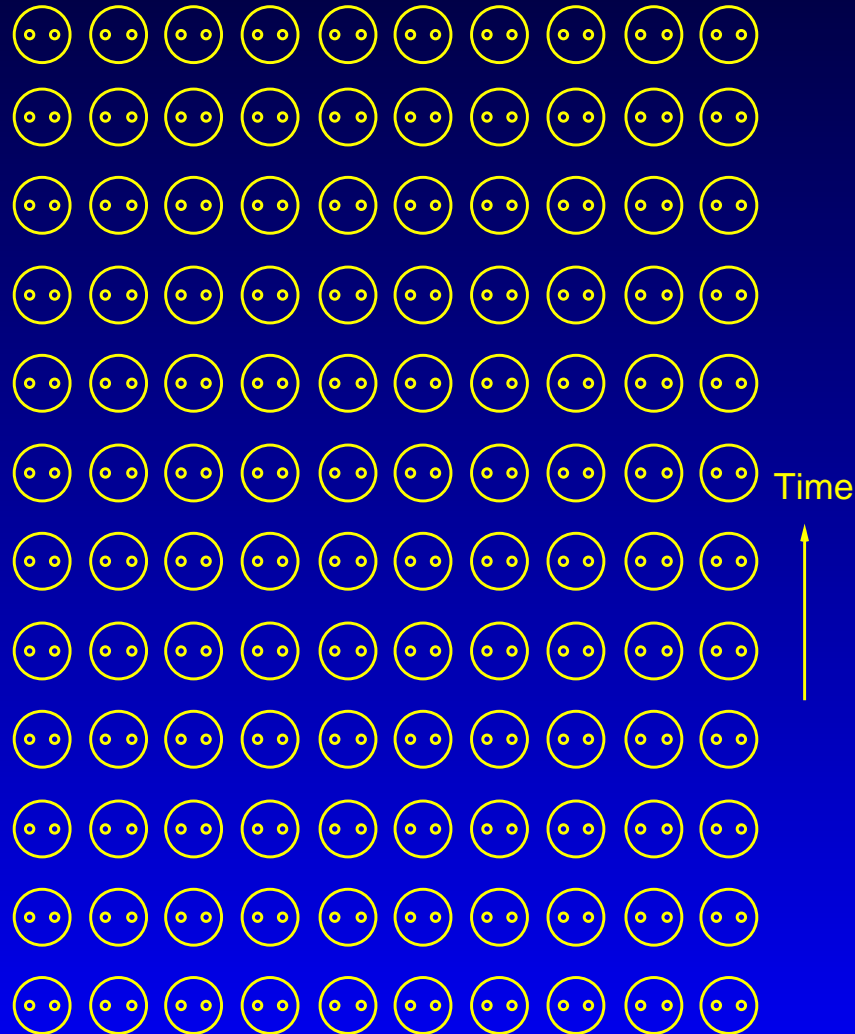
This is the canonical model of genetic drift in populations. It was invented in 1932 and 1930 by Sewall Wright and R. A. Fisher. In this model the next generation is produced by doing this:

- Choose two individuals *with replacement* (including the possibility that they are the same individual) to be parents,
- Each produces one gamete, these become a diploid individual,
- Repeat these steps until N diploid individuals have been produced.

The effect of this is to have each locus in an individual in the next generation consist of two genes sampled from the parents' generation at random, with replacement.

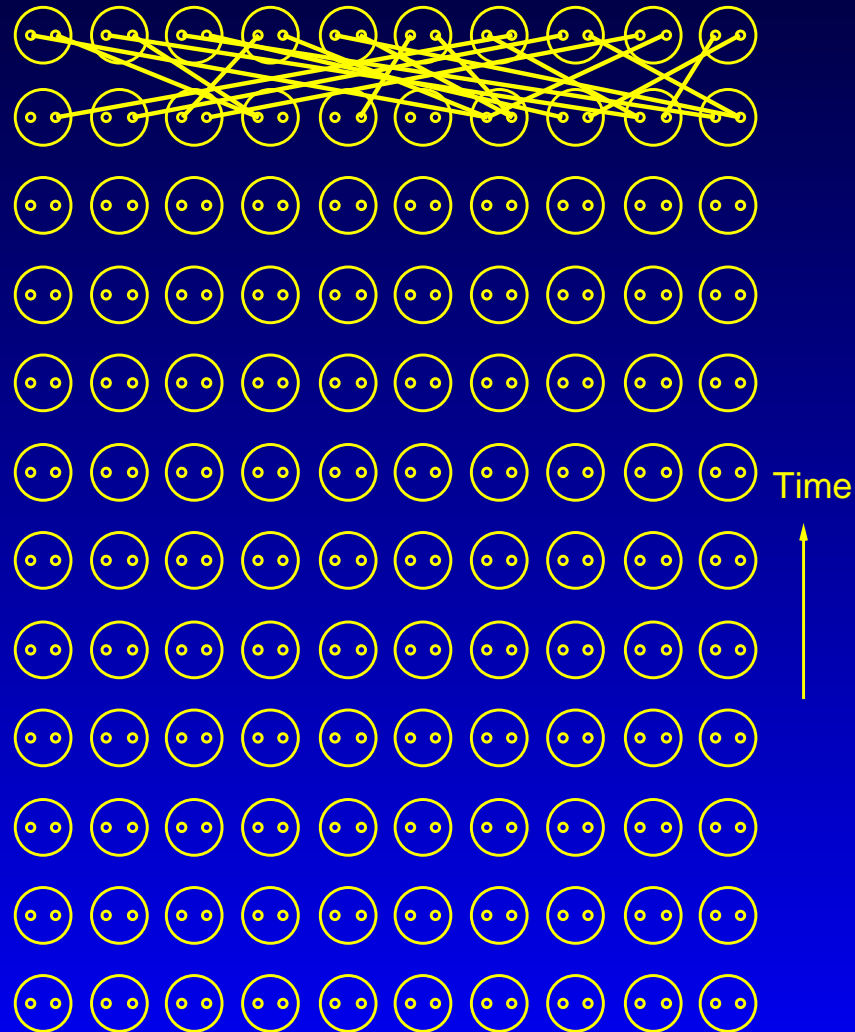
The ancestry of gene copies in a Wright-Fisher model

A random-mating population



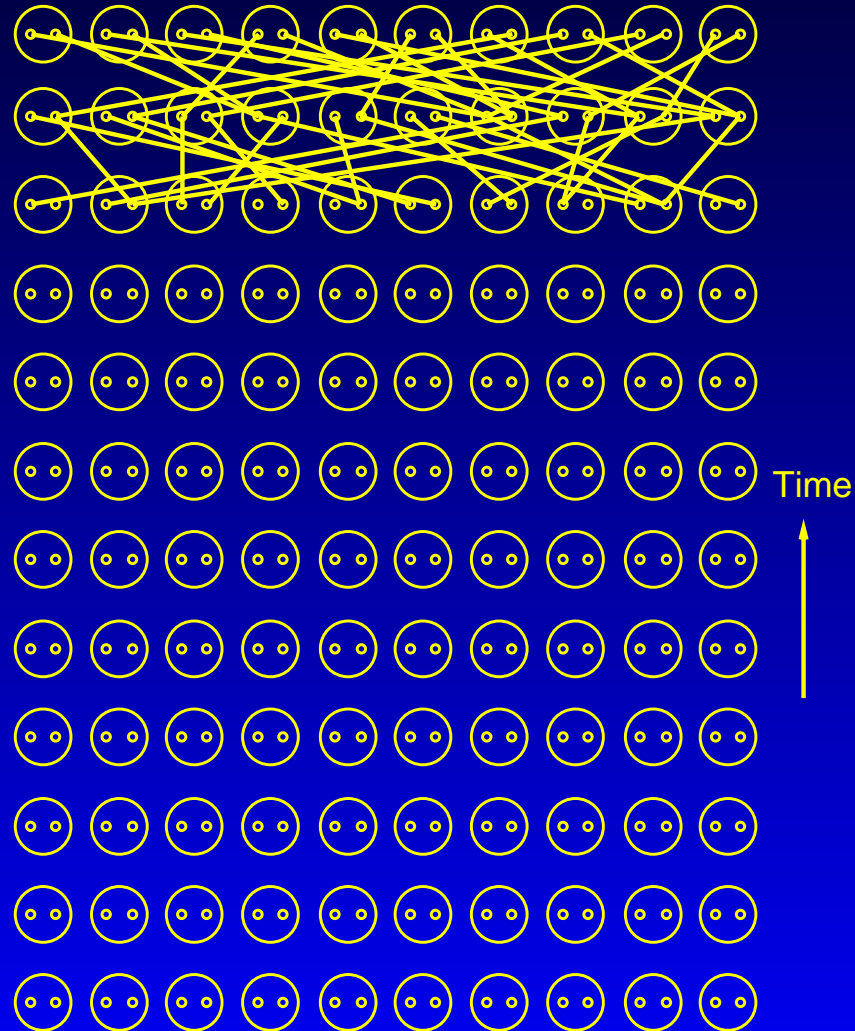
Each gene copy is drawn from a random one in the previous generation

A random-mating population



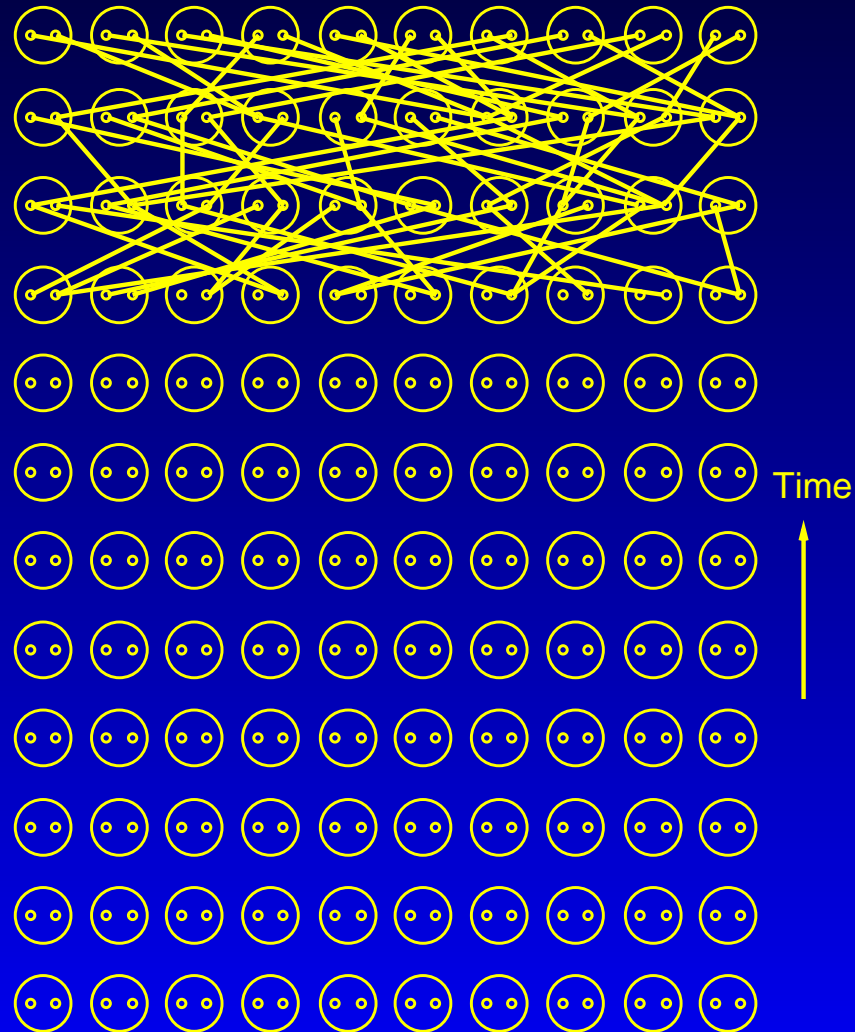
and going further back ...

A random-mating population



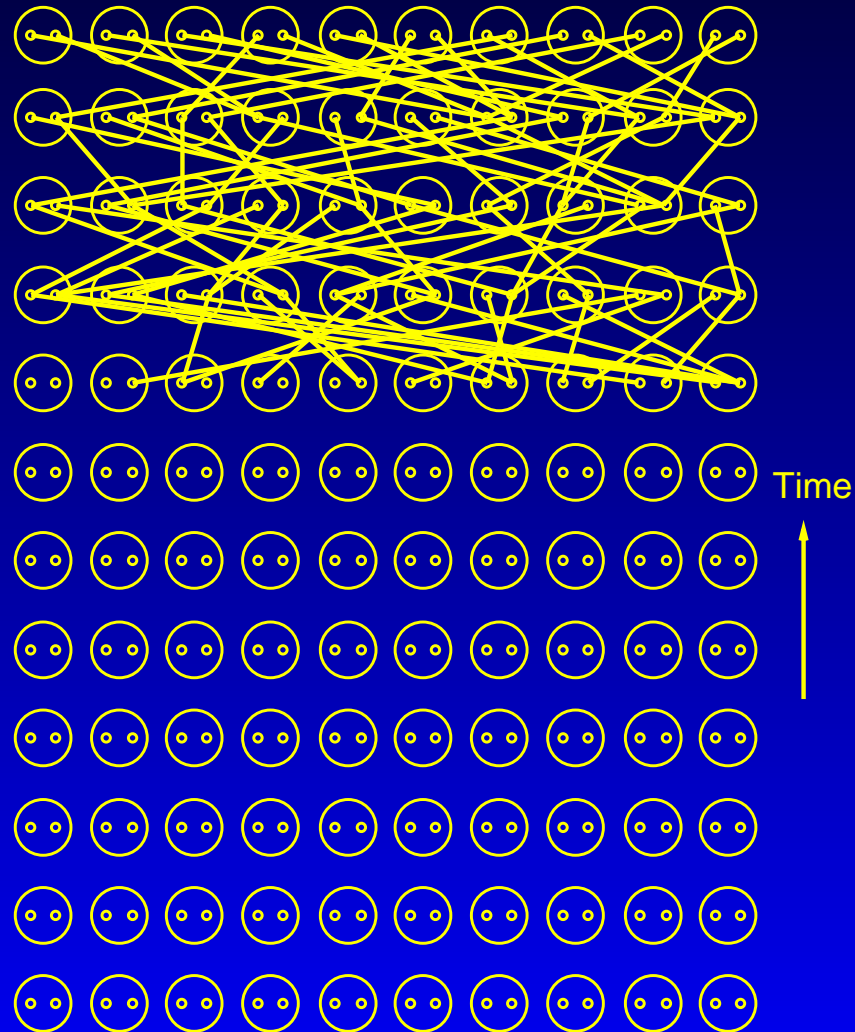
and even further

A random-mating population



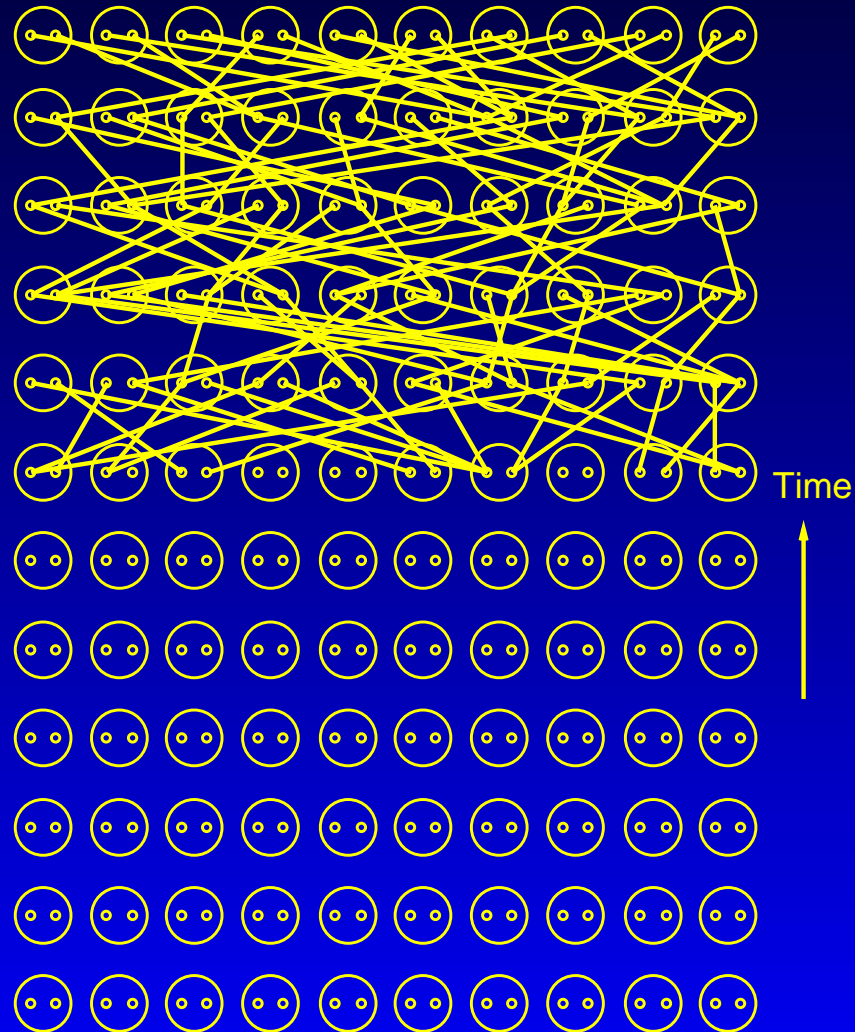
and further

A random-mating population



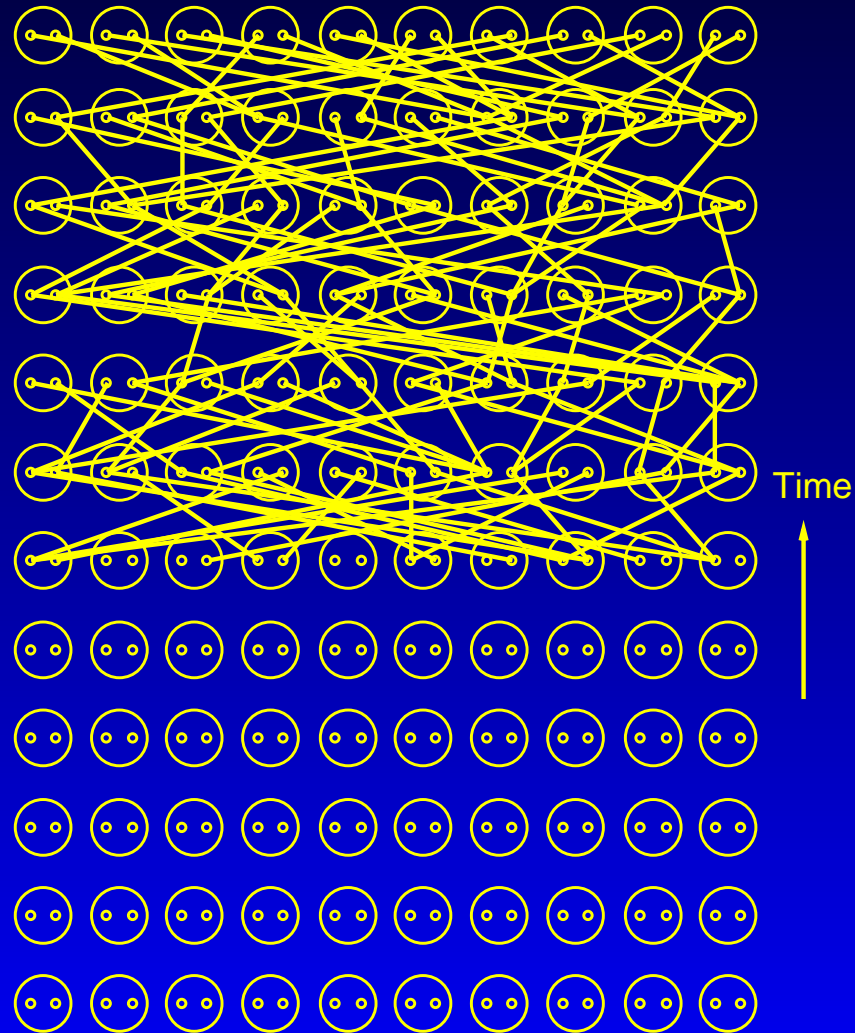
and so on

A random-mating population



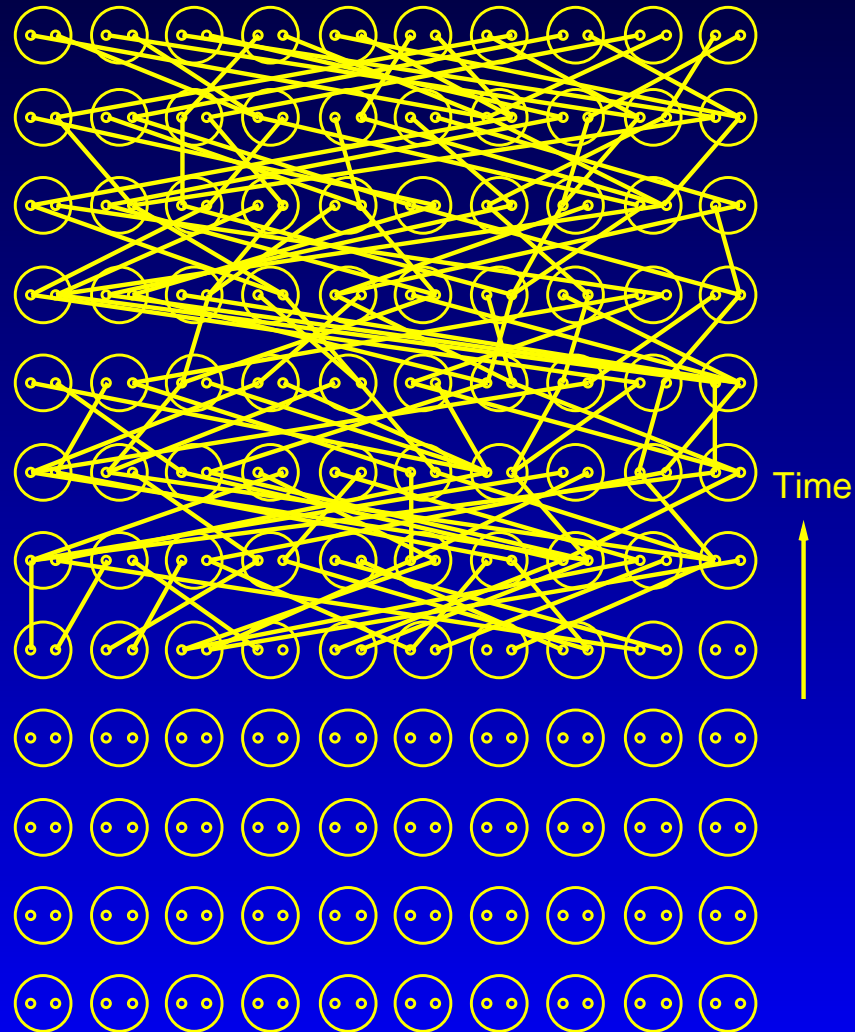
and on

A random-mating population



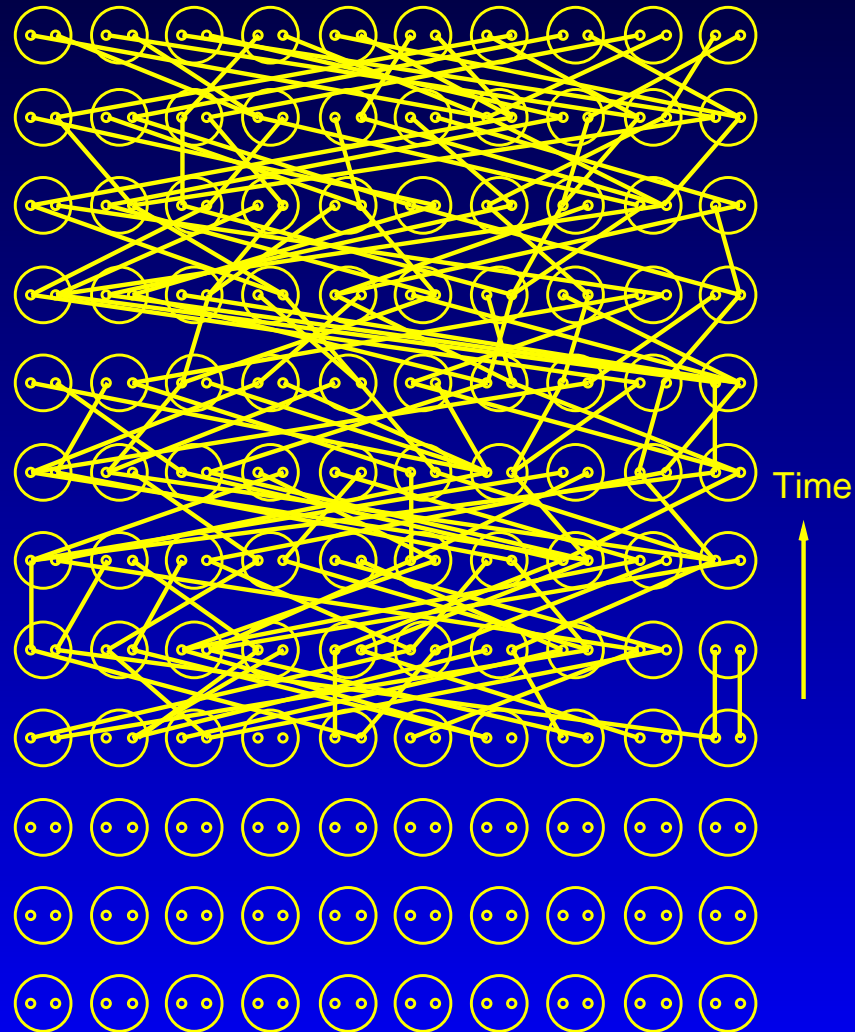
(yawn)

A random-mating population



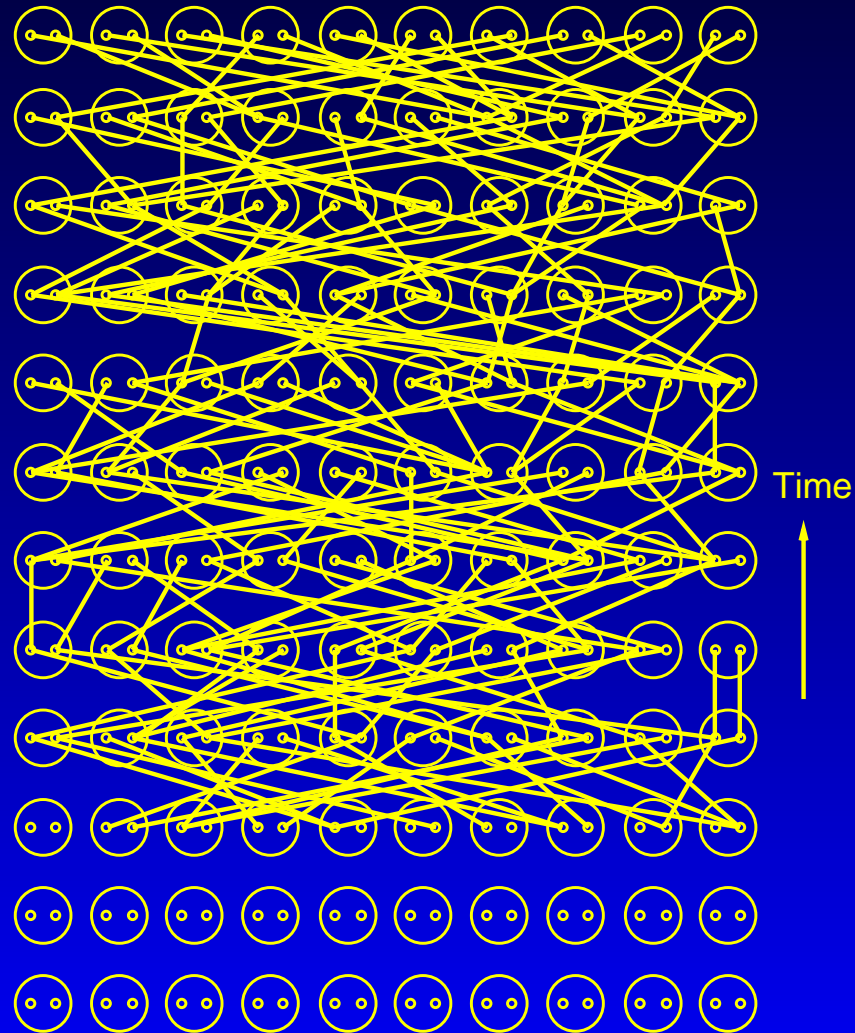
nearly there

A random-mating population



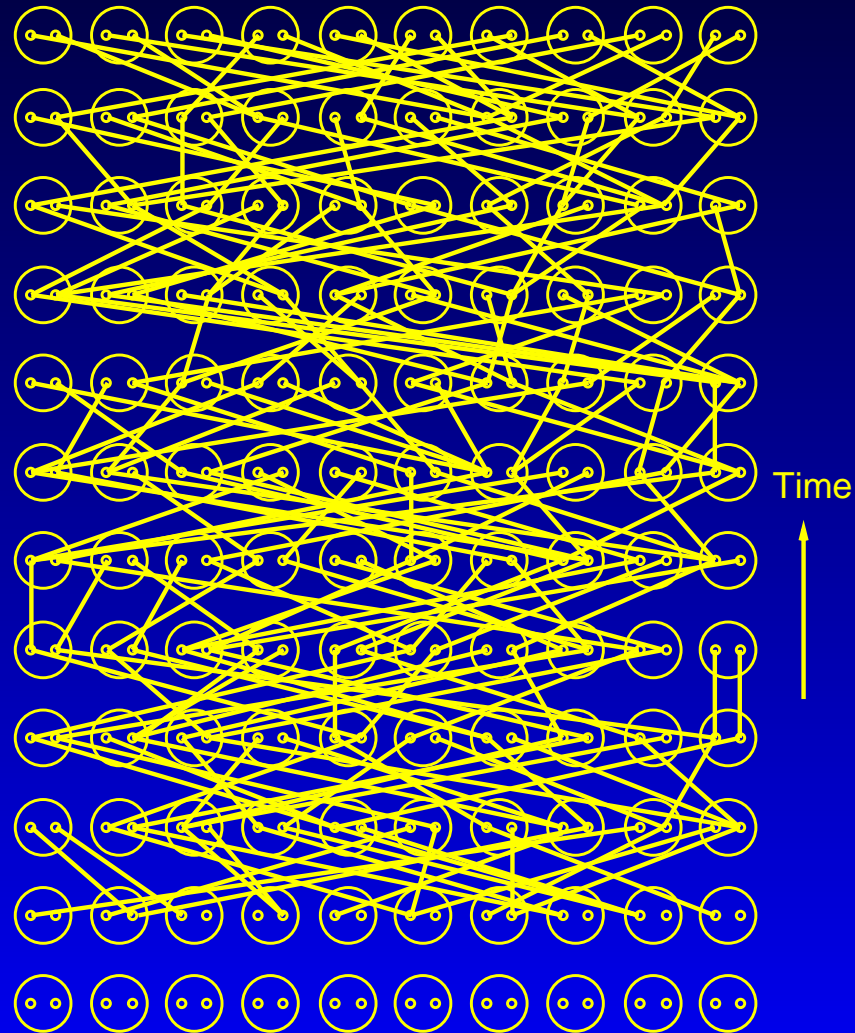
almost!

A random-mating population



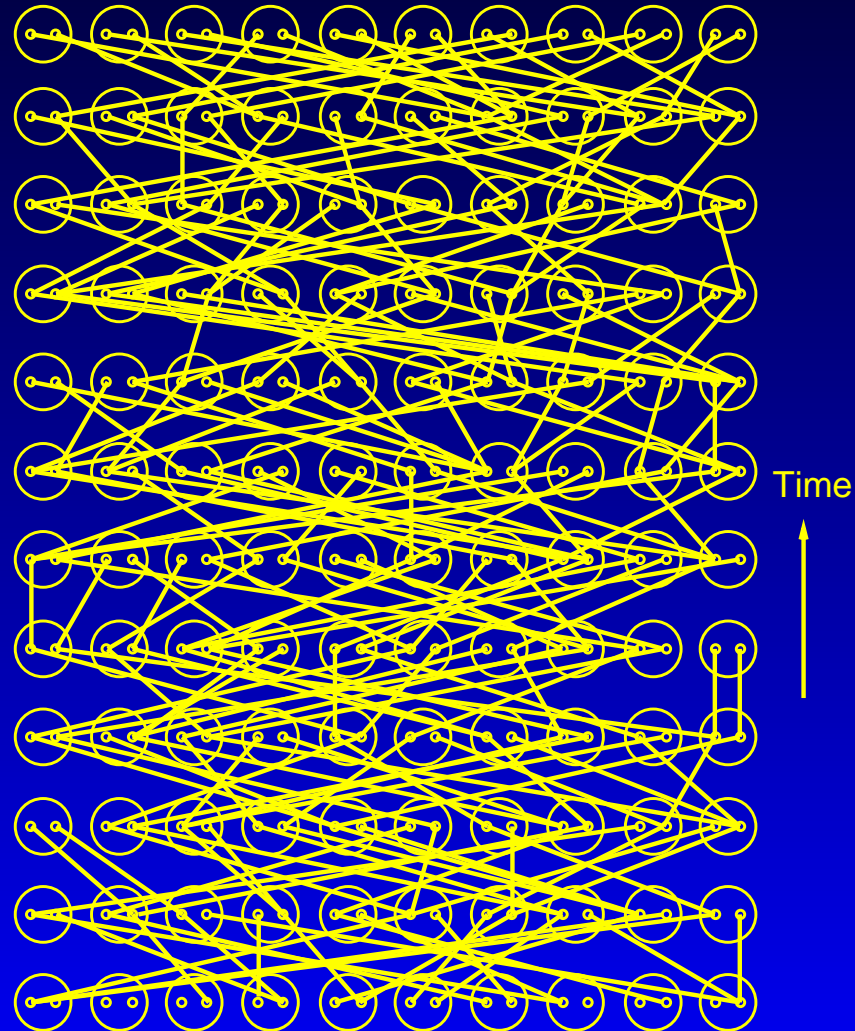
one more after this

A random-mating population



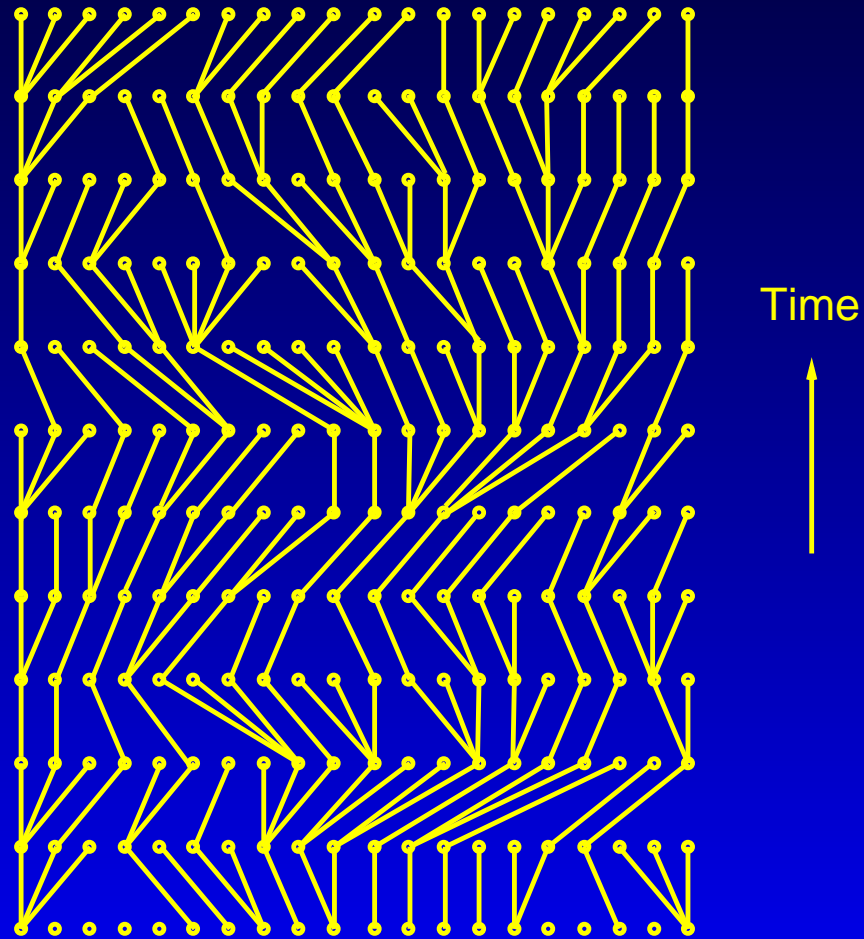
OK, so this is the pedigree of genes

A random-mating population



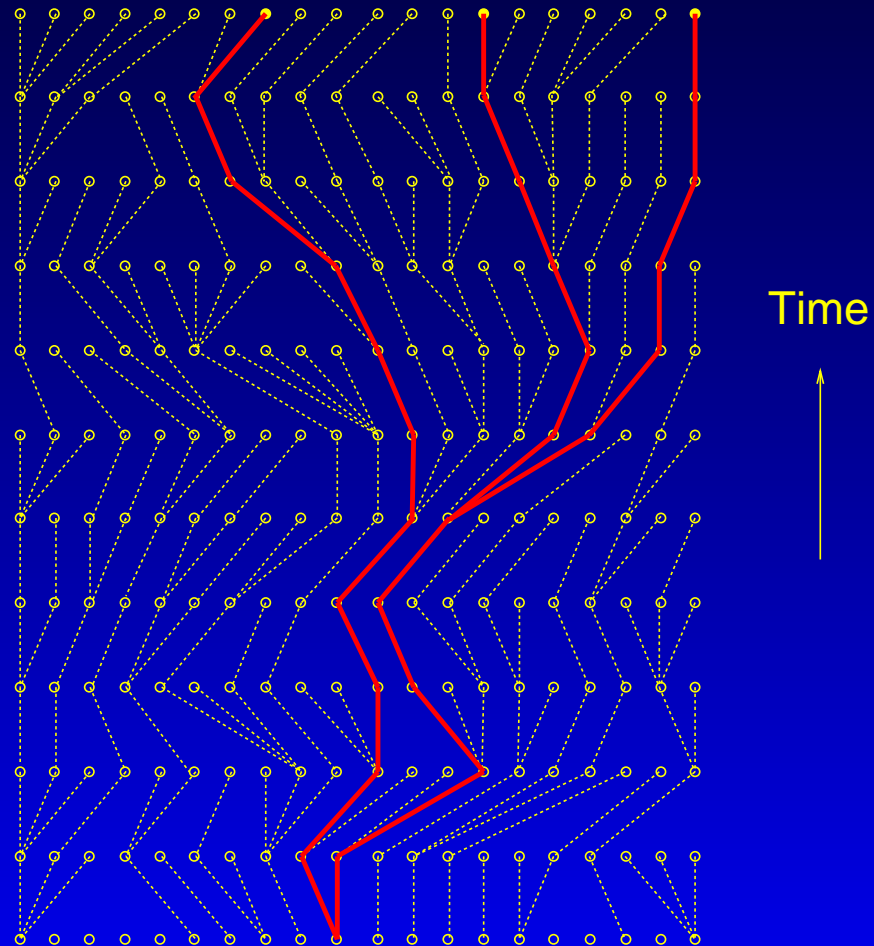
The ancestry of gene copies, untangled

Genealogy of gene copies, after reordering the copies

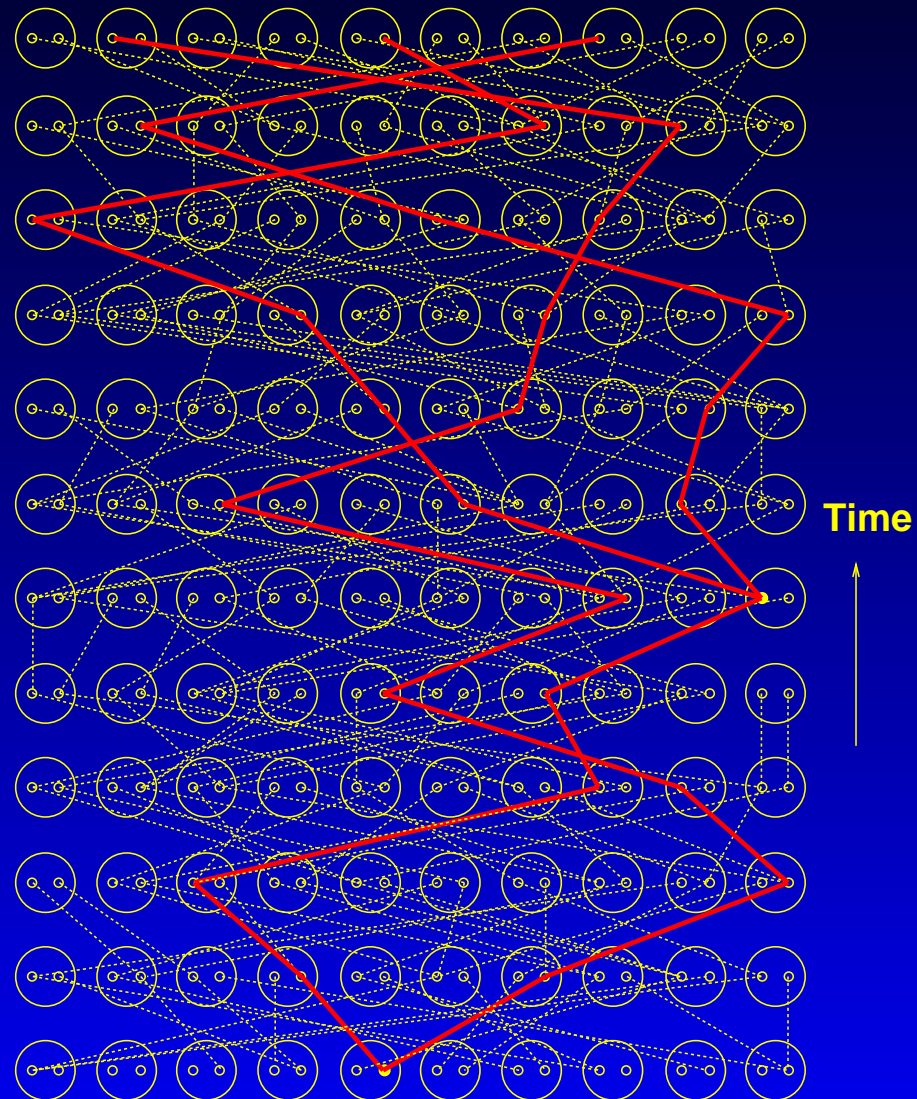


The ancestry of a sample of 3 genes

Genealogy of a small sample of genes from the population



Where the tree of 3 copies is in the genealogy



Kingman's coalescent process

Coalescent trees of gene copies within species (Kingman, 1982)

Random collision of lineages as go back in time (sans recombination)

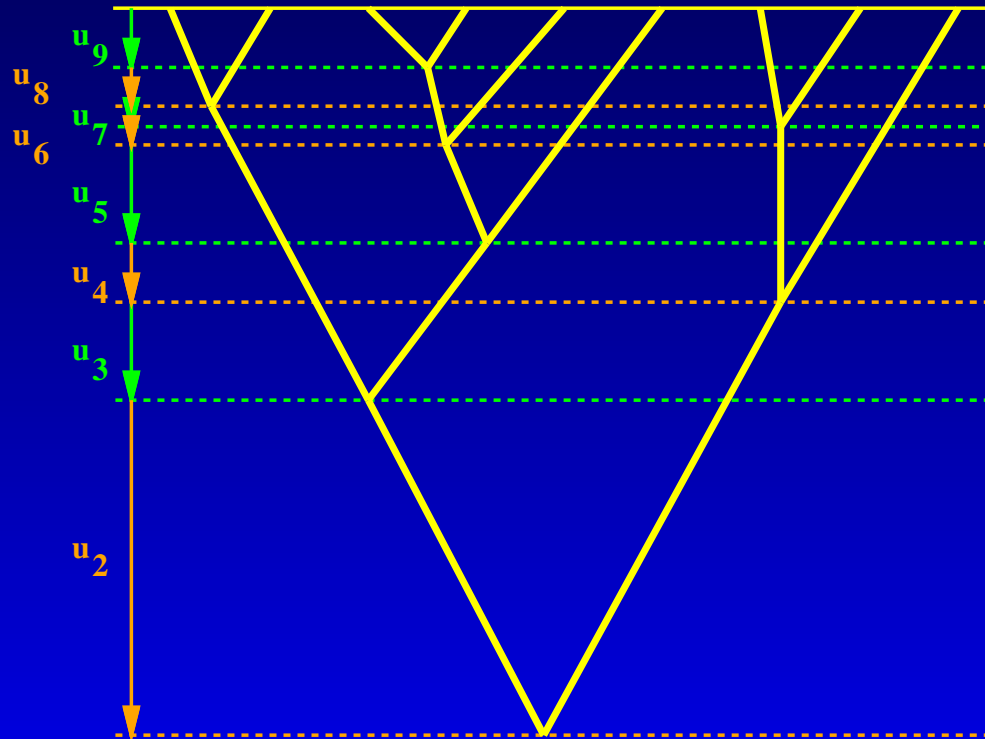
Collision is faster the smaller the effective population size

Average time for
k copies to coalesce to

$$k-1 = \frac{4N}{k(k-1)}$$

Average time for
two copies to coalesce

= 2N generations



In a diploid population of
effective population size 1

Average time for n
copies to coalesce

$$= 4N \left(1 - \frac{1}{n}\right) \text{ generations}$$

The coalescent – a derivation

The probability that k lineages becomes $k - 1$ one generation earlier is (as each lineage “chooses” its ancestor independently):

$$k(k - 1)/2 \times \text{Prob (First two have same parent, rest are different)}$$

(since there are $\binom{k}{2} = k(k - 1)/2$ different pairs of copies)

We add up terms, all the same, for the $k(k - 1)/2$ pairs that could coalesce:

$$= k(k - 1)/2 \times 1 \times \frac{1}{2N} \times \left(1 - \frac{1}{2N}\right) \times \left(1 - \frac{2}{2N}\right) \times \dots \times \left(1 - \frac{k - 2}{2N}\right)$$

so that the total probability that a pair coalesces is

$$= k(k - 1)/4N + O(1/N^2)$$

probability that someone coalesces

Note that the total probability that some combination of lineages coalesces is

$$1 - \text{Prob (all genes have separate ancestors)}$$

$$= 1 - \left[1 \times \left(1 - \frac{1}{2N}\right) \left(1 - \frac{2}{2N}\right) \cdots \left(1 - \frac{k-1}{2N}\right) \right]$$

$$= 1 - \left[1 - \frac{1 + 2 + 3 + \dots + (k-1)}{2N} + O(1/N^2) \right]$$

and since

$$1 + 2 + 3 + \dots + (n-1) = n(n-1)/2$$

(continued)

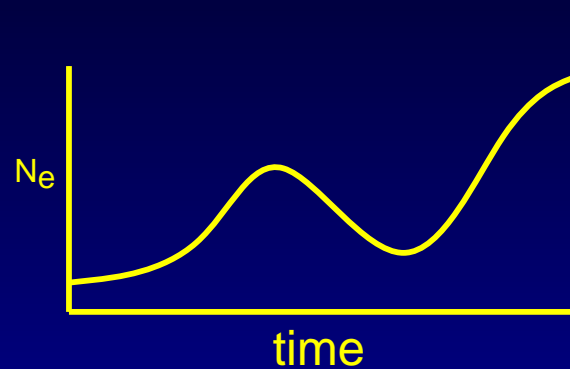
the quantity

$$= 1 - \left[1 - k(k-1)/4N + O(1/N^2) \right] \simeq k(k-1)/4N + O(1/N^2)$$

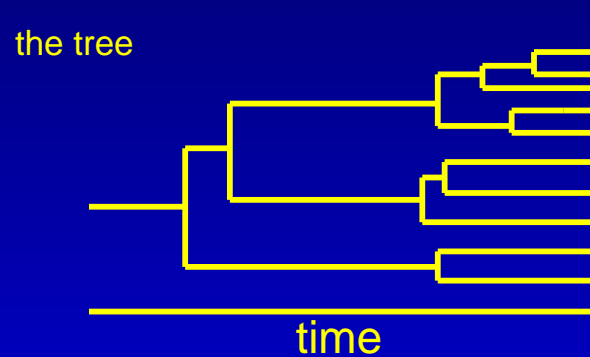
showing that the events involving 3 or more lineages simultaneously coalescing are in the terms of order $1/N^2$ and thus become unimportant if N is large. For example, when $k = 10$ and $N = 100$, there is a 0.7956 chance that there is no coalescence, 0.1874 that one pair coalesces, and only 0.01695 that more than one coalesces.

Effect of varying population size

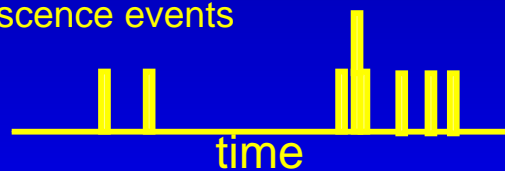
Change of population size and coalescents



the changes in population size will produce waves of coalescence

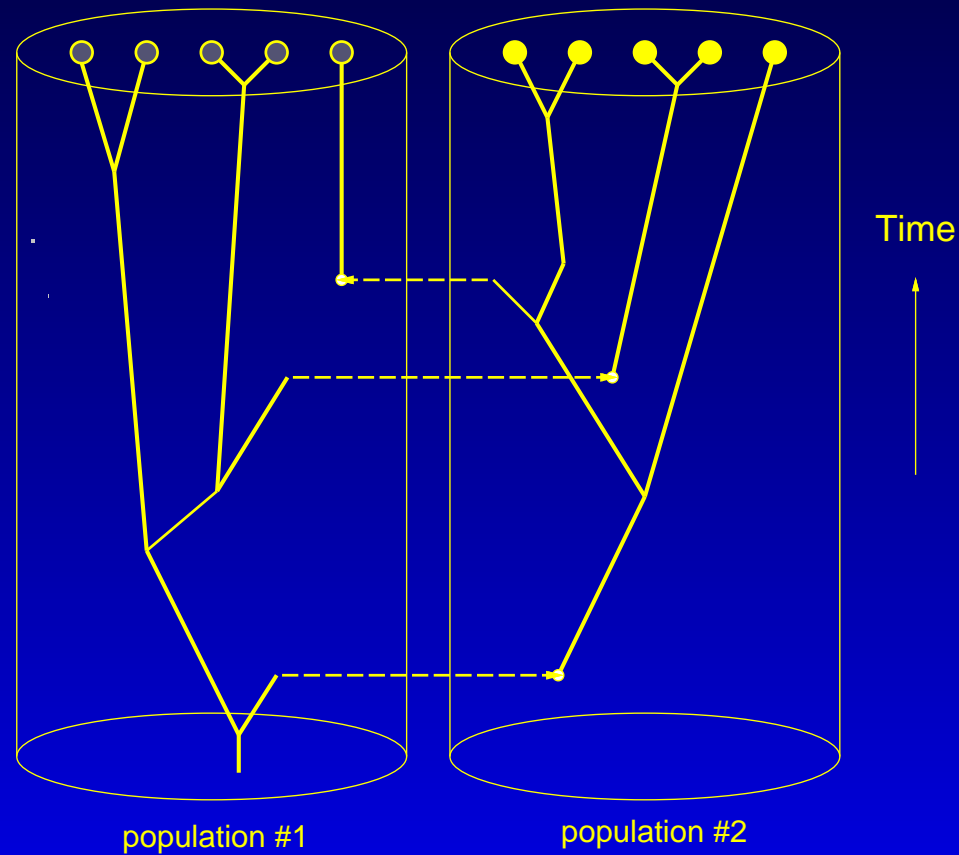


Coalescence events



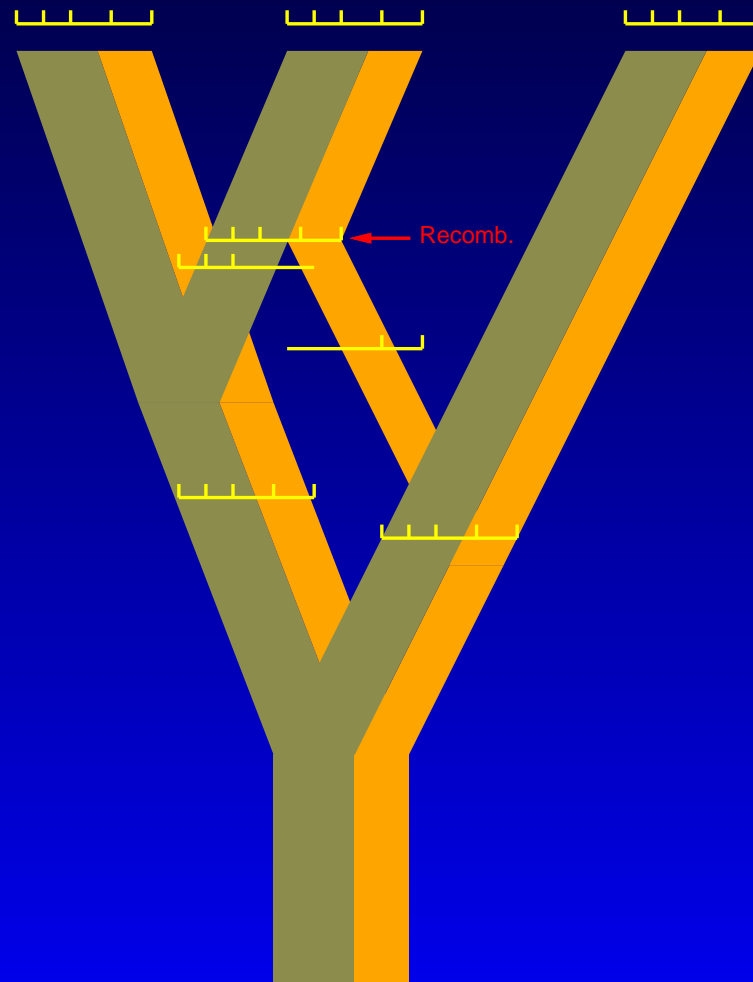
The parameters of the growth curve for N_e can be inferred by likelihood methods as they affect the prior probabilities of those trees that fit the data.

A coalescent with migration (2 populations)



A recombining coalescent

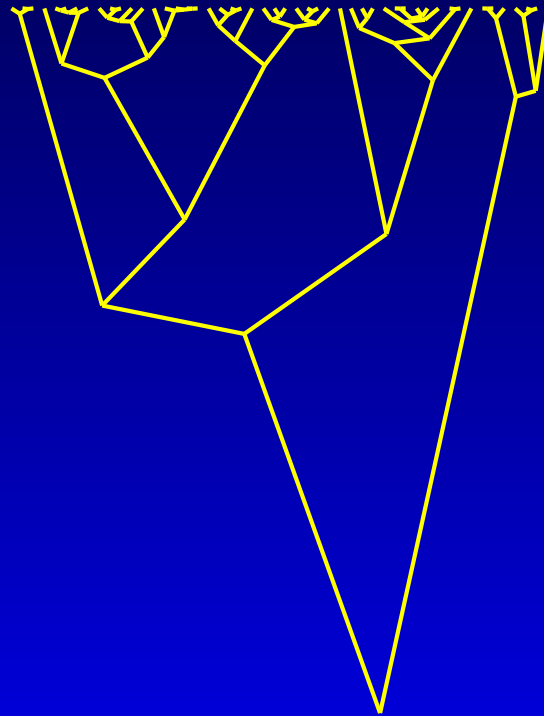
A recombining coalescent



Different markers have slightly different coalescent trees

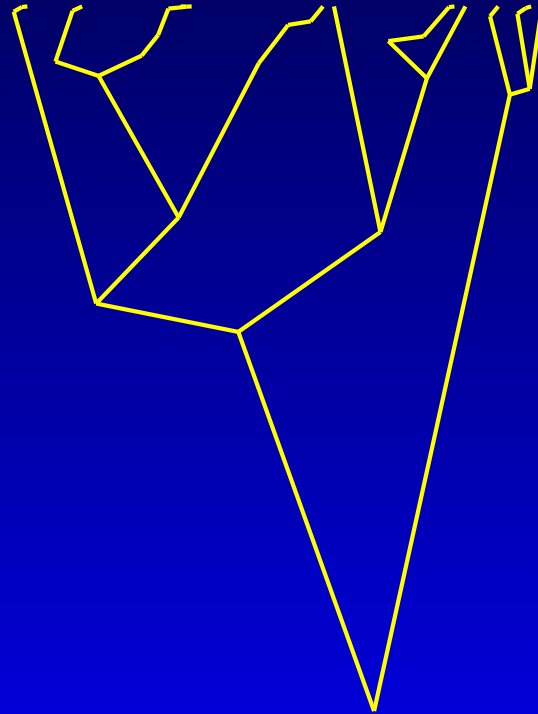
A coalescent of 50 copies

50-gene sample in a coalescent tree



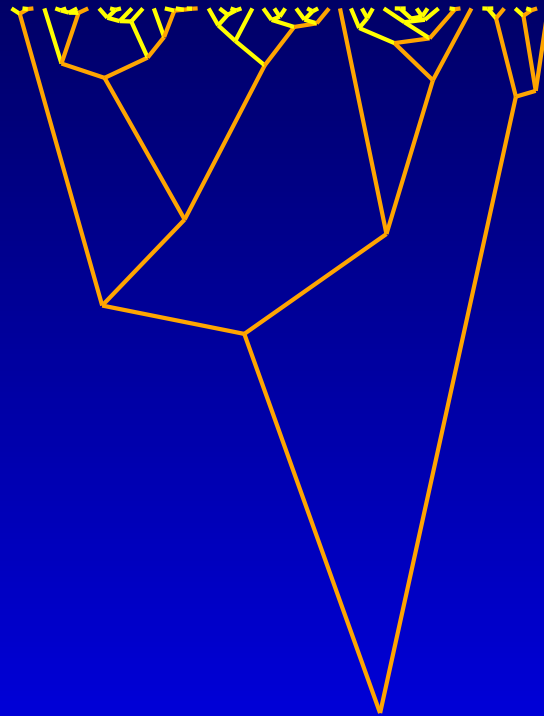
the first 10 copies only

10 genes sampled randomly out of a
10-gene sample in a coalescent tree



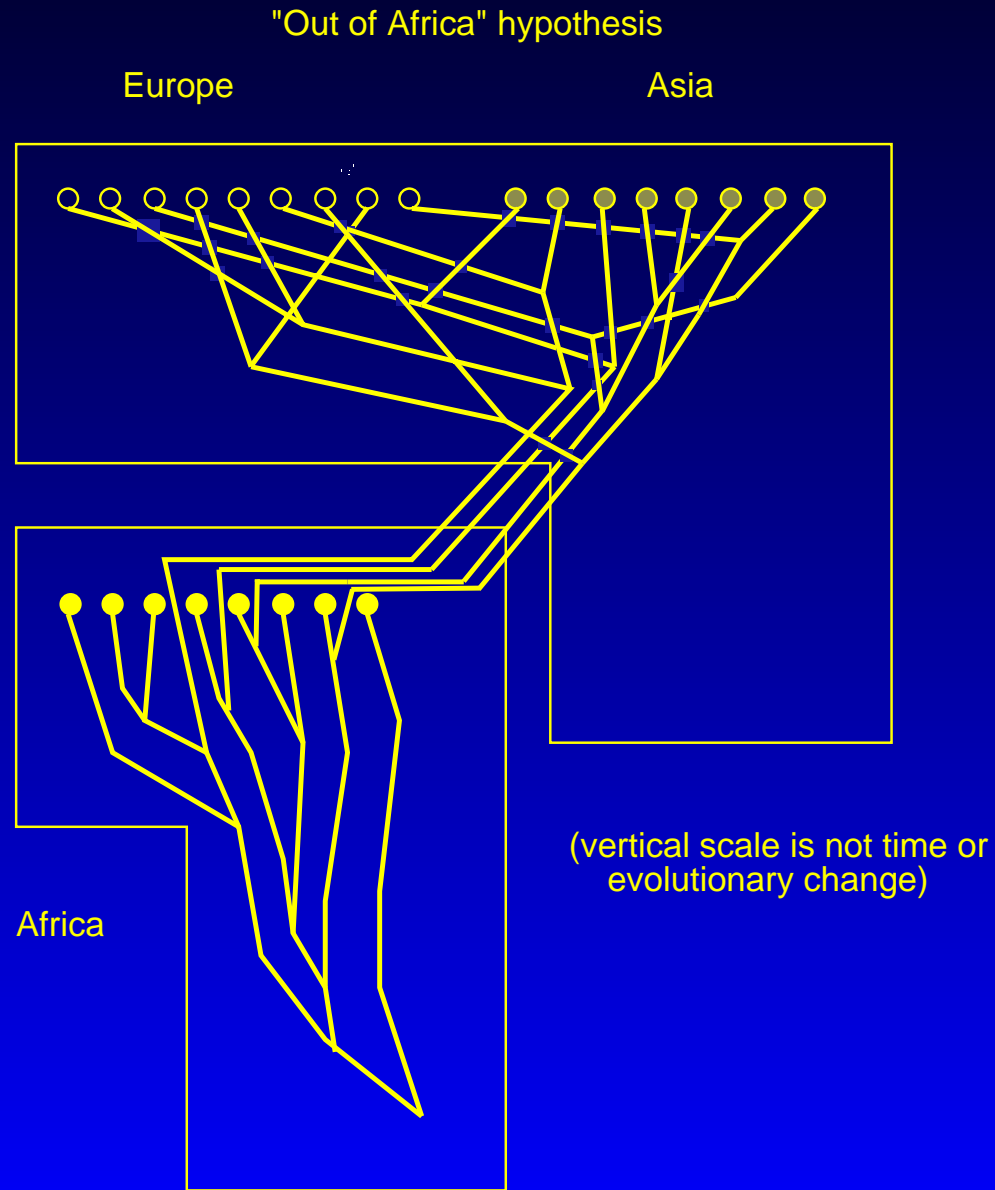
All copies, ancestry of first 10 in orange

10 genes sampled randomly out of a
50-gene sample in a coalescent tree



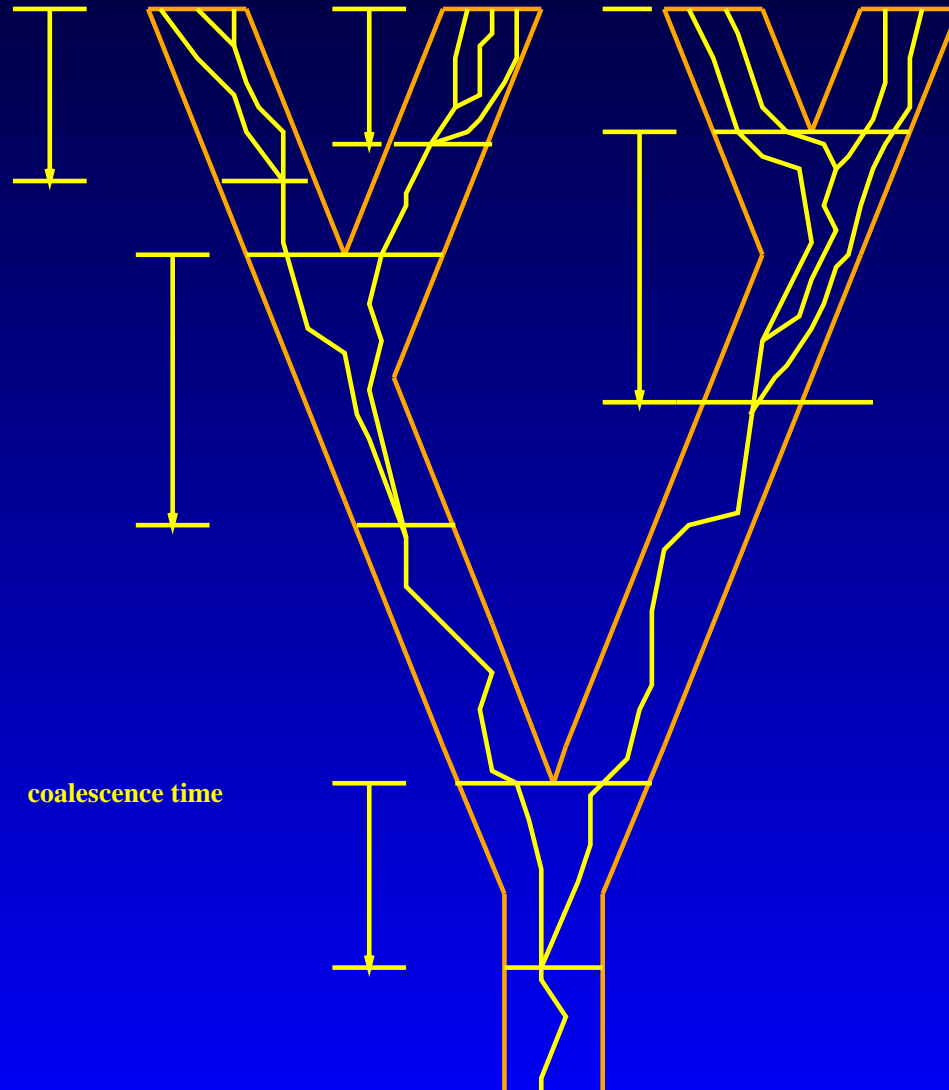
(orange lines are the 10-gene tree)

We ultimately want to treat this case



coalescents in related species

Consistency of gene tree with species tree



References

- Felsenstein, J. 1971. The rate of loss of multiple alleles in finite haploid populations. *Theoretical Population Biology* **2**: 391-403. [**Can be used to derive rates of coalescence**]
- Hudson, R. R. and N. L. Kaplan. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**: 147-164. [**Coalescent with recombination**]
- Hudson, R. R. 1990. Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **7**: 1-44. [**The first major review of coalescents**]
- Kingman, J. F. C. 1982a. The coalescent. *Stochastic Processes and Their Applications* **13**: 235-248. [**One of the papers in which the coalescent is introduced**]

more references

- Kingman, J. F. C. 1982b. On the genealogy of large populations. *Journal of Applied Probability* **19A**: 27-43. [One of the papers in which the coalescent is introduced]
- Krone, S. M. and C. Neuhauser. 1997. Ancestral processes with selection. *Theoretical Population Biology* **51**: 210-237. [A very original extension of the coalescent to allow selection]
- Neuhauser, C., and S. M. Krone. 1997. The genealogy of samples in models with selection. *Genetics* **145**: 519-534. [A very original extension of the coalescent to allow selection]
- Takahata, N. 1988. The coalescent in two partially isolated diffusion populations. *Genetical Research* **52**: 213-222. [Coalescents with migration]

How it was done

This projection produced as a PDF, not a PowerPoint file, and viewed using the Full Screen mode (in the View menu of Adobe Acrobat Reader):

- using the `prosper` style in LaTeX,
- using LaTeX to make a `.dvi` file,
- using `dvips` to turn this into a Postscript file,
- using `ps2pdf` to mill it into a PDF file, and
- displaying the slides in Adobe Acrobat Reader.

Result: nice slides using freeware.