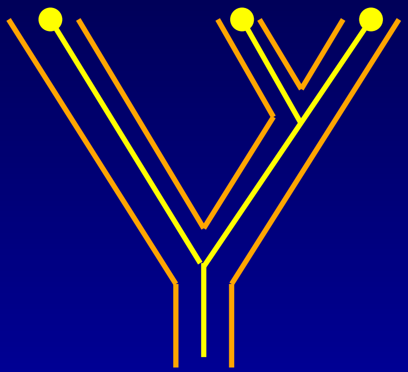
Trees of genes within species

Joe Felsenstein

University of Washington, Seattle

Molecular evolution assumes single-gene populations

In effect, by taking a single copy as representing the species, and the divergence of the copies as representing species divergencea, molecular evolutionists assume population sizes of $N = \frac{1}{2}$.



What happens as we look more closely at genealogies of genes when the species are populations of organisms?

> Random genetic drift (random births and deaths and random Mendelian segregation) forms trees of genes within populations.

The forks in the trees are not speciations, but are copying of one copy into two descendants.

The Wright-Fisher model

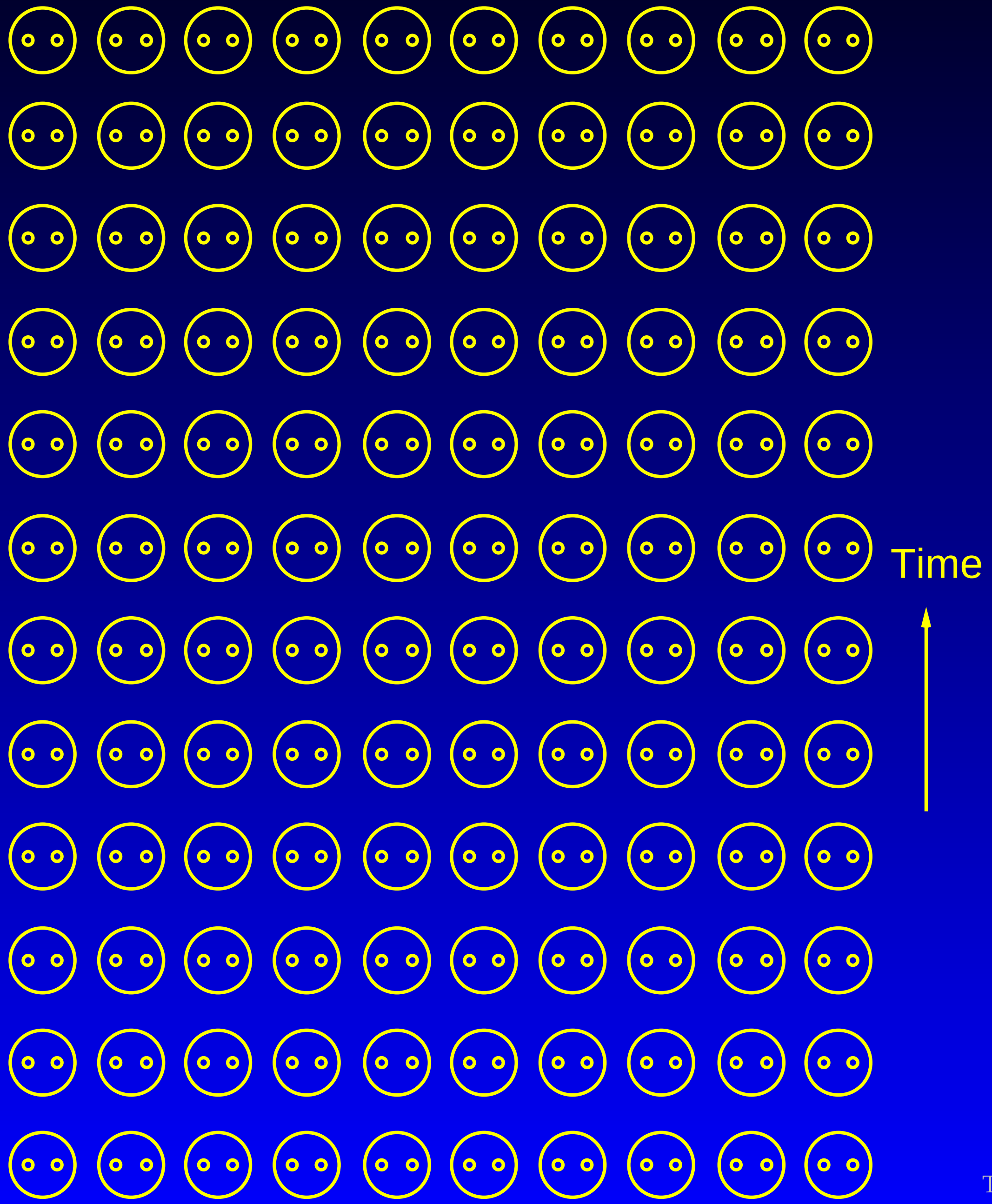
This is the canonical model of genetic drift in populations. It was invented in 1932 and 1930 by Sewall Wright and R. A. Fisher.

In this model the next generation is produced by doing this:

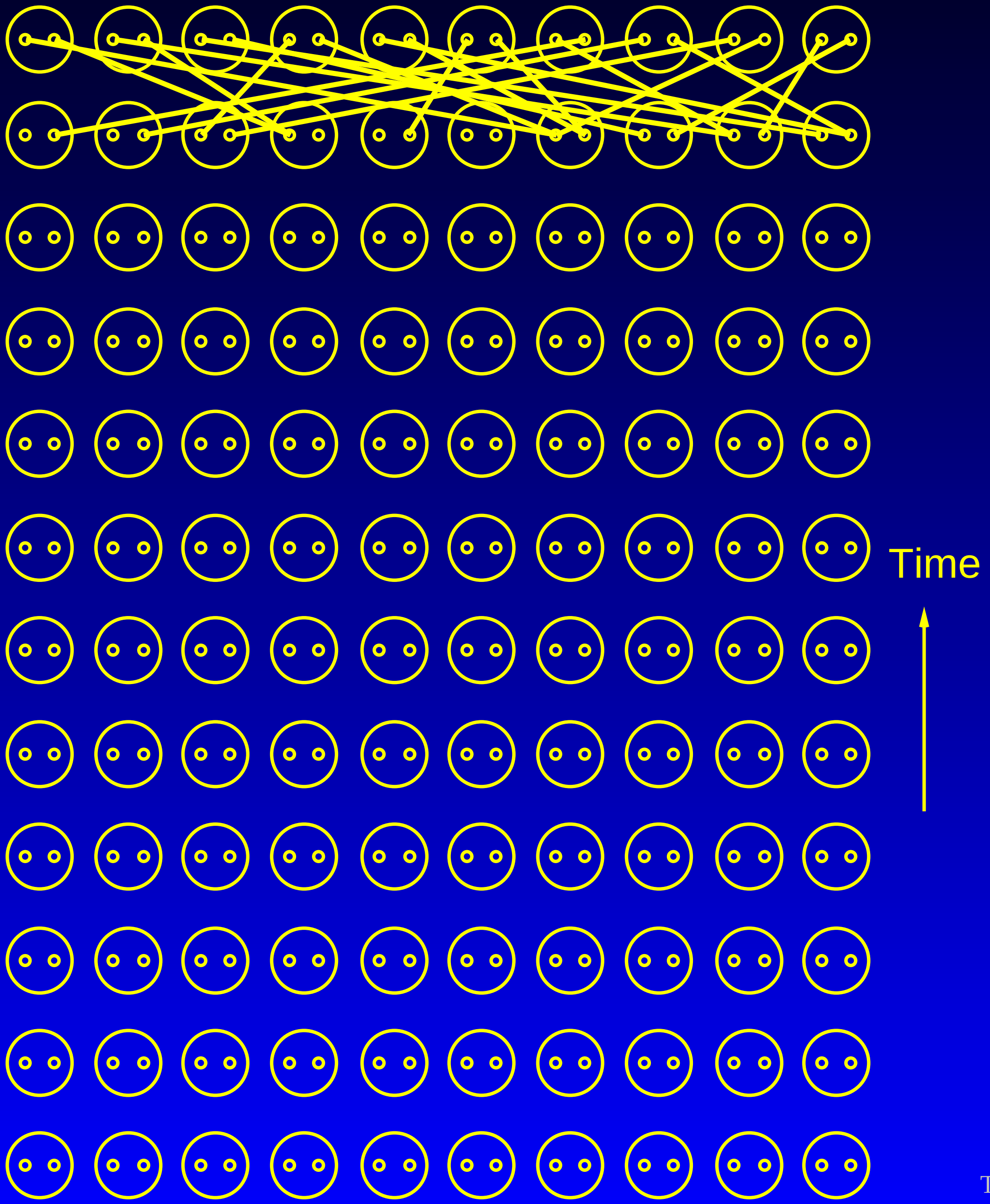
- Choose two individuals *with replacement* (including the possibility that they are the same individual) to be parents,
- Each produces one gamete, these become a diploid individual,
- Repeat these steps until N diploid individuals have been produced.

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.

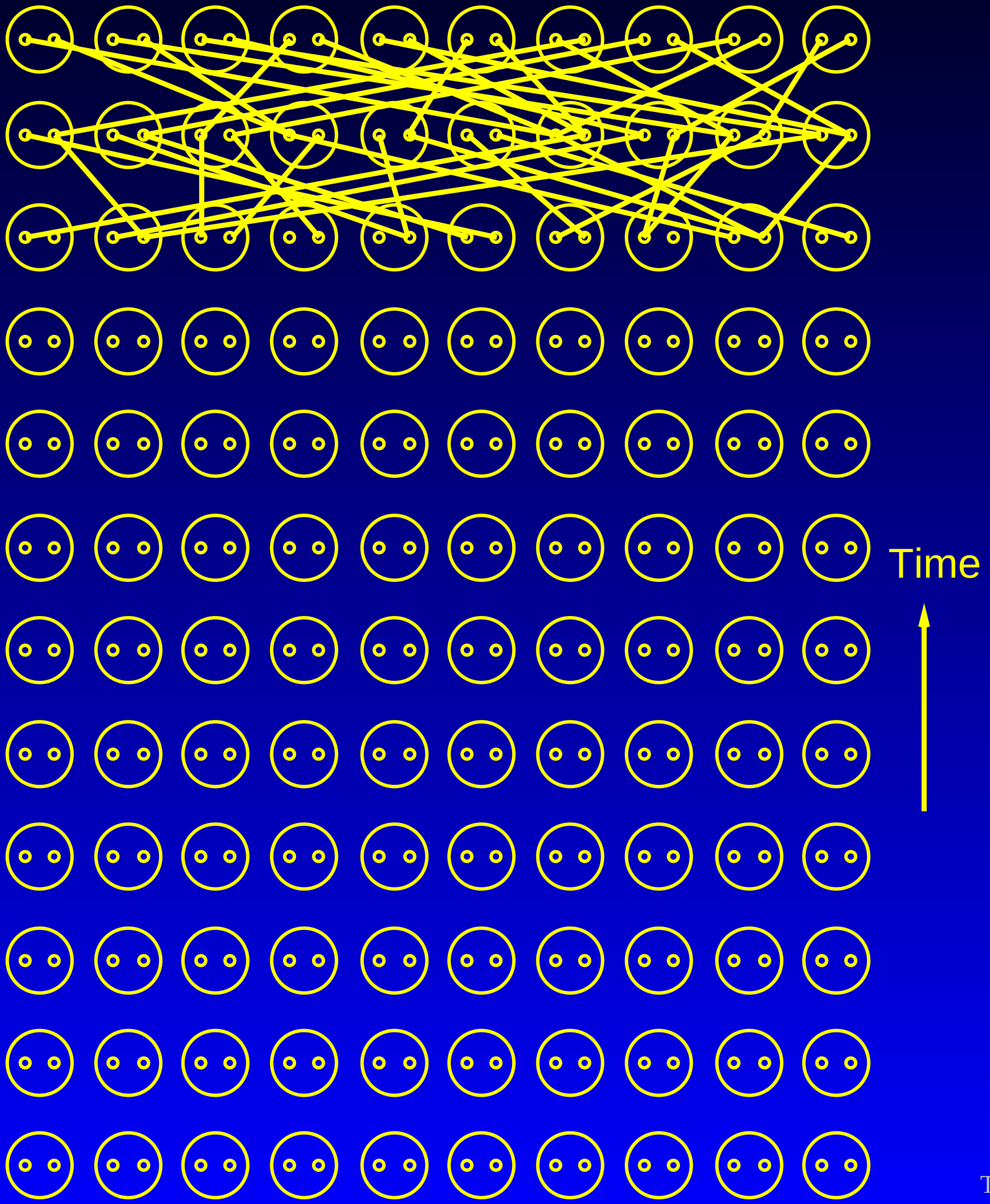
The ancestry of gene copies in a Wright-Fisher model



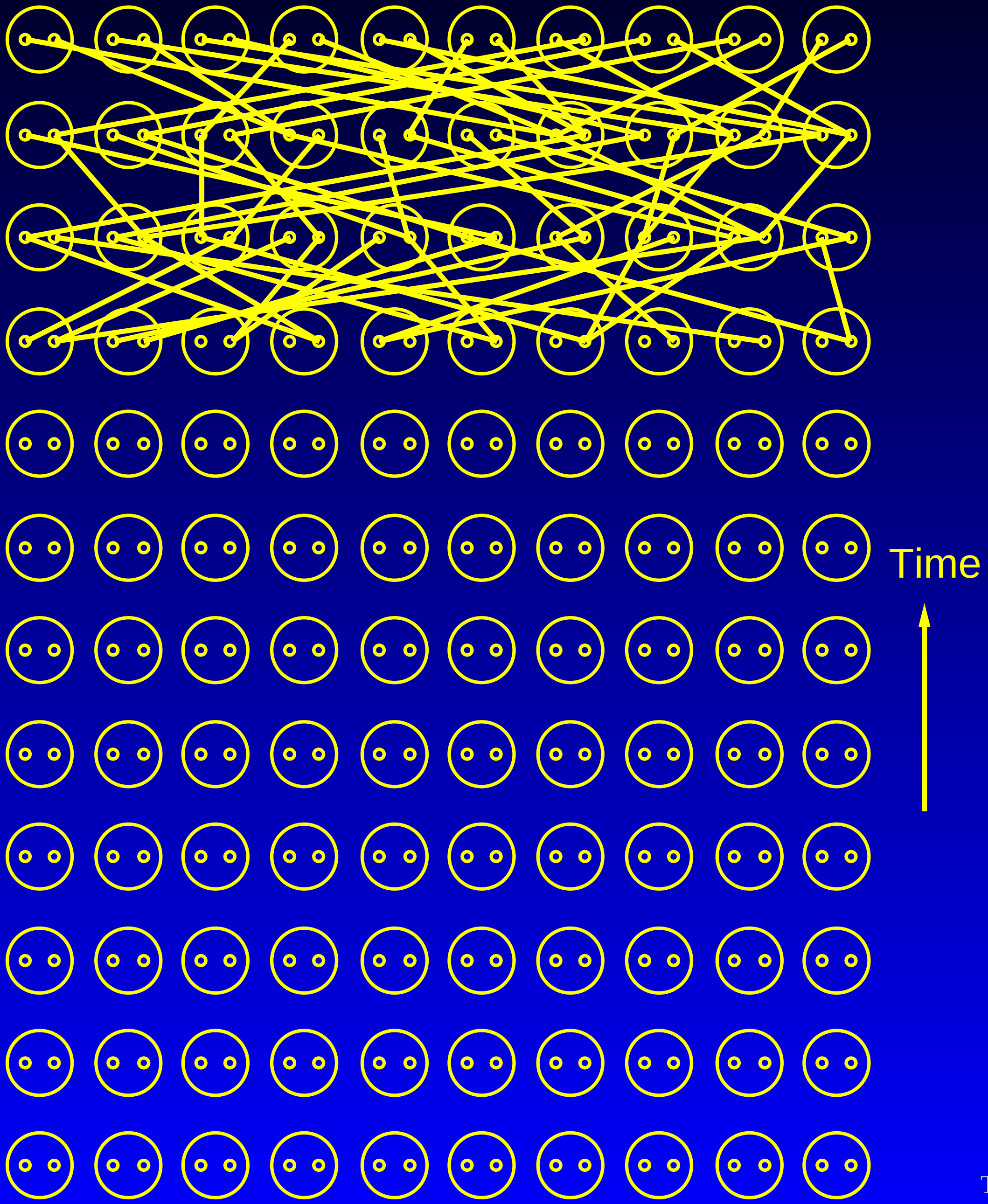
Each drawn from a random one in the previous generation



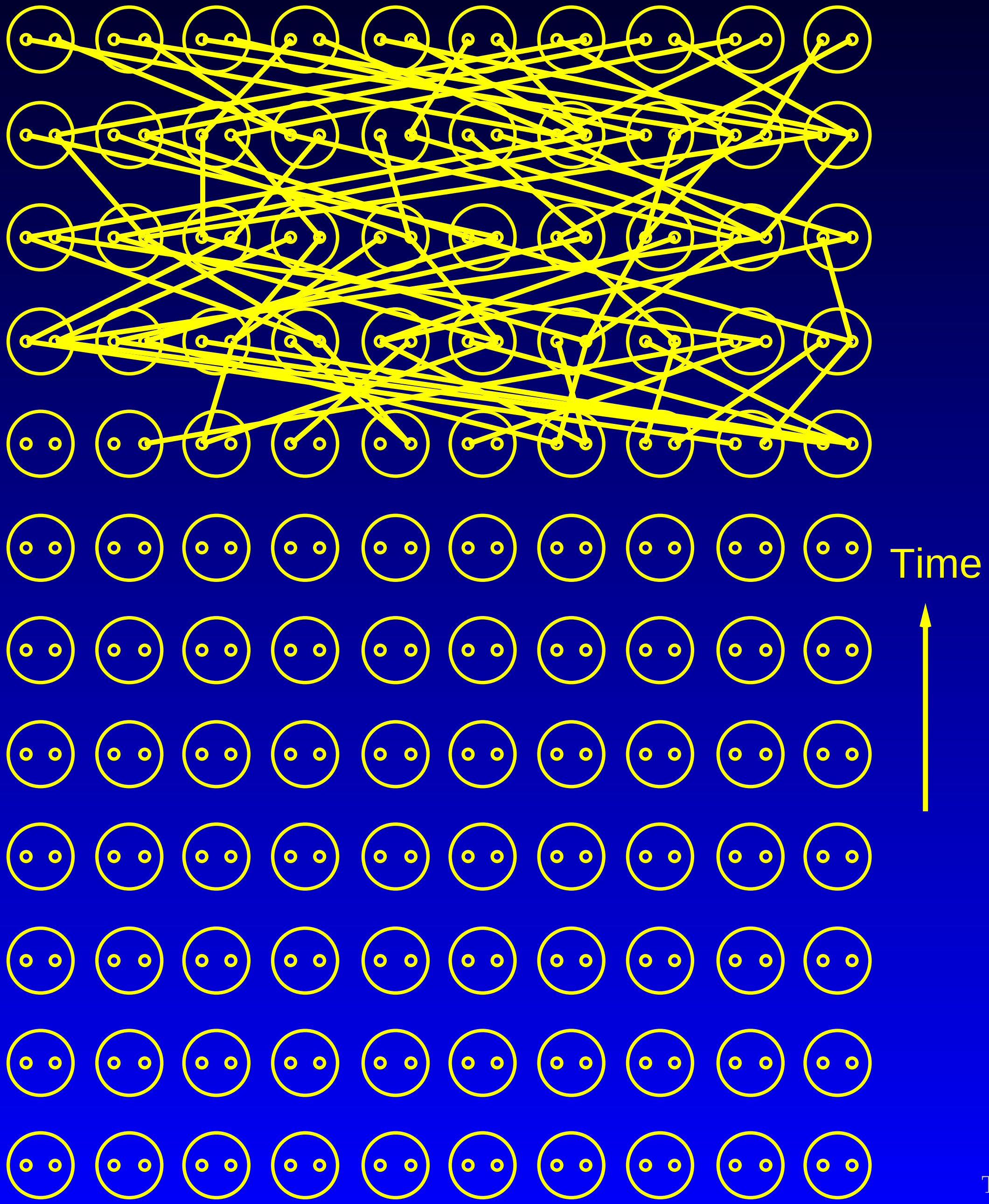
and going further back ...



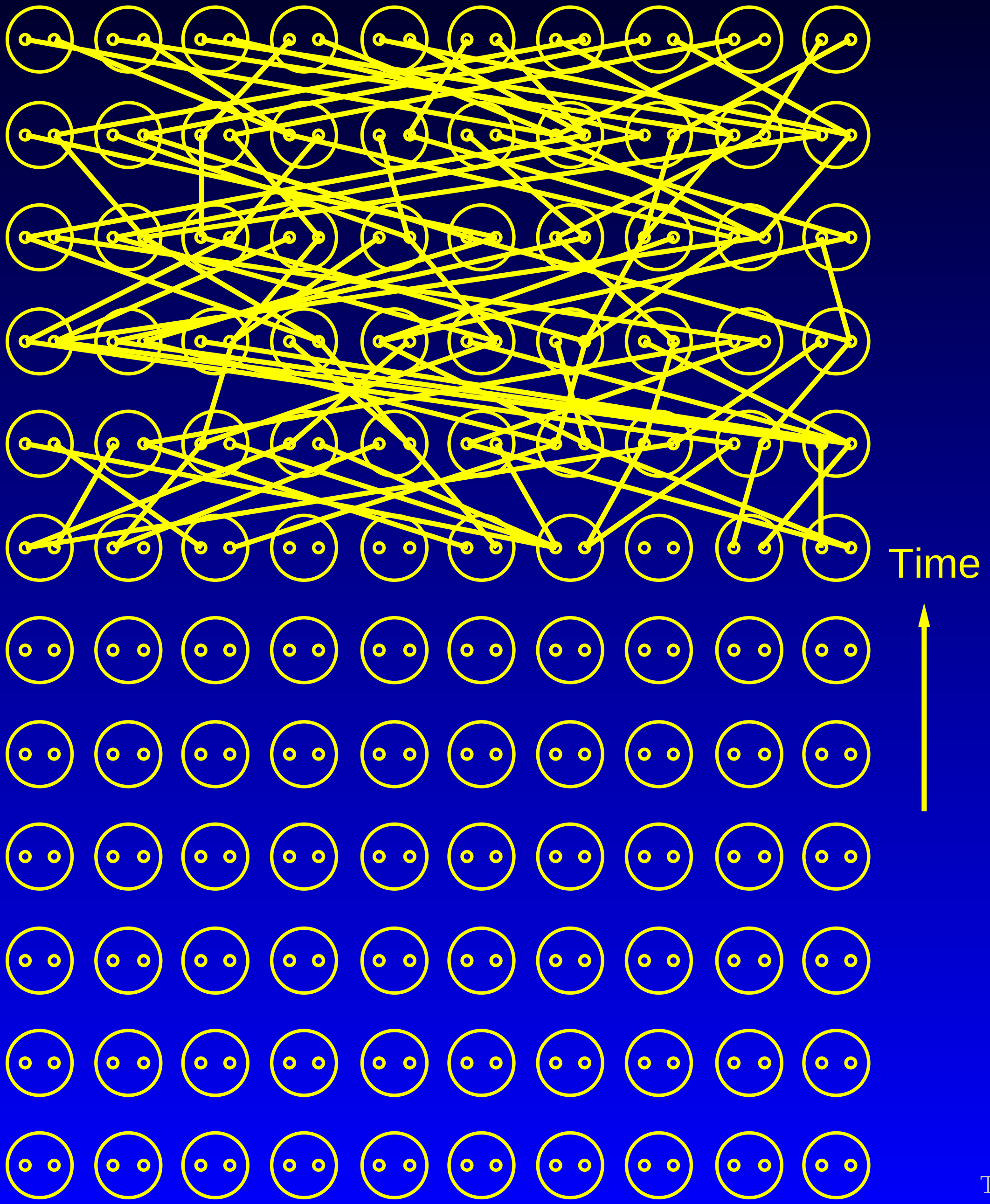
and even further



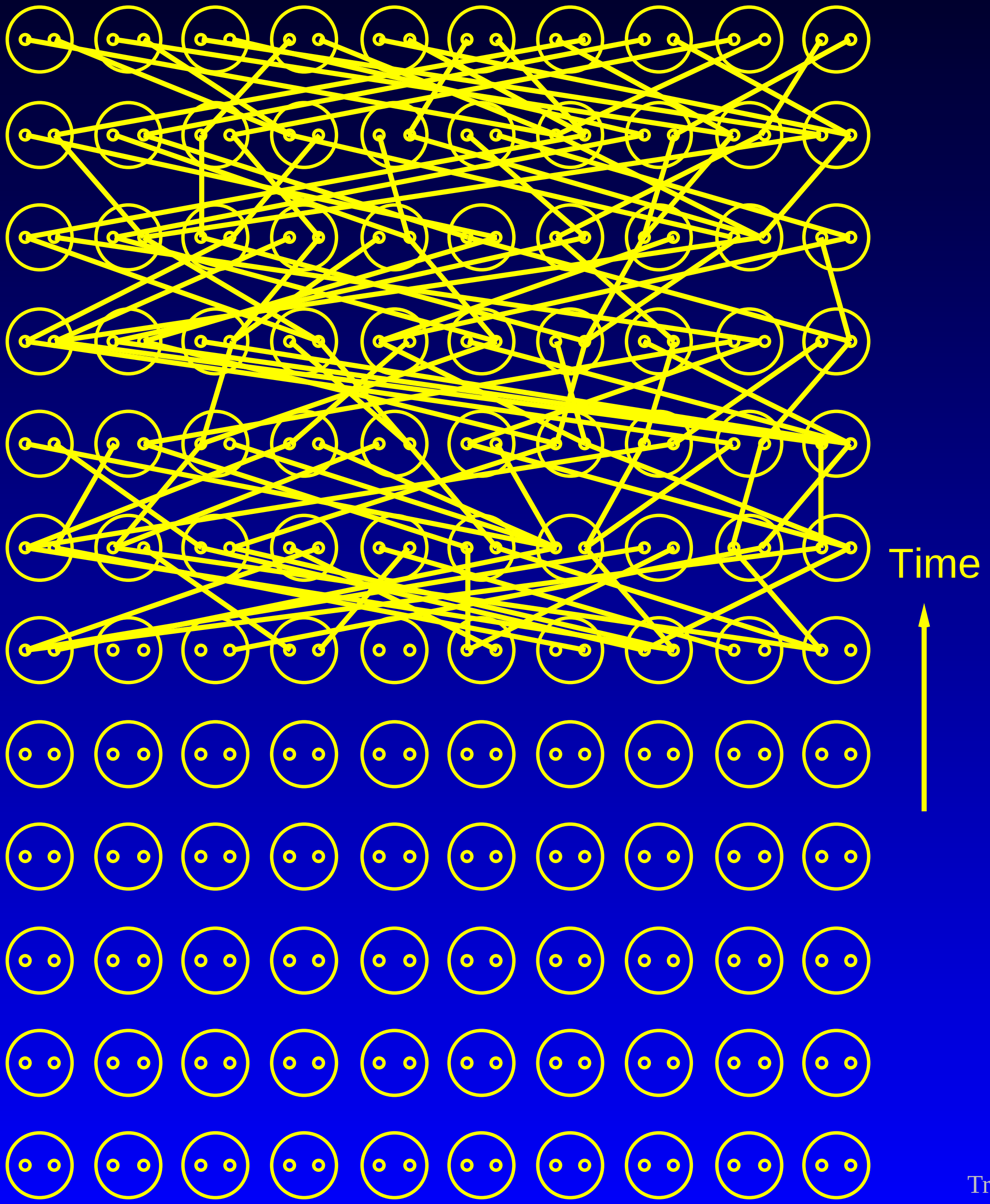
and further



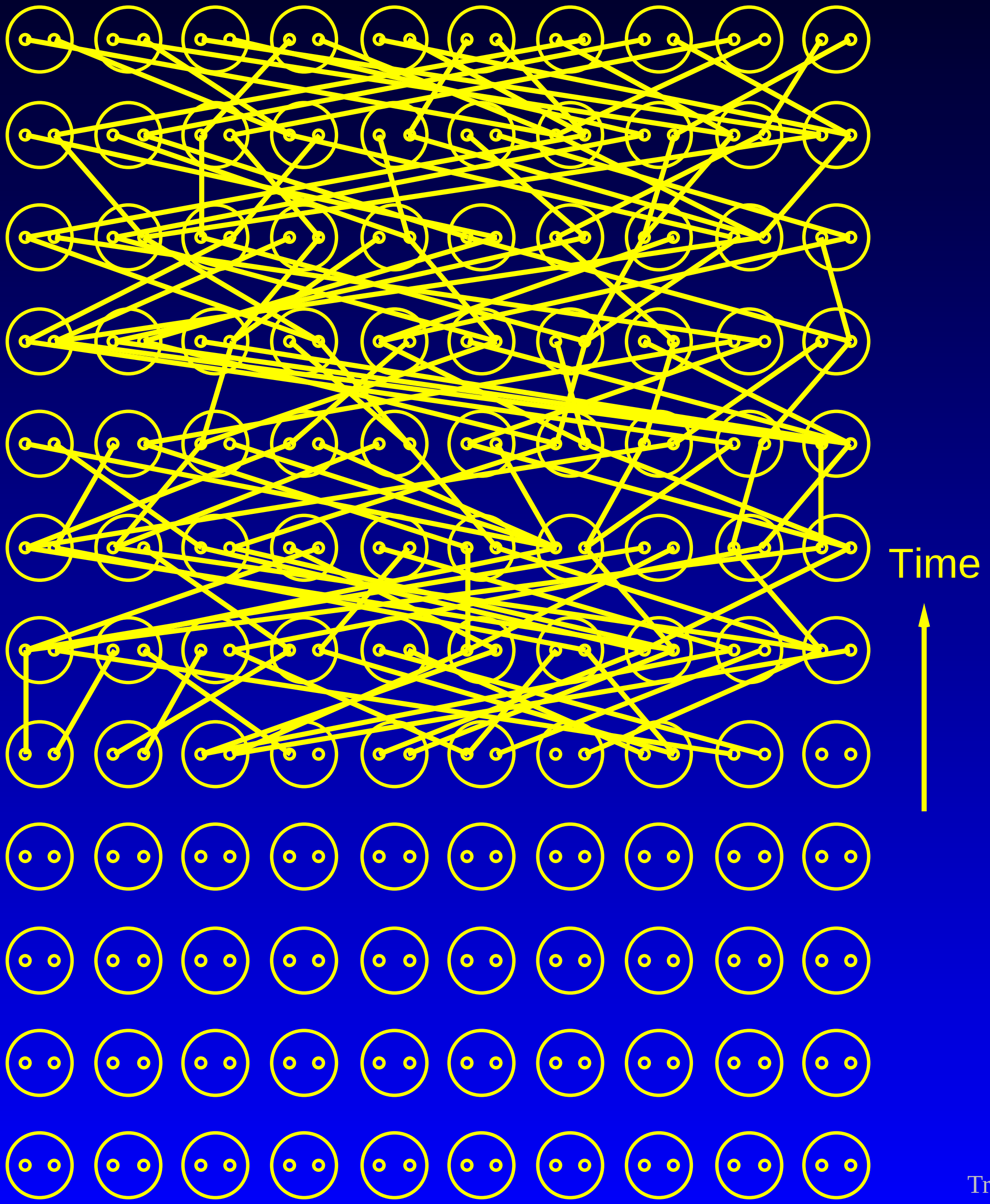
and so on



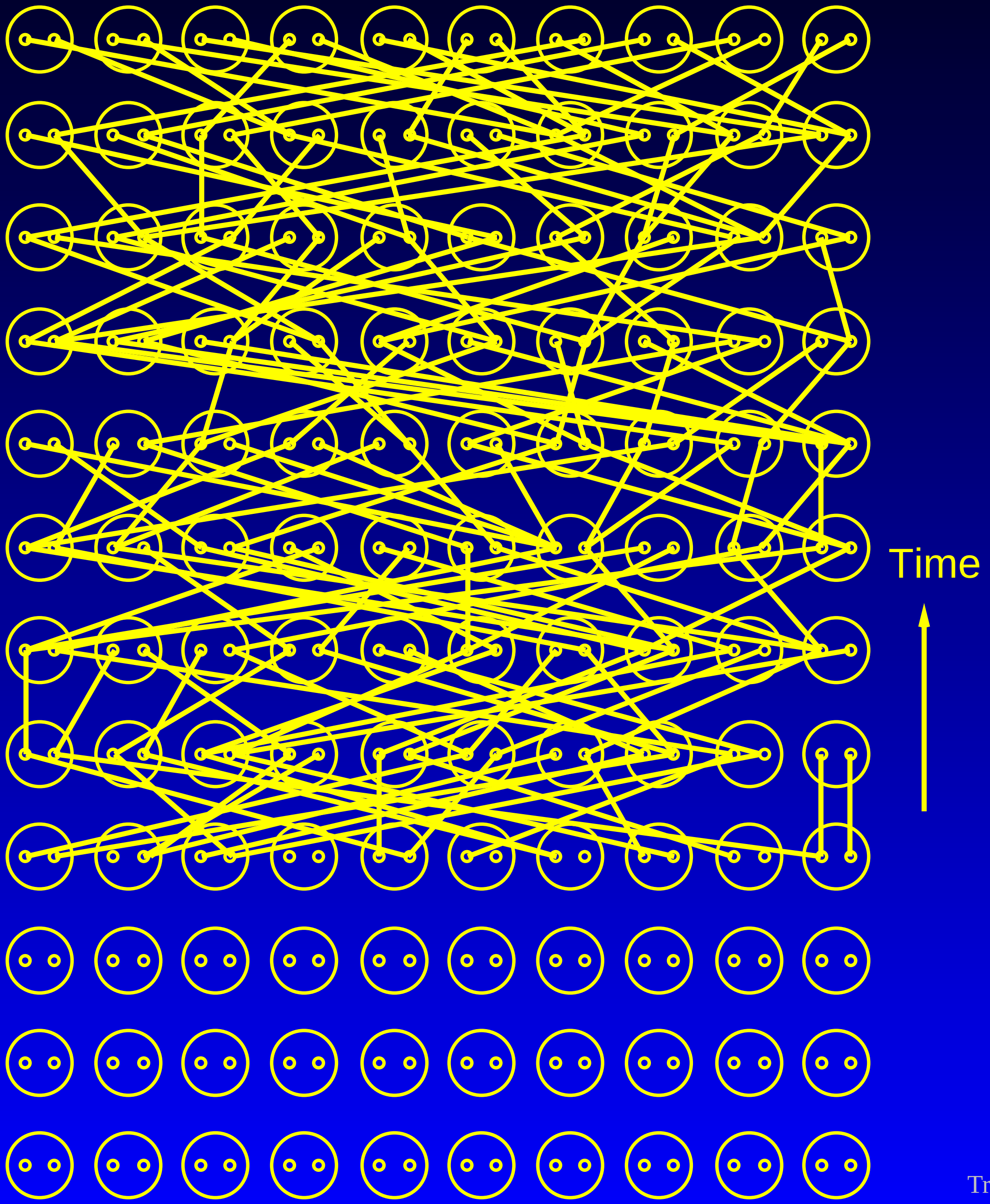
and on



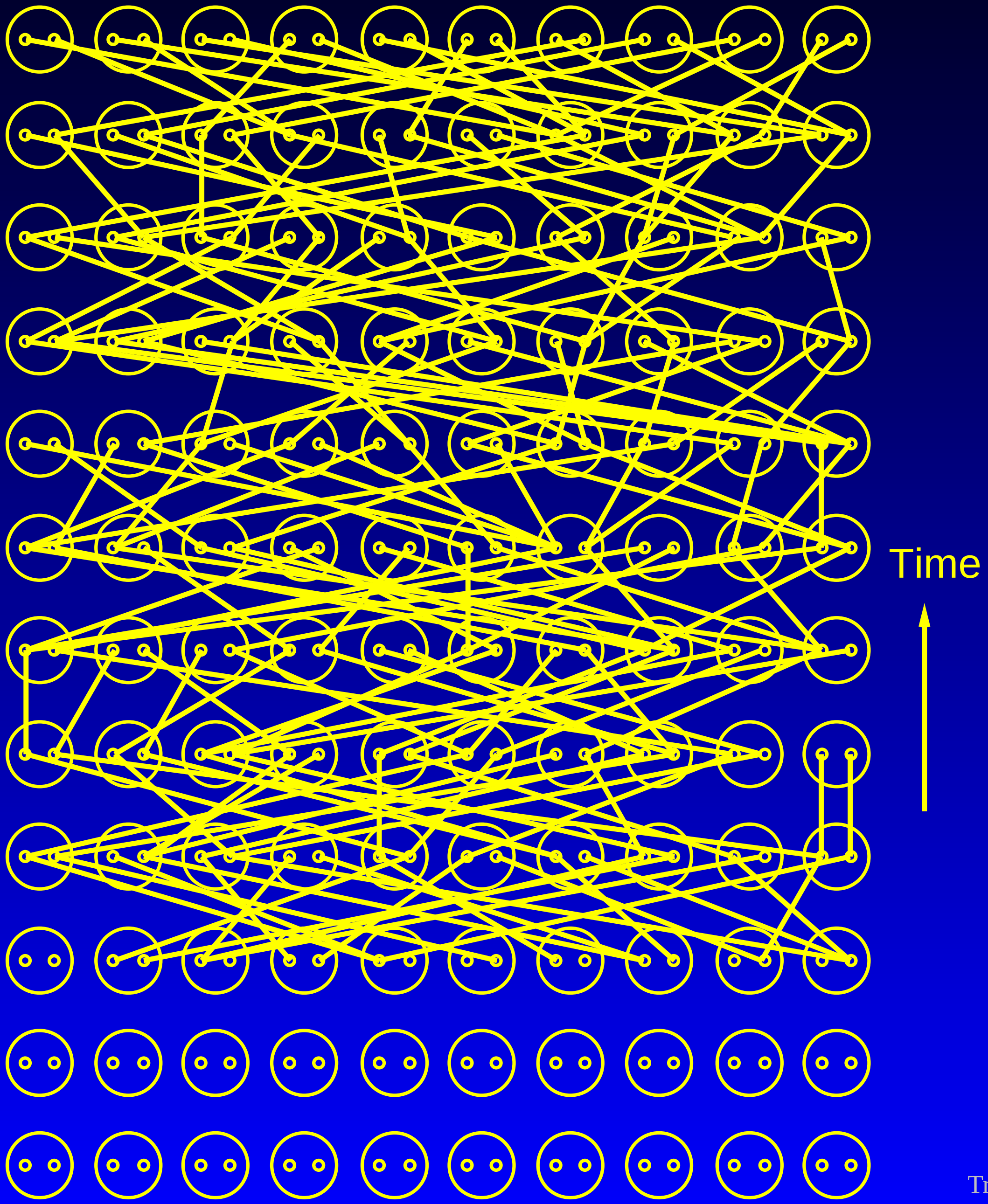
(yawn)



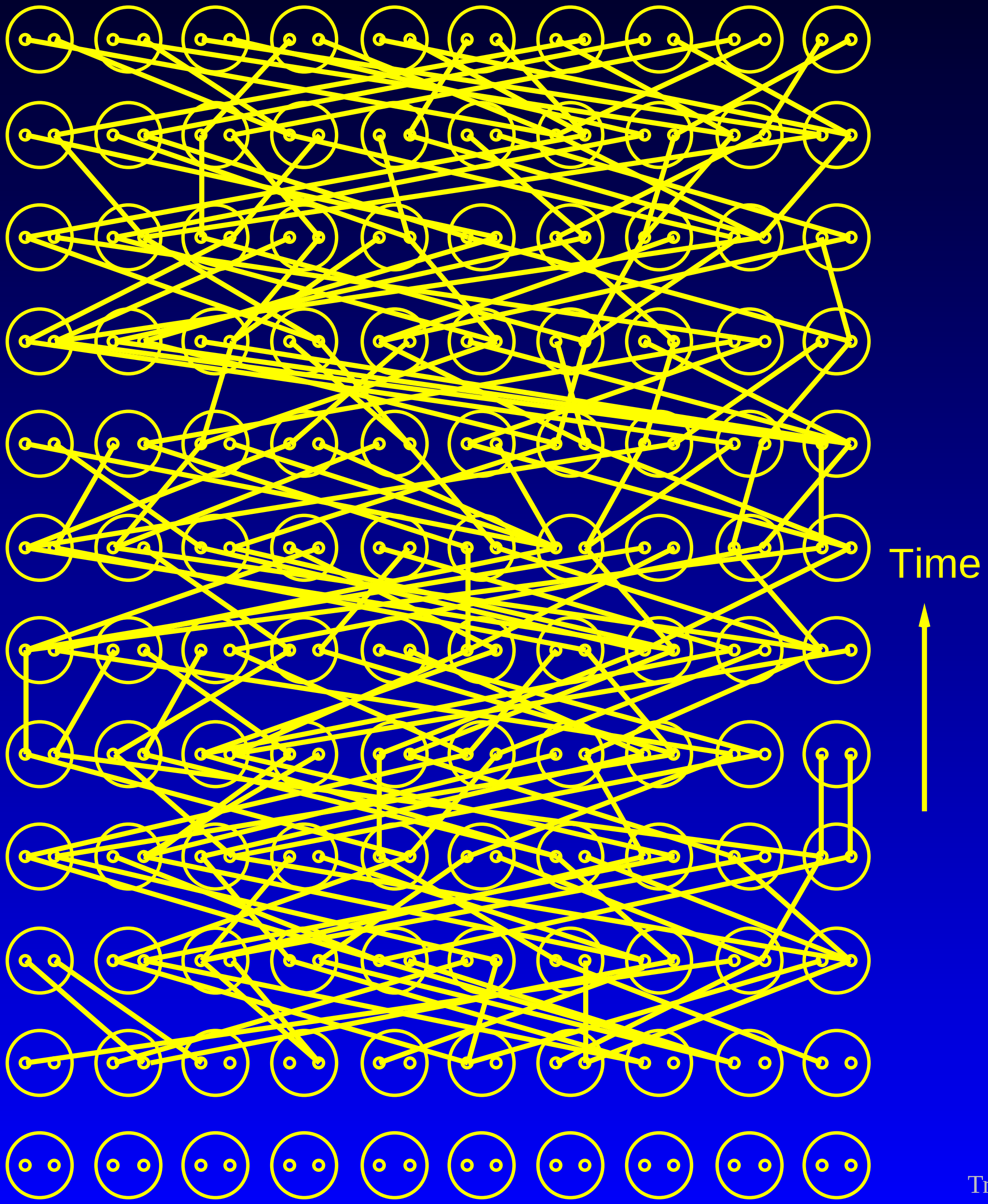
nearly there



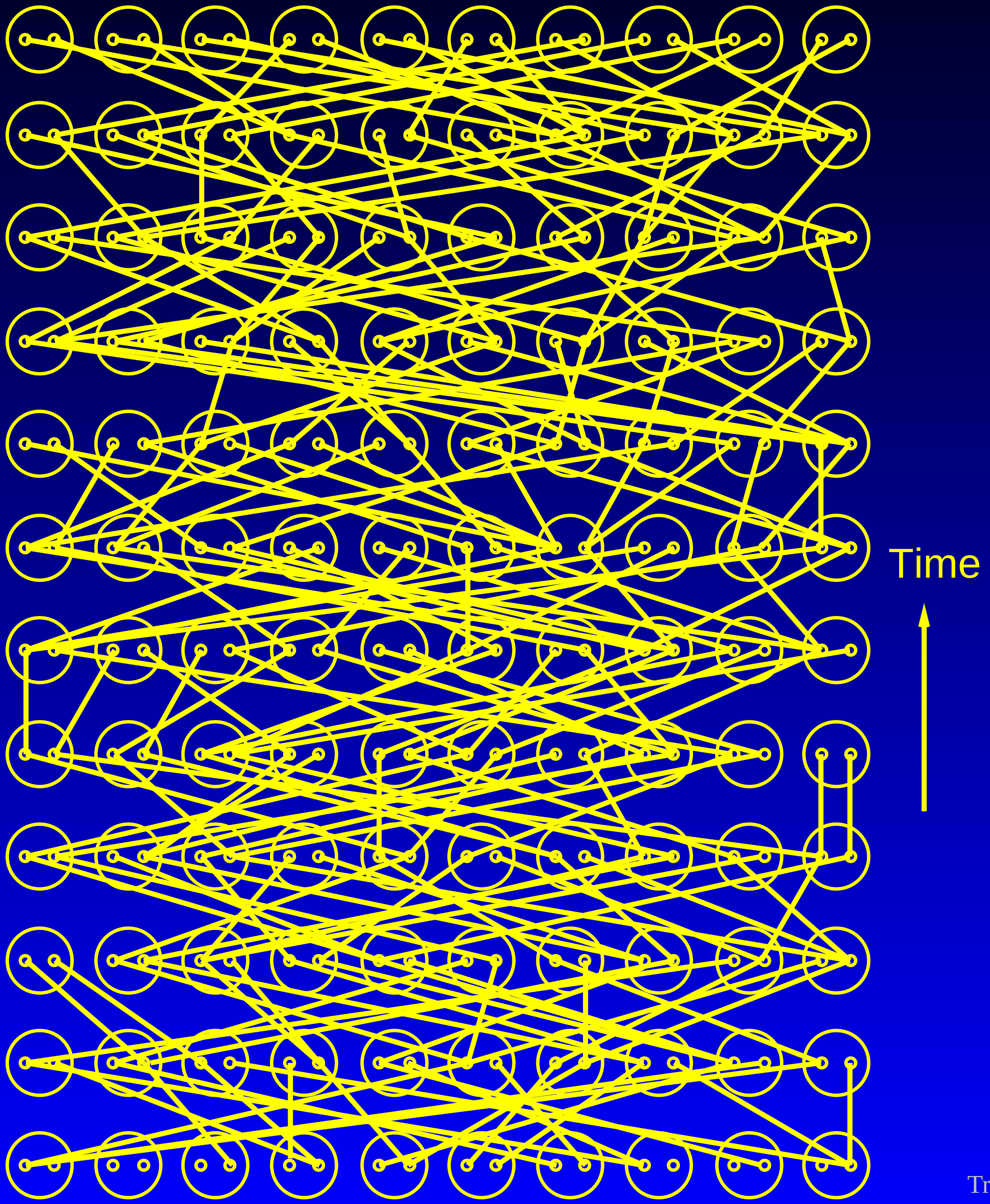
almost!



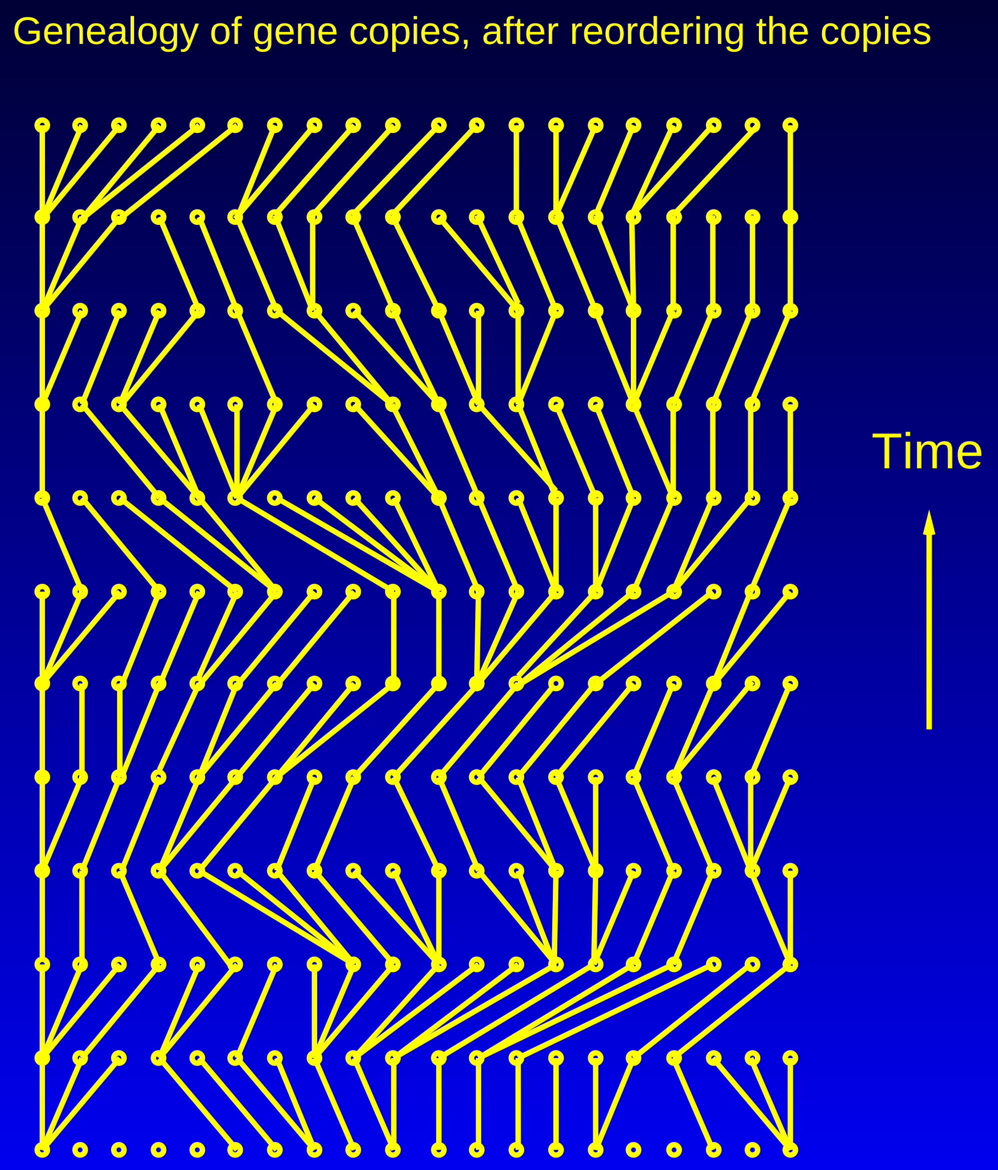
one more after this



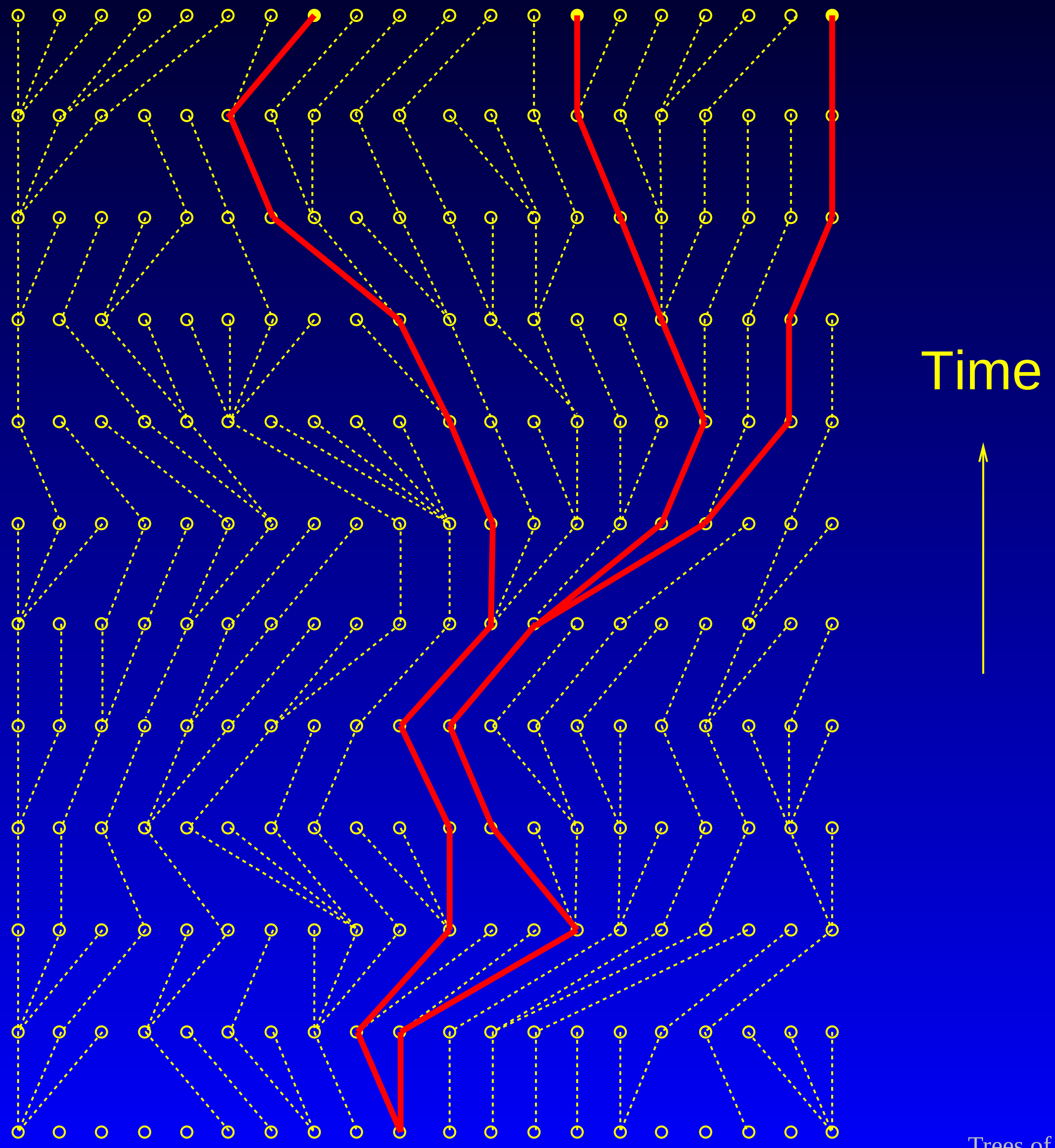
OK, so this is the pedigree of genes



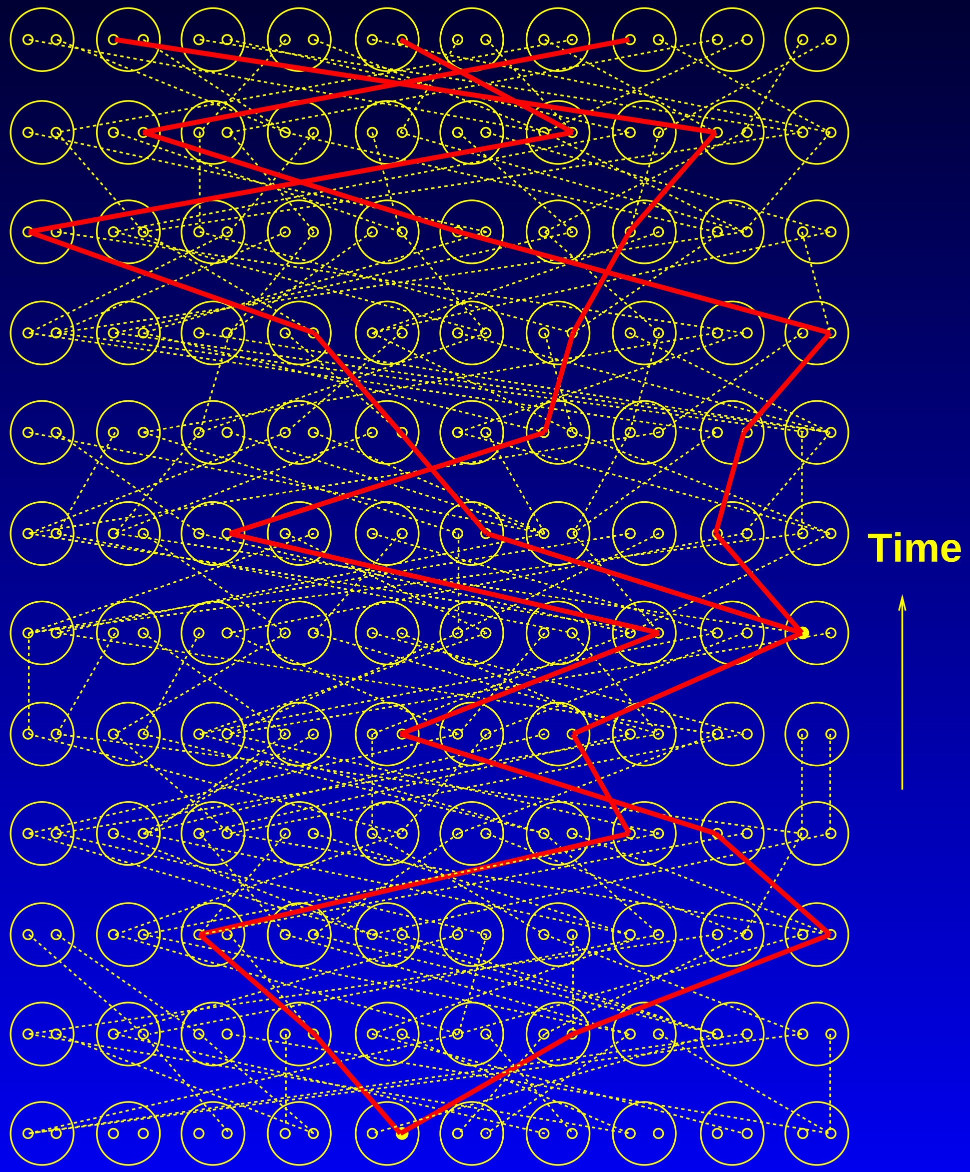
The ancestry of gene copies, untangled


Genealogy of gene copies, after reordering the copies

The ancestry of a sample of 3 genes

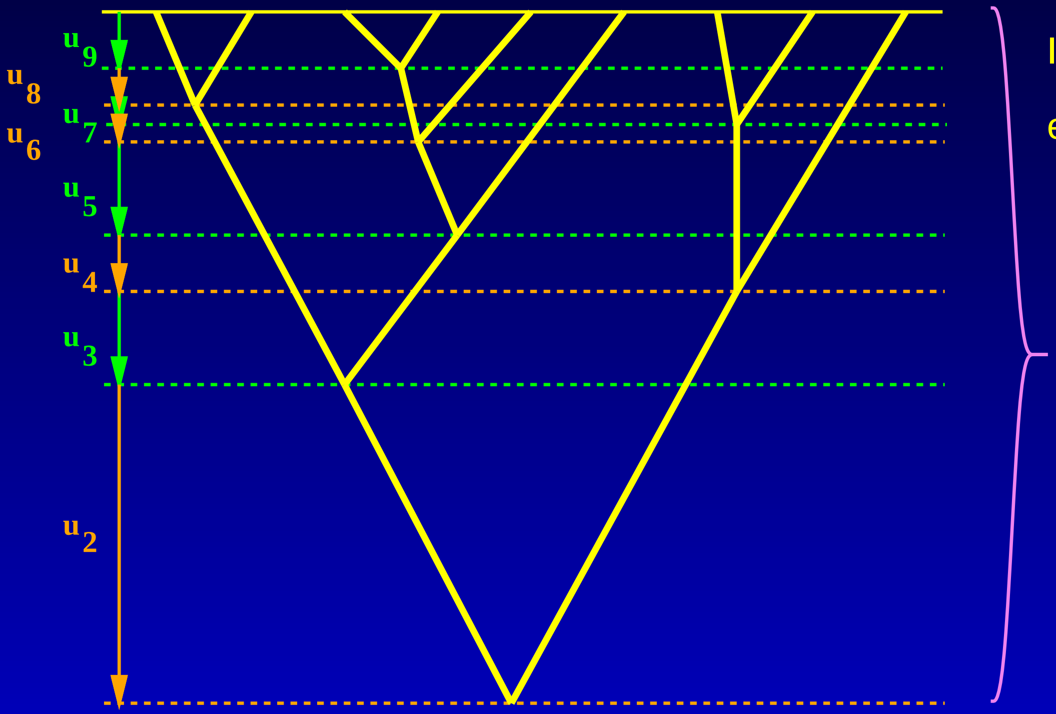


Where the tree of 3 copies is in the genealogy



Kingman’s coalescent process

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

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

Average time for

two copies to coalesce

= 2N generations

In a diploid population of

effective population size N,

Average time for n

copies to coalesce

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

generations

John Kingman

British probabilist and discoverer in 1982 of the coalescent process, an excellent approximation to gene-copy genealogies

The coalescent – a derivation

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

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

(since there are $\binom{\mathsf{k}}{2} = \mathsf{k(k-1)/2}$ different pairs of copies)
We add up terms, all the same, for the $k(k-1)/2$ pairs that could coalesce:

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

so that the total probability that a pair coalesces is

$$= \ \mathsf{k(k-1)/4N + O(1/N^2)}$$

probability that someone coalesces

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

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

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

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

and since

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

(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.

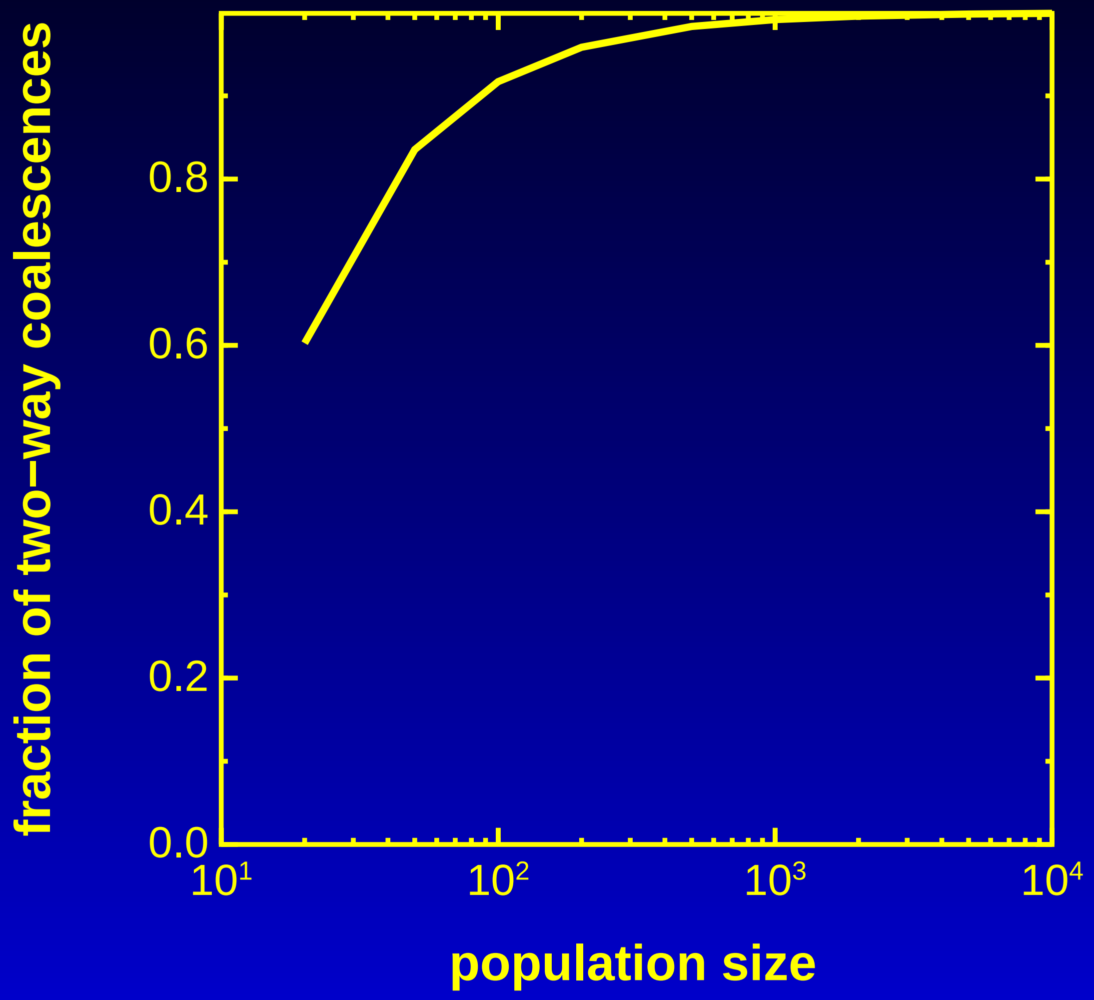
More rigorously ...

We get the coalescent as the limit when we start with k copies and pass through a series of Markov processes, each with population size N, with

- $N \longrightarrow \infty$
- $\mu \longrightarrow 0 \quad \ni \; N\mu$ constant
- (and similarly for other forces such as migration, growth, and recombination rates)
- with time in each process measured so that 1 unit is N generations

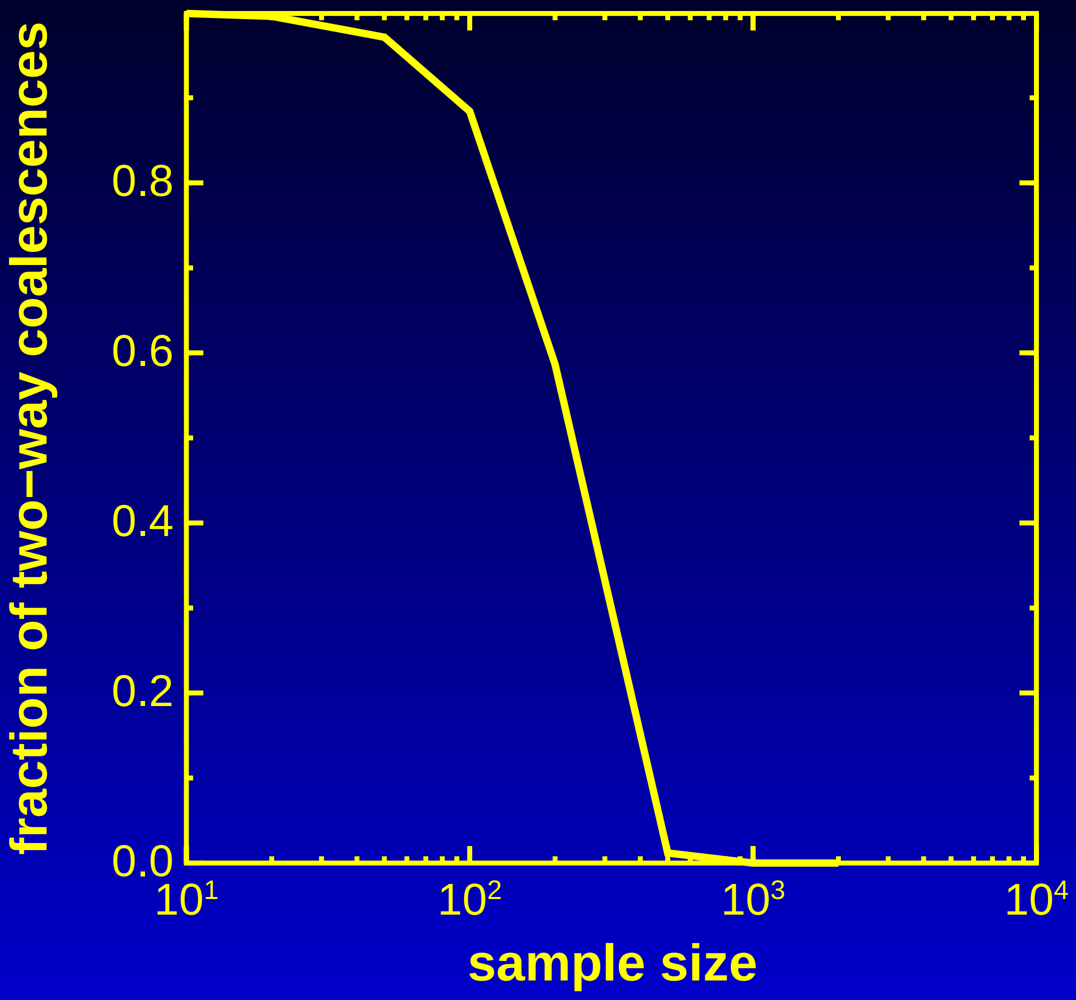
In fact, these are exactly the same conditions for obtaining the standard population genetics diffusion approximations by weak convergence of the Markov processes to the diffusion process.

Accuracy with different population sizes



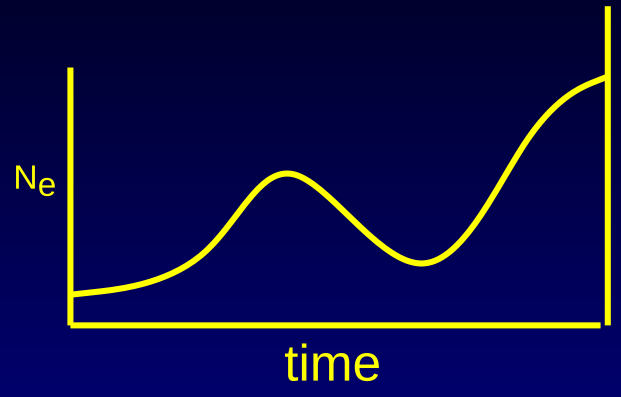
(when sample size = 10)

Accuracy with different sample sizes



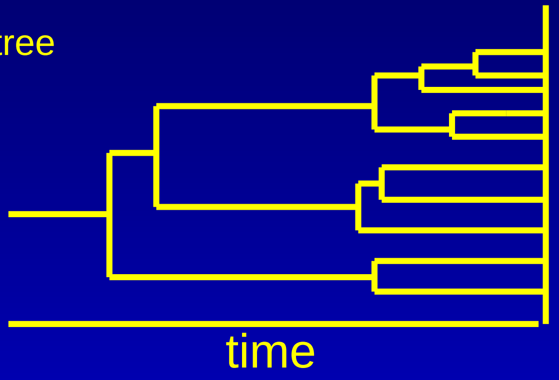
(when population size = 10,000)

Effect of varying population size

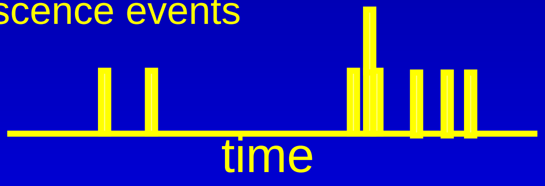


the changes in population size will produce waves of coalescence

the tree

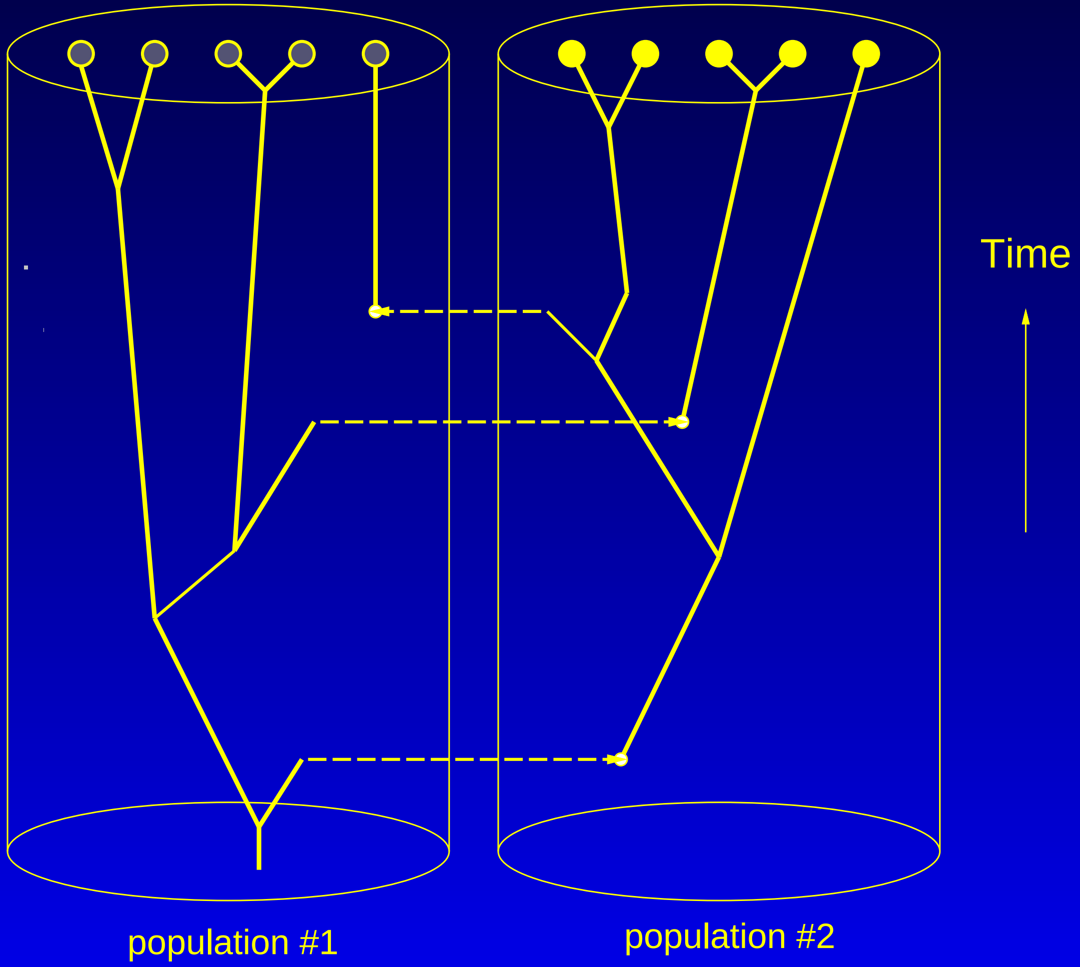


Coalescence events

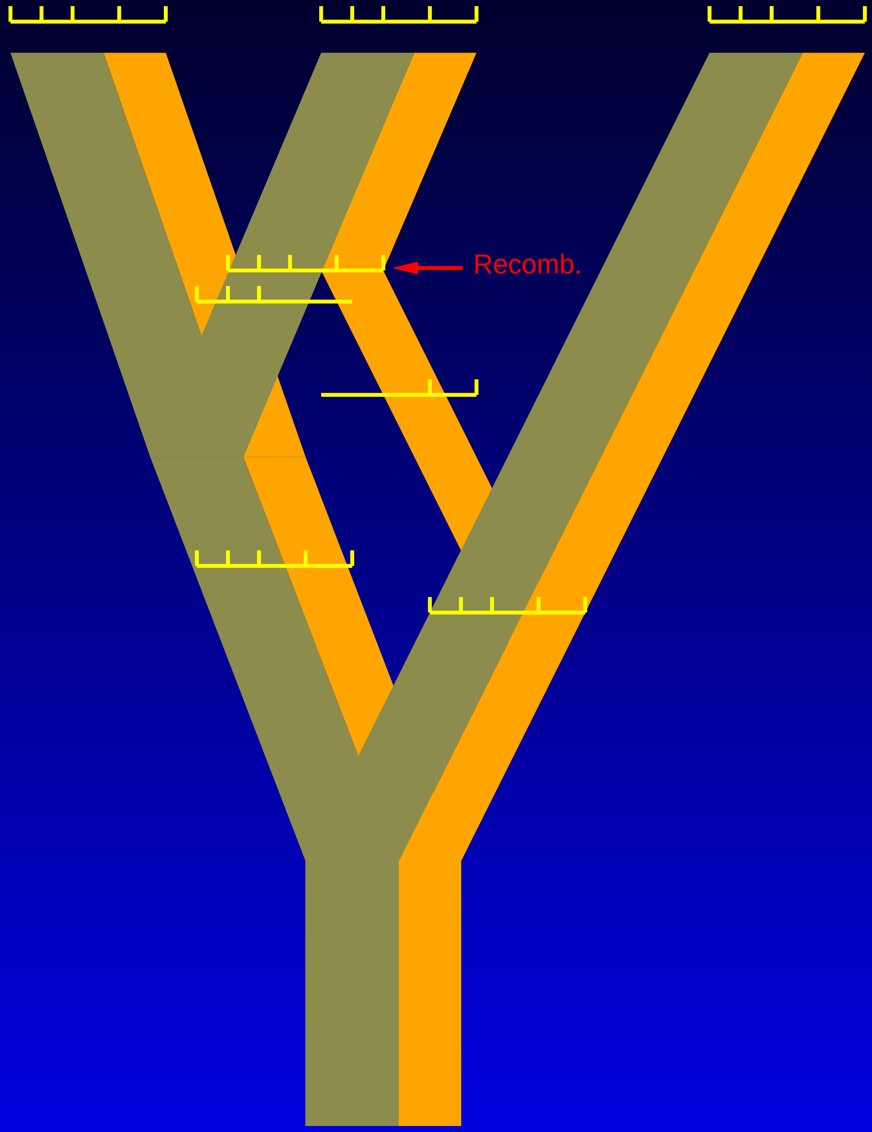


The parameters of the growth curve for N_e can be inferred by likelihood methods as they affect the prior probabilities of those trees that fit the data.

A coalescent with migration (2 populations)

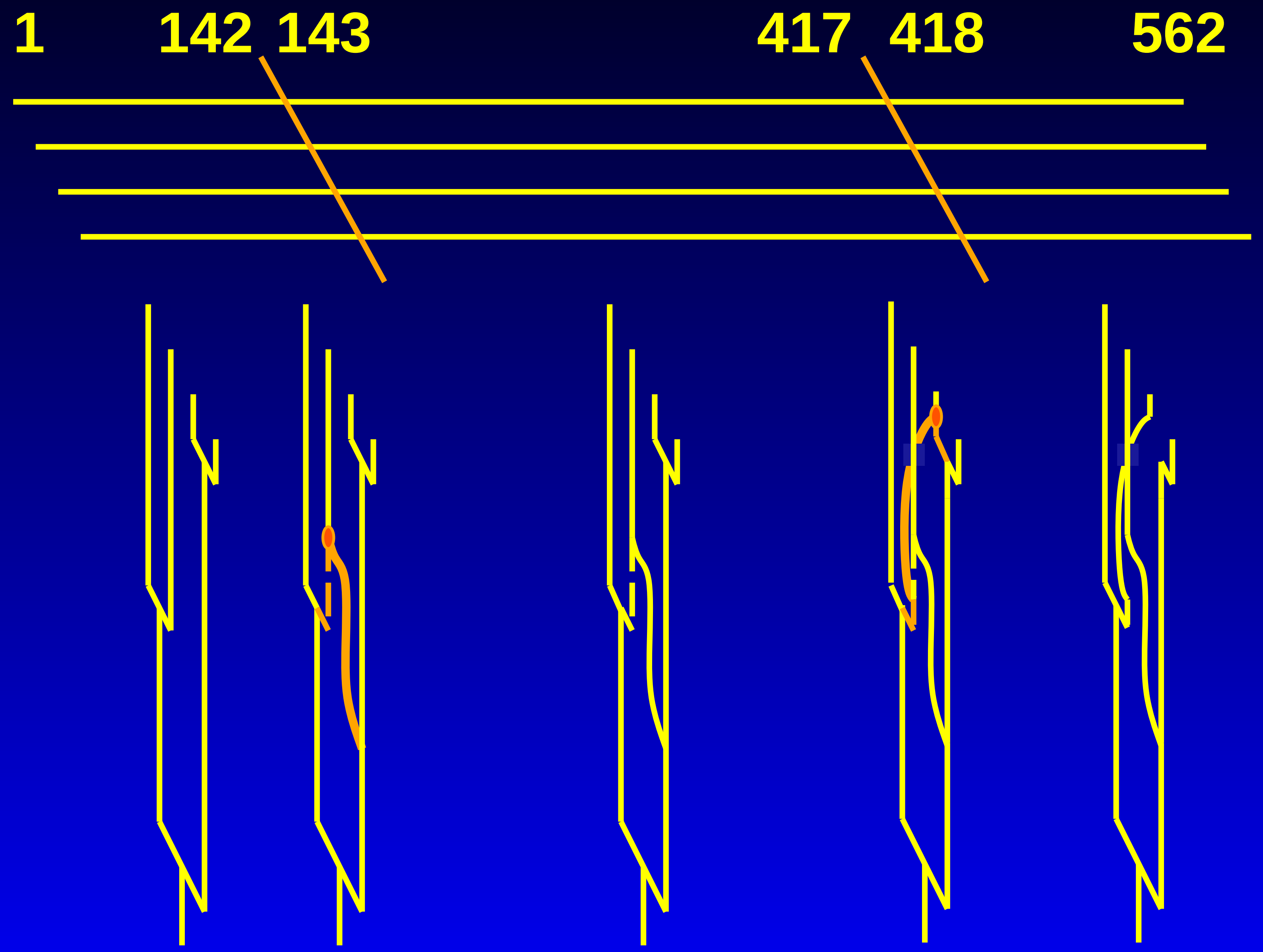


A recombining coalescent

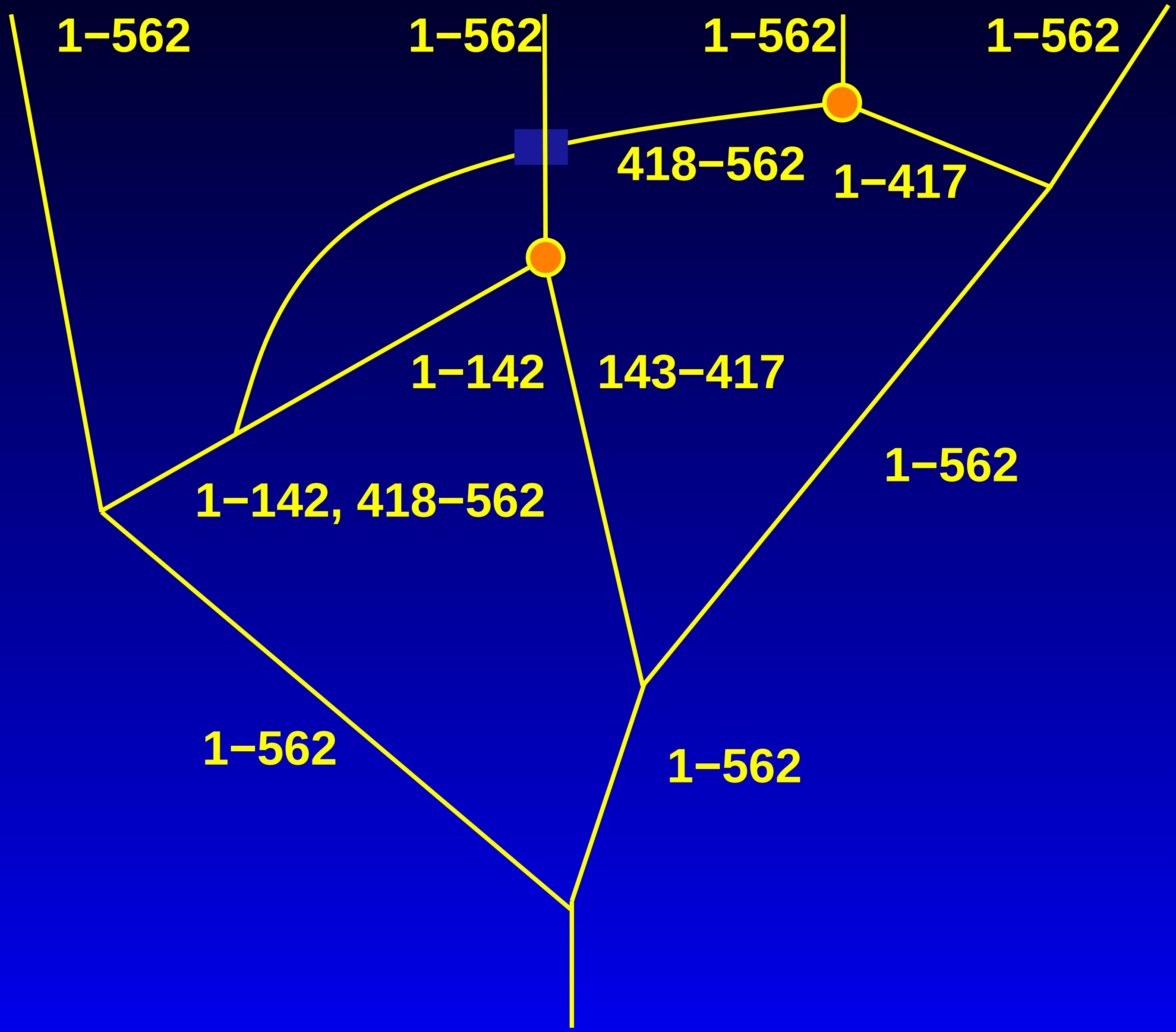


Different markers have slightly different coalescent trees

Coalescents along a genome



Their ancestral recombination graph



How far until a very different tree

- The time back to the root from a tip is about $4N$ generations

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$
- So it is far enough that $r = 1/(4N)$

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$
- So it is far enough that $r = 1/(4N)$
- In a human genome with $N \approx 10^4$ this is $r = 0.000025$

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$
- So it is far enough that $r = 1/(4N)$
- In a human genome with $N \approx 10^4$ this is $r = 0.000025$
- That is about 250 bases (if recombination is evenly spread with no hot spots)

How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$
- So it is far enough that $r = 1/(4N)$
- In a human genome with $N \approx 10^4$ this is $r = 0.000025$
- That is about 250 bases (if recombination is evenly spread with no hot spots)
- If there are strong hot spots with very cold regions between, the distance is the distance between hot spots

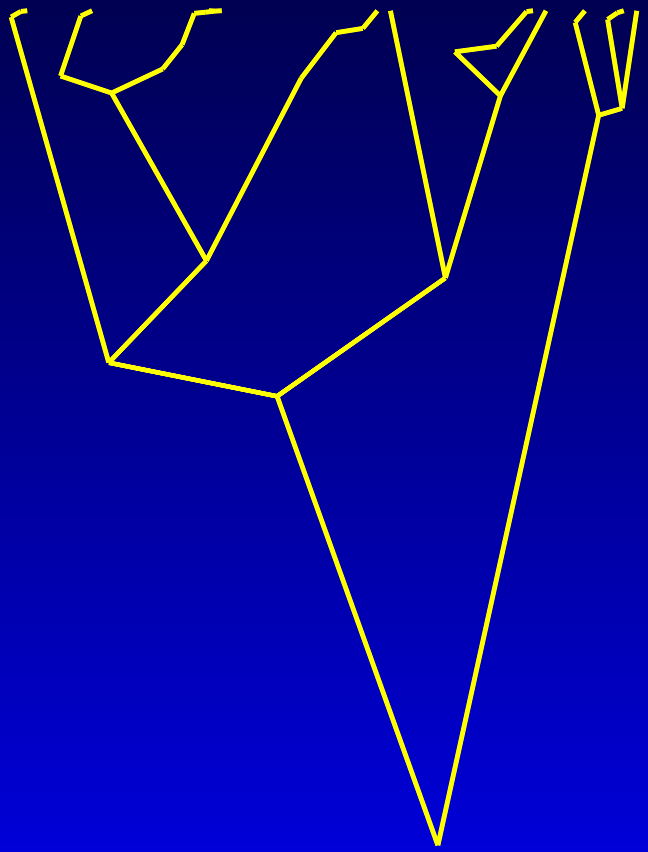
How far until a very different tree

- The time back to the root from a tip is about $4N$ generations
- Get a very different tree if have about 1 recombination in this time
- If recombination rate per base is r this is $4Nr = 1$
- So it is far enough that $r = 1/(4N)$
- In a human genome with $N \approx 10^4$ this is $r = 0.000025$
- That is about 250 bases (if recombination is evenly spread with no hot spots)
- If there are strong hot spots with very cold regions between, the distance is the distance between hot spots
- In any case no more than about 10 kb

A coalescent of 50 copies

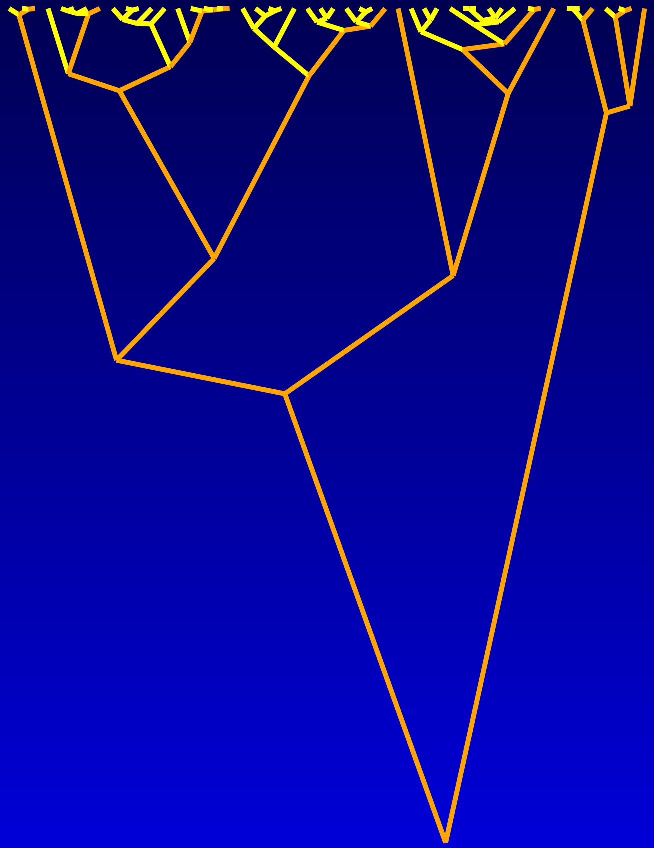
the first 10 copies only

10 gene copies 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)

Some typical data with within-population variation

To infer parameters of evolutionary−genetic models

... we need to compute the likelihood
for a set of genotypes sampled from a population

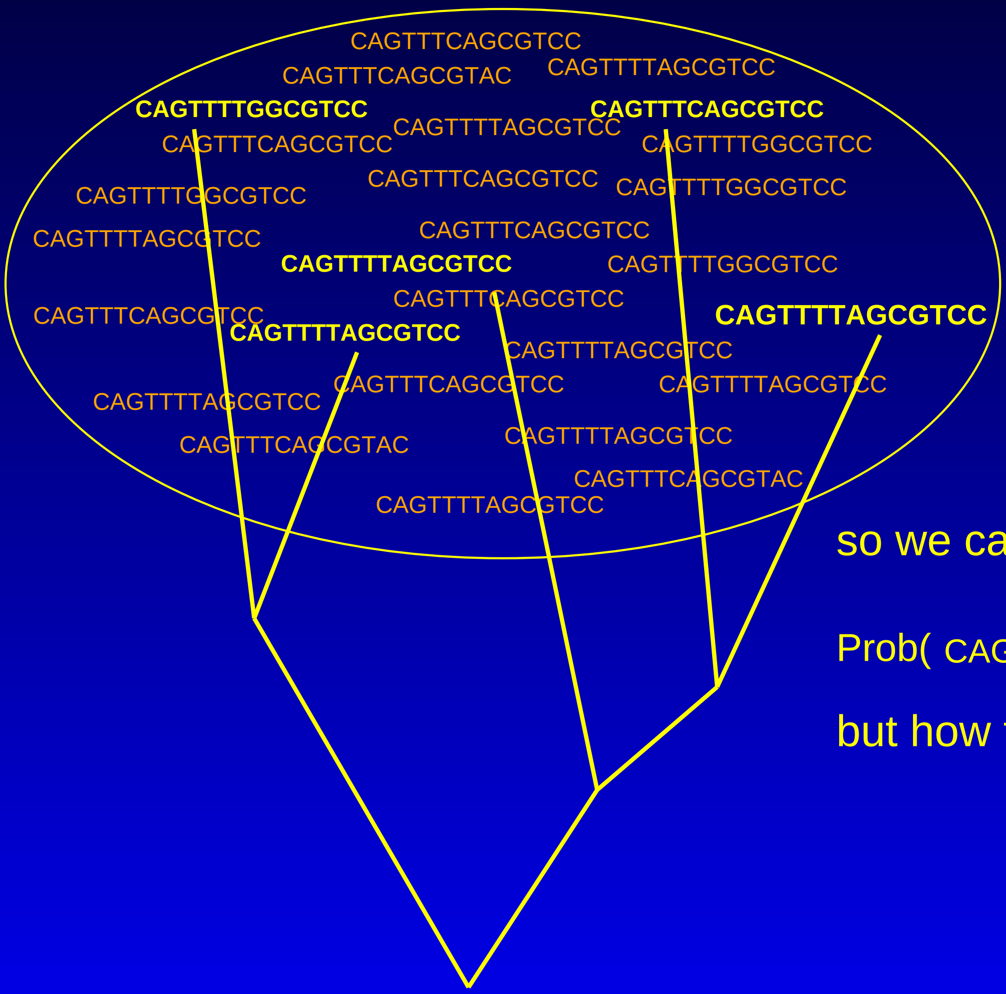
With few exceptions, no expressions for this likelihood exist.

CAGTTTCAGCGTCC
CAGTTTTAGCGTCC
CAGTTTCAGCGTAC
CAGTTTTGGCGTCC
CAGTTTCAGCGTCC
CAGTTTTAGCGTCC
CAGTTTCAGCGTCC
CAGTTTTGGCGTCC
CAGTTTCAGCGTCC
CAGTTTTGGCGTCC
CAGTTTTGGCGTCC
CAGTTTCAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTGGCGTCC
CAGTTTCAGCGTCC
CAGTTTCAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTAGCGTCC
CAGTTTCAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTAGCGTCC
CAGTTTTAGCGTCC
CAGTTTCAGCGTAC
CAGTTTCAGCGTAC
CAGTTTTAGCGTCC

L = Prob (CAGTTTCAGCGTCC , CAGTTTCAGCGTCC , ...) = ??

But there is a way to compute it

If we knew the genealogical tree connecting the haplotypes
we know from work on phylogenies (evolutionary trees) how to
compute the probability of the sample at the tips of that tree



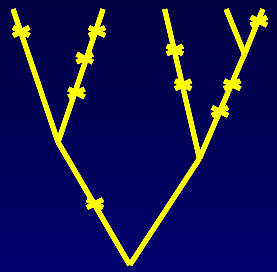
so we can compute

Prob(CAGTTTCAGCGTCC , CAGTTTCAGCGTCC , ... | Genealogy)

but how to compute the overall likelihood from this?

Two sources of variation

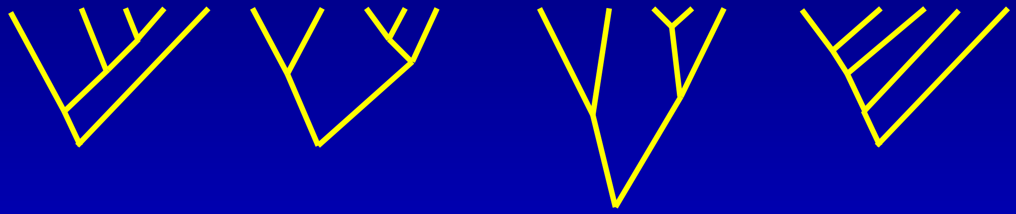
(1) Randomness of mutation



affected by the mutation rate u

can reduce variance of
number of mutations per site per
branch by examining more sites

(2) Randomness of coalescence of lineages



affected by effective population size N_e

coalescence times allow estimation of N_e

can reduce variability by looking at

(i) more gene copies, or

(ii) more loci

The basic equation for coalescent likelihoods

In the case of a single population with parameters

$\mathsf{N_e}$	effective population size
μ	mutation rate per site

and assuming G' stands for a coalescent genealogy and D for the

sequences,

$$\begin{aligned} \mathsf{L} &= \text{Prob}\left(\mathsf{D} \,|\, \mathsf{N_e},\ \mu\right) \\ &= \sum_{\mathsf{G}'} \underbrace{\text{Prob}\left(\mathsf{G}' \,|\, \mathsf{N_e}\right)}_{\text{Kingman's prior}} \ \underbrace{\text{Prob}\left(\mathsf{D} \,|\, \mathsf{G}', \mu\right)}_{\text{likelihood of tree}} \end{aligned}$$

Changing the time scale of branch lengths ...

Rescaling branch lengths of G' so that branches are given in expected

mutations per site, $G = \mu G'$, we get (if we let $\Theta = 4N_e\mu$)

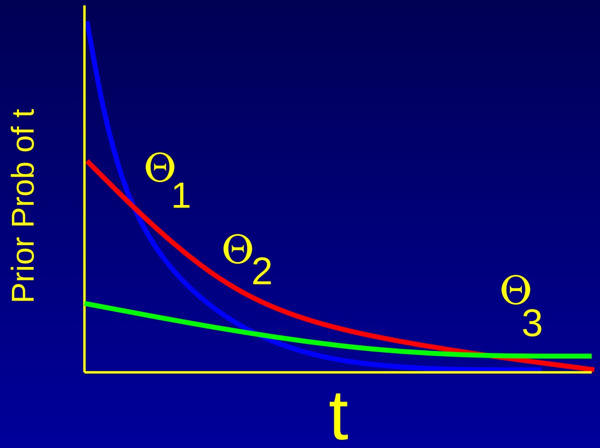
$$L = \sum_G \mathrm{Prob}\,(G \mid \Theta)\ \mathrm{Prob}\,(D \mid G)$$

as the fundamental equation. For more complex population scenarios one simply replaces Θ with a vector of parameters.

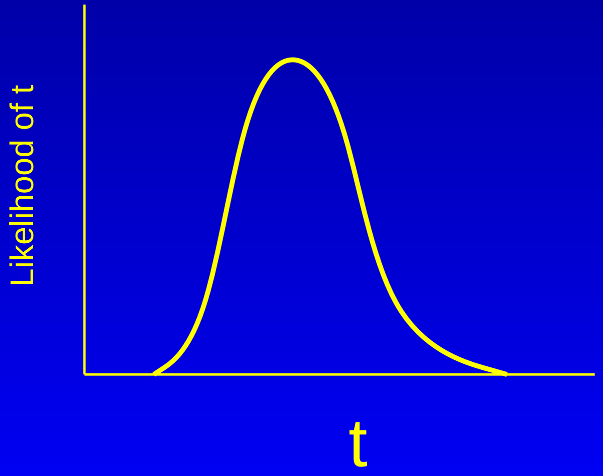
A simple example of the likelihood curve

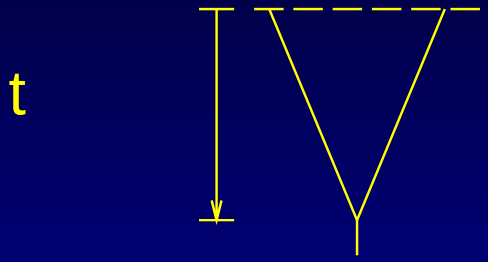
The likelihood calculation in a sample of two gene copies

The product of the prior on t,

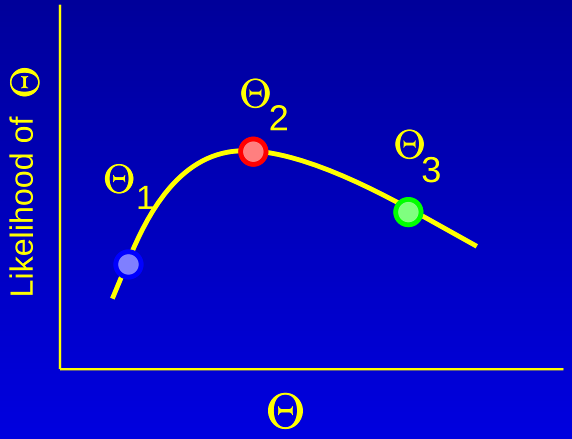


times the likelihood of that t from the data,

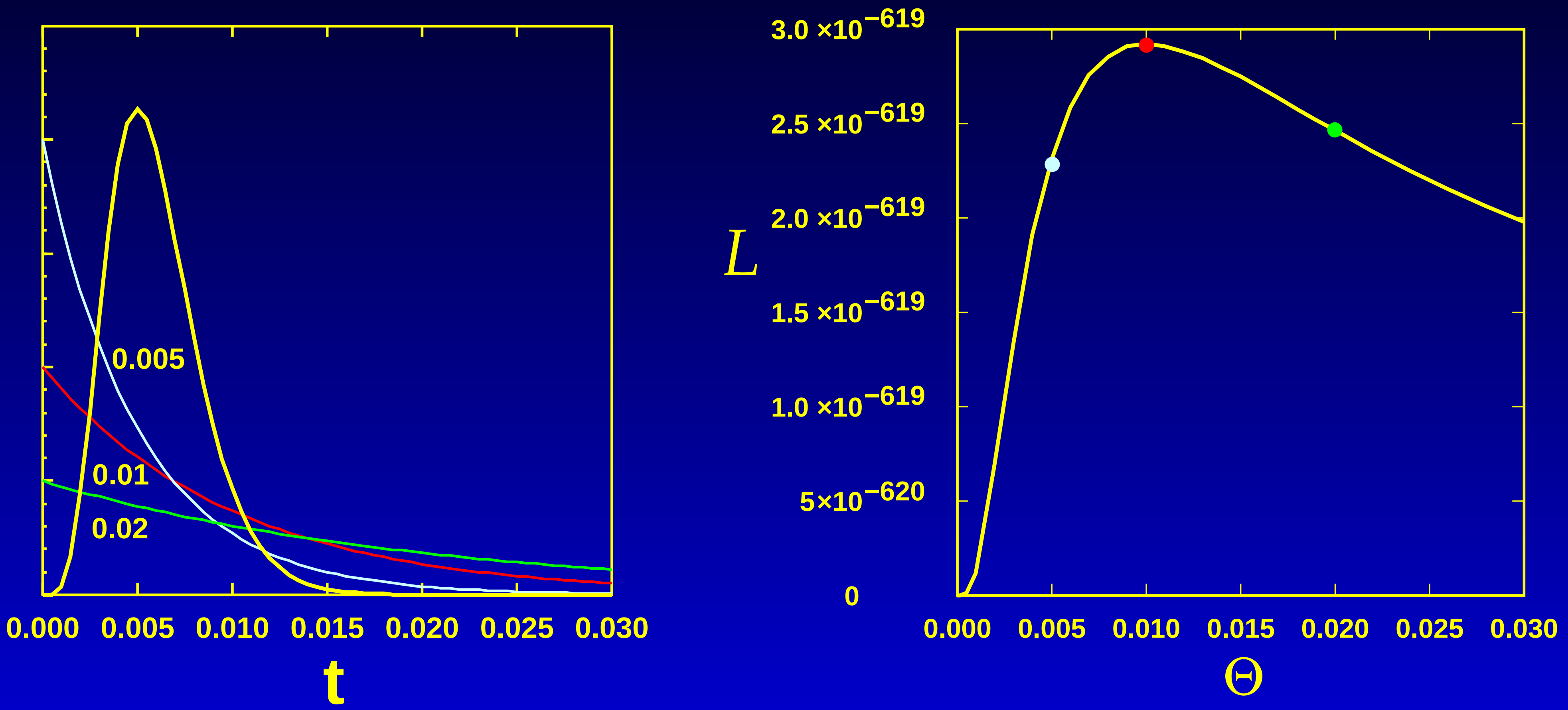




when integrated over all possible t's, gives the likelihood for the underlying parameter Θ



If two sequences with 1000 bases 0.5% different



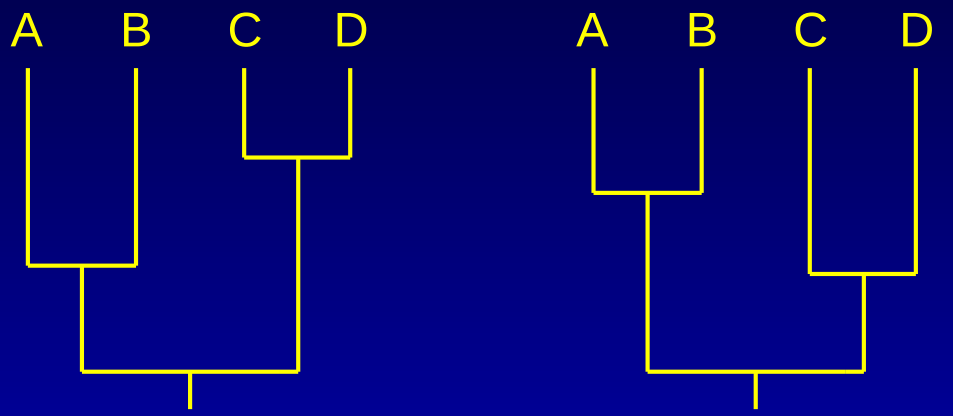
(using a Jukes-Cantor model of base change)

Labelled histories

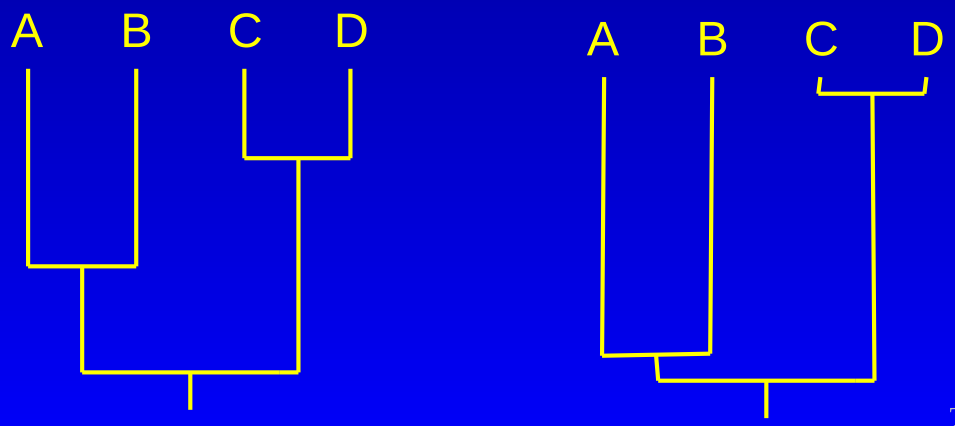
(Edwards, 1970; Harding, 1971)

Trees that differ in the time−ordering of their nodes

These two are different:



These two are the same:



The number of labelled histories

The labelled history is essentally a list of the pairs of lineages that coalesce, in order. So the number of these is

$$\frac{n(n-1)}{2}\ \frac{(n-1)(n-2)}{2}\ \frac{(n-2)(n-3)}{2}\ \cdots\ \frac{2\times 1}{2}$$

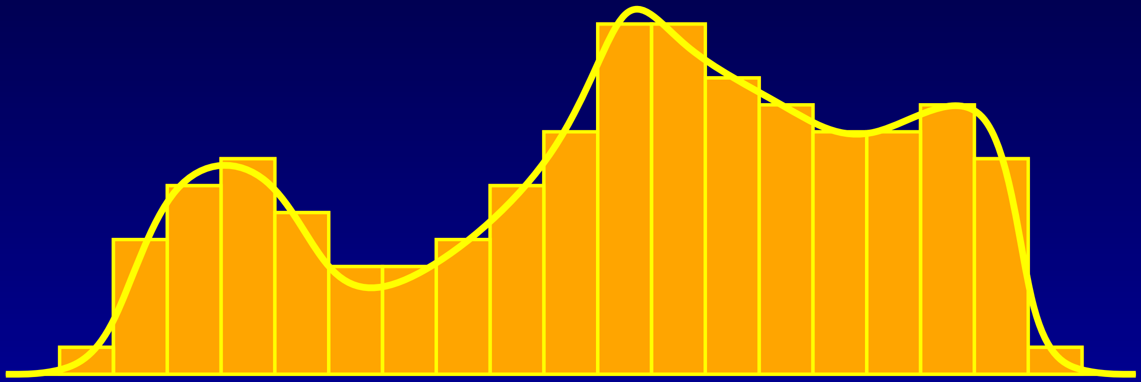
$$= \ \frac{n!(n-1)!}{2^{n-1}}$$

The number of these rises rapidly:

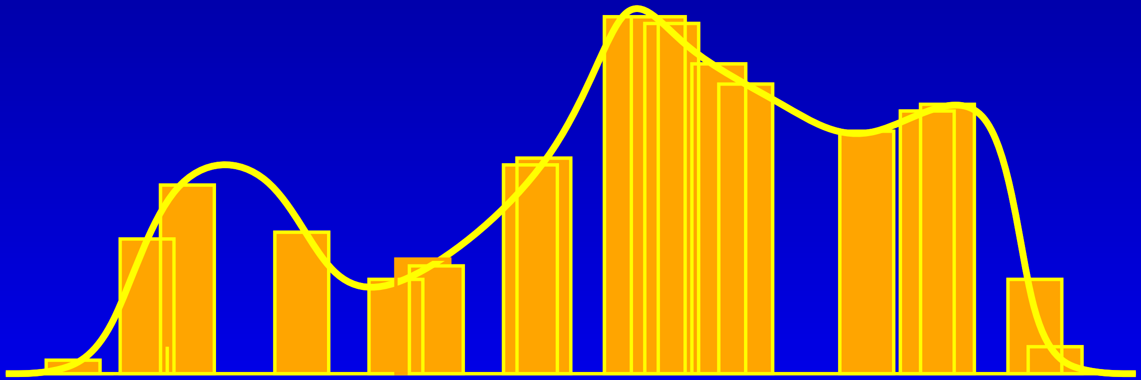
Tips	Labelled histories
2	1
3	3
4	18
5	180
6	2700
7	56,700
8	1,587,600
9	57,153,600
10	2,571,912,000

Monte Carlo integration

To get the area under a curve, we can either evaluate the function ($f(x)$) at a series of grid points and add up heights × widths:



or we can sample at random the same number of points, add up height × width:

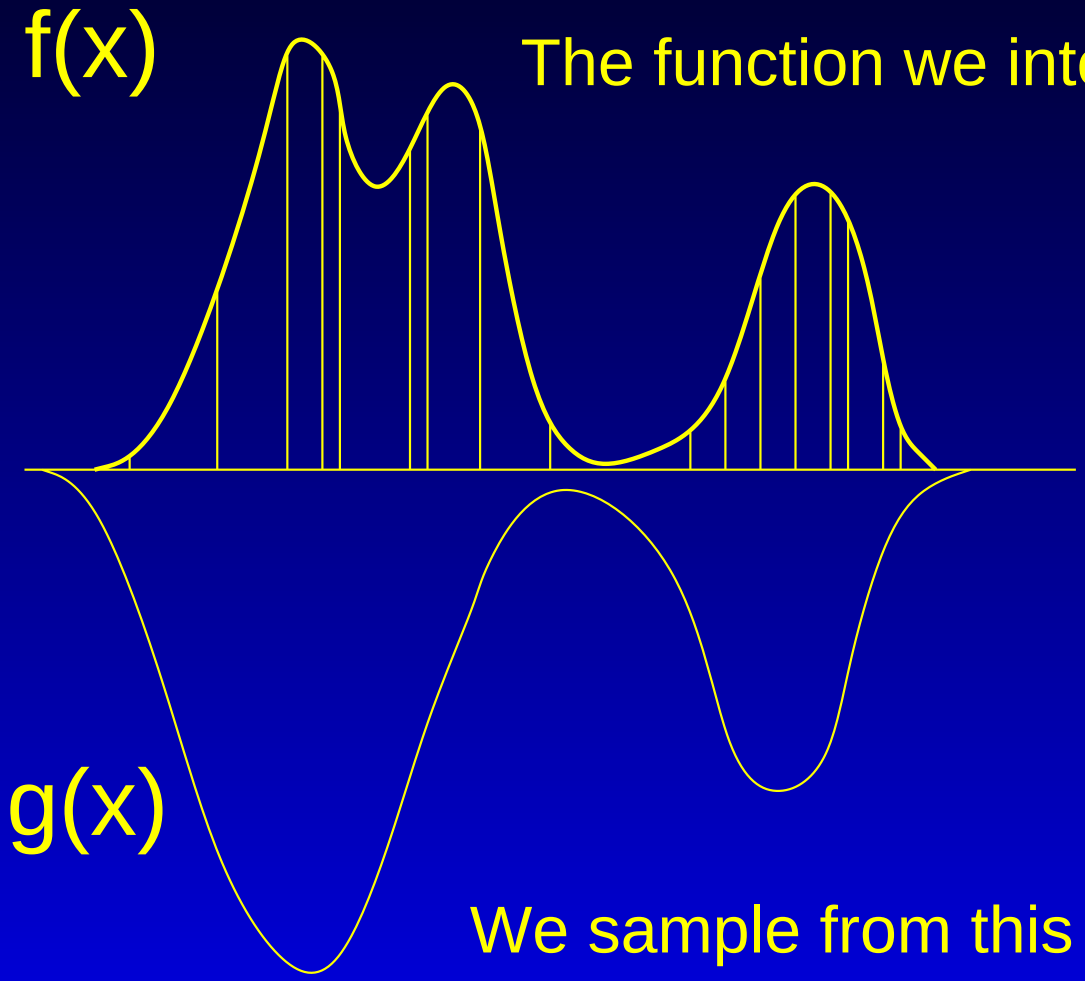


The importance sampling formula

Expectation of a function $h(x)$ over a distribution whose density function is $g(x)$:

$$E_g[h(x)] = \int_x h(x)g(x)\,dx$$

Importance Sampling



The function we integrate

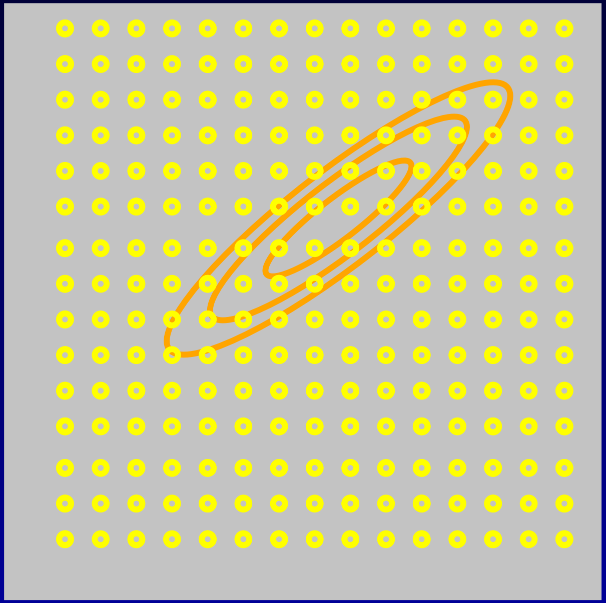
We sample from this density

The integral can be computed as follows:

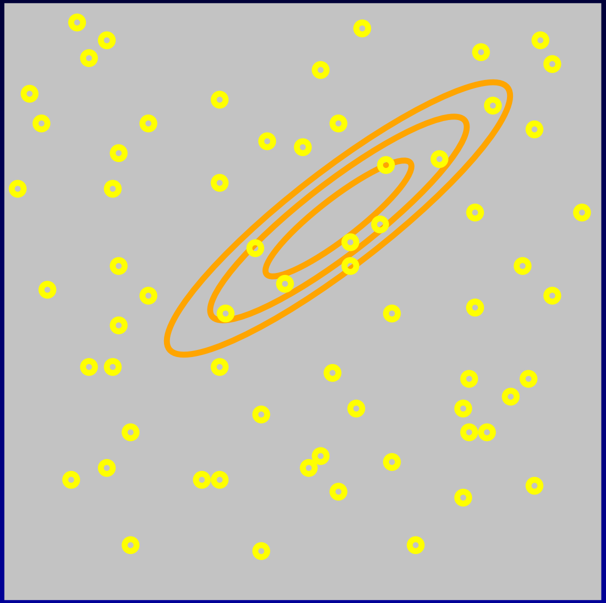
$$\begin{aligned}\int f(x)\,dx &= \int \frac{f(x)}{g(x)}\,g(x)\,dx \\ &= E_g\left[\frac{f(x)}{g(x)}\right] \\ &\simeq \frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i)}{g(x_i)}\end{aligned}$$

(where the sample points x_i are drawn from density $g(x)$)

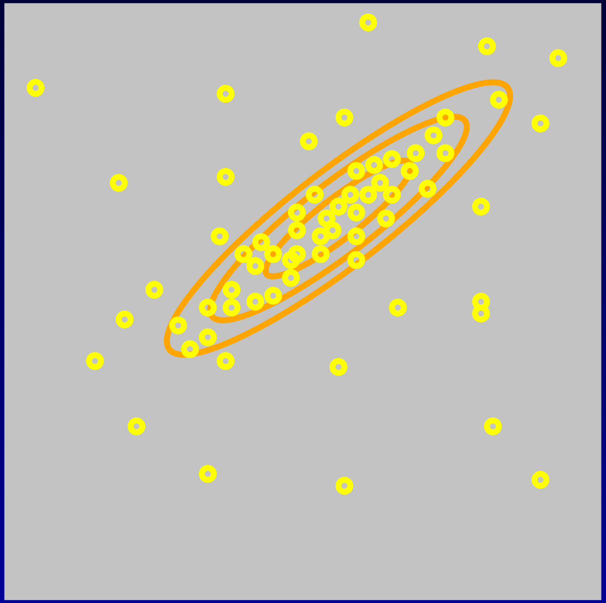
Accuracy is improved by importance sampling



Grid



Monte Carlo



Importance

Transition probabilities that achieve a given distribution

If we desire a particular equilbrium distribution π_i then one way to achieve it is to run a Markov chain that has transition probabilities that achieve *detailed balance*, so that for each pair of states the fraction of cases that move from i to j is the same as the fraction that move from j to i. If P_{ij} is the conditional probability of going from i to j then we achieve this if:

$$\pi_i P_{ij} = \pi_j P_{ji}$$

So if g_i is proportional to the desired distribution,

$$P_{ij}/P_{ji} = g_j/g_i$$

Any choice of P's that satisfies this is OK. To move around as fast as possible, suppose $g_j > g_i$. Then when j is proposed from i, accept it always. When i is proposed from j, accept it with probability g_i/g_j. So we use $P_{ij} = 1$ and $P_{ji} = g_i/g_j$.

MCMC: The Metropolis-Hastings method

To draw a sample $G_1, \dots, G_n$ from a distribution proportional to a function $g(G)$:

(1) Draw a change in G from some “proposal distribution": $x \to y$

(2a) (Metropolis et. al., 1953):
Accept the change if a uniformly-distributed random number R satisfies

$$R < \frac{g(y)}{g(x)}$$

(2b) Hastings (*Biometrika*, 1970) corrected for biases toward some y ’s in the proposal distribution by using instead

$$R < \frac{\mathrm{Prob}(x|y)}{\mathrm{Prob}(y|x)} \frac{g(y)}{g(x)}$$

Repeat many times. If we do this long enough, and various niceness conditions hold, then $G_1, \dots, G_m$ will be a sample from the right distribution.

Computing coalescent likelihoods by MCMC

We want to compute $\int_G \mathrm{Prob}\,(G|\Theta)\mathrm{Prob}\,(D|G)dG$. We use an importance sampling density proportional to the interior of the integral at some trial value Θ_0 of the parameter. Then it is

$$g(G) = \frac{\mathrm{Prob}\,(G|\Theta_0)\mathrm{Prob}\,(D|G)\,dG}{\int_G \mathrm{Prob}\,(G|\Theta_0)\mathrm{Prob}\,(D|G)\,dG}$$

whose denominator is

$$L(\Theta_0) = \int_G \mathrm{Prob}\,(G\,|\,\Theta_0)\mathrm{Prob}\,(D\,|\,G)\,dG$$

The integral is:

$$L(\Theta) \;\simeq\; \frac{1}{n}\sum_{i=1}^{n} \frac{f(G_i)}{g(G_i)}$$

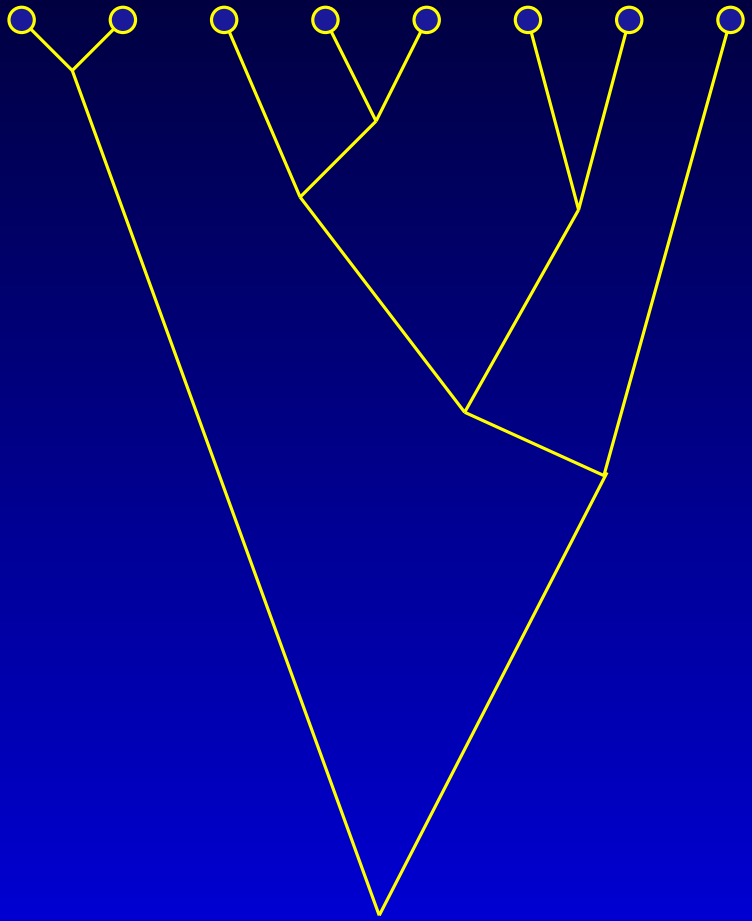
$$\simeq\; \frac{1}{n}\sum_{i=1}^{n} \frac{\text{Prob }(G|\Theta)\text{Prob }(D|G)}{\text{Prob }(G|\Theta_0)\text{Prob }(D|G)/L(\Theta_0)}$$

This leads to

$$\frac{L(\Theta)}{L(\Theta_0)} \simeq \frac{1}{n}\sum_{i=1}^{n} \frac{f(G_i)}{g(G_i)} \simeq \frac{1}{n}\sum_{i=1}^{n} \frac{\text{Prob }(G_i|\Theta)}{\text{Prob }(G_i|\Theta_0)}$$

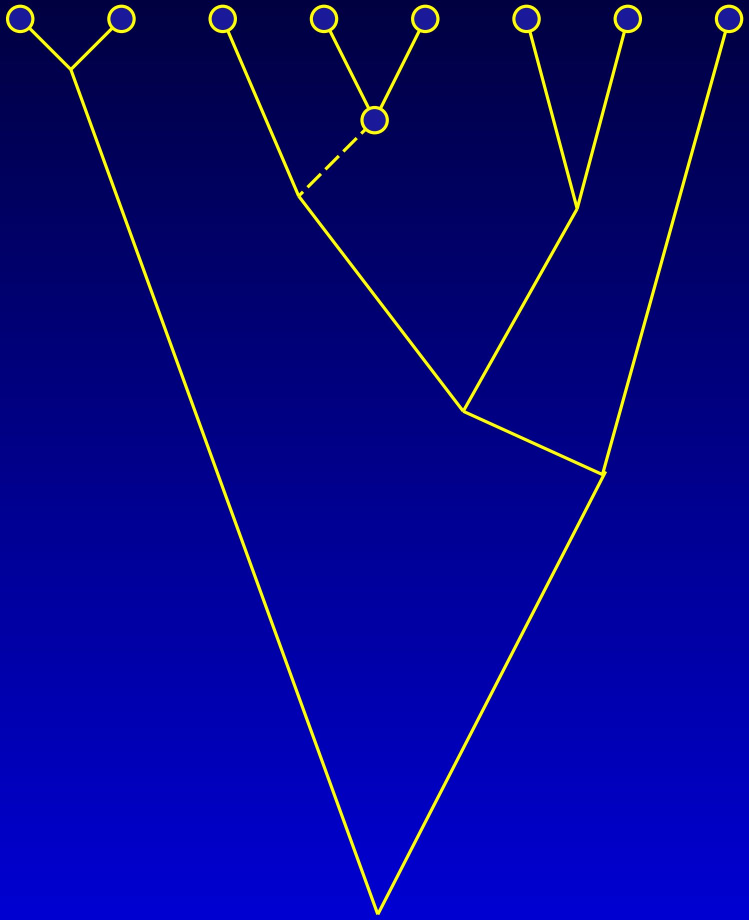
Tree rearrangements proposed:

A conditional coalescent rearrangement strategy



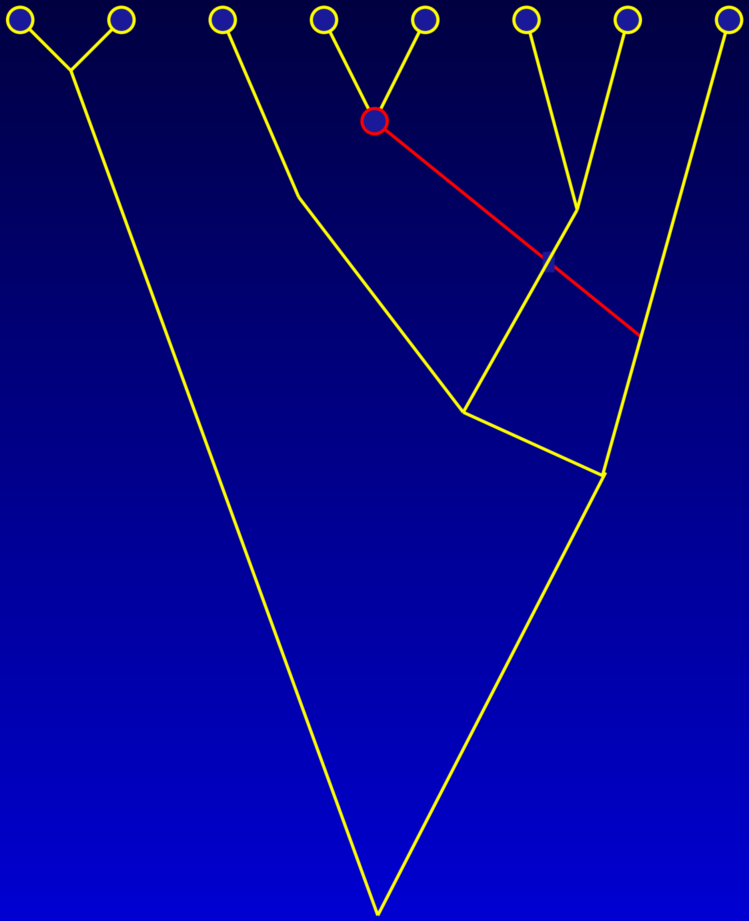
more ...

First pick a random node (interior or tip) and remove its subtree



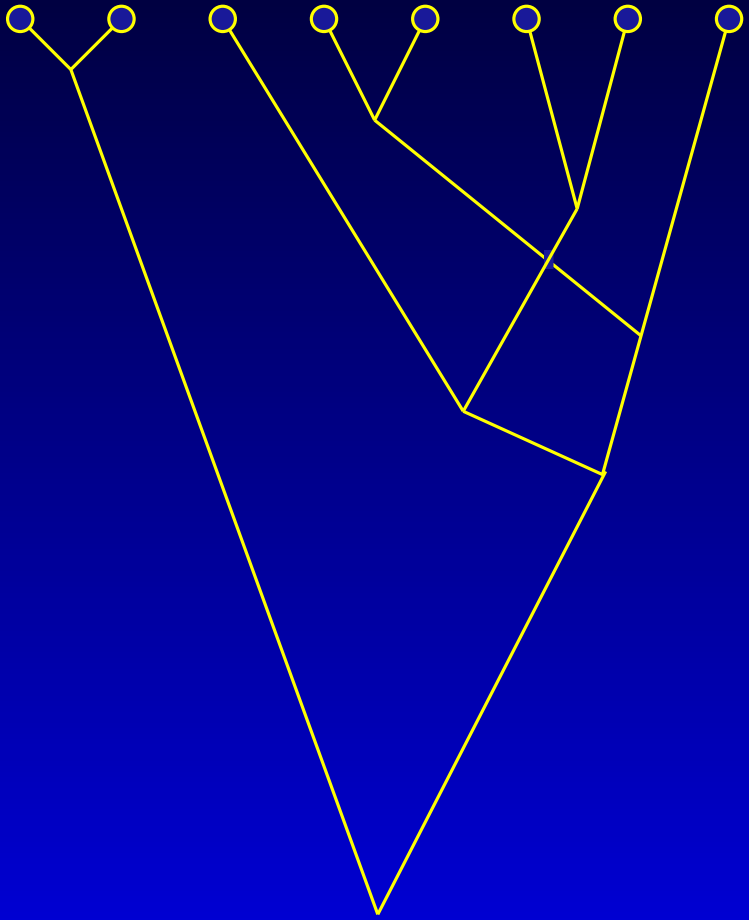
more ...

Then allow this node to re−coalesce with the tree



and finally we get:

The resulting tree proposed by this process

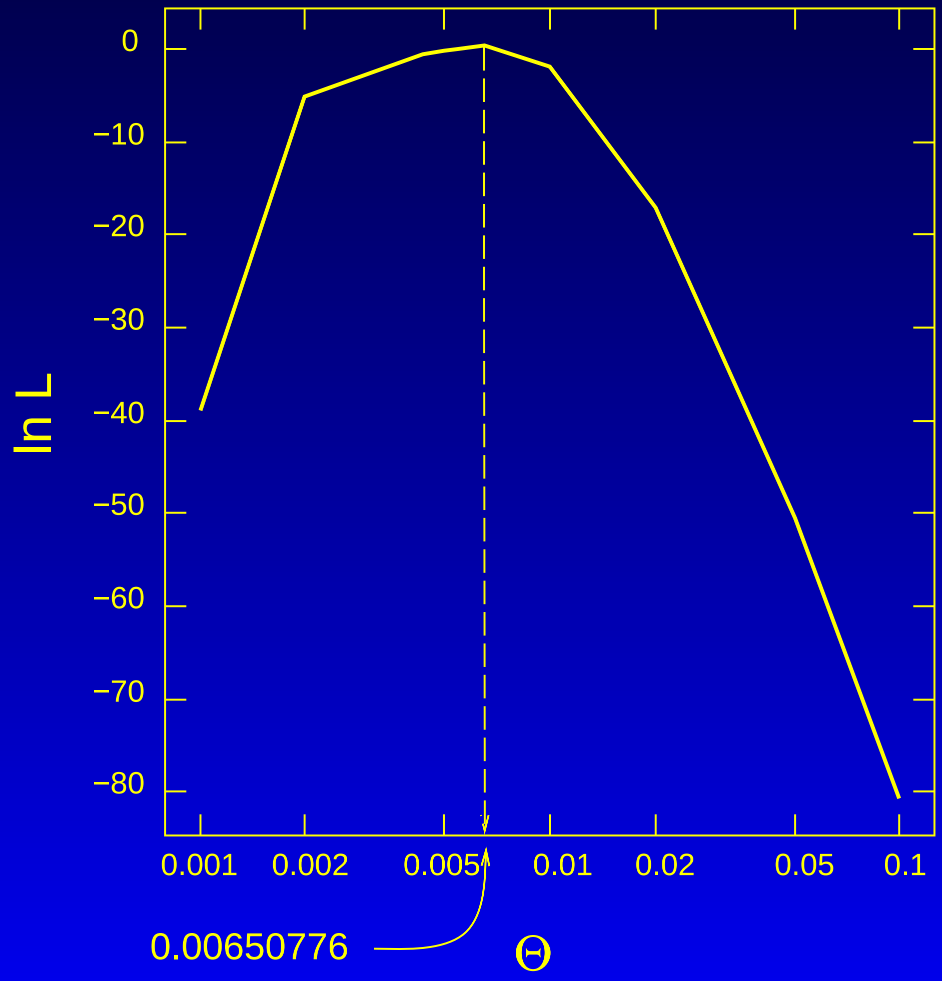


Being left out of the story:

- We choose the rearrangements so that the proposal distribution is a “conditional coalescent".
- We do a Hastings correction given this.
- The end result is a perfect cancellation (which is pleasant rather than essential).
- This leaves us with the rule that we use $\mathrm{Prob}\,(\mathrm{D}\,|\,\mathrm{G})$ as the only function in the Metropolizing.

One ends up with a curve that might look like this:

Results of analysing a data set with 50 sequences of 500 bases
which was simulated with a true value of Θ = 0.01



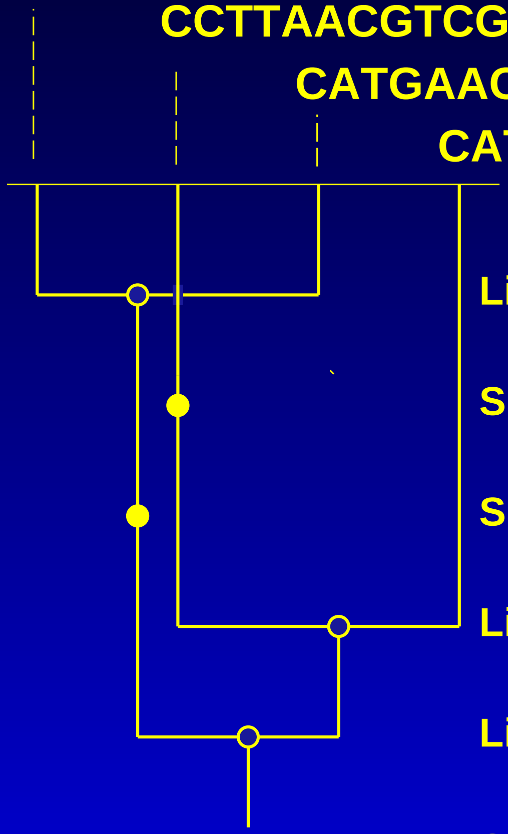
Griffiths’ and Tavaré’s (1994) method

CATGAACGTCG

CCTTAACGTCG

CATGAACGTCG

CATTAACGTCG



Lineages 1 and 3 coalesce

Site 2 in lineage 2 has mutated A−>C

Site 4 in lineage 1 has mutated T−>G

Lineages 2 and 4 coalesce

Lineages 1 and 2 coalesce

CATTAACGTCG

Started with this sequence

They sample sequences of events (coalescences, mutations, etc.)
– these have no times but show mutations explicitly

Griffiths and Tavaré's method as importance sampling

D	the data (sequences)
β	the parameters ($4N_e\mu$ and such)
H_i	the i-th of all possible histories of events
h_{ij}	the j-th event in history H_i
$a_{ijk}(\beta)$	the probability (rate) of the k-th of the possible events that could have happened at stage j of history i (ignoring the data).

Some definitions

$\mathsf{b_{ij}}(\beta)$	the probability (rate) of the one that did happen at stage j of history i
$\mathsf{c_{ijk}}(\beta)$	the probability (rate) of the k-th of the possible events that could have happened at stage j of history i (counting only those that are compatible with the data).

The Griffiths-Tavaré method

$$\mathsf{L} = \sum_{\mathsf{H}} \mathrm{Prob}\ (\mathsf{H}|\beta)\ \mathrm{Prob}\ (\mathsf{D}|\mathsf{H}) = \mathsf{E_f}\ \mathrm{Prob}\ (\mathsf{D}|\mathsf{H})$$

The distribution f is:

$$\mathrm{Prob}\ (\mathsf{H_i}|\mathsf{g}) = \Pi_{\mathsf{j}} \frac{\mathsf{b_{ij}}(\beta)}{\sum_{\mathsf{k}} \mathsf{a_{ijk}}(\beta)} = \frac{\Pi_{\mathsf{j}} \mathsf{b_{ij}}(\beta)}{\Pi_{\mathsf{j}} \left(\sum_{\mathsf{k}} \mathsf{a_{ijk}}(\beta)\right)}$$

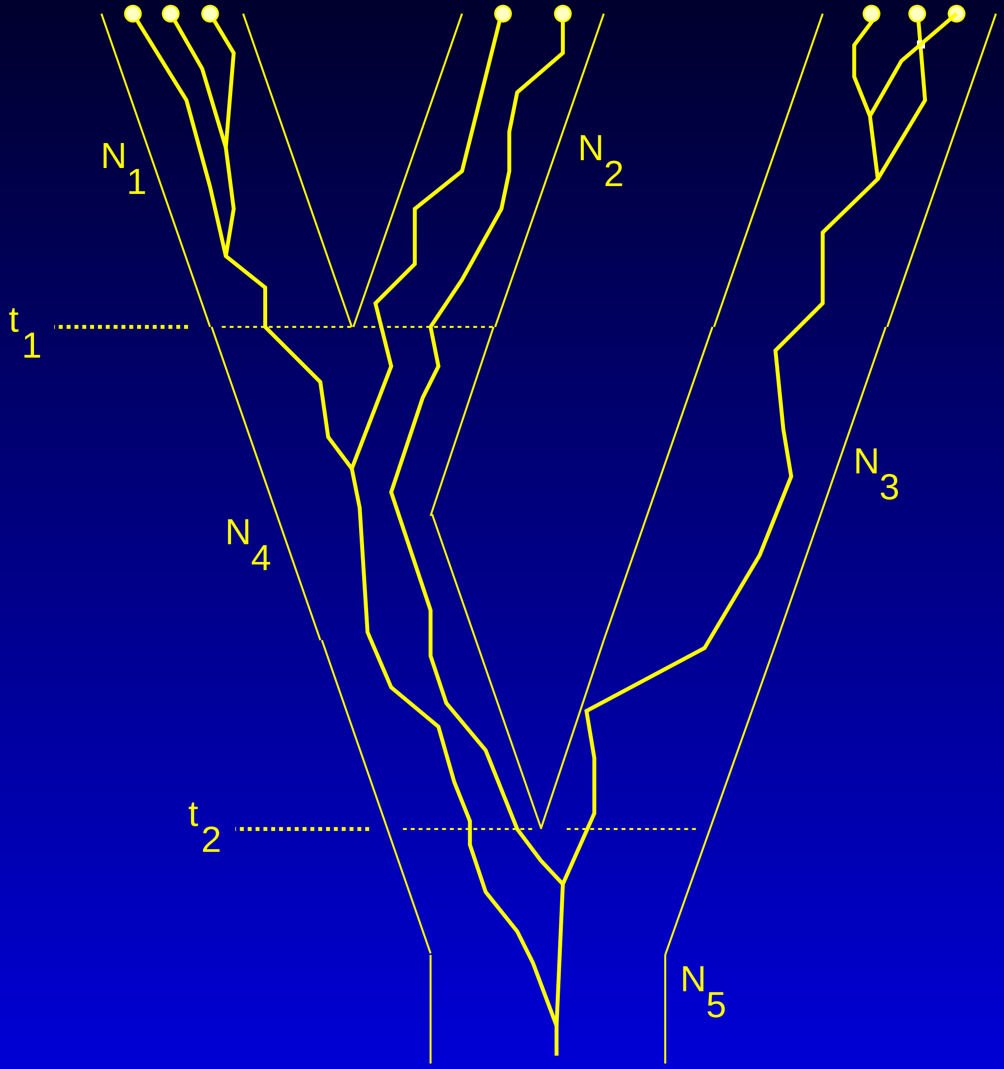
(the distribution g is the same but with c's instead of a's).

$$\mathsf{L}(\beta) = \mathsf{E_f}\ [\mathrm{Prob}\ (\mathsf{D}|\mathsf{H})] = \mathsf{E_g} \left[\frac{\mathsf{f}}{\mathsf{g}} \mathrm{Prob}\ (\mathsf{D}|\mathsf{H}) \right]$$

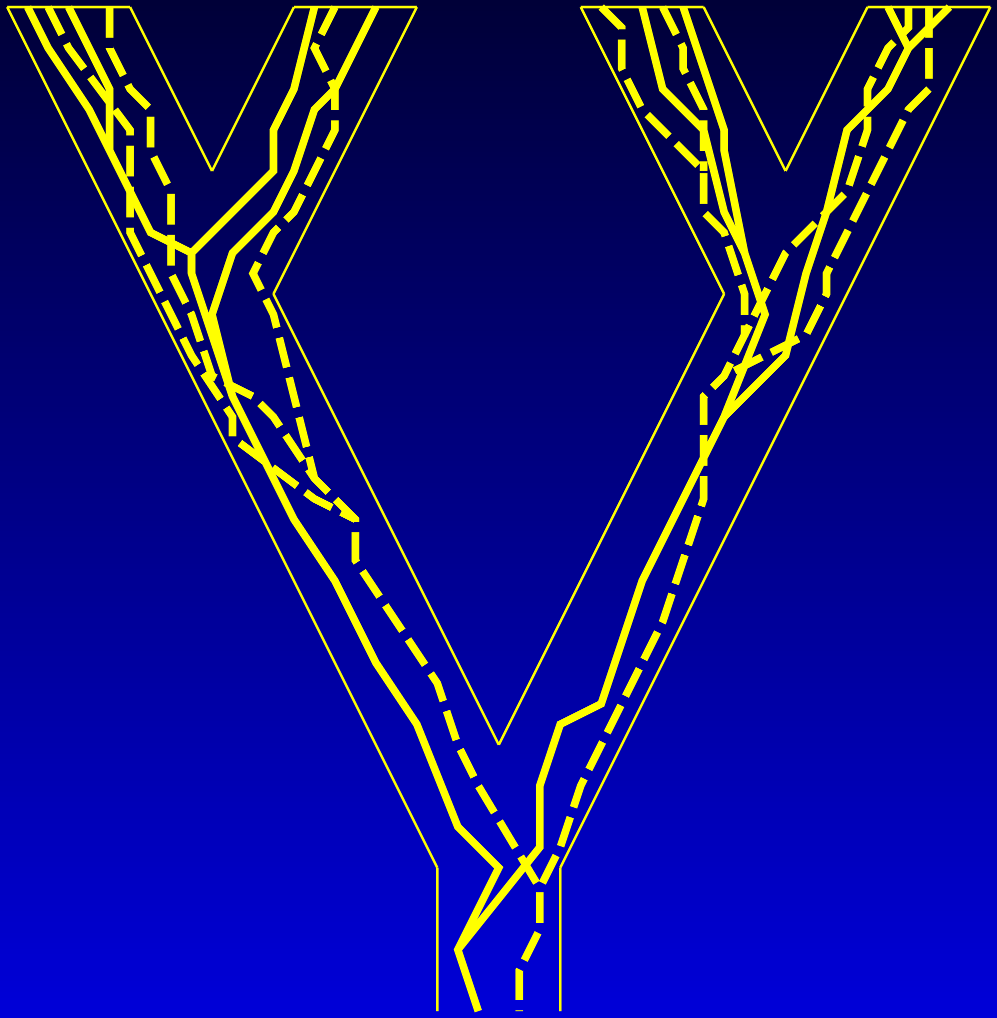
We end up with

$$
\begin{aligned}
L(\beta) &= \mathbb{E}_g \left[\frac{\left(\dfrac{\Pi_j b_{ij}(\beta)}{\Pi_j \left(\sum_k a_{ijk}(\beta) \right)} \right)}{\left(\dfrac{\Pi_j b_{ij}(\beta_0)}{\Pi_j \left(\sum_k c_{ijk}(\beta_0) \right)} \right)} \right] \\
&= \mathbb{E}_g \left[\Pi_j \left(\frac{b_{ij}(\beta)}{b_{ij}(\beta_0)} \right) \Pi_j \left(\frac{\sum_k c_{ijk}(\beta_0)}{\sum_k a_{ijk}(\beta)} \right) \right]
\end{aligned}
$$

A coalescent and a species tree



Different coalescents with one phylogeny



Integrating over all possible genealogies at each locus

$$
\begin{aligned}
L &= \text{Prob}\left(\text{Data} \mid \text{Tree}\right) \\
&= \prod_{i=1}^{\text{loci}} \int_G \text{Prob}\left(\text{coalescent } G \mid \text{Tree}\ \right) \\
&\qquad \times \text{Prob}\left(\text{Data } i \mid \text{coalescent } G\right)
\end{aligned}
$$

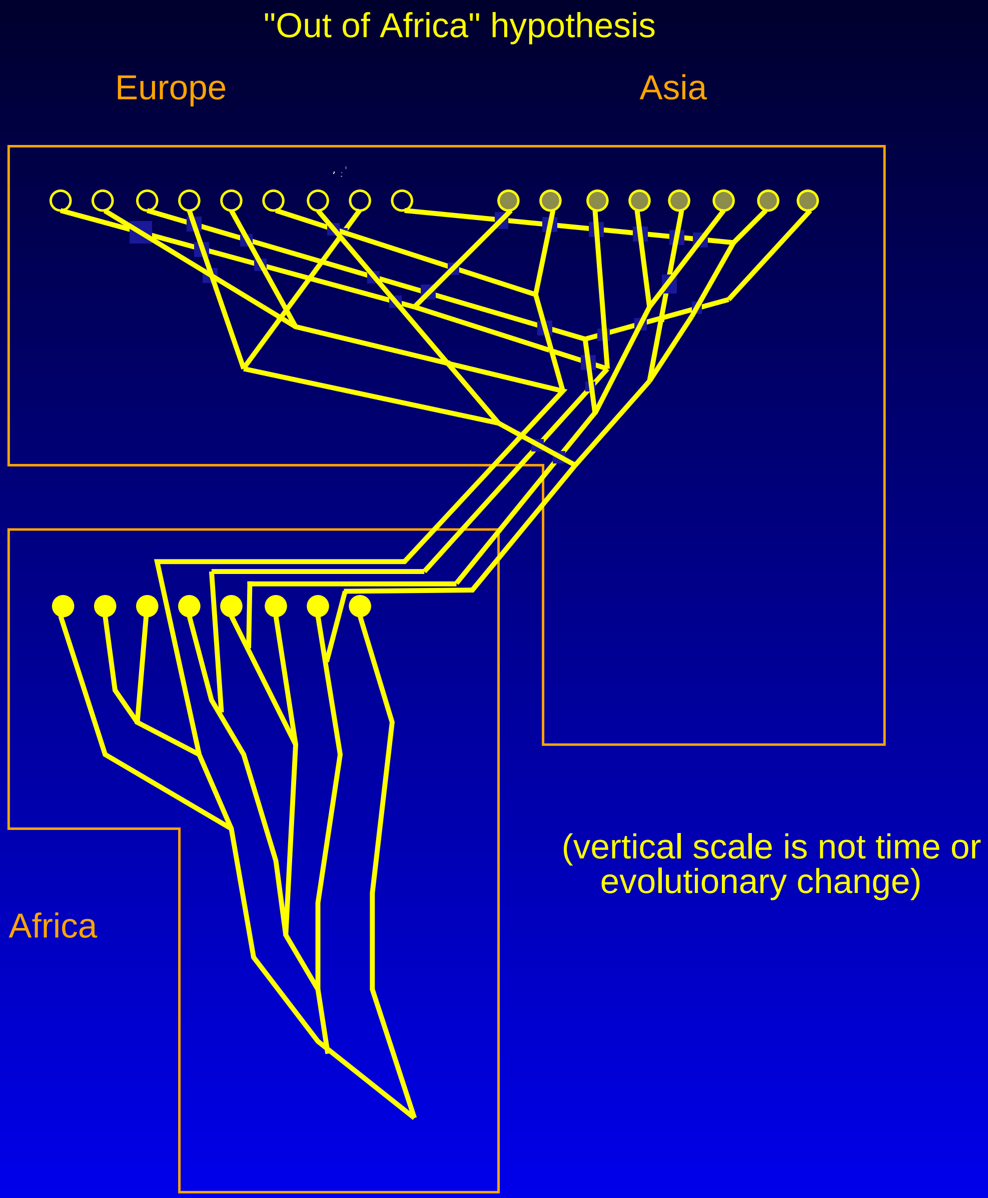
New genetic tools being deployed

Likelihood or Bayesian inference using sampling methods with coalescents

Mig = Migration, Rec = Recombination, Grow = Population growth,
Split = Splittings, Bayes = Bayesian

Program Name	Mig?	Rec?	Grow?	Split?	Bayes?
LAMARC (Kuhner, Beerli et al.)	Y	Y	Y	n	Y
BEAST (Drummond, Rambaut, Pybus)	n	n	Y	Y	Y
Genetree (Griffiths and Bahlo)	Y	n	Y	n	n
Batwing (Wilson and Balding)	n	n	Y	Y	Y
MDIV (Nielsen)	Y	n	n	Y	Y
IM (Hey and Nielsen)	Y	n	n	Y	Y

We ultimately want to treat this case



(vertical scale is not time or evolutionary change)

References

Beerli, P. B. and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152:** 763-773. **[Our approach to migration estimation]**

Edwards, A. W. F. 1970. Estimation of the branch points of a branching diffusion process. *Journal of the Royal Statistical Society B* **32:** 155-174. **[Labelled histories]**

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]**

Felsenstein, J. 1992. Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. *Genetical Research* **59:** 139-147. **[Suggests using the coalescents]**

References

Felsenstein, J., M. K. Kuhner, J. Yamato, and P. Beerli. 1999. Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. pp. 163-185 in *Statistics in Molecular Biology and Genetics*, ed. F. Seillier-Moiseiwitsch. IMS Lecture Notes-Monograph Series, volume 33. Institute of Mathematical Statistics and American Mathematical Society, Hayward, California. **[Overview of my lab's methods]**

Griffiths, R. C. 1989. Genealogical tree probabilities in the infinitely-many-site model. *Journal of Mathematical Biology* **27:** 667-680. **[Summing up over event histories]**

Griffiths, R. C. and S. Tavaré. 1994a. Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Socety of London, Series B (Biological Sciences)* **344:** 403-10. **[Griffiths-Tavaré sampling method]**

Griffiths, R. C. and S. Tavaré. 1994b. Ancestral inference in population genetics. *Statistical Science* **9:** 307-319. **[Griffiths-Tavaré sampling method]**

References

Griffiths, R. C. and P. Marjoram. 1996. Ancestral inferences from samples of DNA sequences with recombination. *Journal of Computational Biology* **3:** 479-502. **[Coalescent likelihoods with recombination]**

Griffiths, R. C. and S. Tavaré. 1997. Computational methods for the coalescent. pp. 165-182 in *Progress in Population Genetics and Human Evolution*, ed. P. Donnelly and S. Tavaré. IMA Volumes on Mathematics and Its Applications, volume 87. Springer, New York. **[Review of their approach]** Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57:** 97-109. **[The Hastings correction]**

Hein, J., M. H. Schierup, and C. Wiuf. 2005. *Gene Genealogies, Variation, and Evolution*. Oxford University Press, Oxford. **[Then first book on the coalescent, emphasizing its population genetics]**

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]**

References

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]**

Kingman, J. F. C. 1982b. On the genealogy of large populations. *Journal of Applied Probability* **19A:** 27-43. **[Another 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]**

Kuhner, M. K., J. Yamato, and J. Felsenstein. 1995. Effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140:** 1421-1430. **[Our MCMC coalescent likelihood method]**

Kuhner, M. K., J. Yamato, and J. Felsenstein 1997. Applications of Metropolis-Hastings genealogy sampling. pp. 183-192 in *Progress in Population Genetics and Human Evolution*, ed. P. Donnelly and S. Tavare. IMA Volumes in Mathematics and its Applications, volume 87. Springer Verlag, Berlin. **[Brief review of our approach]**

References

Kuhner, M. K., J. Yamato and J. Felsenstein. 1998. Maximum likelihood estimation of population growth rates based on the coalescent *Genetics* **149:** 429-434. **[Our approach to growing populations, and describes a bias in estimation]**

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* **21:** 1087-1092. **[The Metropolis algorithm]**

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]**

Nielsen, R. 1997. Maximum likelihood estimation of population divergence times and population phylogenies under the infinite sites model. *Theoretical Population Biology* **53:** 143-151. **[The first coalescent likelihood paper with more than one species]**

Takahata, N. 1988. The coalescent in two partially isolated diffusion populations. *Genetical Research* **52:** 213-222. **[Coalescents with migration]**

Wakeley, J. 2005. *Coalescent Theory, an Introduction*. Roberts and Co., Greenwood Village, Colorado. **[Coming this summer – very readable, with emphasis on the population genetics]**

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 (prosper.sourceforge.net),
- using Latex to make a `.dvi` file,
- using `dvips` to turn this into a Postscript file,
- using `ps2pdf` to make it into a PDF file, and
- displaying the slides in Adobe Acrobat Reader.

Result: nice slides using freeware.