3 main forces in evolution:

I. Genetic Drift

II. Mutation

III. Natural Selection

(recombination?)

One definition of evolution: The change of allele/genotype frequencies in a population

Population Genetics is the study of genetic variation within a population

Hardy-Weinberg Genotype Frequencies ...

Assume gene has allele *A* and allele *a*

p = Frequency of *A*
q = 1-p = Frequency of *a*

Genotype Frequencies with "random mating"

AA: p^2

Aa: 2pq

aa: q^2

Hardy-Weinberg Proportions:

AA - p^2

Aa - $2pq$

Aa - q^2

where p is frequency of allele A and q=1-p is freq. of allele a.

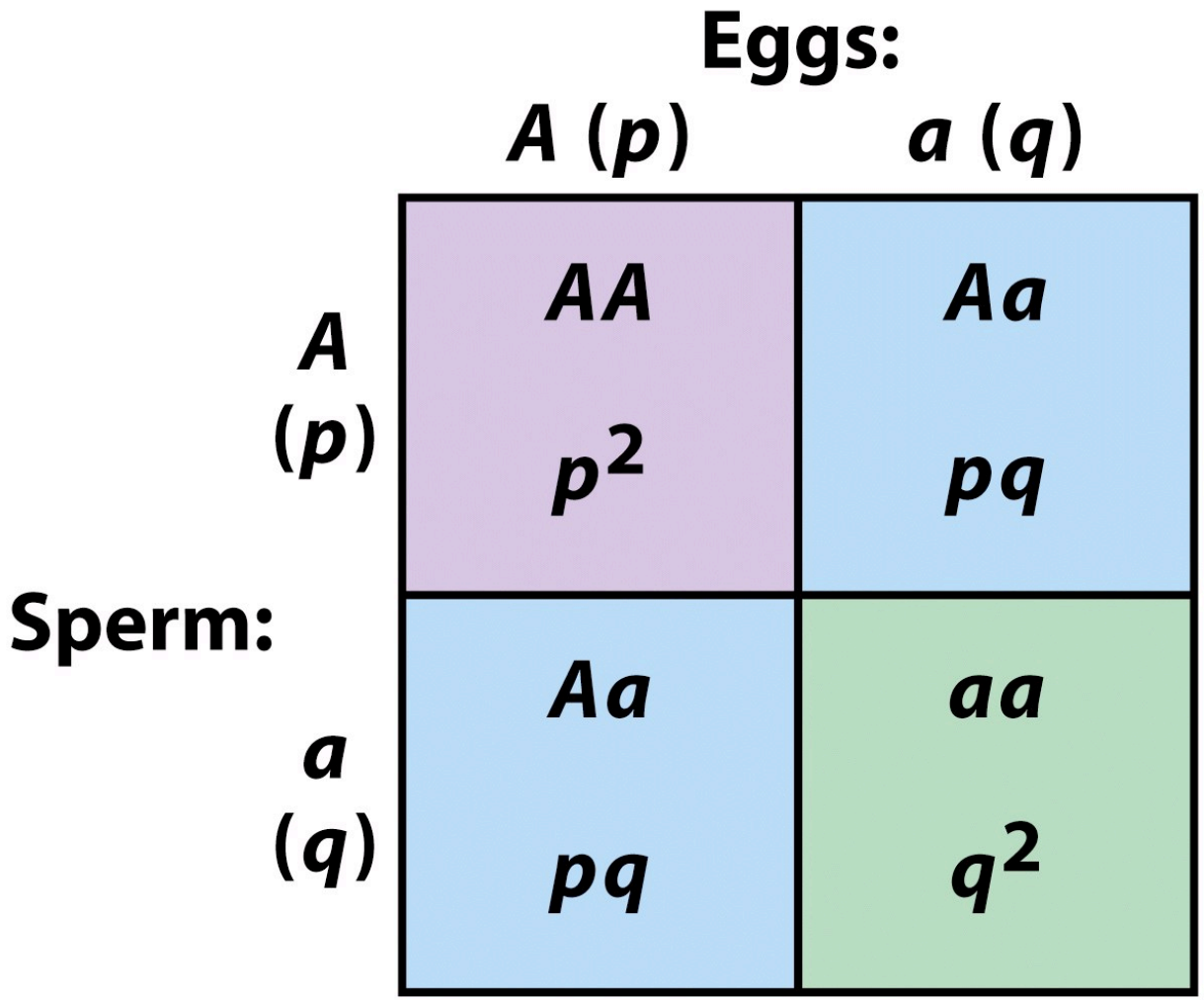


Figure 26-1 Principles of Genetics, 4/e

Reasons for departure from Hardy-Weinberg proportions

1. **Nonrandom mating**
   - **Mating with relatives (i.e., consanguineous mating)**
   - **Mating due to phenotypic similarity (i.e., assortative mating)**

2. **Different survival among genotypes (i.e., fitness differences)**

3. **Population Subdivision / Migration**

Genotype Frequencies with Inbreeding:

AA: $p^2(1-F) + pF$

Aa: $2p(1-p)(1-F)$

aa: $(1-p)^2(1-F) + (1-p)F$

where F is the inbreeding coefficient

Differences in survival (i.e., fitness) among genotypes Can cause post-conception deviations from Hardy-Weinberg proportions.

For example, consider a recessive disease that results in all people with the homozygous recessive disease dying before adulthood. If disease allele "a" has frequency of 0.1, then genotype frequencies of AA, Aa, and aa are respectively 0.81, 0.18, and 0.01 at conception. However, frequencies in adulthood of these 3 genotypes would be 0.81/0.99, 0.18/0.99, and 0.

Hardy-Weinberg deviations due to migration/admixture

Imagine Populations 1 and 2 contribute equal numbers to a newly formed Population 3.

Assume Population 1 frequencies of AA, Aa, and aa are respectively 0.81, 0.18, and 0.01.

Assume Population 2 frequencies of AA, Aa, and aa are respectively 0.36, 0.48, and 0.16.

Then, Population 3 frequencies of AA, Aa, and aa are respectively 0.585, 0.33, and 0.085. Notice these genotype frequencies are not in H-W proportions.

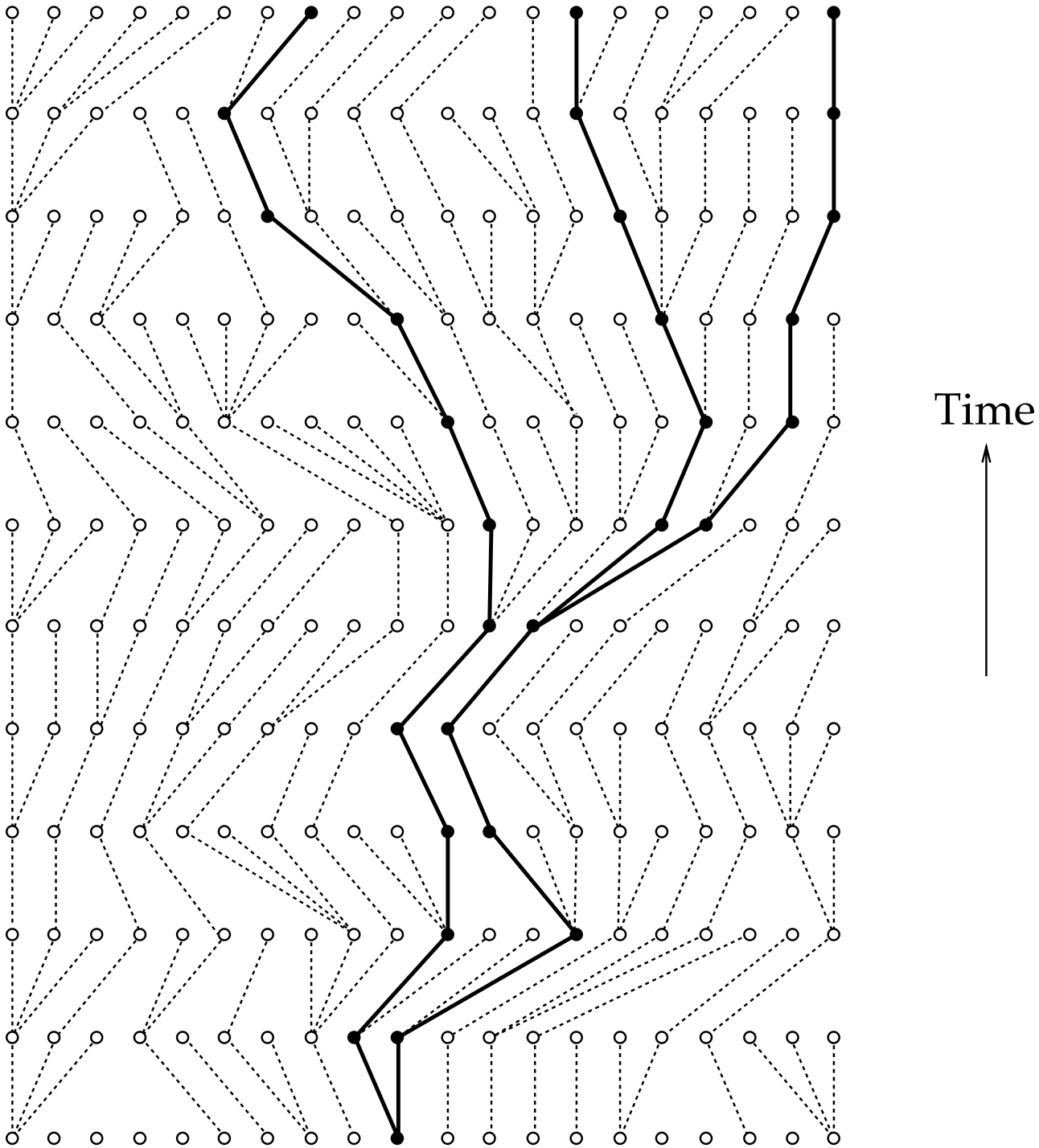
Wright-Fisher Model is an overly simplistic model of how gene frequencies change over time due to genetic drift.

Wright-Fisher Model for diploid creatures:

1. Each generation has N individuals and therefore 2N copies of each gene.

2. Each generation formed from preceding generation by randomly sampling gene copies in preceding generation with replacement.

Most figures in this section were kindly provided by Dr. Joseph Felsenstein of the University of Washington

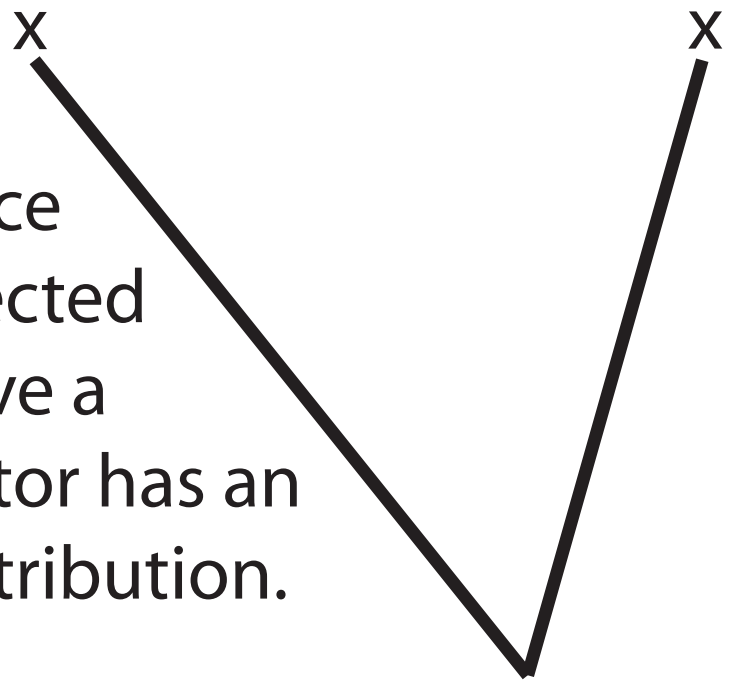

Time

Wright-Fisher model describes how gene frequencies change as time goes from "now" to the future.

Coalescent process approximates Wright-Fisher model but examines genetic relationships with the perspective of time going from "now" to the past.

Because we collect sequence data "now" to examine what has happened in the past with respect to evolution, the coalescent process is very important for data analysis.

With no natural selection and constant diploid population of size N ...



Number of generations since 2 randomly selected gene copies have a common ancestor has an **exponential** distribution.

For diploid population with N individuals (i.e., 2N gene copies), the mean of this exponential distribution is 2N generations.

Coalescent event

Why?

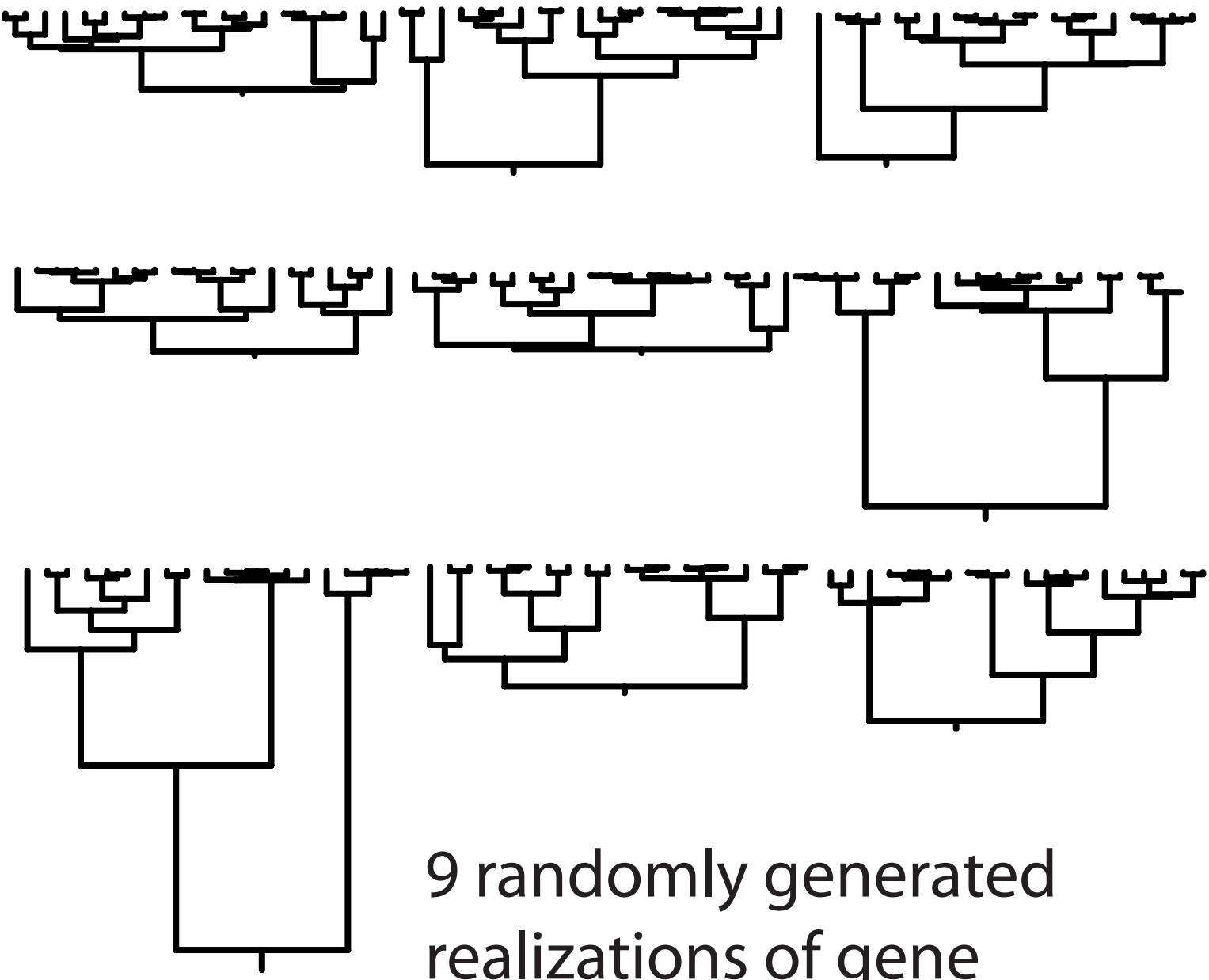
Gen. 0: x x x x ....x x x x x x

(2N total copies)

Gen. 1: x x x x ....x x x x x x

Chance that copy underlined in light blue has same parent in Generation 0 as copy underlined in red is 1/(2N)

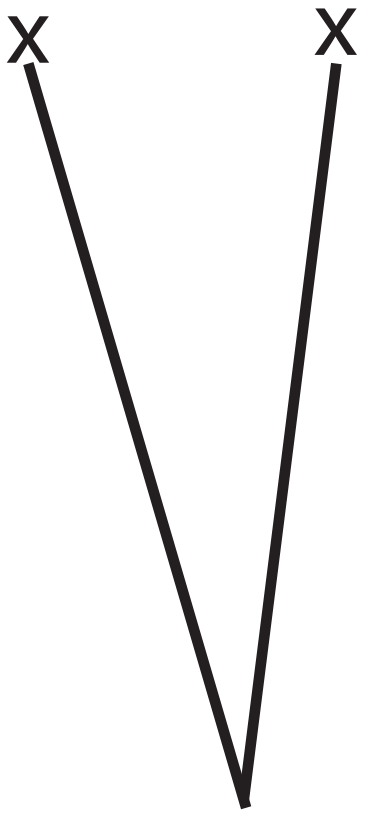
Chance these 2 copies have different parents in Generation 0 is therefore 1- 1/(2N)



9 randomly generated realizations of gene trees from the coalescent process, all with 20 tips & drawn to same scale (Figure 26.5 from "Inferring Phylogenies" book by J. Felsenstein book)

2N gene copies in population

μ is mutation rate per gene copy per generation



How different do we expect 2 randomly selected sequences to be?

We expect they had common ancestor 2N generations ago. Evolution since common ancestor to each sampled copy should then result in $2N\mu$ differences. $2N\mu + 2N\mu = 4N\mu$ is expected number of differences.

N and μ are confounded when sequences compared. Only their product can be estimated! Conventionally, $4N\mu$ is called θ

Much emphasis has been placed on estimating $\theta = 4N_e\mu$ because it is a fundamental parameter of molecular population genetics.

Statistical inference with the coalescent is a relatively advanced topic and work to make the model more realistic has been done to incorporate natural selection, migration, recombination, changes in population sizes, etc.

The coalescent process has some interesting properties.

For a sample of size k, time since most recent common ancestor of all k copies is on average 4N(1 - 1/k). Of this 4N(1-1/k) about 1/2 on average (i.e., 2N) comes from the time between the next-to-last and the last coalescent event (i.e., the time while there are exactly 2 lineages).

MOST of the variance among gene trees in time since k copies had most recent common ancestor comes from variance in time while there are exactly 2 lineages.

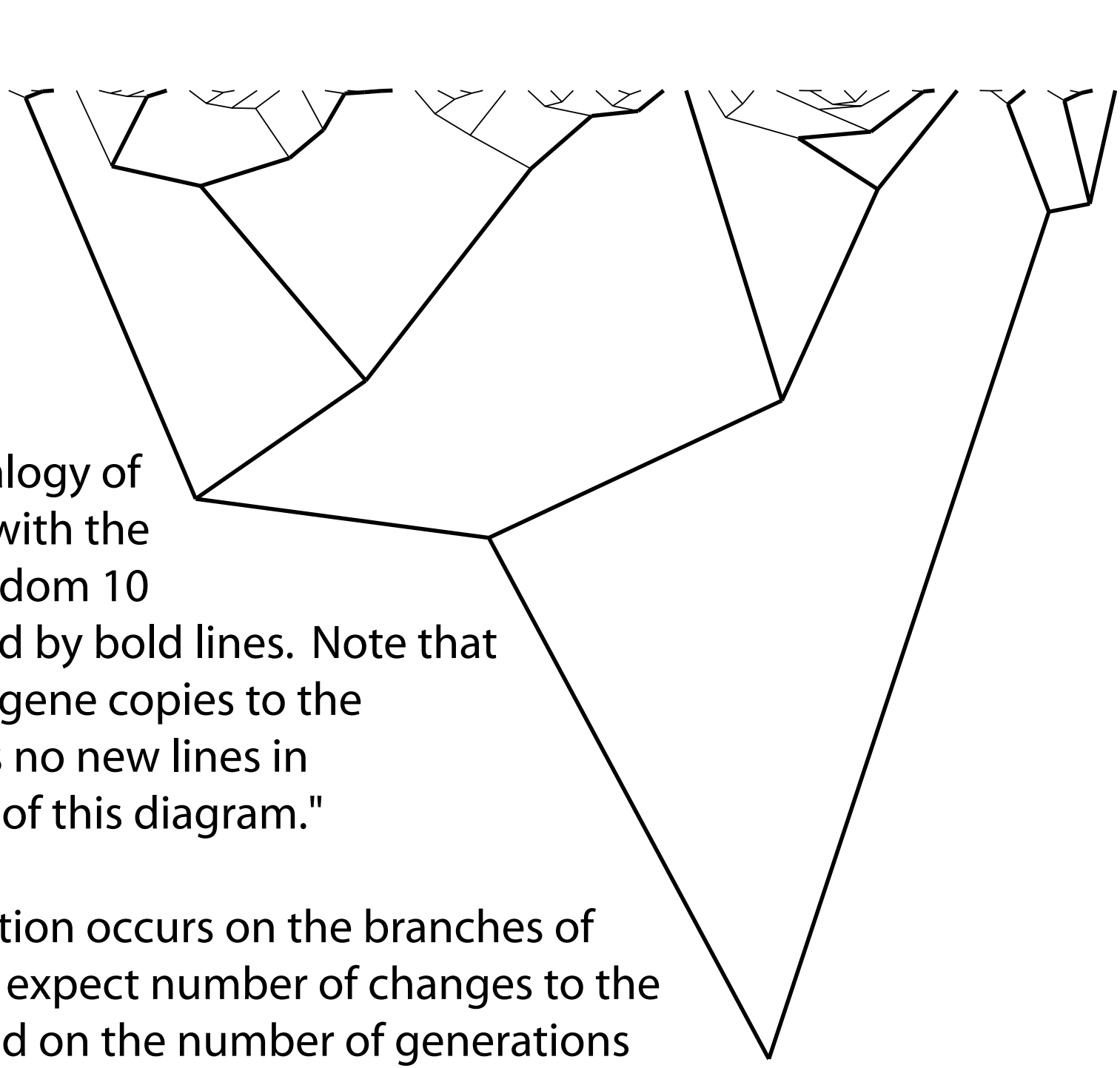


Figure 26.6 from
“Inferring
Phylogenies”
by Felsenstein

"A sample genealogy of 50 gene copies, with the ancestry of a random 10 of them indicated by bold lines. Note that adding 40 more gene copies to the sample discloses no new lines in the bottom part of this diagram."

(sequence evolution occurs on the branches of this tree and the expect number of changes to the sequence depend on the number of generations represented by the branches.)

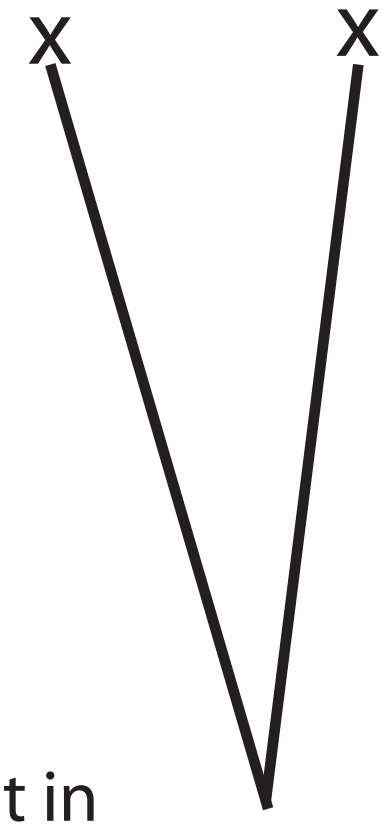
Expectations from the coalescent (based on Table 8.1 from Hein et al.)

Assume: (1) Human Effective Population Size been constant at Ten Thousand
(2) 25 years per human generation
(3) Sample size of 50 human genomes

Time(in 2N gens)	Time (in million yrs)	Prob(TMRCA > t)	#base pairs
1	0.5	0.85	2.6 billion
2	1.0	0.38	1.1 billion
3	1.5	0.14	430 million
4	2.0	0.052	160 million
5	2.5	0.019	58 million
6	3.0	0.007	21 million
8	4.0	0.00097	2.9 million
10	5.0	0.00013	393 thousand
12	6.0	0.000018	53 thousand
16	8.0	0.00000032	973

Notes: (1) Assumes neutrality (2) Implications for phylogenetics

2N gene copies in population



μ is mutation rate per gene copy per generation

How different do we expect 2 randomly selected sequences to be?

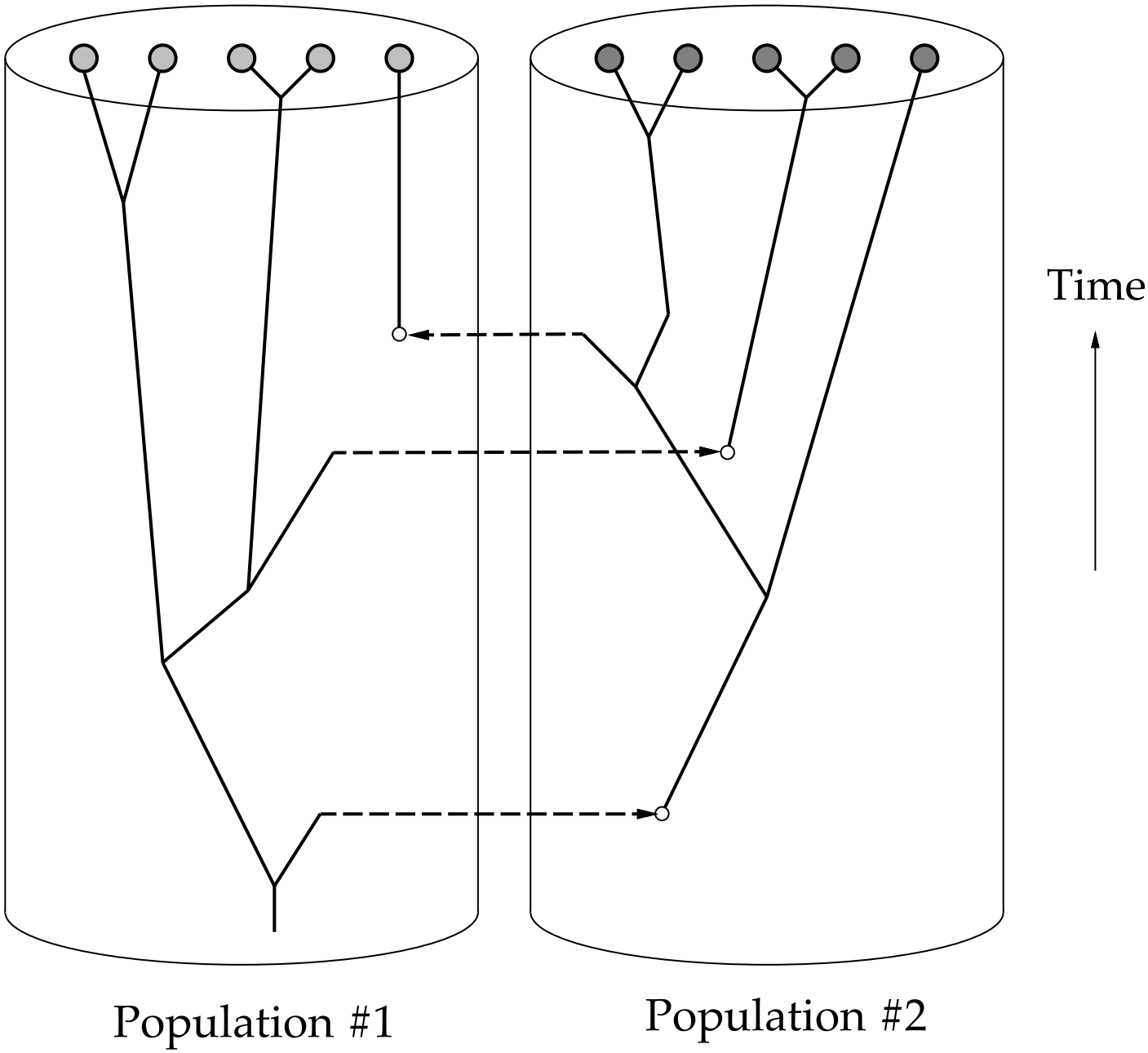
We expect they had common ancestor2N generations ago. Evolution since common ancestor to each sampled copy should then result in $2N\mu$ differences. $2N\mu + 2N\mu = 4N\mu$ is expected number of differences.

N and μ are confounded when sequences compared. Only their product can be estimated!

Conventionally, $4N\mu$ is called θ

Much emphasis has been placed on estimating $\theta=4N\mu$ because it is a fundamental parameter of molecular population genetics.

Statistical inference with the coalescent is a relatively advanced topic and work to make the model more realistic has been done to incorporate natural selection, migration, recombination, changes in population sizes, etc.



Coalescent trees with migration (Figure 26.7 from Felsenstein book)

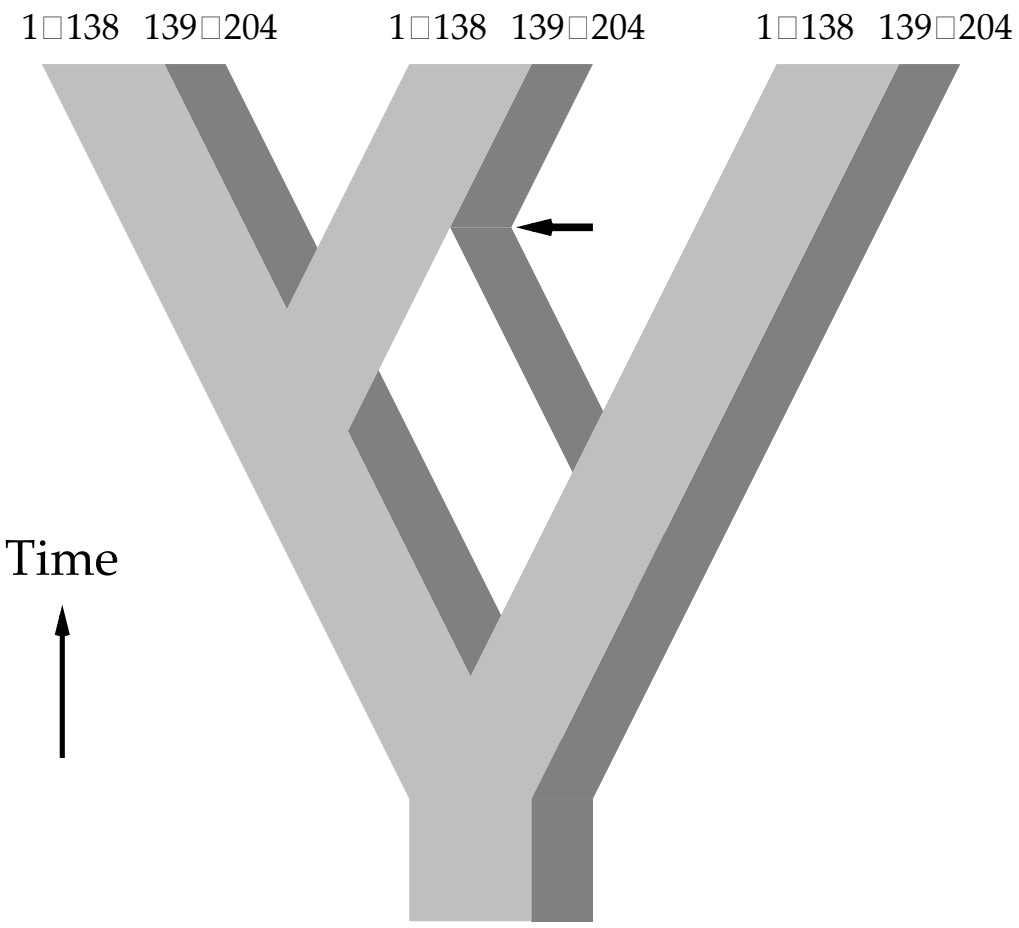
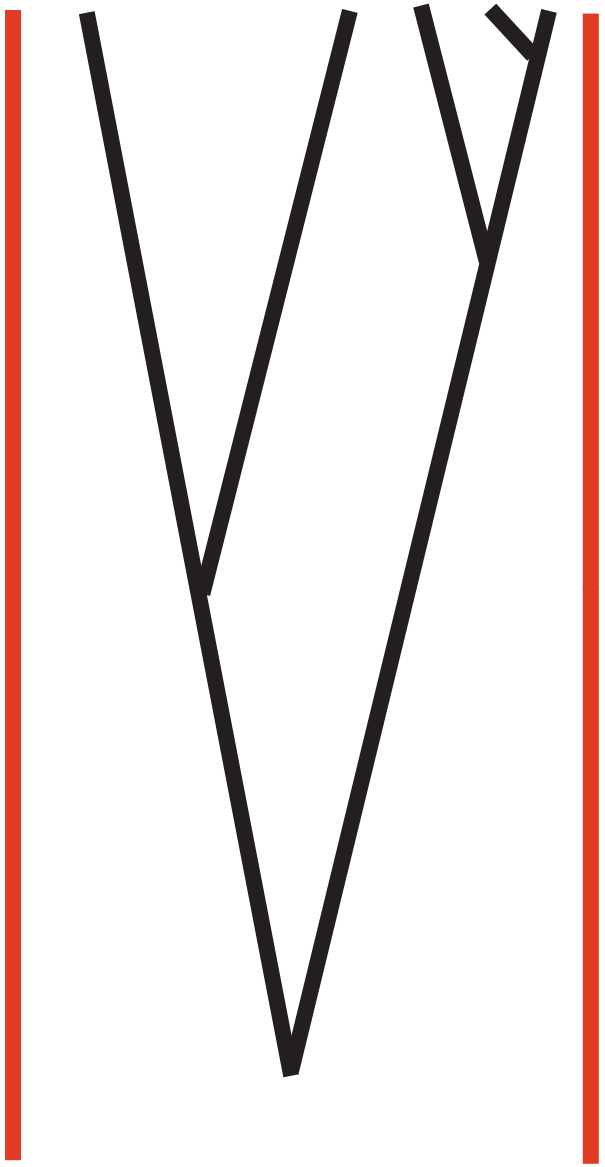
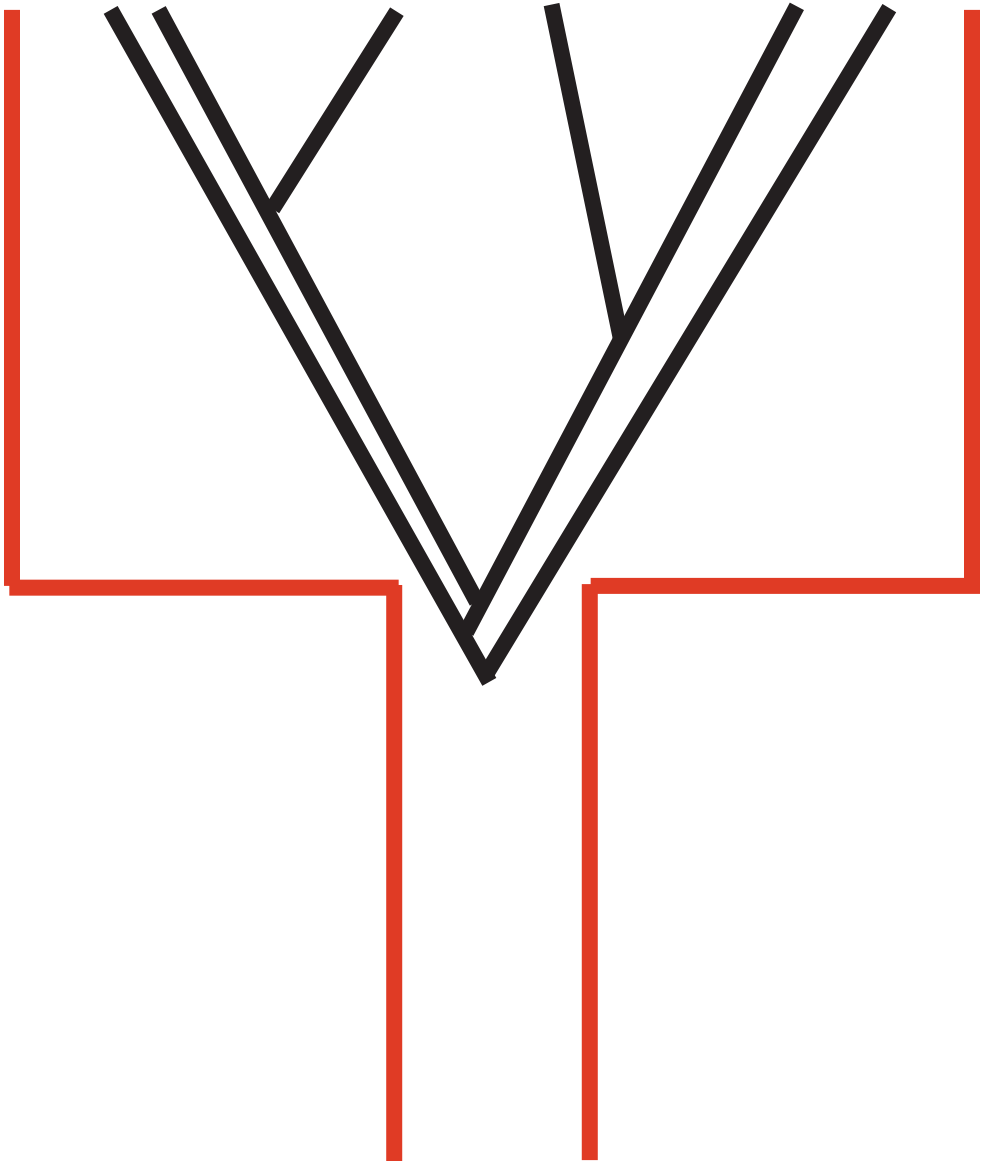


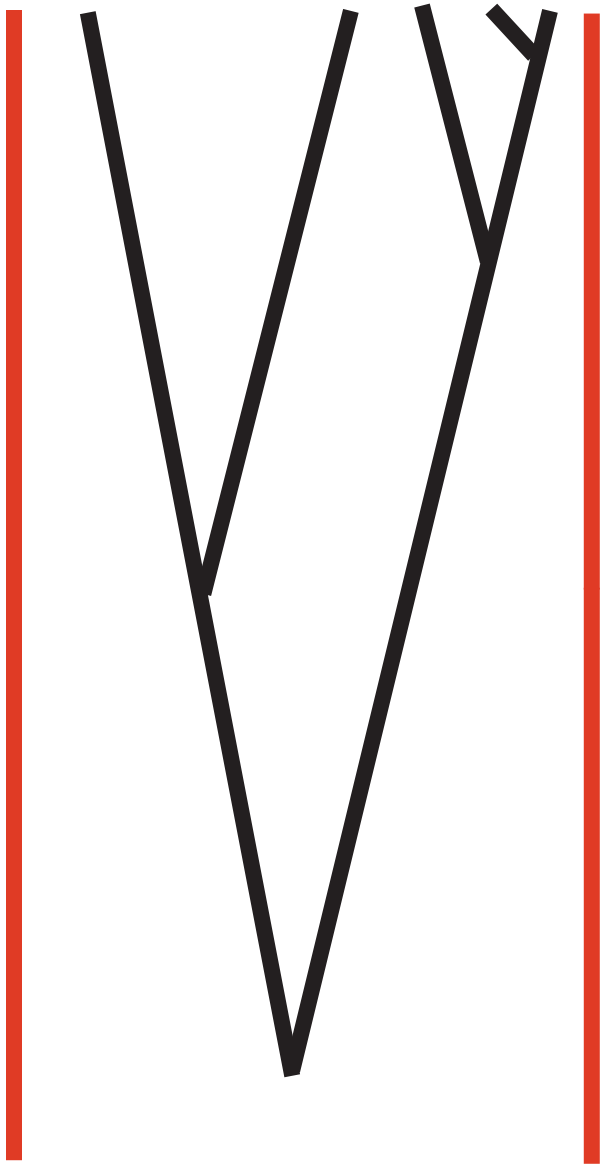
Figure 28.8 from Felsenstein: Coalescent with recombination. Horizontal arrow shows recombination event between positions 138 and 139 of a sequence.



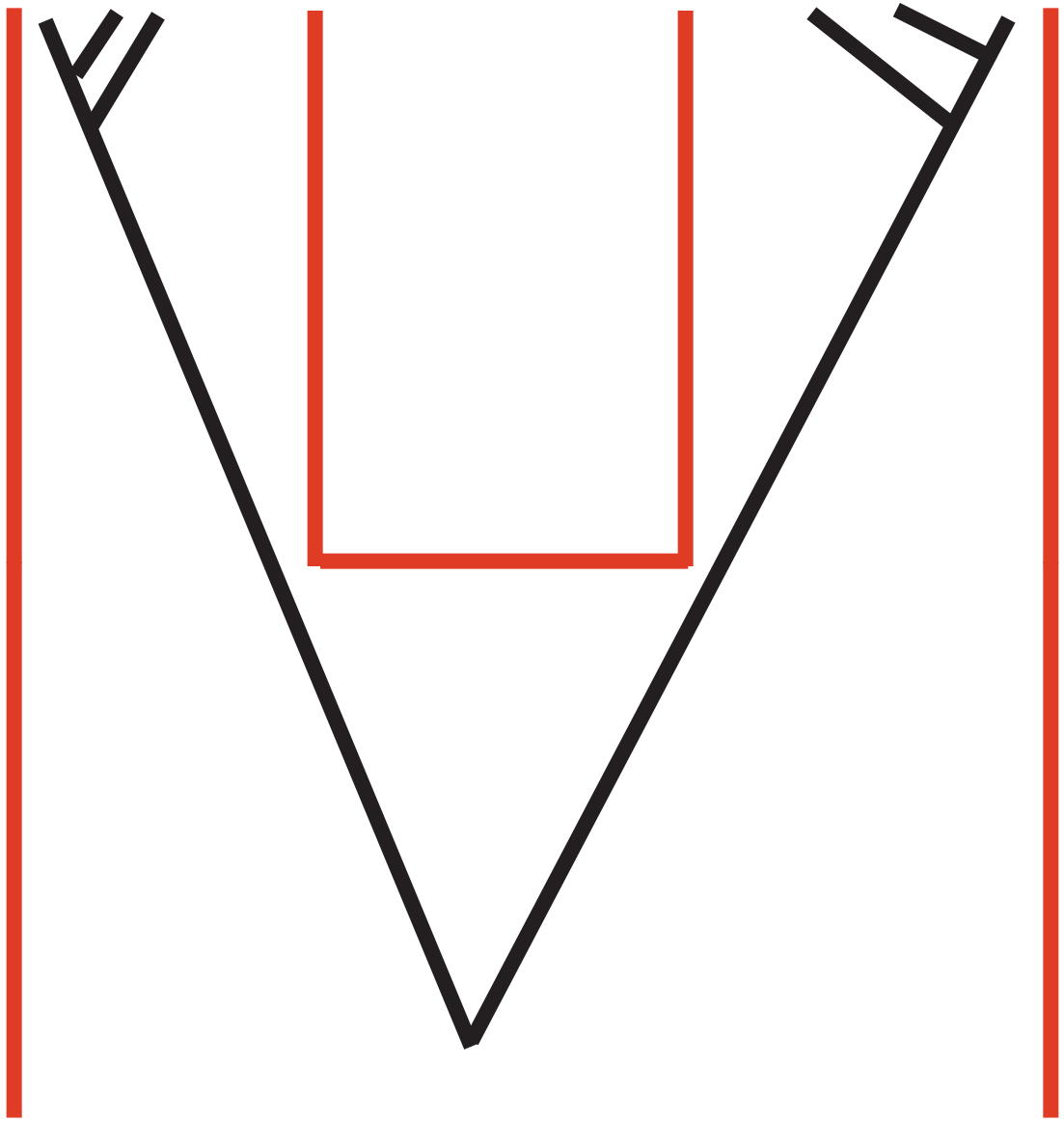
"Typical" gene tree with constant population size



"Typical" gene tree with quick expansion of pop. size (or recent selective sweep)



"Typical" gene tree with constant population size



"Typical" gene tree with Pop'n subdivision or balancing selection