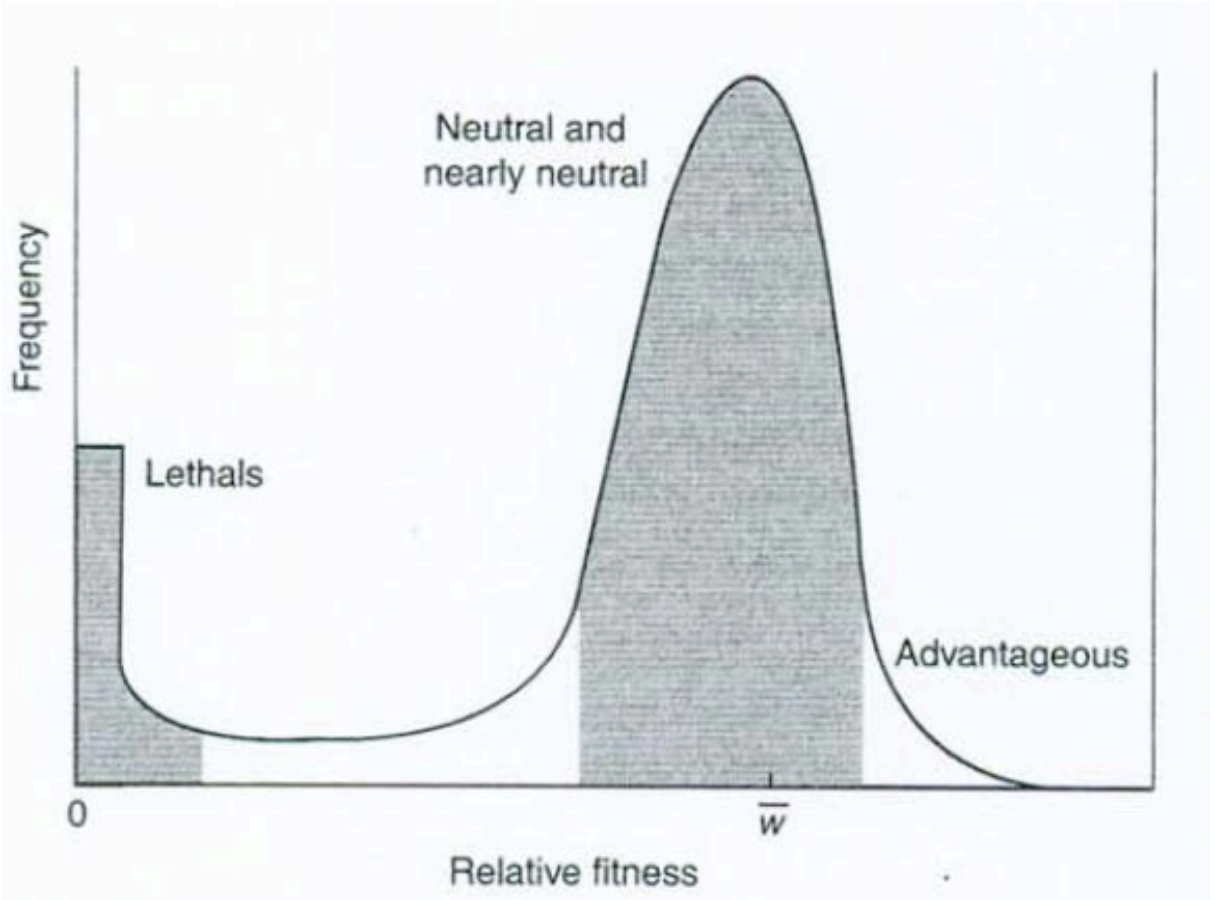
Lecture Handout 18: **Neutral Theory or Neutrality** (Kimura, 1968): Genetic variation is primarily influenced by mutation generating variation, and genetic drift eliminating it. Note: does NOT dispute the role of selection.



Neutral theory is consistent with a **molecular clock** (definition): there is a constant rate of substitution for molecular variants, equal to the locus's mutation rate, and constant over time.

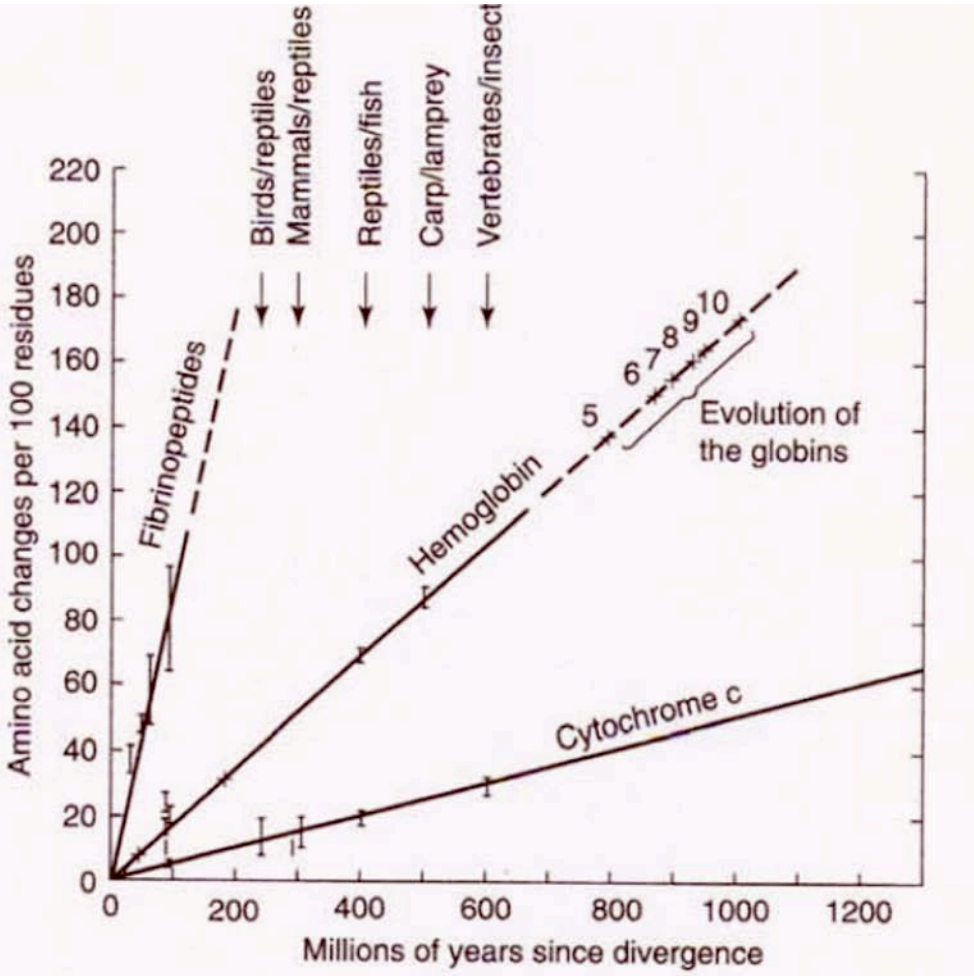


Figure 8.3. The amount of amino acid substitution for three proteins having different rates of substitution. The horizontal axis gives the times since divergence of various organisms in millions of years (after Dickerson, 1971).

The amount of divergence, K, between two nucleotide sequences is: K = 2*u*t

So t = K/(2u) (t is number of generations)

And the expected time between neutral substitutions is 1/*u*

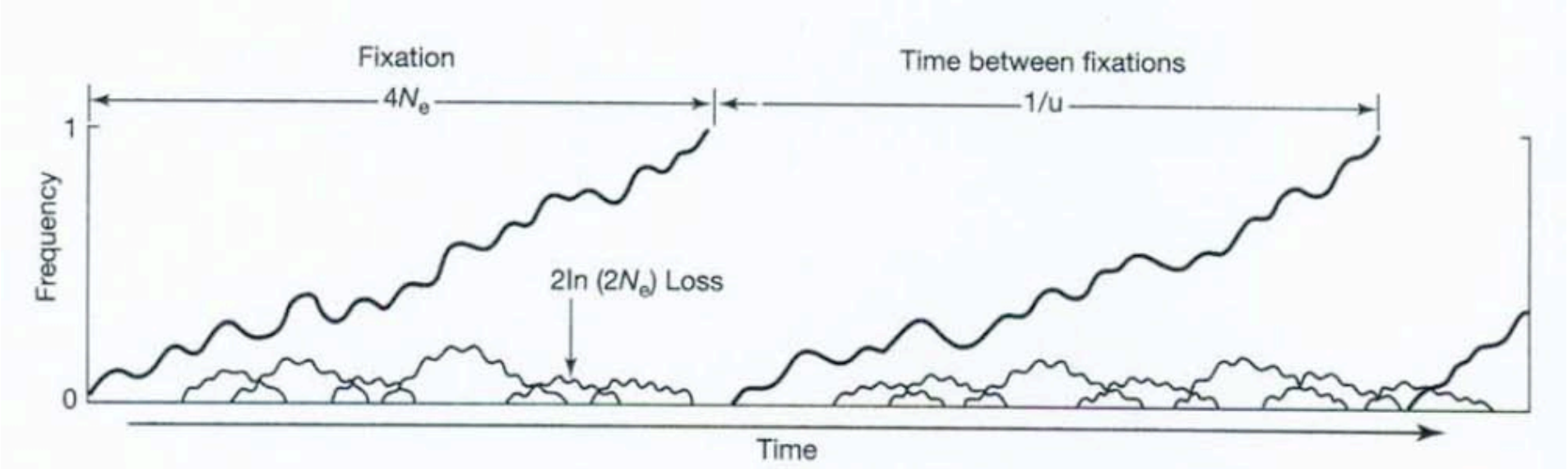
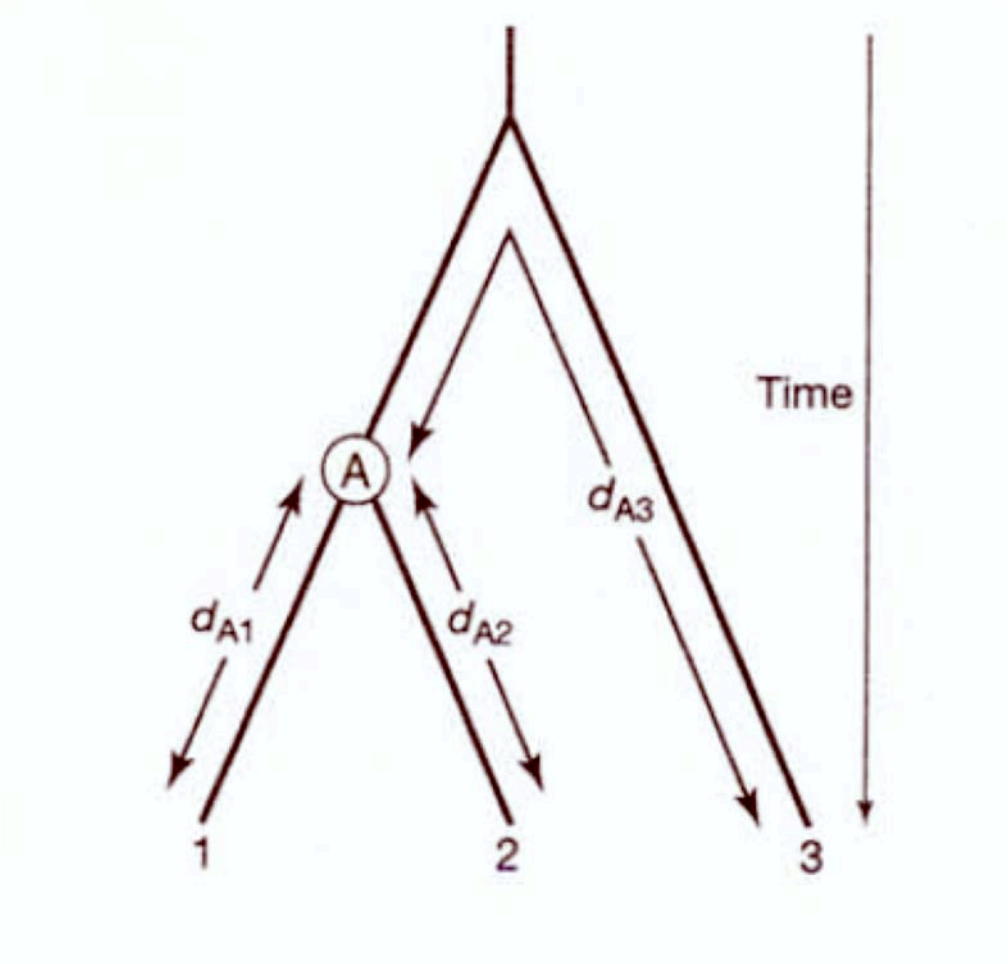
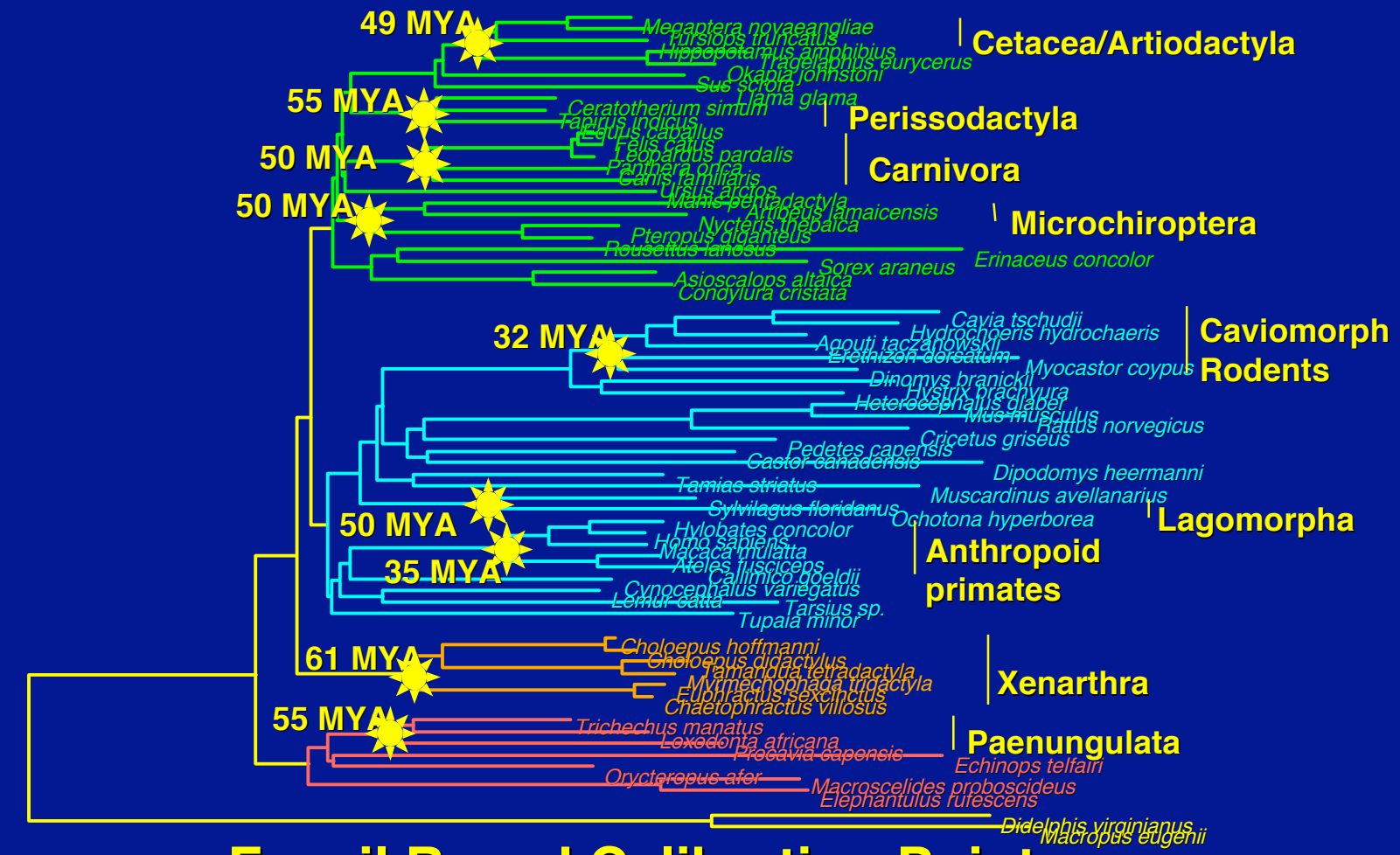


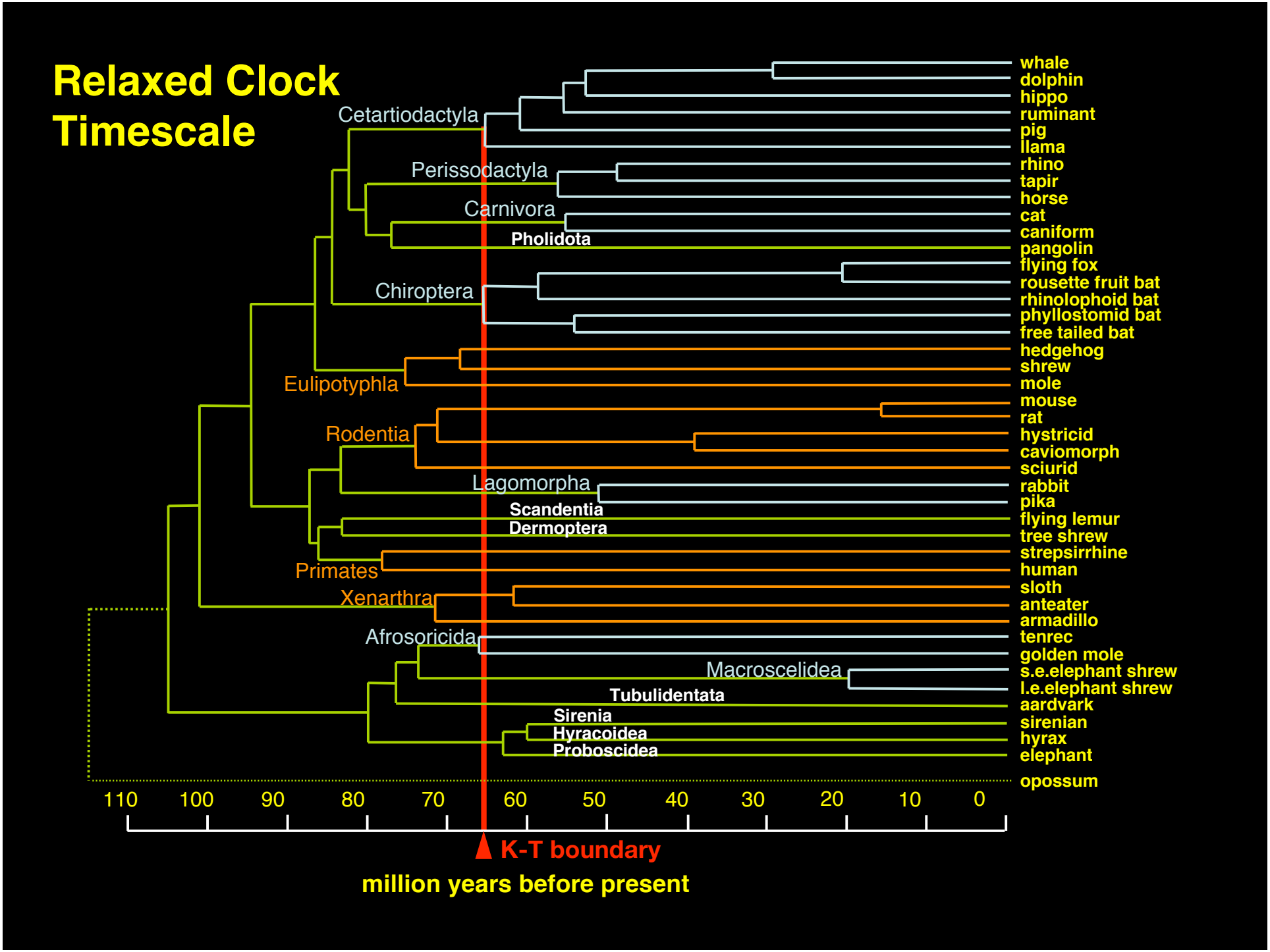
Figure 8.2. The change in frequency of new neutral mutants over time (after Kimura, 1983). The two mutations that eventually become fixed are represented by thick lines, whereas the many mutations that are lost are represented by thin lines. In general, the time is smallest for loss of mutants, larger for fixation of mutants, and longest for the time between fixation of mutants, with the expected values a function of N_e and u.

Relative rate tests: determine whether different lineages are evolving at the same rate (previously required to estimate molecular divergence dates)



Recent Bayesian methods can conduct molecular date estimates for lineages with different evolutionary rates.




Relaxed Clock Timescale
Cetartiodactyla
Perissodactyla
Carnivora
Pholidota
Chiroptera
Eulipotyphla
Rodentia
Lagomorpha
Scandentia
Dermoptera
Primates
Xenarthra
Afrosoricida
Macroscelidea
Tubulidentata
Sirenia
Hyracoidea
Proboscidea
whale
dolphin
hippo
ruminant
pig
llama
rhino
tapir
horse
cat
caniform
pangolin
flying fox
rousette fruit bat
rhinolophoid bat
phyllostomid bat
free tailed bat
hedgehog
shrew
mole
mouse
rat
hystricid
caviomorph
sciurid
rabbit
pika
flying lemur
tree shrew
strepsirrhine
human
sloth
anteater
armadillo
tenrec
golden mole
s.e.elephant shrew
l.e.elephant shrew
aardvark
sirenian
hyrax
elephant
opossum
110
100
90
80
70
60
50
40
30
20
10
0
K-T boundary
million years before present

Nearly neutral model (Ohta, 1973): rate of substitution increases with increasing population size, since detrimental variants are effectively neutral where $s < 1/(2N)$

Tests of selection (or neutrality):

1. Ewens-Watterson test: based on calculation of theta (expected number of effective alleles less one), θ , defined as $4N_e u$

Compares expected homozygosity under Hardy-Weinberg proportions to the equilibrium homozygosity under neutral theory. (Note: NOT the same as observed vs expected)

Ewens-Watterson test: Compares expected homozygosity under Hardy-Weinberg proportions to the equilibrium homozygosity under neutral theory. (Note: NOT the same as observed vs expected)

TABLE 8.5 Allele frequencies and the observed Hardy–Weinberg homozygosity, f_e, from three different hypothetical samples with four alleles and a sample size of 200.

	Allele frequencies					
Sample	A_1	A_2	A_3	A_4	f_e	*Comments*
1	0.35	0.30	0.20	0.15	0.275	Too even
2	0.76	0.17	0.06	0.01	0.610	Neutrality
3	0.965	0.025	0.005	0.005	0.932	Too uneven

Tests of selection (or neutrality):

2. Tajima test: also based on calculation of theta (expected number of effective alleles less one), θ , defined as $4N_e u$

Calculates θ using the nucleotide diversity π to estimate variation; and also calculates θ using number of sites segregating, S. Compares them to calculate “Tajima’s D”, which is zero under neutrality.

Tests of selection (or neutrality):

3. Hudson-Kreitman-Augade, HKA, test: requires data from two or more genetic regions in two or more species. Compares within-species variation and between-species divergence for multiple regions. Under neutrality, levels of variation should be similar across genetic regions.

Tests of selection (or neutrality):

4. McDonald-Kreitman (MK) Test: Similar to HKA test (ie uses within- and between-species comparisons) but examines synonymous to non-synonymous fixed differences between species, along with synonymous to non-synonymous polymorphisms within species. Under neutrality, both ratios should be similar.

Tests of selection (or neutrality):

5. Synonymous versus silent substitutions. Tests use either the rate of synonymous substitution (d_S) and rate of non-synonymous substitutions (d_N); or the number of synonymous substitutions per synonymous site (K_S) and the number of non-synonymous substitutions per non-synonymous site (K_N or K_A)

Purifying selection $d_N < d_S$ is more common than Darwinian selection, eg MHC $d_S < d_N$

Codon bias: unequal usage of synonymous codons is common. Codon bias in a species is positively correlated with expression levels for a gene, and with abundance of the tRNA for the codon.

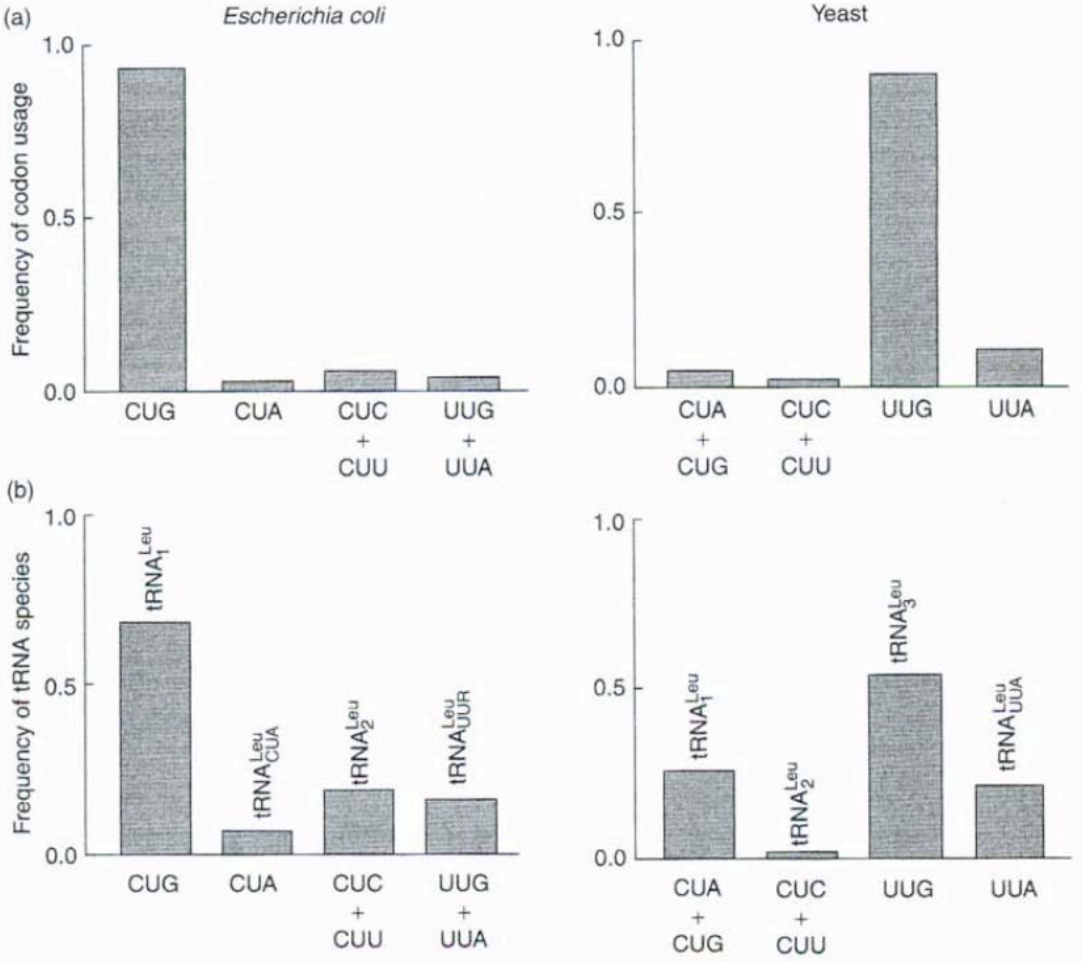


Figure 8.10. (a) The frequencies of codon usage for leucine codons in highly expressed genes and (b) the frequencies of the different leucine tRNA species in *E. coli* (left) and yeast (right) (Li and Graur, 1991).

Coalescence: the point at which common ancestry for two alleles occur.

Model: only genetic drift, not selection, assumed to affect the alleles. Sequences assumed to diverge by mutation.

For DNA sequences, the generation of gene trees allows the **most recent common ancestor (MRCA)** of different sequences to be determined

Examples:

allelic ancestry

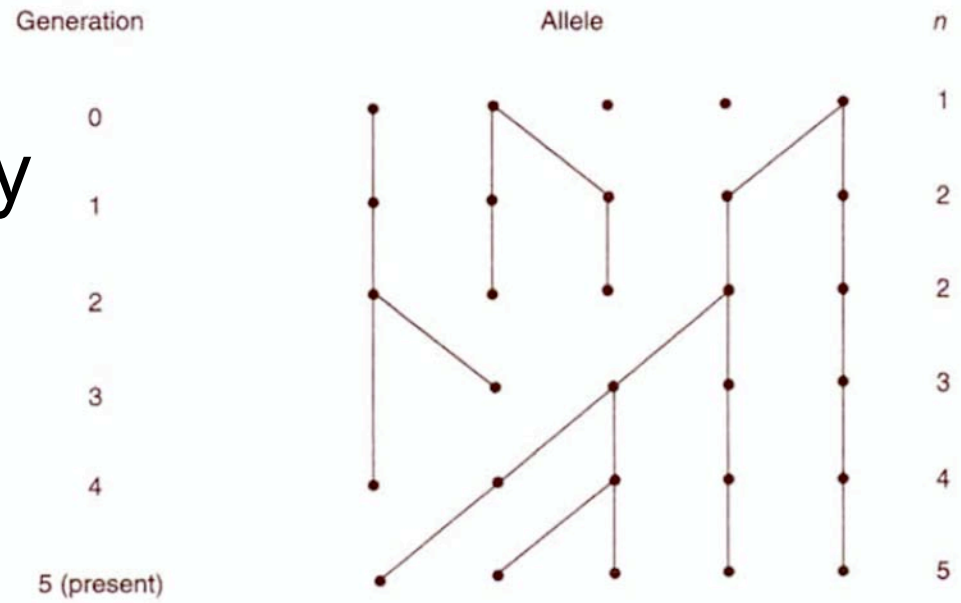
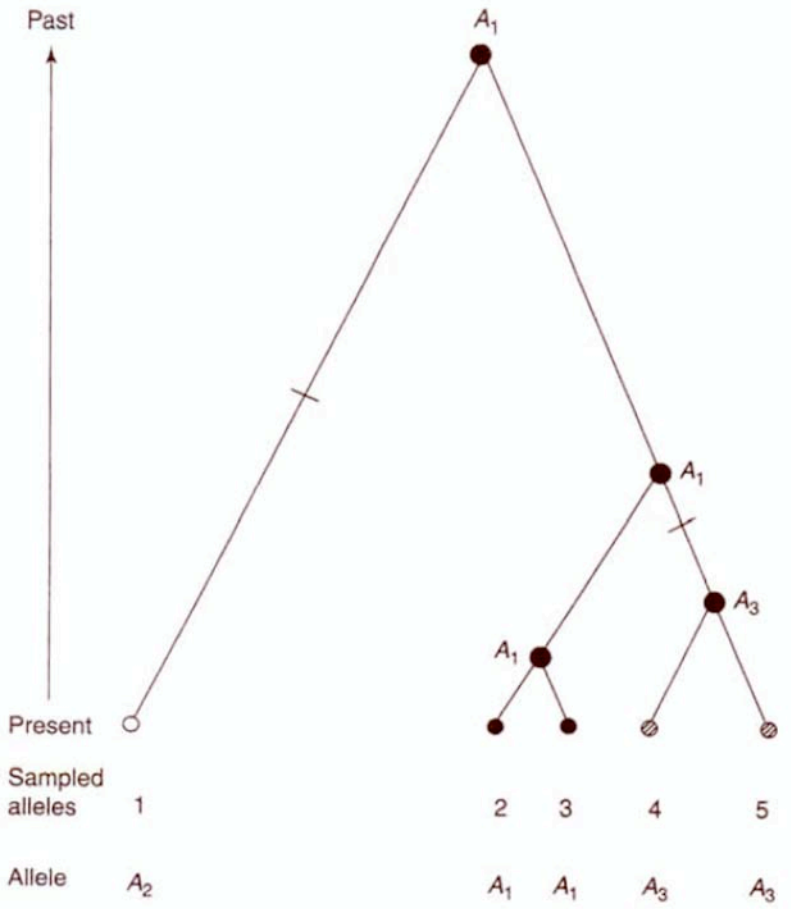
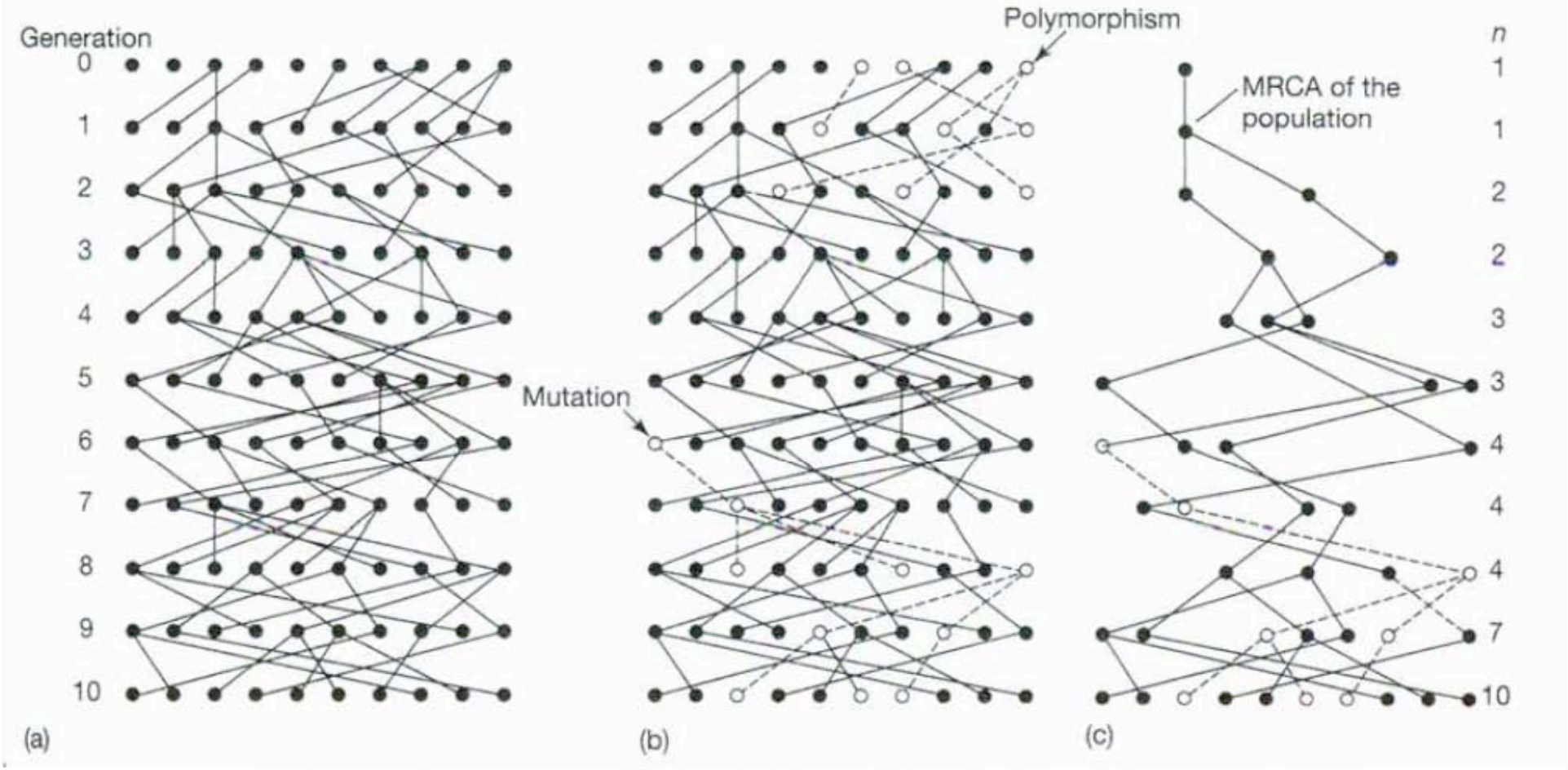


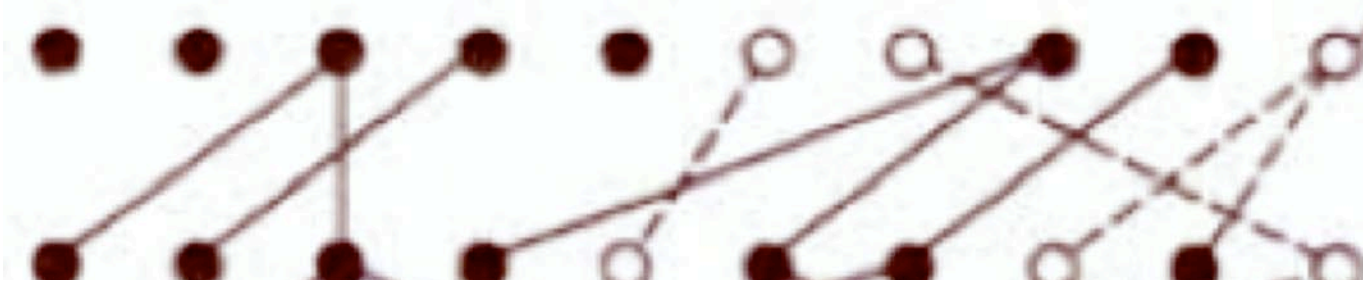
Figure 8.12. A hypothetical example illustrating the ancestry of five alleles sampled in the present generation. If we go forward in time, top to bottom, then we see the effects of genetic drift resulting in the fixation of a single allele in generation 5. If we go back in time, bottom to top, then we see the coalescence of the five sampled alleles in the right-hand ancestral allele in generation 0. Here n is the number of ancestral alleles from which the five alleles sampled in the present generation are descended.

gene tree



Genetic drift and mutation across a genealogy





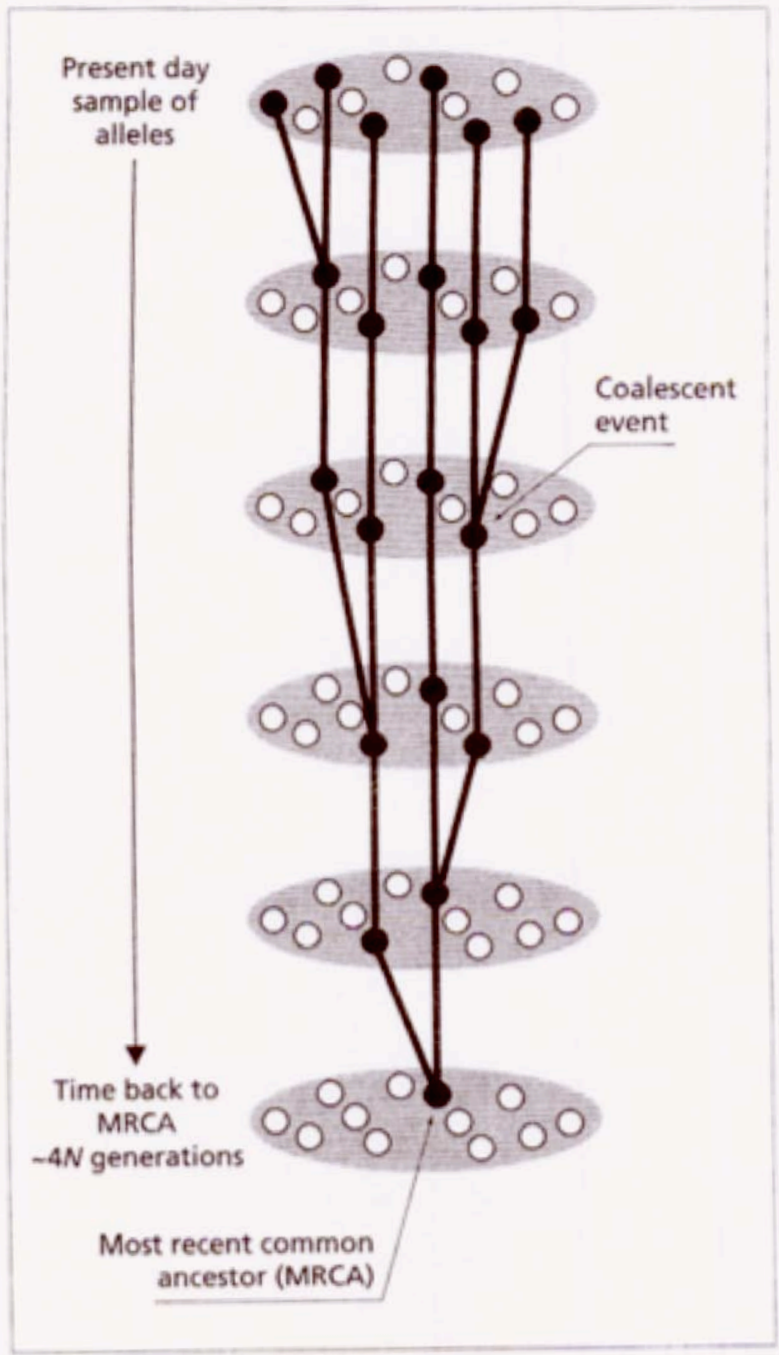


Fig. 4.16 A gene genealogy linking a sample of six individuals (black circles) from a larger population observed at different time points. We can trace the lineages connecting these individuals back through time to the point when they last shared a common ancestral allele—coalescent events. The most recent common ancestor (MRCA) for all individuals in the sample will be reached after an average of $4N$ generations under genetic drift (although this time is reduced if the number of individuals sampled is very small). Compare this with Fig. 4.6 which describes the fixation of alleles by genetic drift.

Coalescence can be used to infer evolutionary events

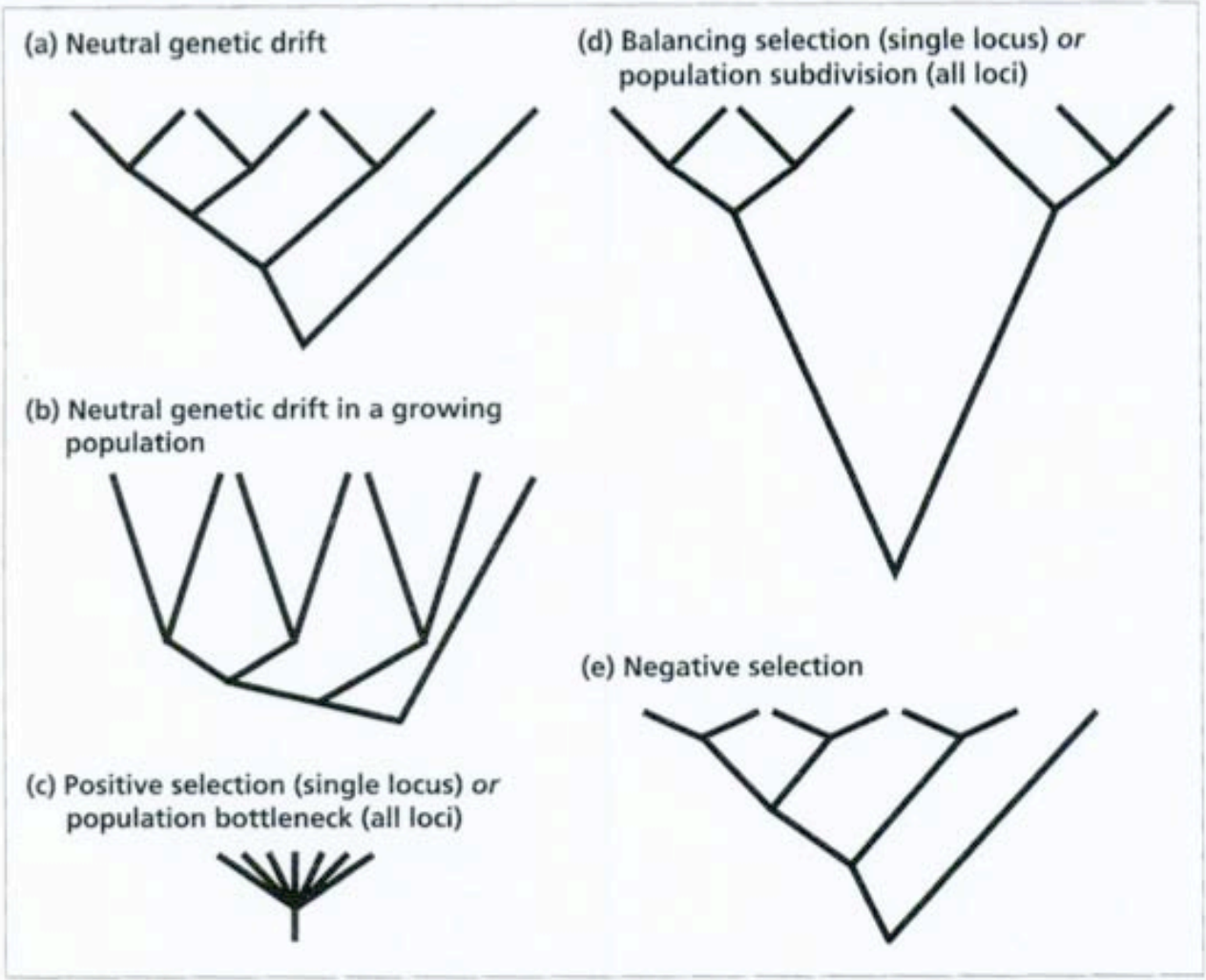


Fig. 4.17 How evolutionary processes leave different signatures in the shape of gene genealogies and the distribution of coalescence times.

Coalescence and neutral theory has been used to infer effective population size. Ne = θ / (4u)

For mtDNA: Nef = θ / (2u)

TABLE 8.15 The level of mtDNA diversity (θ), the estimated generation length (T), and the estimated female effective population sizes (N_{ef}) for three species of baleen whales (Roman and Palumbi, 2003). From these estimates, the estimated census number before hunting is extrapolated and is compared with current census estimates.

				Estimated census numbers	
Species	θ	T	N_{ef}	Before hunting ($N_e \times 7$)	Current
Humpback whale	0.0216	18	34,300	240,100	10,700
Fin whale	0.0430	25	49,100	344,000	56,000
Minke whale	0.0231	17	38,800	271,600	215,500