Mutation Models, Neutral Theory

TABLE 3

Proportion of loci, out of 18, polymorphic and proportion of the genome estimated to be heterozygous in an average individual for each population studied

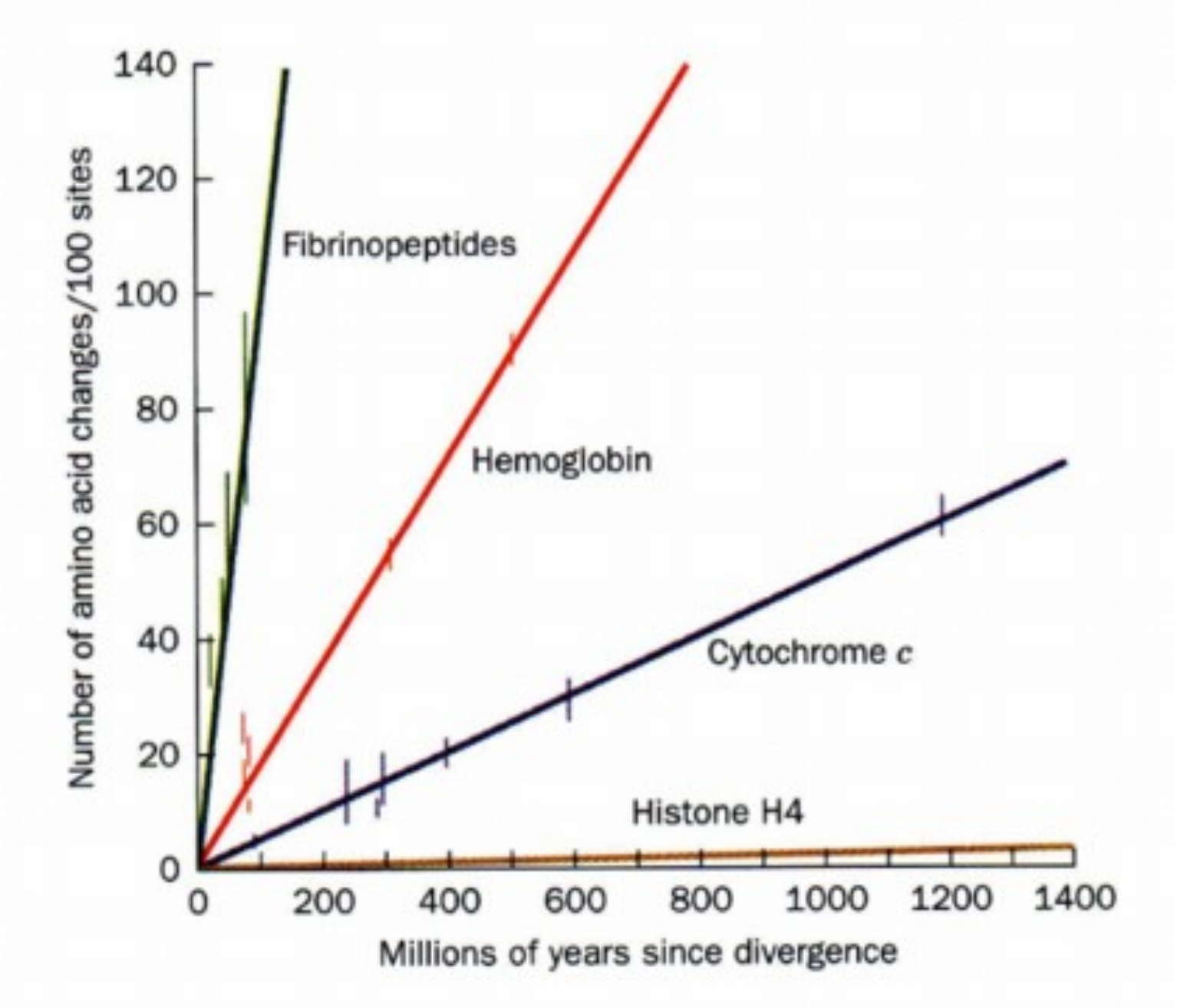
Population	No. of loci polymorphic	Proportion of loci polymorphic	Proportion of genome heterozygous per individual	Maximum proportion of genome heterozygous
Strawberry Canyon	6	.33	.148	.173
Wildrose	5	.28	.106	.156
Cimarron	5	.28	.099	.153
Mather	6	.33	.143	.173
Flagstaff	5	.28	.081	.120
Average		.30	.115	.155

Lewontin and Hubby 1966

Table 1.3
The heterozygosity for 71 allozyme loci in humans (Harris and Hopkinson, 1972).

Locus	*Heterozygosity (H)*
51 monomorphic loci	0.00
Peptidase C	0.02
Peptidase D	0.02
Glutamate-oxaloacetate transaminase	0.03
Leucocyte hexokinase	0.05
6-Phosphogluconate dehydrogenase	0.05
Alcohol dehydrogenase-2	0.07
Adenylate kinase	0.09
Pancreatic amylase	0.09
Adenosine deaminase	0.11
Galatase-1-phosphate uridyl transferase	0.11
Acetyl cholinesterase	0.23
Mitochondrial malic enzyme	0.30
Phosphoglucomutase-1	0.36
Peptidase A	0.37
Phosphoglucomutase-3	0.38
Pepsinogen	0.47
Alcohol dehydrogenase-3	0.48
Glutamate-pyruvate transaminase	0.50
RBC acid phophatase	0.52
Placental alkaline phosphatase	0.53

Molecular Clock



Irreversible Mutation

- 1 locus, 2 alleles
  - A, a (frequencies p, q)
- Let μ = A to a mutation rate (per generation)
  - Pr(A mutates to a) = μ

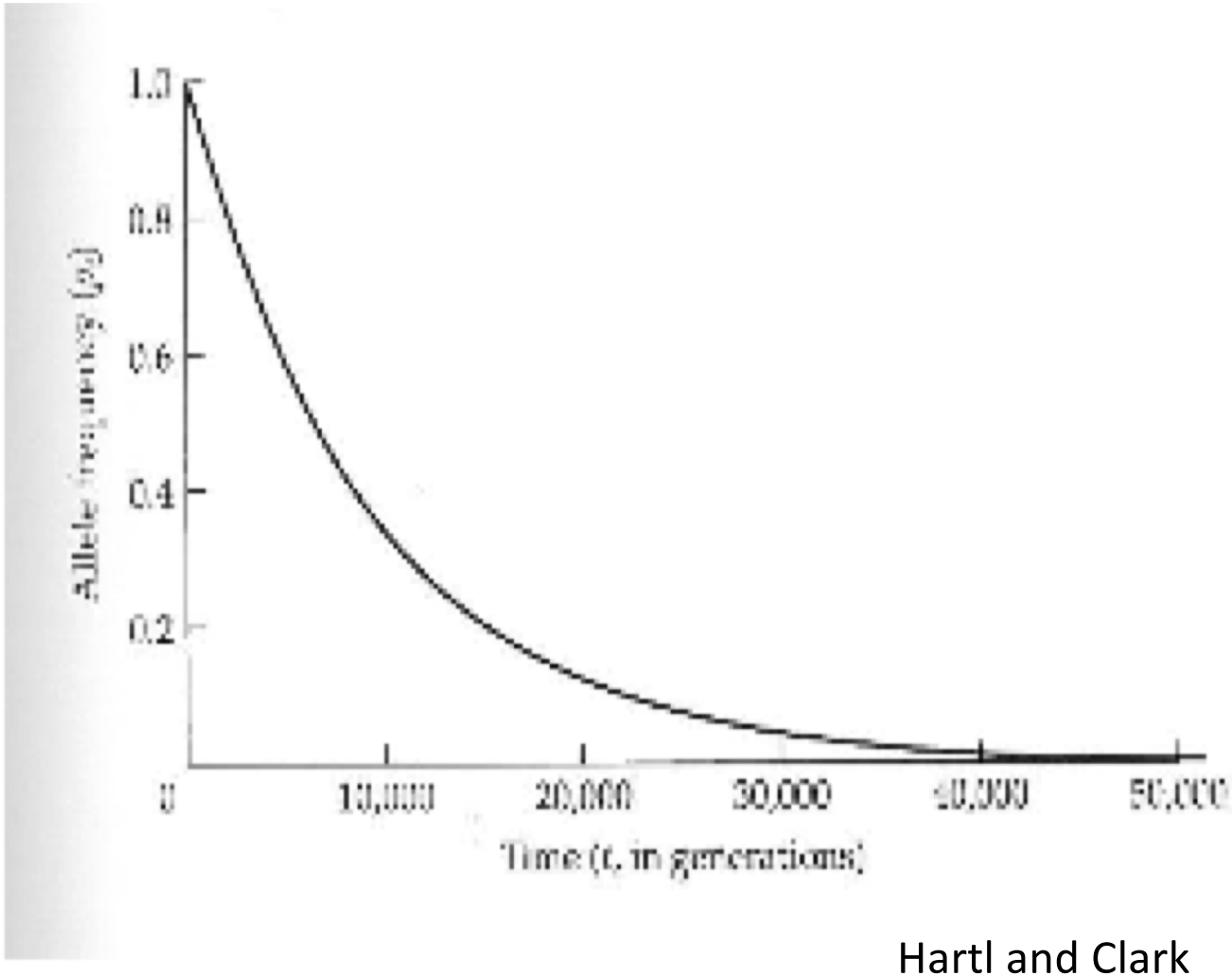
$$p_t = p_{t-1}(1-\mu)$$

$$p_t = p_{t-2}(1-\mu)^2$$

$$p_t = p_0(1-\mu)^t$$

What does p_t approach as $t \to \infty$?

Irreversible Mutation



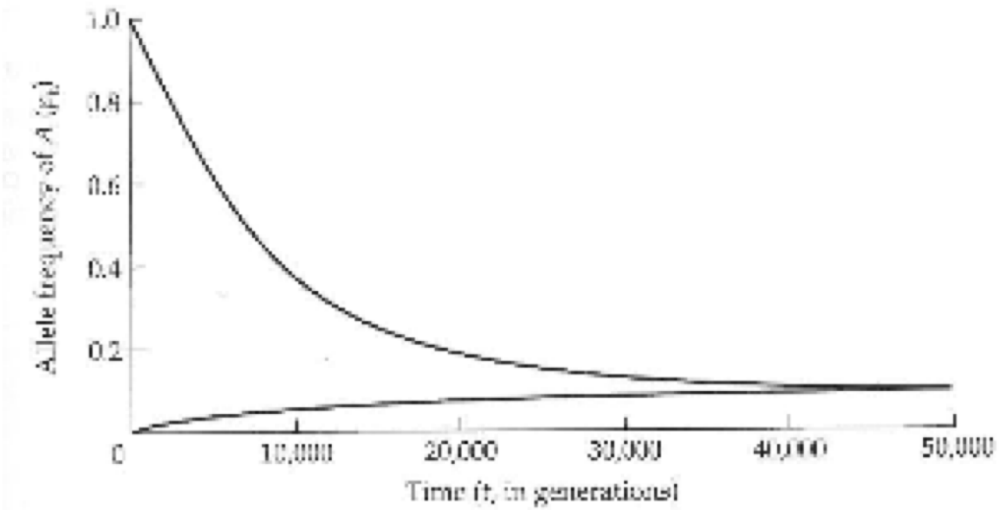
Hartl and Clark

Reversible Mutation

- 1 locus, 2 alleles
  - A, a (frequencies p, q)
- Let μ = A to a mutation rate (per generation)
- Let ν = a to A mutation rate (per generation)

$$p_t = p_{t-1}(1-\mu) + (1-p_{t-1})\nu$$

$$\hat{p} = \frac{\nu}{\mu + \nu}$$



Hartl and Clark

Summary so far

- Random Mating
  - Discrete Generations
- Hardy Weinberg
  - Infinite pop size

Allele frequency constant, genetic variation maintained

- Wright Fisher
  - Mutational model

Allele frequency changes, genetic variation lost

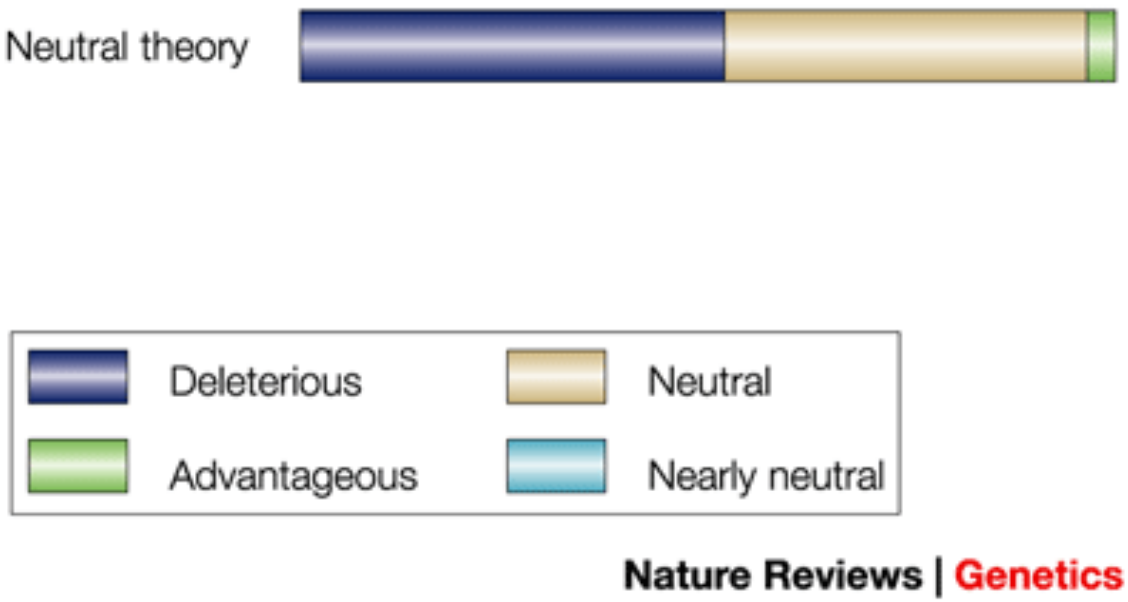
Allele frequency changes SLOWLY, genetic variation lost, maintained

- Neutral model

Motoo Kimura

Neutral Theory

- Intersection of mutation with drift
- Most mutations selectively neutral
- Drift determines allele frequencies

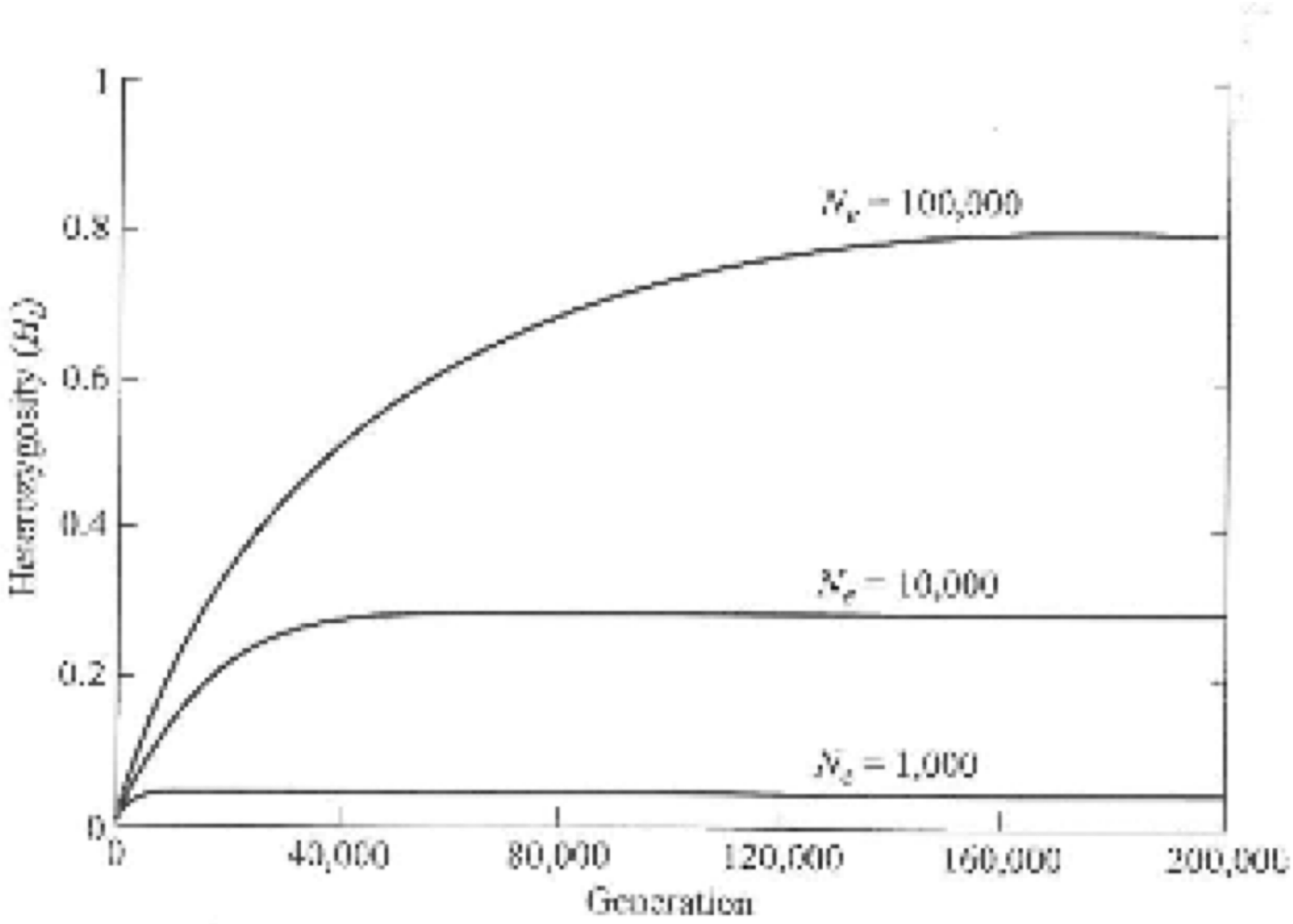


Infinite Alleles Model

- Each mutation creates **new** allele
  - 2 alleles with identical sequence MUST be IBD
- To measure homozygosity, we can measure Pr(IBD)

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1}$$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N}(1-\mu)^2 + \left(1 - \frac{1}{2N}\right)(1-\mu)^2 F_{t-1}$$


1
0.8
0.6
0.4
0.2
0
Heterozygosity (H_t)
N_e = 100,000
N_e = 10,000
N_e = 1,000
0
40,000
80,000
120,000
160,000
200,000
Generation

Hedrick

Infinite Alleles Model

- Each mutation creates **new** allele
  - 2 alleles with identical sequence MUST be IBD
- To measure homozygosity, we can measure Pr(IBD)

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N}(1-\mu)^2 + (1 - \frac{1}{2N})(1-\mu)^2 F_{t-1}$$

$$\hat{F} = \frac{1}{1 + 4N_e\mu} \qquad \hat{H} = 1 - \hat{F} = \frac{4N_e\mu}{1 + 4N_e\mu} \qquad \theta = 4N_e\mu$$

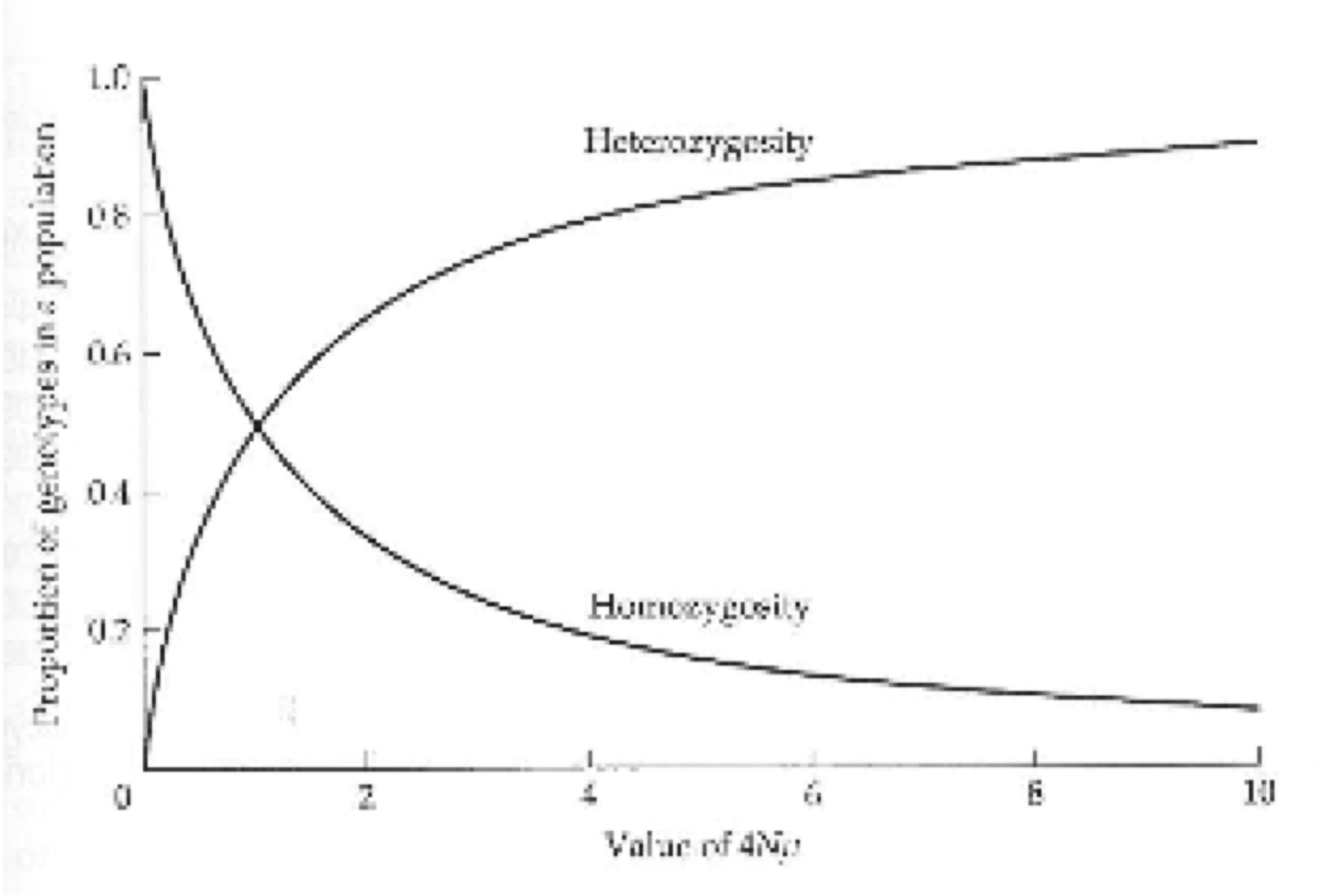
Infinite Alleles Model

- Each mutation creates **new** allele
  - 2 alleles with identical sequence MUST be IBD
- To measure homozygosity, we can measure Pr(IBD)

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N} + (1 - \frac{1}{2N})F_{t-1}$$

$$\Pr(\text{IBD}_t) = F_t = \frac{1}{2N}(1-\mu)^2 + (1 - \frac{1}{2N})(1-\mu)^2 F_{t-1}$$

$$\hat{F} = \frac{1}{1+\theta} \qquad \hat{H} = 1 - \hat{F} = \frac{\theta}{1+\theta} \qquad \theta = 4N_e\mu$$



Hartl and Clark

At equilibrium:

- Steady-State under infinite alleles:
  - H= $\theta/1+\theta$
  - # alleles stationary

$$E(k) = \sum_{i=1}^{n} \frac{\theta}{\theta + (i-1)}$$

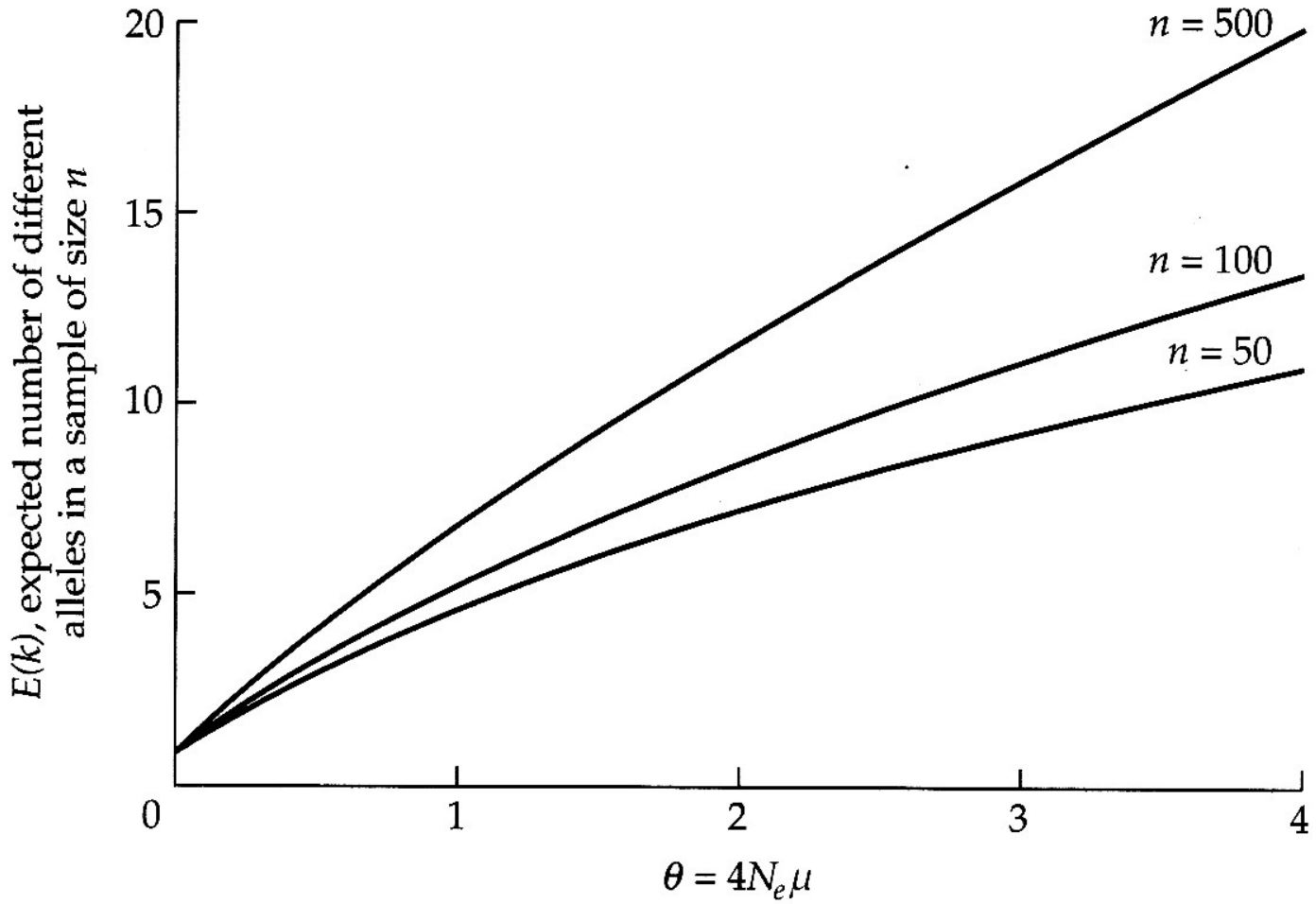


FIGURE 4.8 Relation between θ, the expected number of alleles, and the sample size according to the Ewens sampling theory of a population in steady state under the infinite-alleles model of neutral mutation.

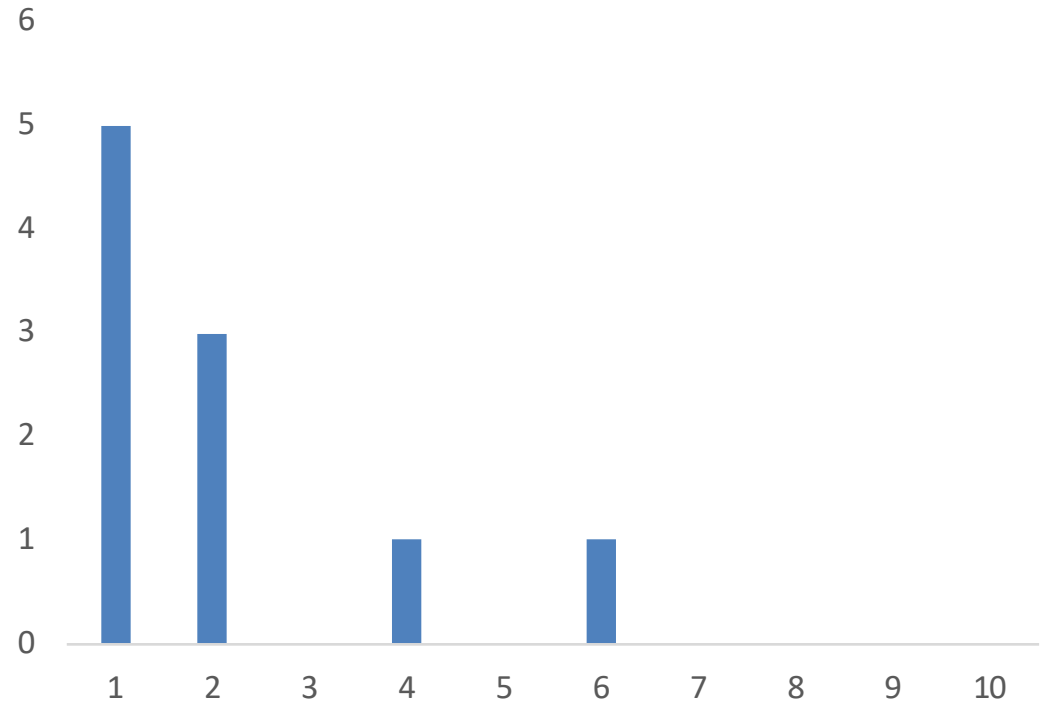
Hartl and Clark

At equilibrium

- Stationary distribution of allele frequencies
  - Allele frequency spectrum
    - (Unique) alleles 1…*k*
    - Allelic configuration (frequencies p_1, p_2…p_k)
      - Allele frequency spectrum

Allele Frequency Spectrum

- Sample size $n = 20$, $k = 10$ unique alleles
  - $p_1 = 6$
  - $p_2 = 4$
  - $p_3 = p_4 = p_5 = 2$
  - $p_6 = p_7 = p_8 = p_9 = p_{10} = 1$



Implications

- If we know *n*, θ , we can write down *E(k)*
- If we know *n*, *k*, we can generate expected allele frequency distribution under neutrality
- We can use neutral expectations as null models to test for deviations from neutrality

Summary

- Neutral model is intersection of mutation, drift
- Mutations introduced through a population
- Once there, alleles are subject to drift and are ultimately fixed or lost
- At equilibrium there is a balance between drift and mutation
  - Every allele introduced by mutation is exactly balanced by allelic loss through drift

Controversial implications

- Allele frequency changes driven by drift, not selection
- Most polymorphisms have nothing to do with adaptation

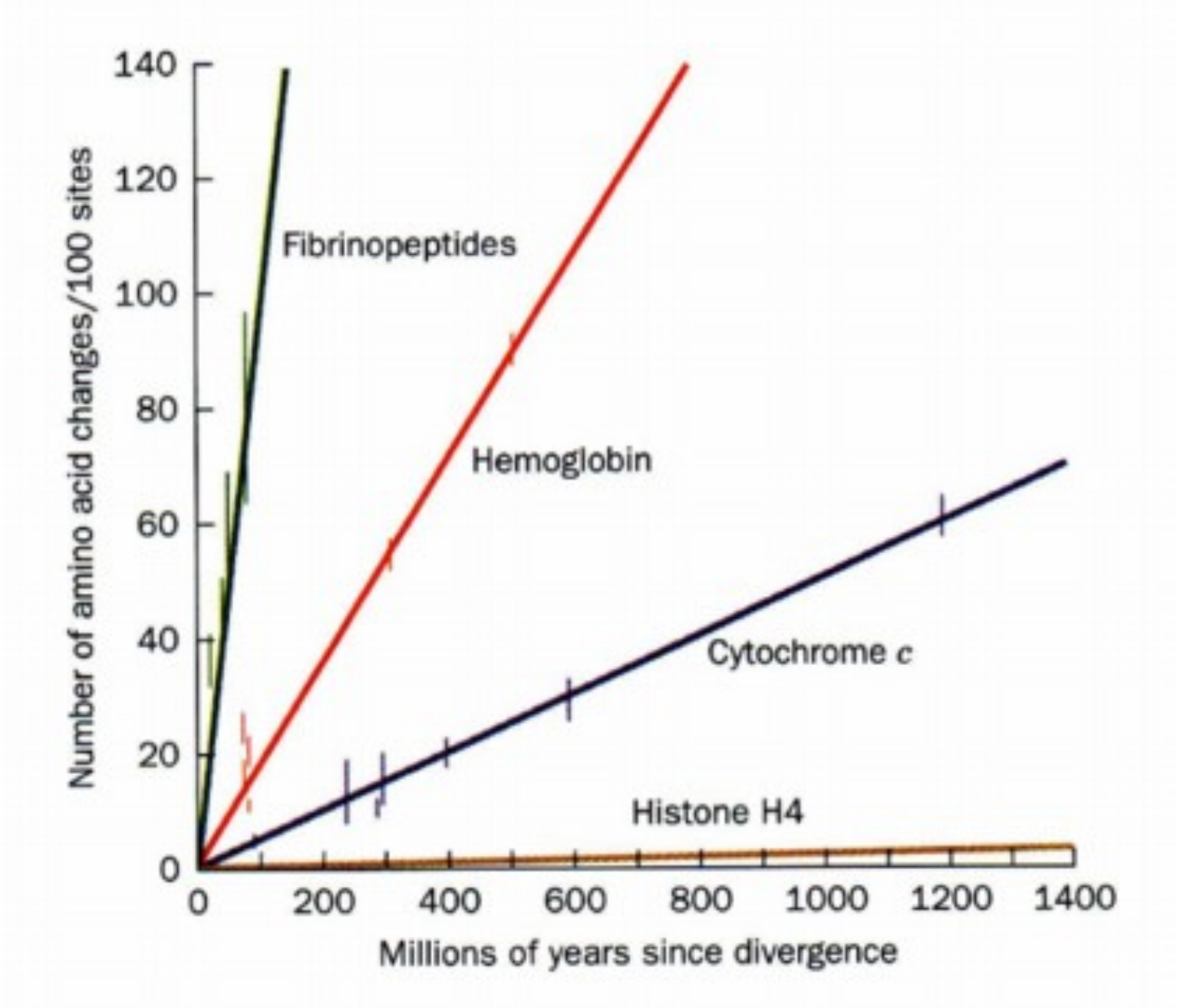
Molecular Evolution

$$k = 2N_e\mu \Pr(\textit{fixation})$$

$$k = 2N_e\mu\left(\frac{1}{2N}\right)$$

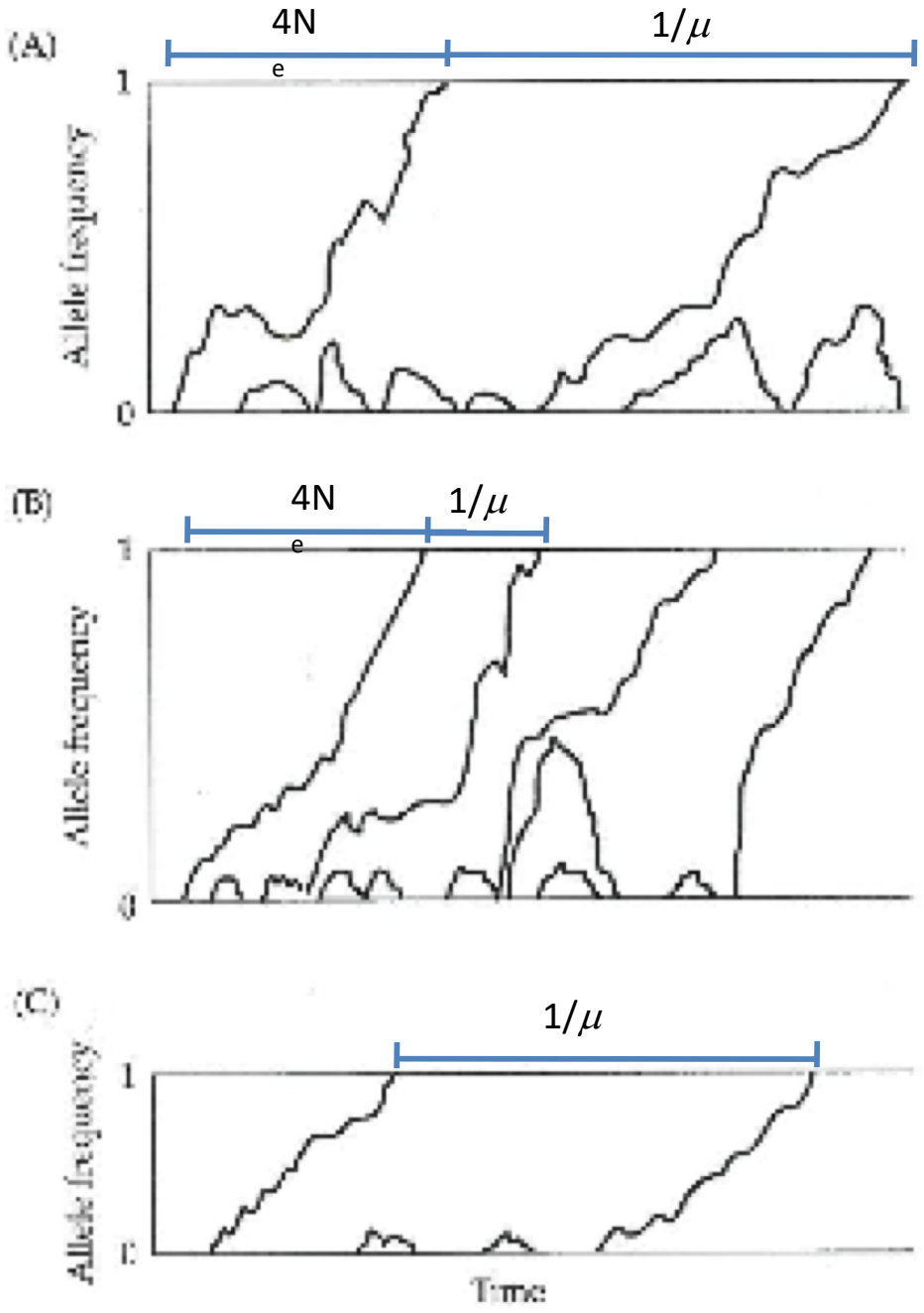
$$k = \mu$$

Molecular Clock



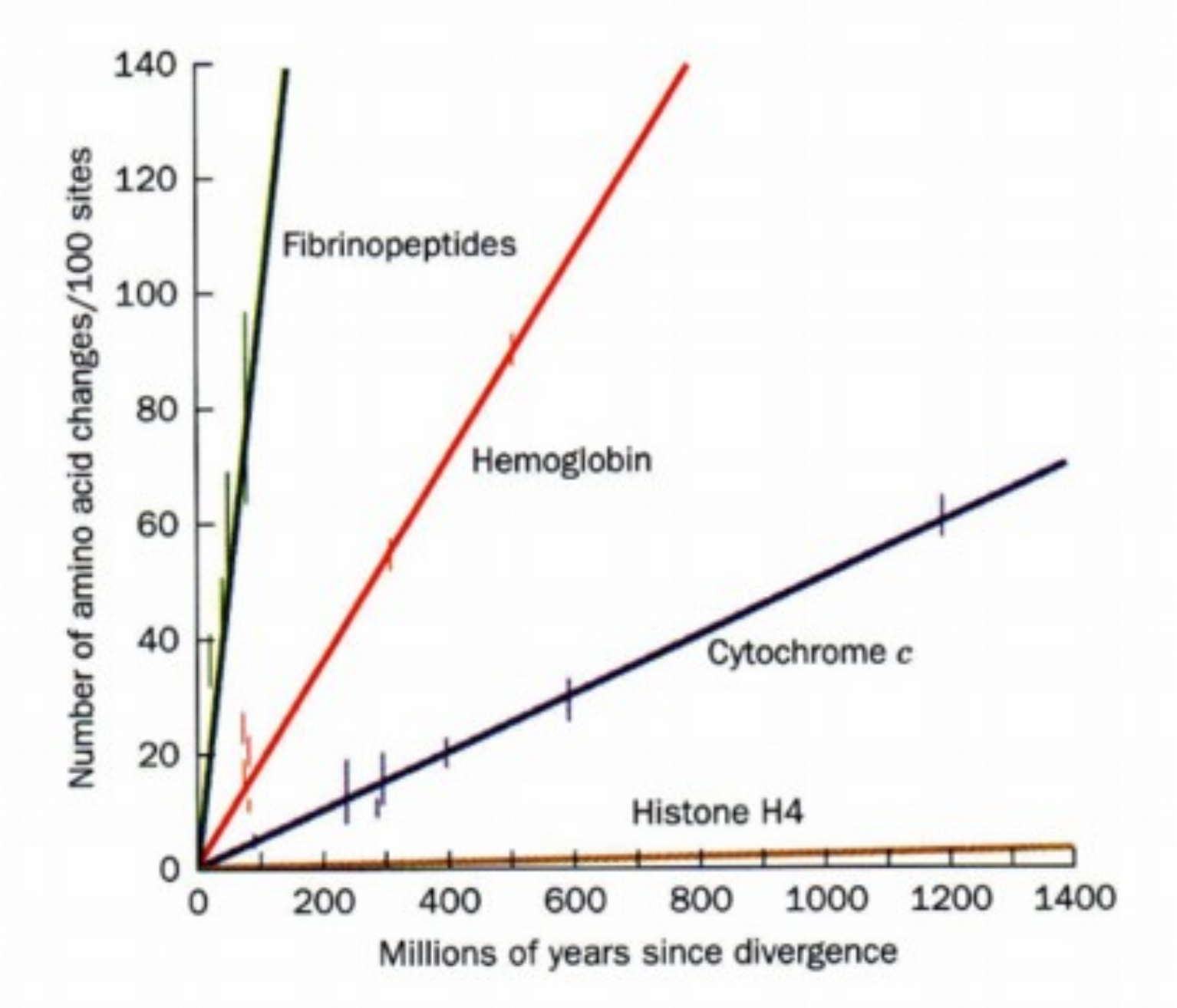
Molecular Evolution

- $k = \mu$
- Expected time b/t substitutions is $1/\mu$
- $K = 2\mu t$
- For $p = 1/2N$, $t_{fix} \approx 4N_e$
- For $p = 1/2N$, $t_{loss} \approx 2\ln(2N_e)$



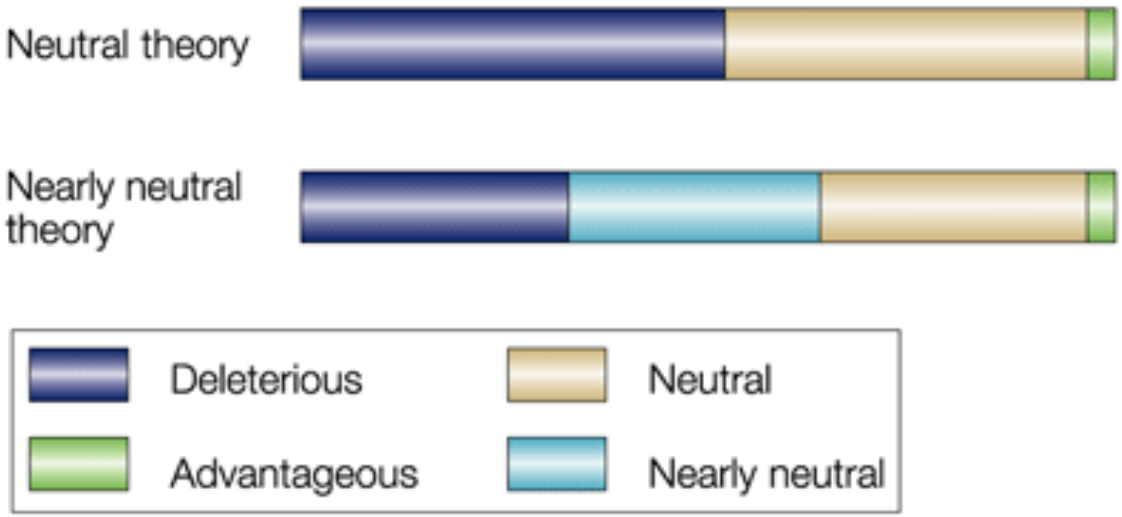
Hartl and Clark

Molecular Clock



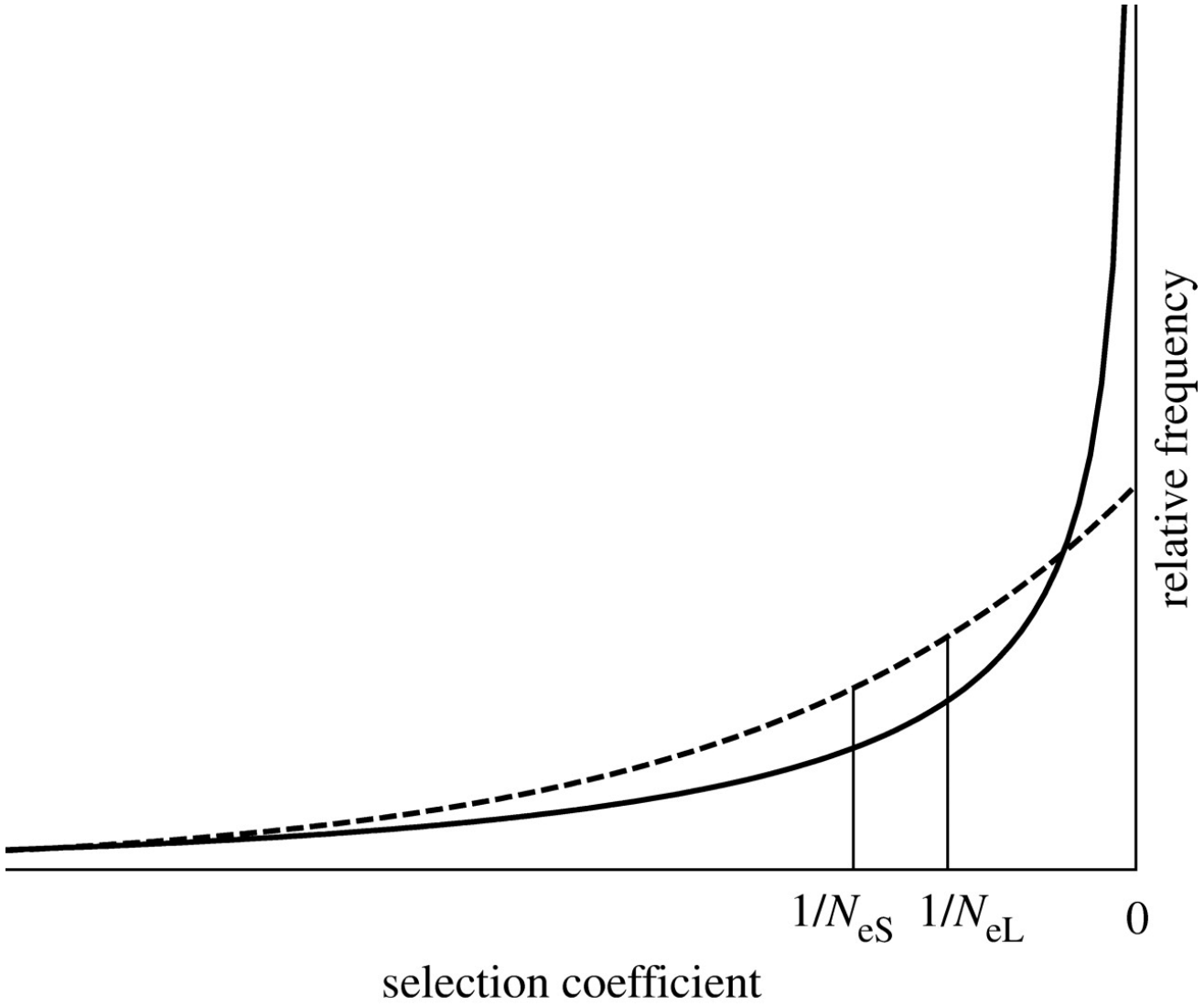
Nearly Neutral Theory

- Considers ‘slightly deleterious’ mutations
  - $0 < |N_e s| < 1$
- Nearly neutral mutations
  - $|N_e s| < 1$



Nature Reviews | Genetics

The distributions of fitness effects modelled by Ohta (1977) (exponential or gamma with β=1, dashed curve) and Kimura (1979) (gamma with β=0.5, solid curve).



Woolfit M Biol. Lett. 2009;5:417-420

biology letters

Molecular Clock