Mutation

PopGen 2011

17.11.2011

Original slides by Outi Savolainen

Mutation

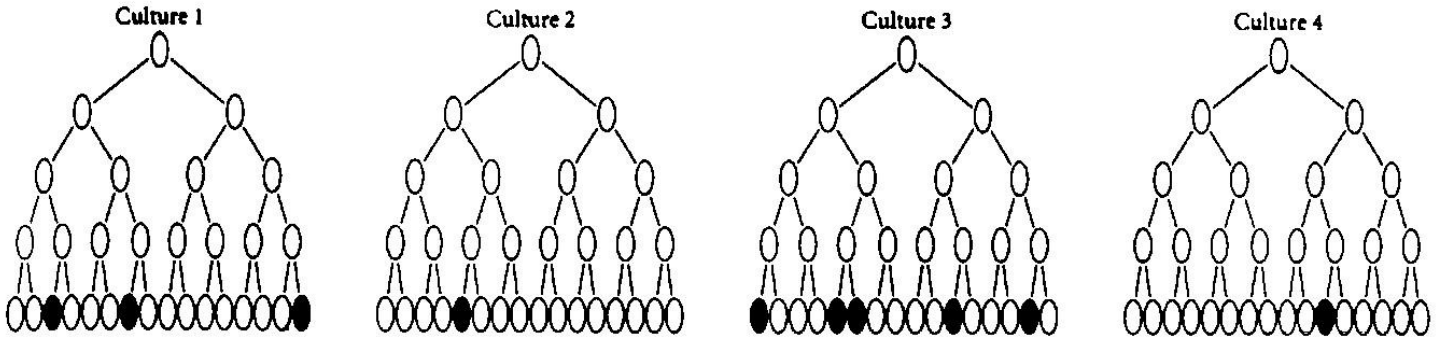
1. Random: spontaneous and induced mutation
2. Nature of mutation: different models
3. Estimating mutation rates, molecular
4. Mutation and back mutation
5. Mutation and drift
6. Estimating genome-wide mutation rates *U* and selective effects
7. Mutation and selection

from Futuyma 1998

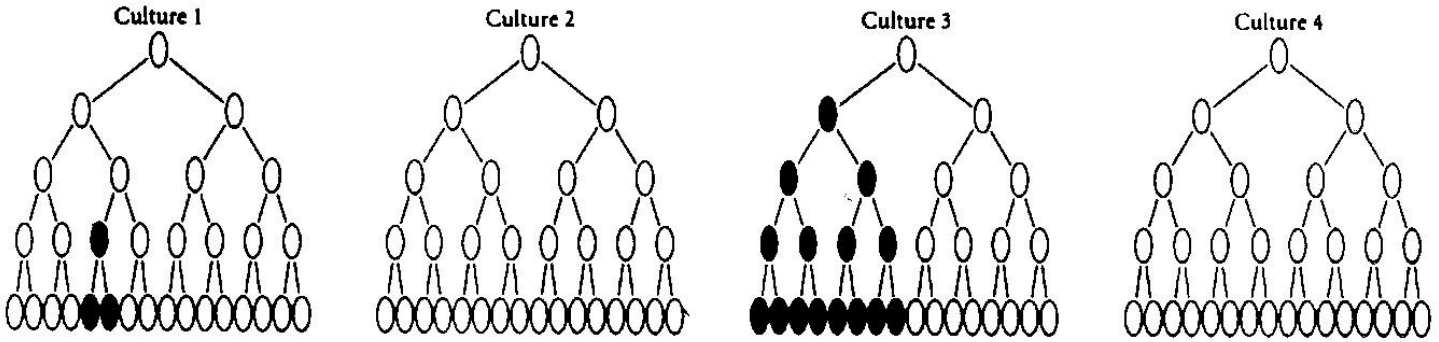
Nature of mutation: mutation is random

FIGURE 10.12 Two hypotheses for the origin of mutations for phage resistance in bacteria. Luria and Delbrück's experiment was designed to determine which hypothesis is correct. In both A and B, the pedigrees of bacteria in several cultures, each derived from a single cell, are shown. Bacteria are exposed to phage in the fourth generation. Shaded cells carry a mutation for phage resistance. It is assumed (and experimentally demonstrated) that a small minority of cells are resistant. (A) Under the hypothesis that exposure to phage directly induces mutations for resistance, the few resistant cells should be distributed at random among the cultures. (The mathematical description of the expected frequency distribution is called the Poisson distribution). (B) Under the hypothesis that mutations for resistance occur at random, before exposure to phage, resistant cells will be "clumped" in certain cultures and absent in all others, because mutations in ancestral cells are inherited by multiple descendants. (After Stent and Calendar 1987.)

(A) Mutation induced by environment



(B) Random mutation



Epigenetic mutation Cubas *et al*. 1999. Nature.

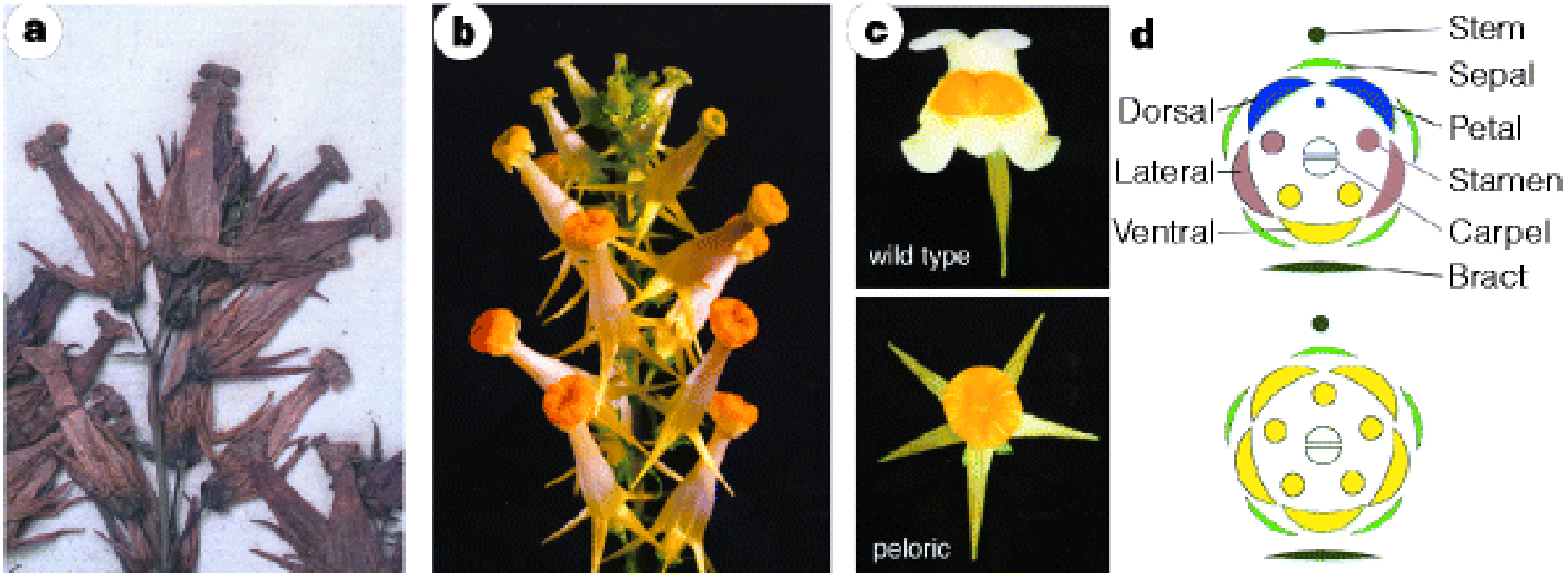


Figure 1 Wild-type and peloric *Linaria vulgaris* flowers. **a**, Original herbarium specimen of peloric *Linaria* inflorescence collected by Linneaus, London. **b**, Peloric *Linaria* inflorescence from a living specimen. **c**, Face view of a wild-type *Linaria* flower compared to a peloric mutant. **d**, Floral diagrams of wild-type (top) and peloric (bottom) flowers showing the relative positions of different organs, with identities indicated by colours: blue (dorsal) brown (lateral) yellow (ventral). The wild-type flower has an axis of dorsoventral asymmetry orientated such that the upper part is nearer the stem whereas the lower part is nearer to the subtending leaf. The peloric flower is radially symmetrical.

-*Lcyc* gene of the mutant is highly methylated, inherited, sometimes reverted.

Mol pop gen mutation models

- *Infinite-allele model (IAM)*

  -consider the entire allele

  -each new mutation generates a new allele, different alleles are not related to each other

  -gene with 1000 bp => 4^{1000} possible alleles

  -molecular basis not necessarily known (e.g. enzyme genes)

  -simplified model $\theta = \frac{4N_e\mu}{4N_e\mu + 1}$

- *Stepwise mutation model (SMM)*

  -special case

  -insertions and deletions, alleles are related, new allele will be similar in length, differing by one or two repeats

  -mutation can give rise to pre-existing alleles: generation of variation and level of heterozygosity less than for IAM

  -microsatellites

- *Infinite-sites model*

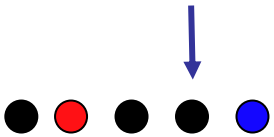
  -assume an infinite number of nucleotide sites, each mutation hits a new site and generates a new allele

  $$\theta = 4N_e\mu$$

- *Finite-sites model*

  -finite number of sites, mutations can hit the same site multiple times

Infinite Allele



-only identity, non-identity is determinable- mutation creates a new allele

Infinite Site

Sequence 1	Sequence 2
…CATGGATCTT…	…CATGGATCTT…
…CATcGATCTT…	…CATGGATCTT…
…CATcGATCTg…	…gATGGATaTT…

-mutation always hits a new position

Finite Site

Sequence 1	Sequence 2
CATGGATCTT	CATGGATCTT
CATcGATCTT	CATGGATCTa
CATgGATCTg	gATGGATaTa

Θ

-mutation changes nucleotide at chosen positions

Mutation Leeflang *et al*. 1999

-26 individuals, >3500 sperm, starting phenotypes 27-62 repeats (15 normal)

-mutation frequency increased with repeat number (82->92%)

-mean change in allele size increased with repeat number

-mitotic mutations

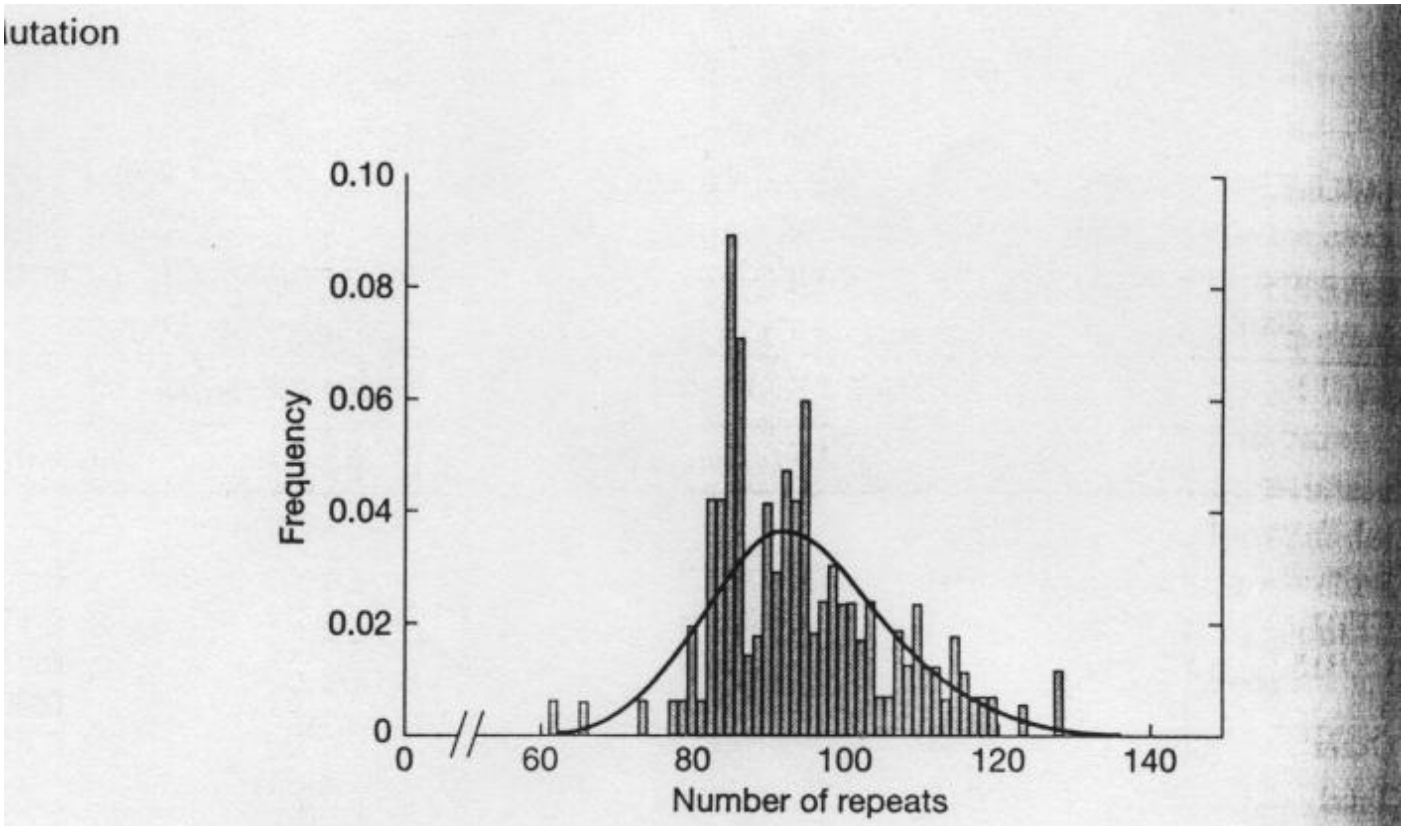


Figure 7.15. The observed frequencies of repeat mutations of different sizes in 168 sperm from an individual with a Huntington disease allele with 62 repeats (the smoothed line gives a theoretical distribution). (Courtesy of Leeflang, E.P., S. Tavare, P. Marjoram, *et al.* 1999. Analysis of germline mutation spectra at the Huntington's disease locus supports a mitotic mutation mechanism. *Hum. Molec. Genet.* 8:173–183.)

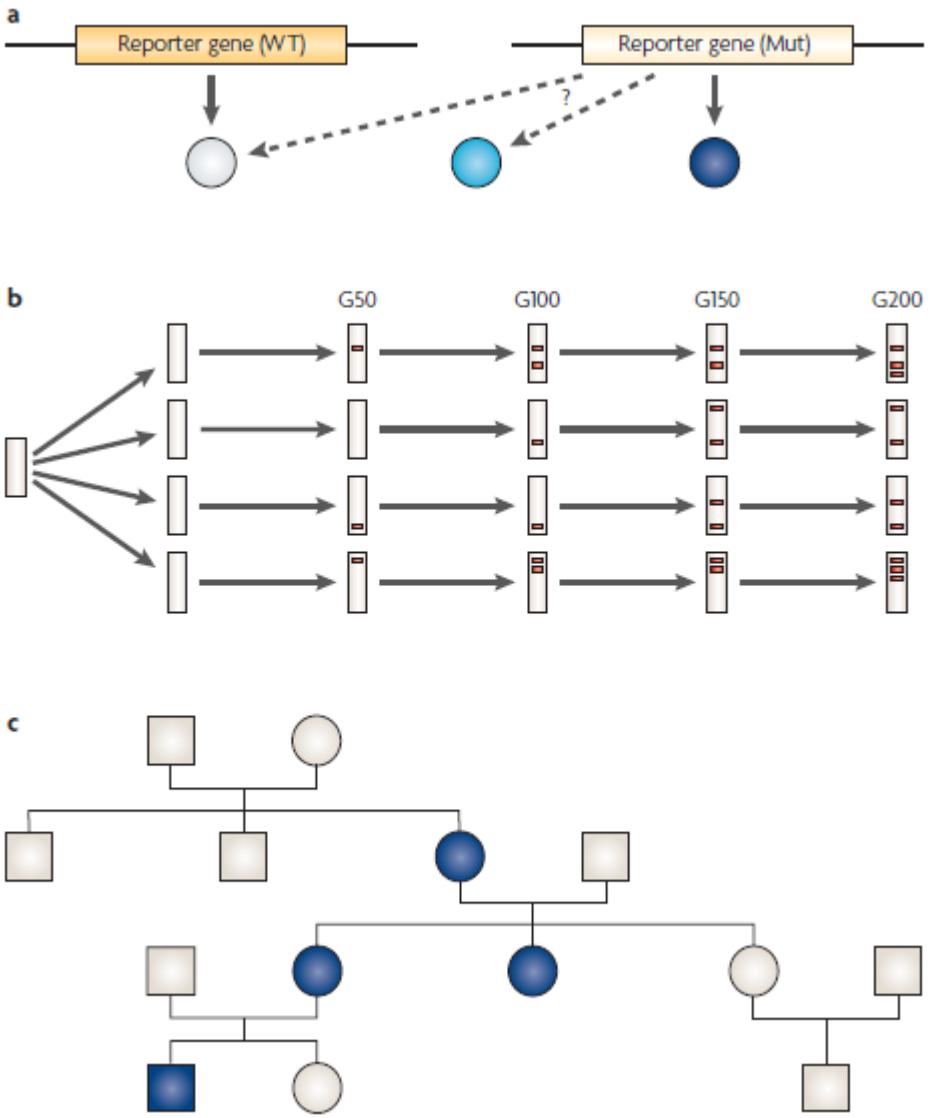


Figure 2 | **Experimental methods for molecular mutation rate estimation.** Three methods are shown for reporter genes (**a**), mutation-accumulation lines (**b**) and pedigree approaches (**c**). Some mutations might yield incomplete or subtle phenotypes that might be missed in reporter gene screens (**a**). The rectangles in panel **b** represent chromosomes that accumulate mutations (red) across generations in the mutation-accumulation line approach. Individuals who carry a new mutation in the pedigree approach (**c**) are represented by a blue square or circle. Mut, mutant; G, generation; WT, wild type.

Estimating neutral mutation rates

d. Estimation from neutral divergence between species

$r = K/(2T) = \mu$

Baer *et al*. 2007 NatRevGenet

Mutation rates from molecular evolution

Expected substitution rate for neutral nucleotide sites $r = \mu$ (/site/year)

Example: Average divergence rate at neutral sites between two species Ks =0.03. They diverged 5 MYA.

Substitution rate

$r = K_s/(2T)$ equals the mutation rate

$r = 0.03/(2x5x10^{6)} = 3x10^{-9}$

Visible mutation

-mutations at individual loci rare, large-scale experiments necessary to estimate rates

-forward and backward spontaneous mutations at five coat-colour loci in mice ($7x10^6$ ind.)

TABLE 8.1 Spontaneous mutation rates at five specific coat-color loci in mice (Schlager and Dickie, 1971).

Locus	*Number of gametes tested*	*Number of mutations*	*Mutation rate* ($\times 10^{-6}$)	95% *confidence limits* ($\times 10^{-6}$)
		Mutations from wild type (forward)		
Nonagouti	67,395	3	44.5	9.2–130.1
Brown	919,699	3	3.3	0.7–9.5
Albino	150,391	5	33.2	10.8–77.6
Dilute	839,447	10	11.9	5.2–21.9
Leaden	243,444	4	16.4	4.5–42.1
Total	2,220,376	25	11.2	7.3–16.6
		Dominant mutations (backward)		
Nonagouti	8,167,854	34	4.2	2.9–5.8
Brown	3,092,806	0	0	0–1.2
Albino	3,423,724	0	0	0–1.1
Dilute	2,307,692	9	3.9	1.8–11.1
Leaden	266,122	0	0	0–13.9
Total	17,236,978	43	2.5	1.8–3.4

Hedrick 2000

Mutation rates per genome

from Hedrick 2000

TABLE 8.8 The mutation rate per genome standardized by the effective size of the genome (G_e) compared to the size of the total genome (G) and the number of cell replications per sexual generation (Rep./Gen.) (Drake *et al.*, 1998).

			Mutation rate			
			Sexual generation		*Replication*	
Species	G_e/G	*Rep./Gen.*	G	G_e	G	G_e
C. elegans	0.225	9.1	0.16	0.036	0.018	0.004
D. melanogaster	0.094	25	1.5	0.14	0.058	0.005
Mouse	0.030	62	30	0.9	0.49	0.014
Human	0.025	400	64	1.6	0.16	0.004

Ge/G, proportion of functional genes in the genome
Rep/Gen, number of cell divisions per sexual generation

Note: constancy of last column Ge/rep

Estimating mutation rates – ”post genomic”

Table 3. A comparison of pre- and post-genomic estimates of mutation rate parameters (μ_n, U_D, U_T) in model species

	Pre-genomic estimate		Post-genomic estimate		
Organism	μ_n	U_D	μ_n	U_D	U_T
S. cerevisiae	2.2×10^{-10}	10^{-5}–10^{-3}	3.3×10^{-10}		0.32
C. elegans	2.3×10^{-10}	0.005	2.1×10^{-8}	0.48	2.1
D. melanogaster	3.4×10^{-10}	0.02–0.1	8.4×10^{-9}	1.2	

μ_n, mutation rate per base; U_D, deleterious mutation rate per genome based on fitness; U_T, total mutation rate per genome

-post-genomic rate estimates higher than pre-genomic
-mutation rate affected by sequence context and genome position: Rate variation within genome

Nishant *et al*. 2009 Bioessays

4. Mutation and back mutation

A mutates to a with frequency μ, back mutation a to A with v

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

-equilibrium allele frequency

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

Change of frequency under mutation

from Hartl and Clark 1997

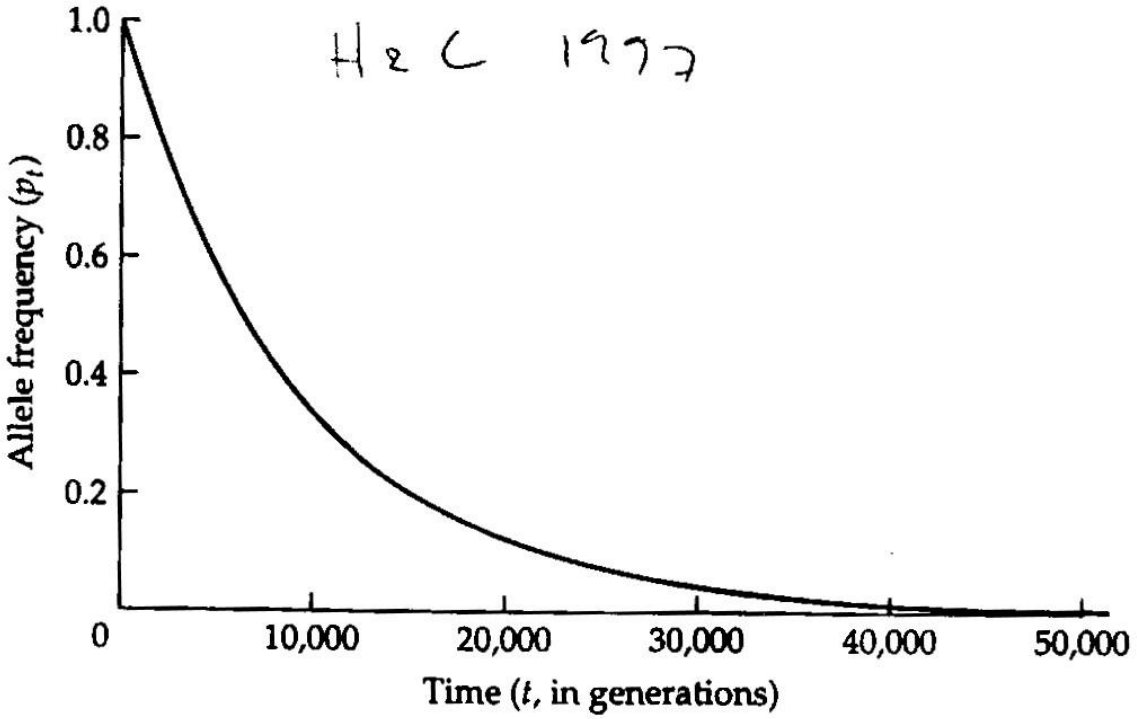
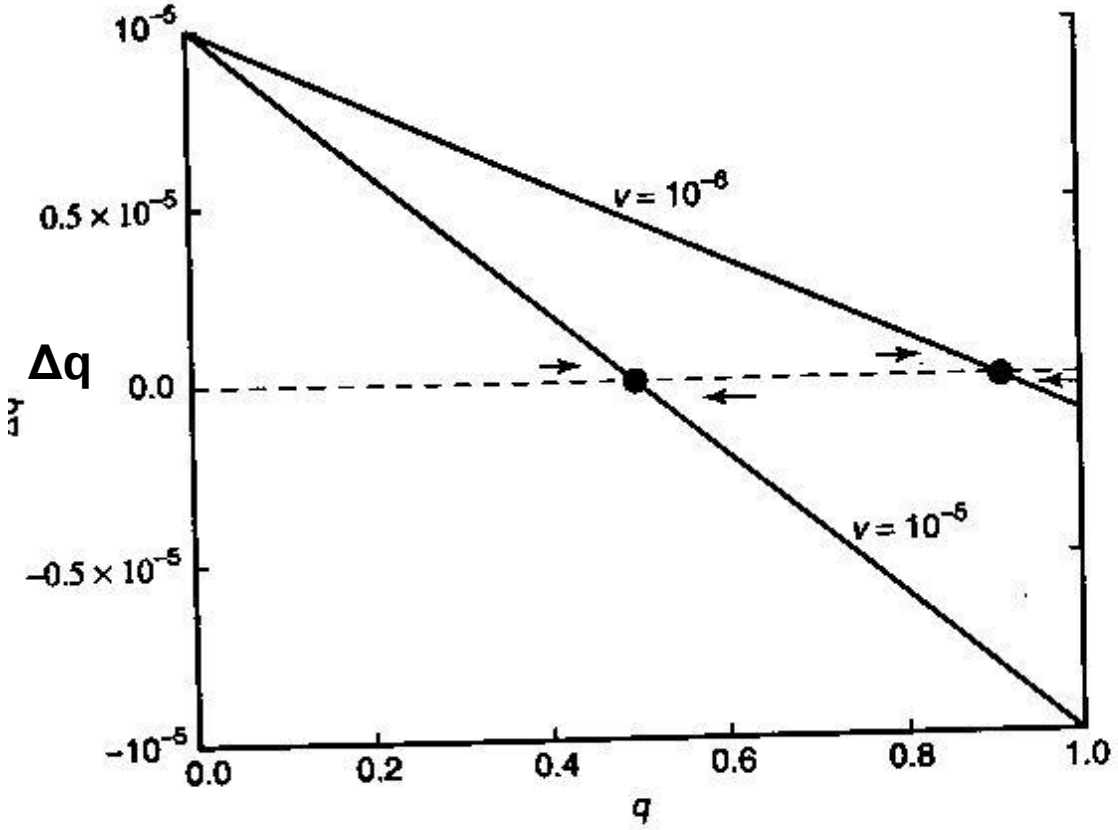


Figure 5.1 Change in frequency under mutation pressure. In this example, an allele A mutates to a at a rate of $\mu = 1 \times 10^{-4}$ per generation; p_t is the allele frequency of A in generation t. We assume that $p_0 = 1$. With the given value of μ, the allele frequency decreases by half every 6931 generations.

Mutation-back mutation balance
(from Hedrick 2005)

I. Allele Frequency Change Caused by Mutation



q_e = 0.50 or 0.91

Figure 7.3. The change in allele frequency due to mutation when $u = 10^{-5}$ and v is either 10^{-5} or 10^{-6}. The closed circles indicate allele frequency equilibria.

5. Mutation and drift

N_e and expected heterozygosity – infinite-allele model

$$H_e = \frac{4N_e\mu}{4N_e\mu + 1}$$

-neutral variation, in a large population drift less efficient than in small

$$f_t = \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right) f_{t-1})\right](1-\mu)^2$$

$(1-\mu)^2$ probability that both alleles do not mutate

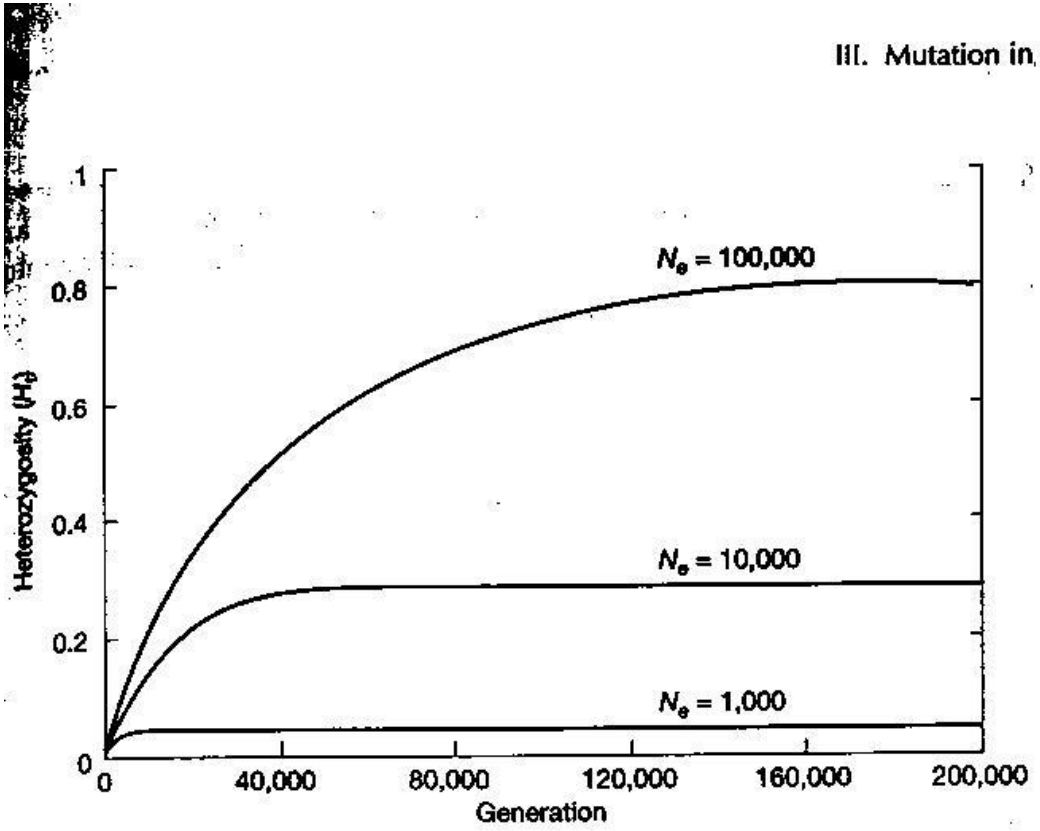
-in mutation-drift equilibrium $f_t = f_{t-1} = f_e$, ignore μ^2

$$f_e \approx \frac{1-2\mu}{4N_e\mu + 1 - 2\mu}$$

ignore 2μ

$$f_e \approx \frac{1}{4N_e\mu + 1}$$

Mutation and population size



igure 7.9. The expected change in heterozygosity from mutation and genetic rift when $H_0 = 0$ and $u = 10^{-5}$ for three different effective population sizes.

Different final result
Different time scale

6. Estimating fitness effects and genomewide rates of mutation, *U* and *s*

- Estimation of μ (per locus, or per nucleotide)
  - Progeny assays
  - Molecular evolution
  - Mutation accumulation
- Estimation of *U*, deleterious genomic mutation rate
  - Make chromosomes homozygous
  - Mutation accumulation
- Estimating *s*
  - Mutation accumulation (estimation of mutation rate and fitness effects simultaneously)

Estimating rates and effects of deleterious and beneficial mutations

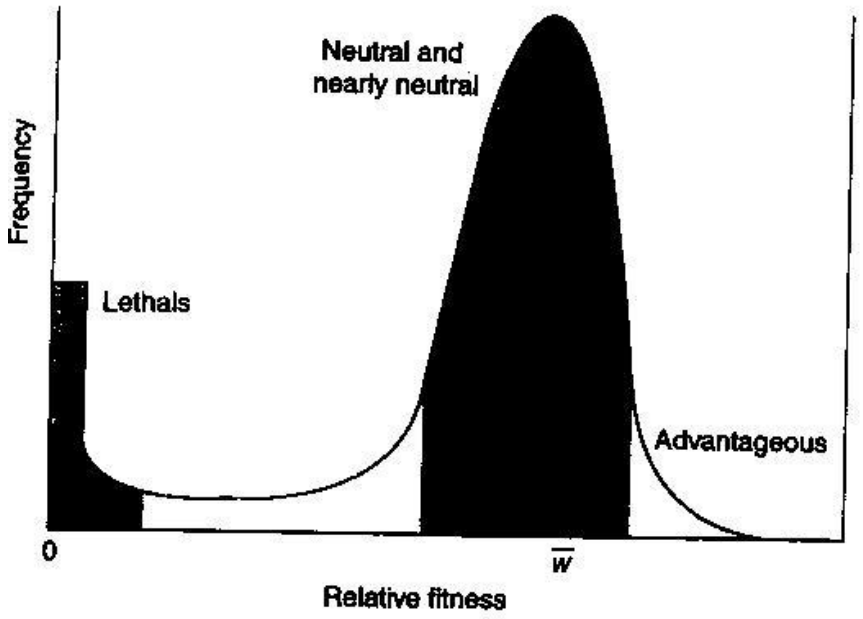


Figure 7.2. A hypothetical distribution of relative-fitness values for new mutants, where $\overline{w}$ is the mean population fitness.

...ates of the shape of this distribution and th...

from Hedrick 2005

Fate of mutations

-large population, probability of loss in one generation e^{-1}

from Hedrick 2000

TABLE 8.3 The probability of loss and of survival of a new mutant when there is neutrality or a 1% selective advantage (after Fisher, 1930).

	Neutral		$s = 0.01$	
Generation	*Loss*	*Survival*	*Loss*	*Survival*
1	0.368	0.632	0.364	0.636
2	0.532	0.468	0.526	0.474
3	0.626	0.374	0.620	0.380
4	0.688	0.312	0.681	0.319
5	0.732	0.268	0.725	0.275
⋮	⋮	⋮	⋮	⋮
15	0.887	0.113	0.878	0.122
⋮	⋮	⋮	⋮	⋮
127	0.985	0.015	0.973	0.027
⋮	⋮	⋮	⋮	⋮
∞	1.0	0.0	0.98	0.02

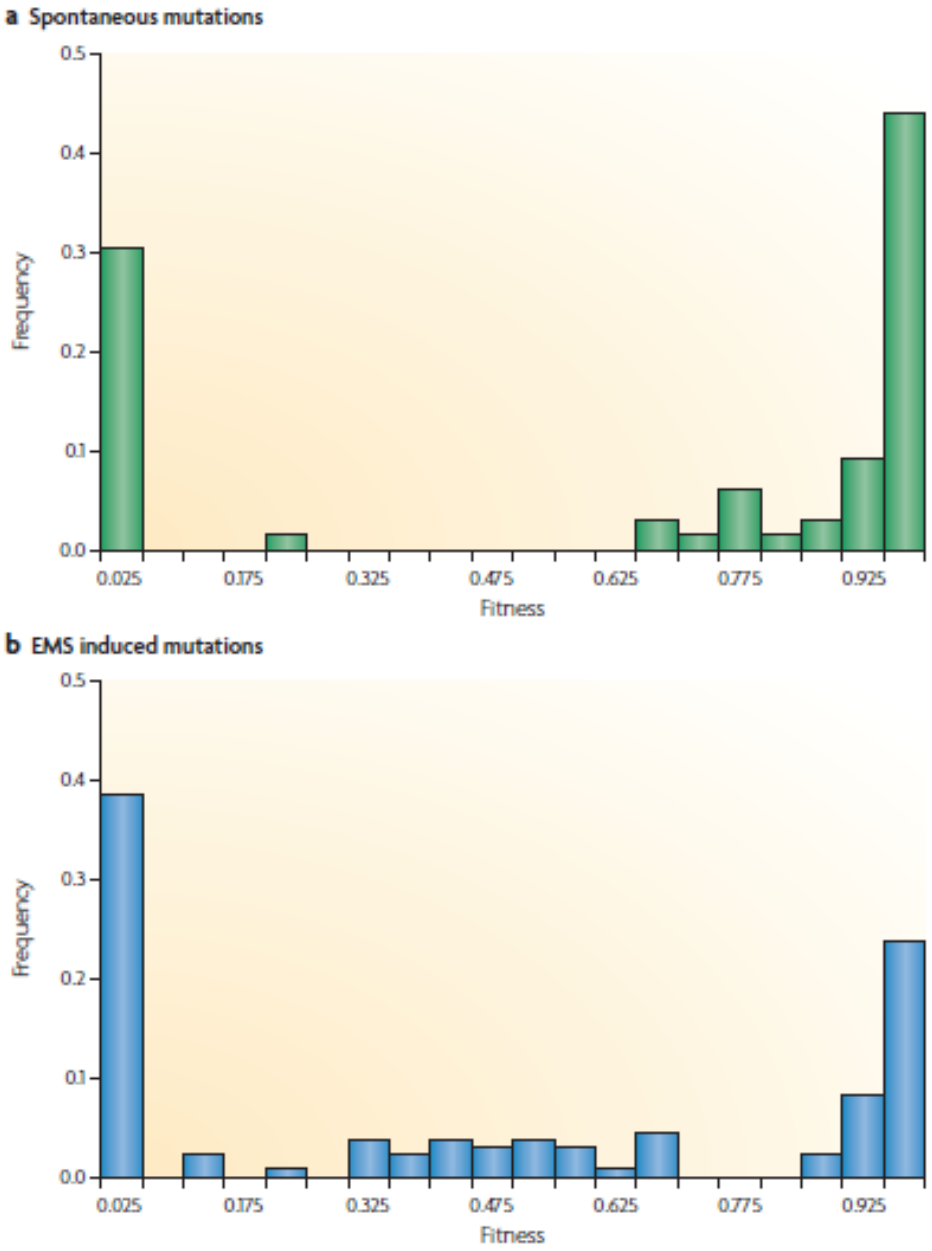


Figure 2 | **The distribution of fitnesses among yeast lines.** Diploid yeast lines were either allowed to accumulate spontaneous mutations (panel **a**) or were subject to chemical mutagenesis using ethylmethane sulphonate (EMS) (panel **b**). After a period of inbreeding, the cells were made to undergo meiosis and the growth of the meiotic products was measured. Data from REF. 17.

Fitness consequences of mutations in yeast

-distributions for mutation accumulation and mutagenesis differ
-bimodal
-30-40% lethals
-mutations with small effects undetected?

Eyre-Walker and Keightley 2007

Mutation accumulation, estimate *U* and *s*

$$U \geq \frac{(\Delta M)^2}{\Delta V}$$

$$\bar{s} \leq \frac{\Delta V}{\Delta M}$$

Mutation accumulation lines:

-start a large number of inbred lines
-include large controls, no selection in controls
-follow change in mean *M* (fitness), Δ*M*
-change in variance *V* in viability of lines, Δ*V*
-results allow estimation of lower limit
of genome-wide mutation

Mutation accumulation (Hedrick 2005)

TABLE 7.8 The estimated values of decline in mean viability per generation (ΔM) and the increase in variance among lines (ΔV) are given for four different studies. By using these values and expressions 7.15*a* and 7.15*b*, the rate of mutation affecting viability per haploid second chromosome per generation (U) and the average effect of the mutations ($\bar{s}$) are calculated. Also given is the estimated per-zygote mutation rate U_Z, which is obtained by multiplying U by 5.

Study	ΔM	ΔV	U	$\bar{s}$	U_Z
Mukai (1964)	0.0038	0.00010	0.14	0.027	0.70
Mukai *et al.* (1972)	0.0040	0.000094	0.17	0.023	0.85
Onishi (1977)	0.0017	0.000051	0.058	0.030	0.29
Fry *et al.* (1999)	0.0024	0.00027	0.021	0.113	0.10

-early studies by Mukai: frequent slightly deleterious,
some recent studies: rare but more deleterious

7. Mutation - selection balance

-recessive detrimental allele, fitnesses AA, Aa, aa 1,1,1-*s*
-mutation adds alleles, selection removes them

$$\Delta q_s = -\frac{sq^2 p}{1 - sq^2}$$

-assume that *q* is small, forget denominator, increase in allele frequency due to mutation $\Delta q_\mu = \mu p$

$$\Delta q_\mu + \Delta q_s = 0$$

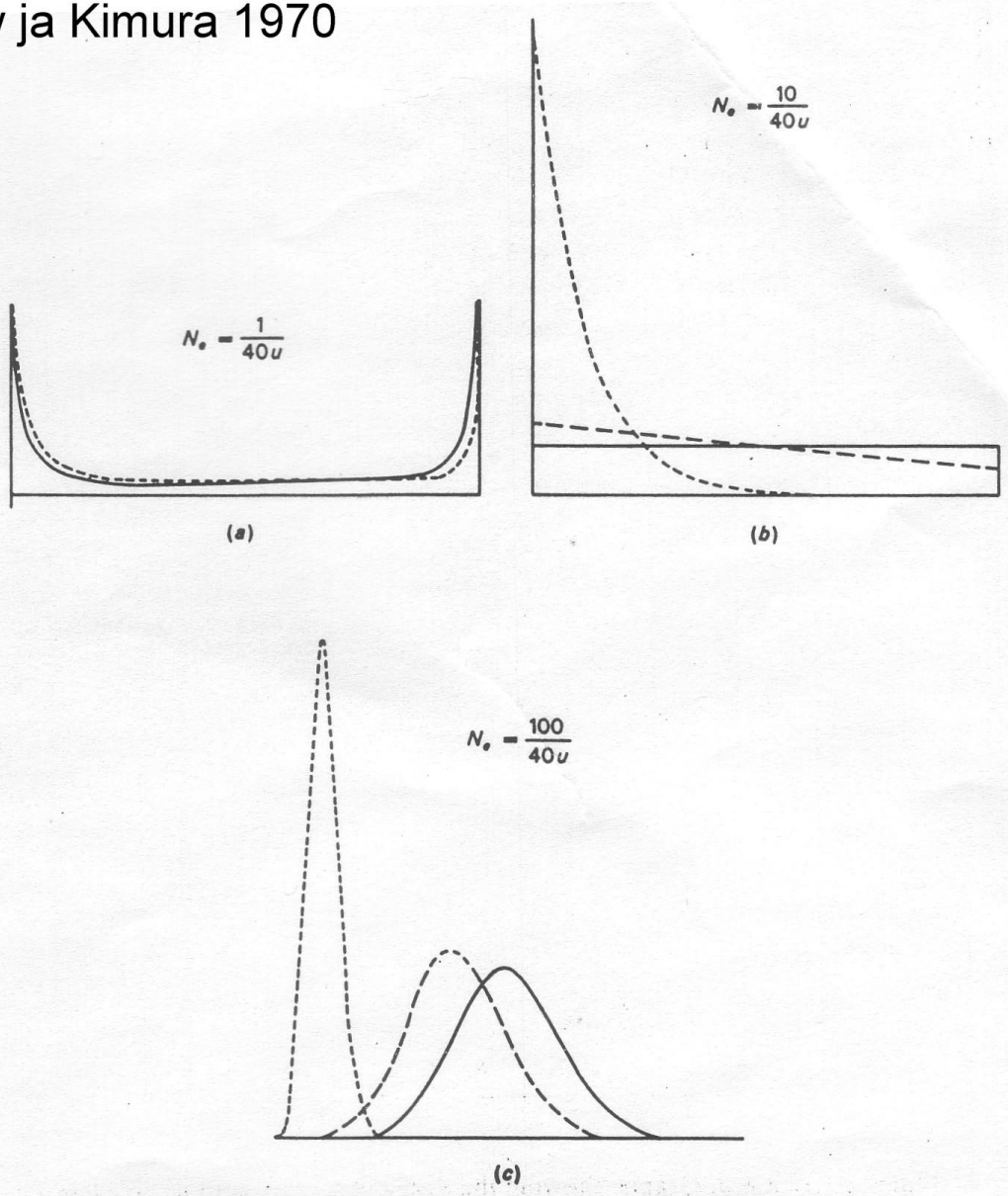
$$\Rightarrow \quad \mu p = \frac{sq^2 p}{1 - sq^2}$$

$q_e^2 = \mu/s \Rightarrow q_e = \sqrt{\mu/s}$

For dominant mutations:
AA, Aa, aa 1,1-s,1-s
$q_e = \mu/s$

AA, Aa, aa 1,1-hs,1-s
$q_e = \mu/(hs)$

Crow ja Kimura 1970



Figures 9.3.1*a,b,c*. Graphs showing the frequency distribution of a deleterious gene ($s < 0$) in a small (*a*), an intermediate (*b*), and a large (*c*) population, assuming equal mutation rates in both directions ($u = v$). In each figure, curves with solid, broken, and dotted lines represent cases with $s = -u/10$, $-u$, and $-10u$, respectively. (From Wright, 1937.)

Balance of selection and drift
drift is effective in small populations, also for rare alleles

Mutation, selection and dominance

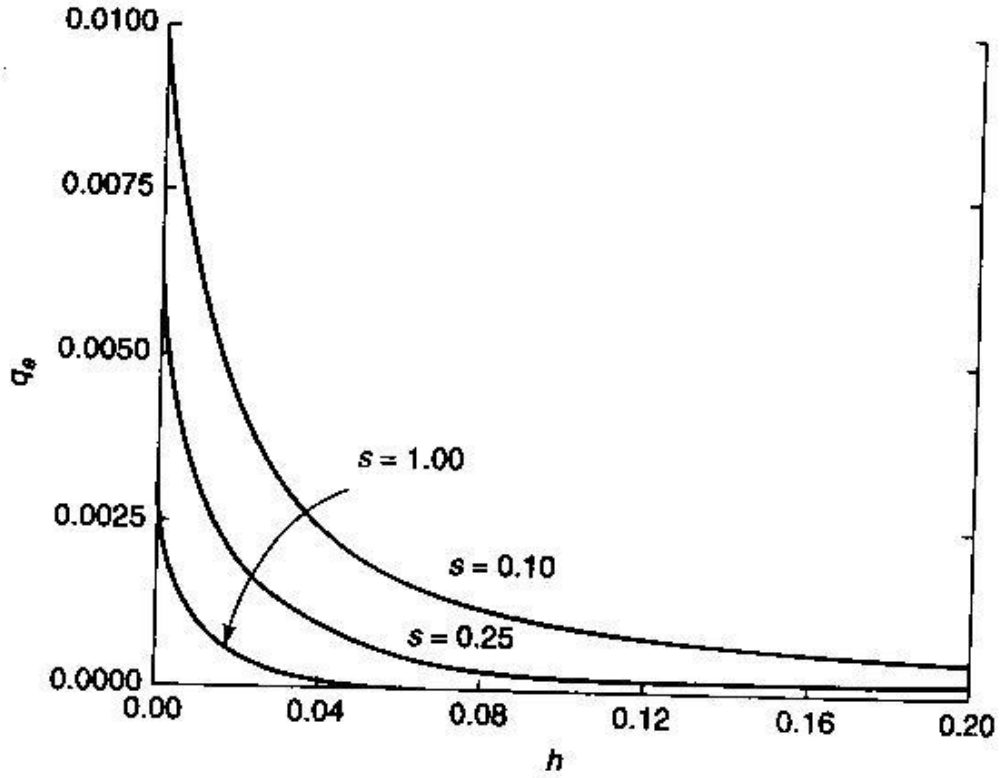


Figure 7.6. The mutation–selection equilibrium frequency for different levels of dominance and three levels of selection ($u = 10^{-5}$).

Deleterious mutations are kept at a low frequency if they are partly dominant, *h* measures dominance

Inbreeding – selection: effect on deleterious frequency (Hedrick 2005)

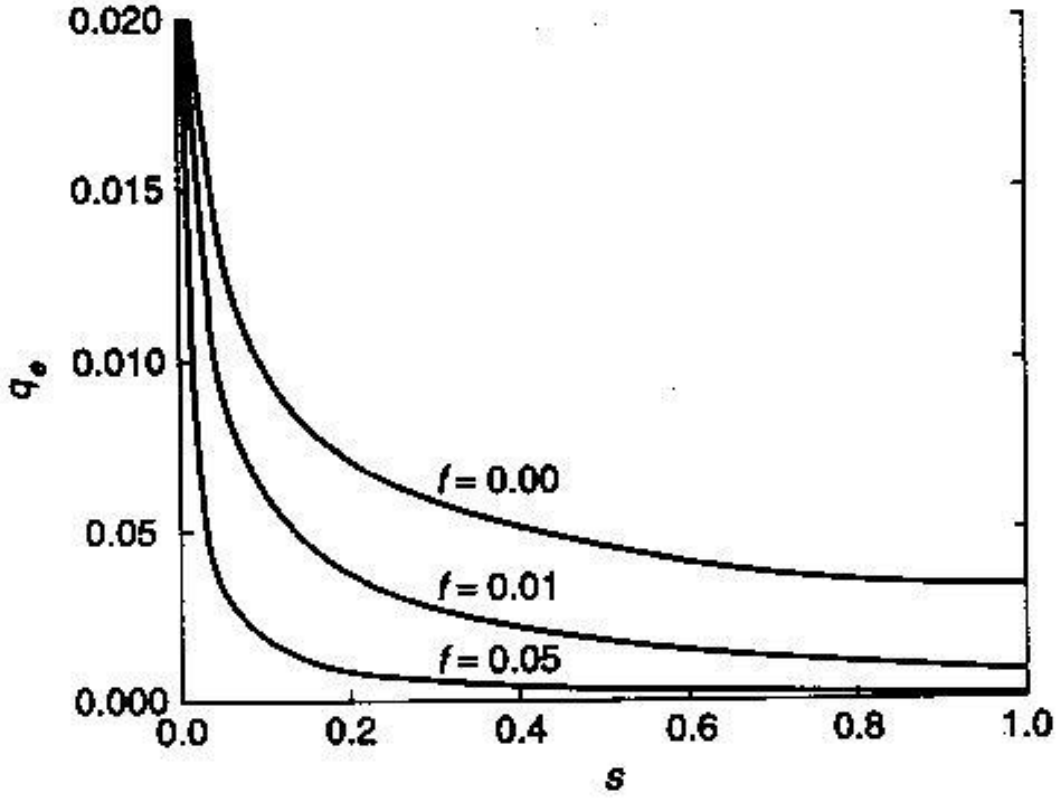


Figure 7.5. The mutation–selection equilibrium allele frequency for different levels of selection and three values of inbreeding ($u = 10^{-5}$).

High inbreeding levels result in low frequency of deleterious recessives