Lecture 11: The Impact of Genetic Drift on Selected Alleles. (Ch12)

Population genetic PCB4553/6685

# Stochastic loss of strongly selected alleles

- **Even strongly beneficial alleles can be lost from the population when they are sufficiently rare.**

- Reason:

  - number of offspring left by one individual to the next generation is fundamentally stochastic

  - Initial frequency in the pop $= 1/2N$

- A selection coefficient of $s = .01$ is a strong selection coefficient

- On average only 1% more children than their peers

Probability of eventual loss of selected allele in large haploid populations

- Probability an allele is *eventually* lost
  from a population p_L

- P_k, the probability that an individual
  carrying selected allele has k offspring.

- Apply the law of total probability

$$p_L = \sum_{k=0}^{\infty} P_k p_L^k$$

-
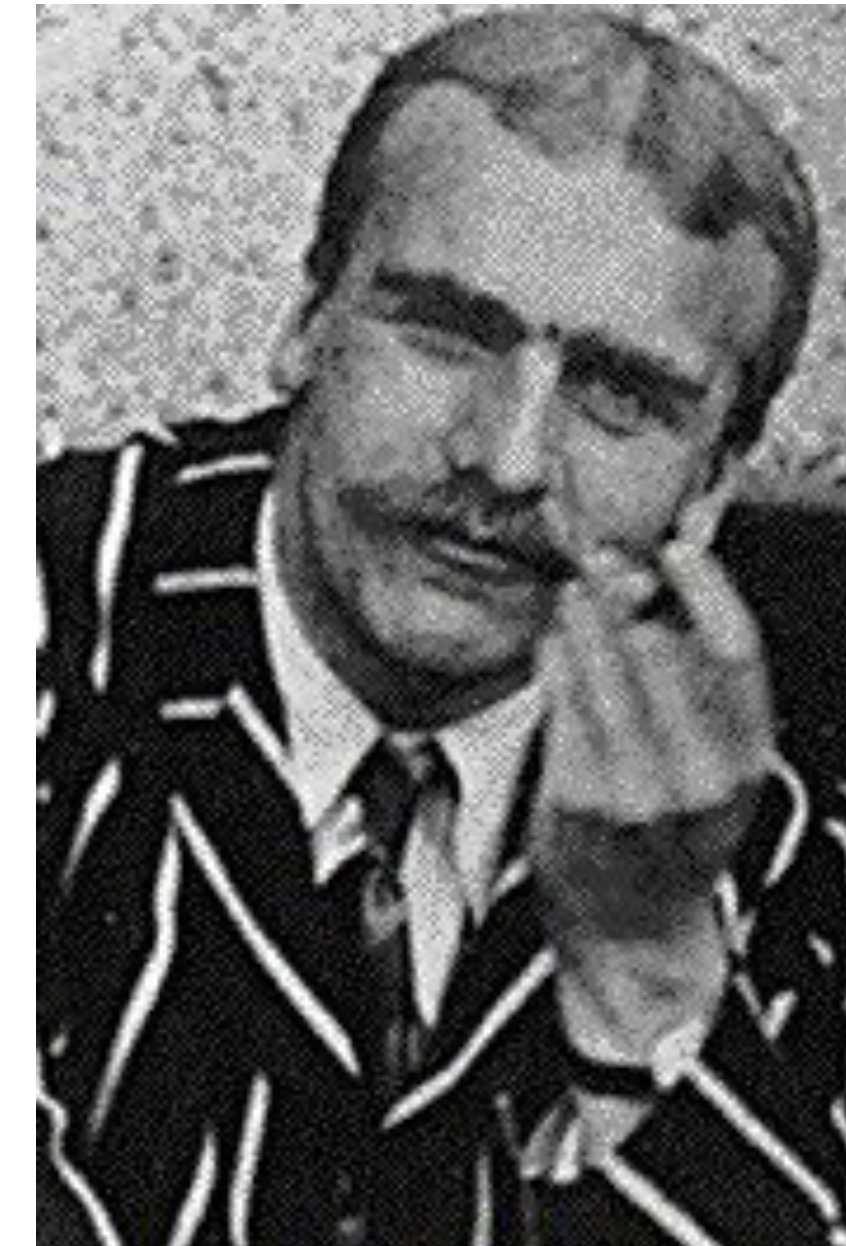
JBS Haldane

Probability of eventual loss of selected allele in large haploid populations

- $$p_L = \sum_{k=0}^{\infty} P_k p_L^k$$

- P_k: Poisson with mean $1 + s$

- $$P_i = \frac{(1+s)^i e^{-(1+s)}}{i!}$$

- $$p_L = \sum_{k=0}^{\infty} \frac{(1+s)^k e^{-(1+s)}}{k!} p_L^k$$
  $$= e^{-(1+s)} \left(\sum_{k=0}^{\infty} \frac{(p_L(1+s))^k}{k!} \right)$$

- $p_L = e^{(1+s)(p_L - 1)}$

Probability of eventual loss of selected allele in large haploid populations

- Solve for $p_L = e^{(1+s)(p_L-1)}$

  - $p_F = 1 - p_L$

  - $1 - p_F = e^{-p_F(1+s)}$

  - $1 - p_F \approx 1 - p_F(1+s) + p_F^2(1+s)^2/2$

- $p_F \approx 2s$

- *Thus even an allele with a 1% selection coefficient has a 98% probability of being lost when it is first introduced into the population by mutation.*

Rate of fixation of beneficial mutations

- mutation rate towards our advantageous allele is μ

- $N\mu$ advantageous mutations arise per generation

- $N\mu p_F$: number of advantageous mutations arising per generation that will eventually fix in the population

- $1/N\mu p_F$: waiting time for a mutation that will fix to arise

- Polymorphism is simply a phase of substitution
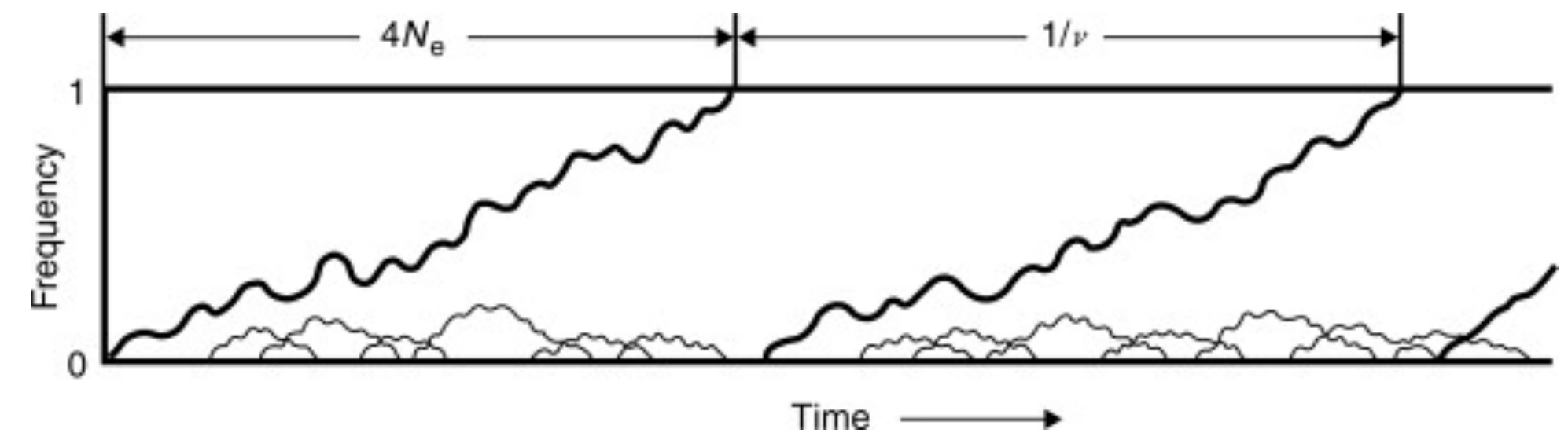
Ohta 2001

Diploid model of stochastic loss of strongly selected alleles

- Heterozygote fitness: $1 + hs$

- Probability allele 1 is not lost, starting from a single copy in the population

  - $p_F = 2hs$

- Only depends on the relative fitness advantage of the heterozygote

- Because when the allele is rare it is usually present in heterozygotes

Question 1.

'Haldane's sieve' is the name for the idea that the mutations that contribute to adaptation are likely to be dominant or at least co-dominant.

A) Briefly explain this argument with a verbal model relating to the results we've developed in the last two chapters.

B) Haldane's sieve is thought to be less important for adaptation from previously deleterious standing variation, than adaptation from new mutation. Can you explain the intuition behind of this idea?

C) Haldane's sieve is likely to be less important in inbred, e.g. selfing, populations. Why is this?

Interaction between genetic drift and *weak* selection

- For *strongly* selected alleles, once the allele has escaped initial loss at low frequencies, its path will be determined deterministically by its selection coefficients.

- if selection is *weak* compared to genetic drift, the stochasticity of reproduction can play a role in the trajectory an allele takes even when it is common in the population.

- If selection is *sufficiently weak* compared to genetic drift, then drift will dominate the dynamics of alleles and they will behave like they're effectively neutral.

- Thus, the extent to which selection can shape patterns of molecular evolution will depend on the relative strengths of selection and genetic drift.

Interaction between genetic drift and weak selection

- How weak (strong) must selection on an allele be for drift to overpower selection?

- Qualitative model

  - Assumptions

  - $h = 1/2$, and $s \ll 1$, so that $\overline{w} \approx 1$

Interaction between genetic drift and weak selection

- Qualitative model

  - Variance in our allele frequency change across one generation due to drift

  $$Var(\Delta p) = Var(p_1 - p) = Var(p_1) = \frac{p'(1 - p')}{2N} \approx \frac{p(1 - p)}{2N}.$$

  - Mean expected change across one generation due to selection

    - $E(\Delta p) = \frac{s}{2}p(1 - p)$ (Chapter 10)

  - When does selection dominate genetic drift?

    - $E(\Delta p) \gg Var(\Delta p)$, i.e. when $|Ns| \gg 1$

  - When does drift dominate selection? $|Ns| \ll 1$

Interaction between genetic drift and weak selection - Fixation probability

- Fate of alleles with selection coefficients of the order $1/N$ ($Ns \approx 1$)?

- Fixation probability $p_F(p)$: probability that a weakly selected mutation is eventually fixed

- *Rename* it as $Q(p)$

- Applying law of total probability

- $$Q(p) = \int Q(p + \Delta p) P(\Delta p) d(\Delta p)$$

Interaction between genetic drift and weak selection - Fixation probability

- $$Q(p) = \int Q(p + \Delta p)P(\Delta p)d(\Delta p)$$

- Apply Taylor expansion of $Q(p)$

- $$Q(p + \Delta p) \approx Q(p) + \Delta p\frac{dQ(p)}{dp} + \frac{1}{2}(\Delta p)^2\frac{d^2Q(p)}{dp}$$

- Taking the expectation over Δp

- $$Q(p) \approx Q(p) + \mathbb{E}(\Delta p)\frac{dQ(p)}{dp} + \frac{1}{2}\mathbb{E}(\Delta p)^2\frac{d^2Q(p)}{dp}$$

Interaction between genetic drift and weak selection - Fixation probability

- $\mathbb{E}(\Delta p) = \dfrac{s}{2} p(1 - p)$

- $Var(\Delta p) = E[(\Delta p)^2] - E^2[\Delta p]$

- Assuming $s \ll 1$

  - $\mathbb{E}(\Delta p) \approx 0$

  - $Var(\Delta p) \approx E[(\Delta p)^2] = \dfrac{p(1 - p)}{2N}$

- Putting everything together

  - $$Q(p) \approx Q(p) + \mathbb{E}(\Delta p)\frac{dQ(p)}{dp} + \frac{1}{2}\mathbb{E}(\Delta p)^2\frac{d^2Q(p)}{dp}$$

  - $$\frac{s}{2}p(1-p)\frac{dQ(p)}{dp} + \frac{p(1-p)}{4N}\frac{d^2Q(p)}{dp} = 0$$

Solving the equation

- $$\frac{s}{2}p(1-p)\frac{dQ(p)}{dp} + \frac{p(1-p)}{4N}\frac{d^2Q(p)}{dp} = 0$$

The solution

- Solution:

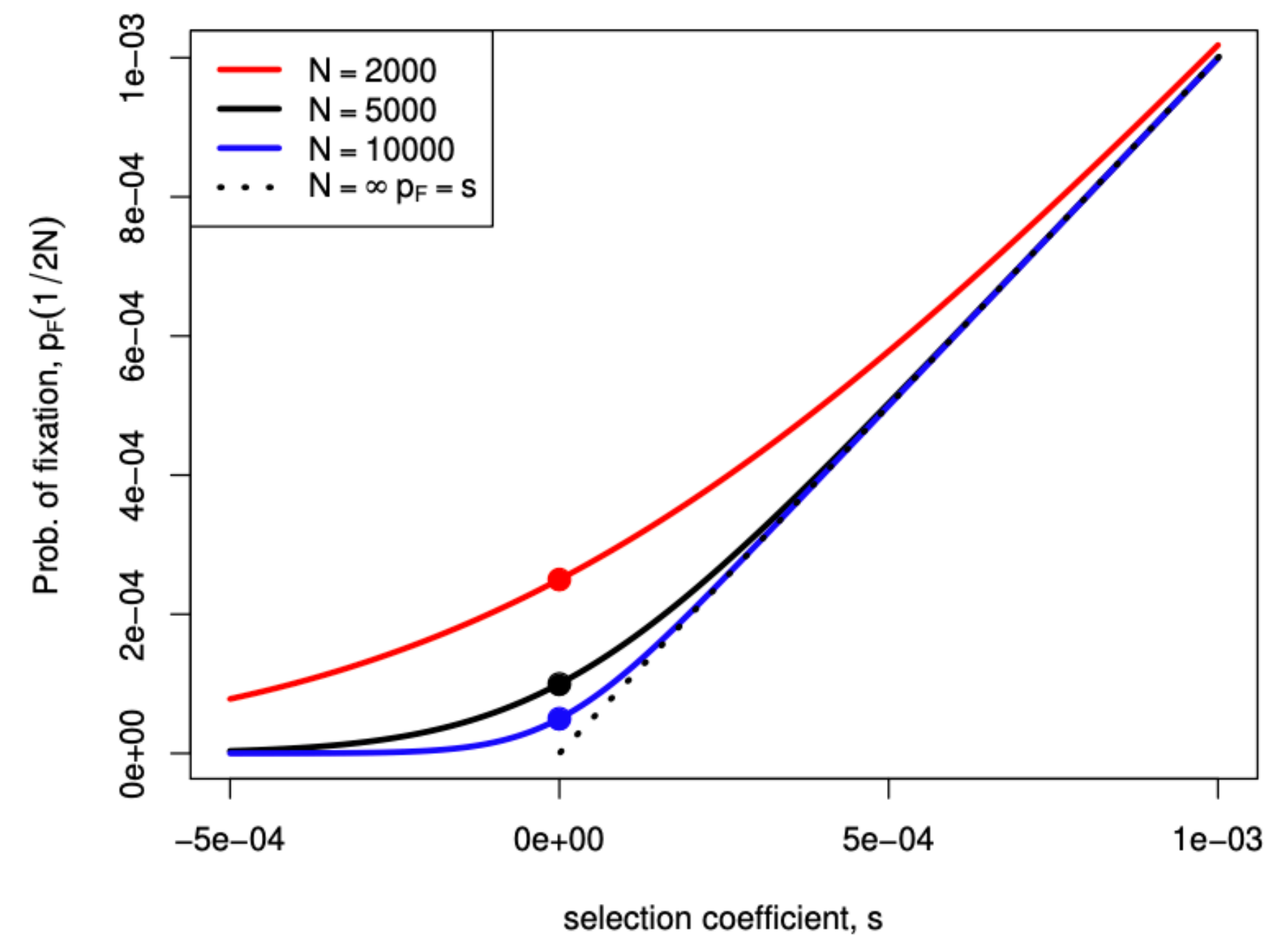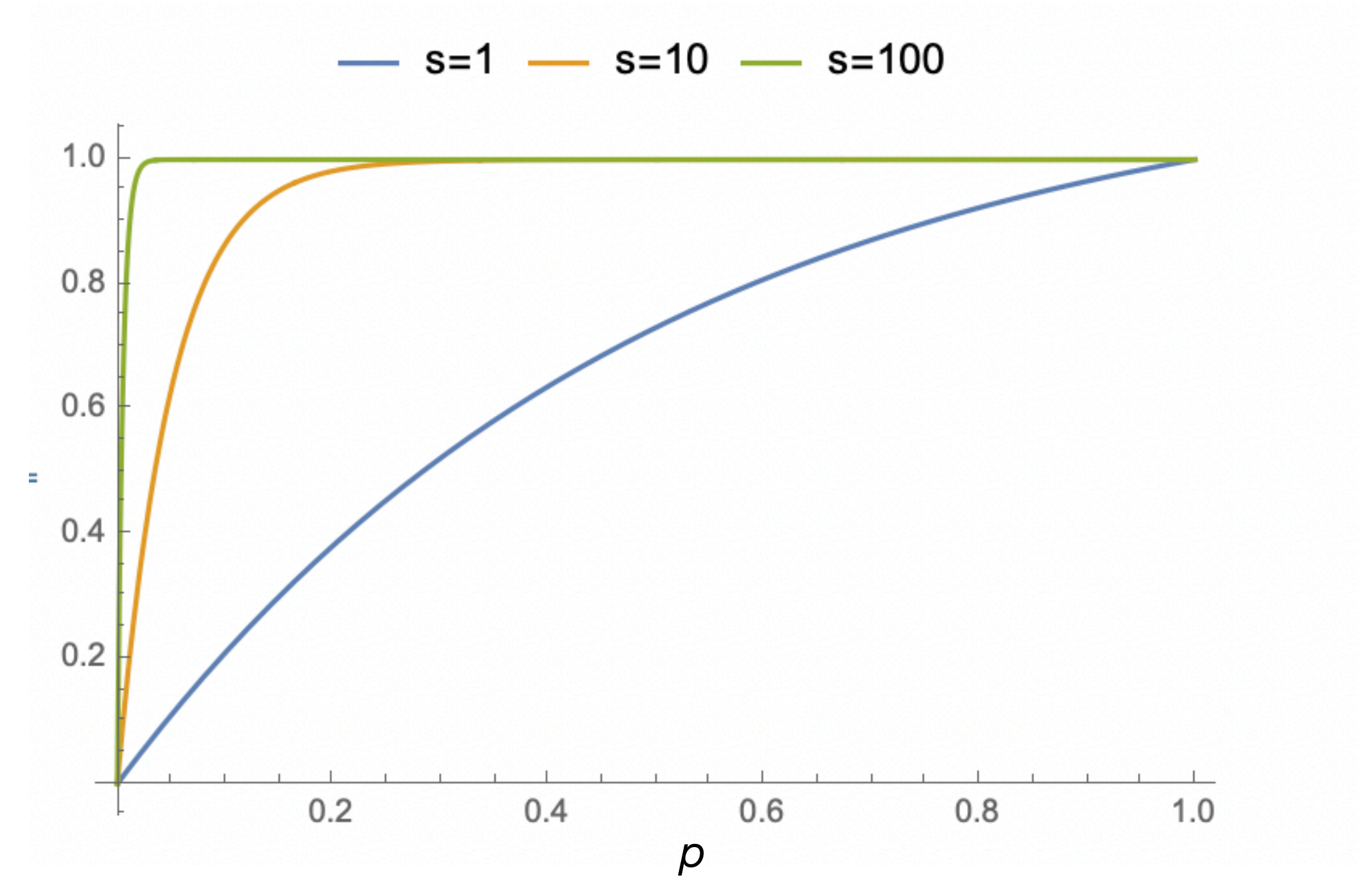- $$Q(p) = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}$$

- $$Q(\frac{1}{2N}) = \frac{1 - e^{-s}}{1 - e^{-2Ns}}$$

- If $s \ll 1$ but $Ns \gg 1$ then

  - $Q(1/2N) \approx s$

- If Ns close to one, then

- $$Q(1/2N) = \frac{s}{1 - e^{-2Ns}}$$

Effectiveness of weak selection

- For selection to operate on an allele, we need the selection coefficient to satisfy $|Ns| \gg 1$, or $|s| \gg 1/N$.

- Selection coefficients on the order of 10^{-5} to 10^{-6} can be effectively selected upon

- "codon usage bias" likely reflects the combined action of weak selection and mutational pressure, pushing the codon composition of the genome and tRNA abundances towards an adaptive compromise.
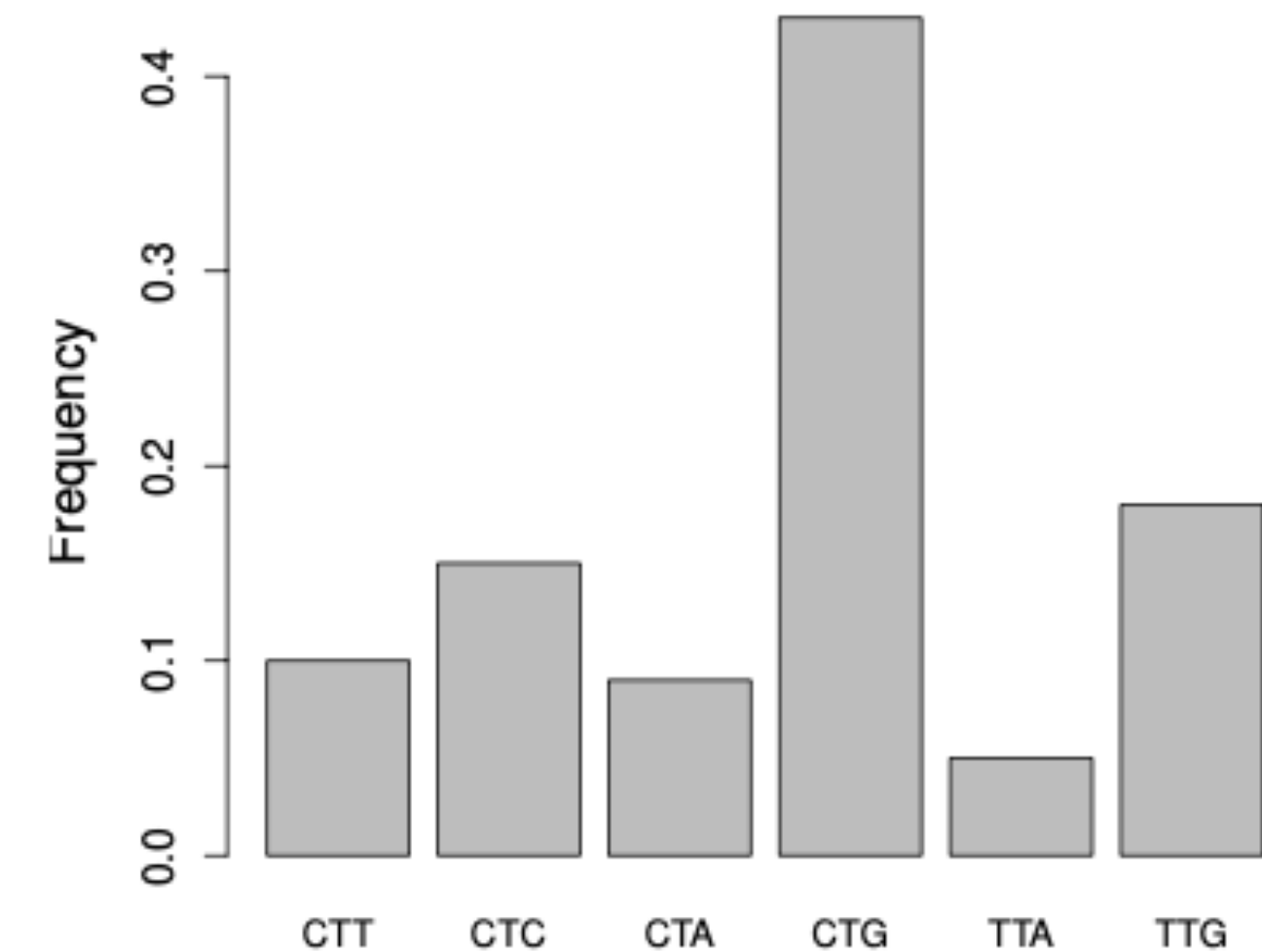

Figure 12.9: Data from *Drosophila melanogaster* on the frequency of different codons for Leucine. Data from Genscript. Code here.

Fixation of deleterious mutations

- Selection coefficient $-s < 0$ against allele 2

- $Q(\dfrac{1}{2N}) = \dfrac{1 - e^s}{1 - e^{2Ns}}$

- $Q(\dfrac{1}{2N}) \approx \dfrac{s}{e^{2Ns} - 1}$

- Deleterious alleles can fix within populations (albeit at a low rate)