

## Outline

- Lec 1
1. Forces of evolution
  2. Genetic drift
  3. Mutations
  4. Mutations + drift

- Lec 2
1. Selection
  2. Selection + mutations + drift.

### 1) Forces of evolution

Definitions: Genotype; Locus, Allele, Phenotype

AGGCATTA  $1P$  {A,T,G,C}

↑

A G,C,T

diploid: two chromosomes for each locus

$A_1, A_2 \leftarrow$  alleles

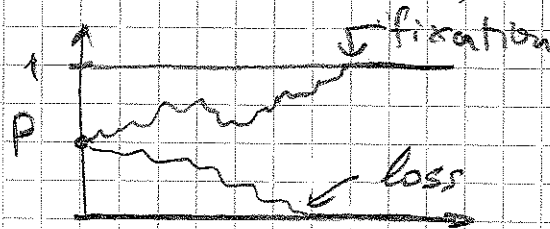
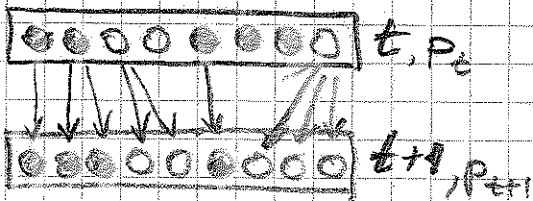
$p, 1-p \leftarrow$  allele frequency

### 2) Genetic drift

Consider a population of  $N$  organisms ( $2N$  alleles)

Mating:   
 1) Choose an allele at random   
 2) Make its copy (exact)   
 3) Place into a new generation,   
 repeat  $2N$  times

$A_1$   $A_2$



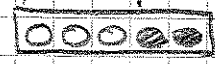
Let's calculate change in  $p$  ( $2Np_t$  time) in one generation.

$$P_j = p_t^j (1-p_t)^{2N-j} \binom{2N}{j}$$

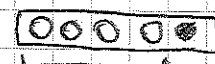
$$\langle j \rangle = p_t \cdot 2N \Rightarrow \langle P_{t+1} \rangle = \frac{j}{2N} = p_t$$

$$\text{Var}(j) = \langle j^2 \rangle - \langle j \rangle^2 = p_t(1-p_t) \cdot 2N$$

$$\text{Var}(p_t) = \text{Var}(j) / (2N)^2 = p_t(1-p_t) / 2N$$



$P_t$   $2Np_t - \# \text{ of } A_2$  in current generation

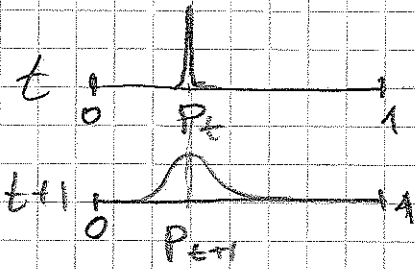


$P_{t+1}$   $j \leftarrow \# \text{ of } A_1$  in the next generation

$$P_{t+1} \equiv \frac{j}{2N}$$

$$\langle P_{t+1} \rangle = P_t$$

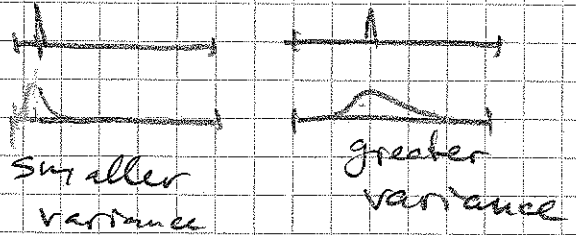
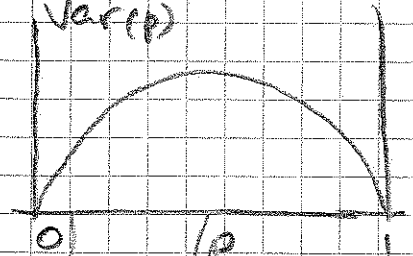
$$\text{Var}(P_{t+1}) = \frac{P_t(1-P_t)}{2N}$$



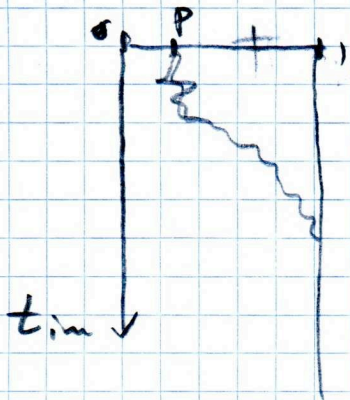
Conclusions:

- 1) Random mating doesn't change the mean frequency
- 2) The variance of  $p$  increases

↑ diffusion!  
 But  $D$  is not the same for all  $p$   
 $\text{Var}(p)$



$\pi(p)$  Probability of fixation, starting at the frequency  $p = p_0$



$$\pi(p_0) = \sum \text{Prob}(\Delta p) \pi(p_0 + \Delta p)$$

Expand:  $\pi(p_0 + \Delta p) \approx \pi(p_0) + \pi'(p_0) \Delta p + \pi''(p_0) \frac{\Delta p^2}{2}$

By plugging into  $\pi(p_0)$  we obtain

$$\begin{aligned} \pi(p_0) &= \sum_{\Delta p} \text{Prob}(\Delta p) \left[ \pi(p_0) + \pi'(p_0) \Delta p + \pi''(p_0) \frac{\Delta p^2}{2} \right] \\ &= \pi(p_0) + \pi'(p_0) \sum_{\Delta p} \text{Prob}(\Delta p) \Delta p + \pi''(p_0) \sum_{\Delta p} \text{Prob}(\Delta p) \frac{\Delta p^2}{2} \\ &= \pi(p_0) + \pi'(p_0) \langle \Delta p \rangle + \pi''(p_0) \frac{\langle \Delta p^2 \rangle}{2} \end{aligned}$$

$$\boxed{\pi'(p_0) \langle \Delta p \rangle + \pi''(p_0) \frac{\langle \Delta p^2 \rangle}{2} = 0}$$

$$\boxed{\pi'(p_0) \langle \Delta p \rangle + \pi''(p_0) \frac{\langle \Delta p^2 \rangle}{2} = 0}$$

$$\boxed{\pi(0) = 0 ; \pi(1) = 1}$$

Kolmogorov  
Backward  
equation

Let's obtain probability of fixation under drift only  
(no mutations, no selection)

$$\langle \Delta p \rangle = 0$$

$$\langle \Delta p^2 \rangle = \text{Var}(\Delta p) = \frac{p(1-p)}{2N}$$

$$\pi''(p_0) p_0(1-p_0) = 0$$

$$\pi''(p_0) = 0$$

$$\pi'(p_0) = c$$

$$\pi(p_0) = a p_0 + b$$

$$\pi(p_0) = p_0$$

using boundary  
conditions

### 3. Estimating the timescale of genetic drift

Consider a randomly mating population of size  $N$

1. Choose an allele at random among  $2N$  alleles
2. Make its exact copy
3. Place into new generation
4. Repeat until  $2N$  alleles are in the generation.

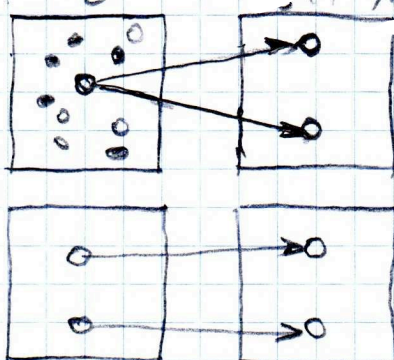
Evolution of  $p$



Evolution of  $H$

$G_{t+1}$  given  $G_t$

$G_t$  = Probability to pick two identical alleles in the population



$$\frac{1}{2N}$$

$$\left(1 - \frac{1}{2N}\right) G_t$$

$$G_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G_t$$

$$G_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G_t$$

$$1 - H_{t+1} = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) (1 - H_t)$$

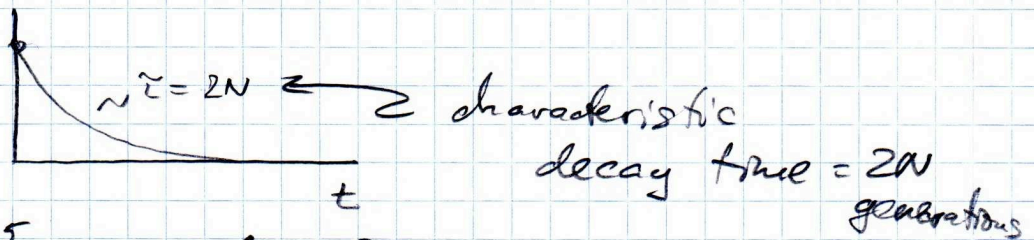
$$H_{t+1} = \left(1 - \frac{1}{2N}\right) - \left(1 - \frac{1}{2N}\right) + H_t \left(1 - \frac{1}{2N}\right)$$

$$H_{t+1} = H_t \left(1 - \frac{1}{2N}\right)$$

$$H_t = H_0 \left(1 - \frac{1}{2N}\right)^t \approx H_0 e^{-\frac{t}{2N}}$$

drift

$$\Delta H = -\frac{H_t}{2N}$$



E.g.  $N \approx 10^5$  gen. time 20 years

$\tau = 40 \cdot 10^5$  years = 4 million years  $\approx$  [common ancestor with chimp.]

$\Rightarrow$  Drift is very slow in large populations.

#### 4. Mutations and drift

: Infinite allele model = no backward mutations

$\mu$  - mutation rate = mutation probability

$A_1 \rightarrow A_2 \rightarrow A_3 \rightarrow A_4$  etc

per locus  
per generation

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

drop  $\mu^2$  terms

drop  $\frac{\mu}{N}$  terms

$$G_{t+1} = (1 - 2\mu) \left( \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G_t \right) = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) G_t - 2\mu G_t$$

$$H_{t+1} = 1 - G_{t+1} = \left(1 - \frac{1}{2N}\right) - \left(1 - \frac{1}{2N}\right) (1 - H_t) + 2\mu H_t + 2\mu$$

$$H_{t+1} = \left(1 - \frac{1}{2N}\right) H_t + 2\mu(1 - H_t)$$

$$\Delta H = H_{t+1} - H_t = -\frac{H_t}{2N} + 2\mu(1 - H_t)$$

Steady state  $\Delta H = 0$

$$\frac{H_t}{2N} = 2\mu(1 - H_t)$$

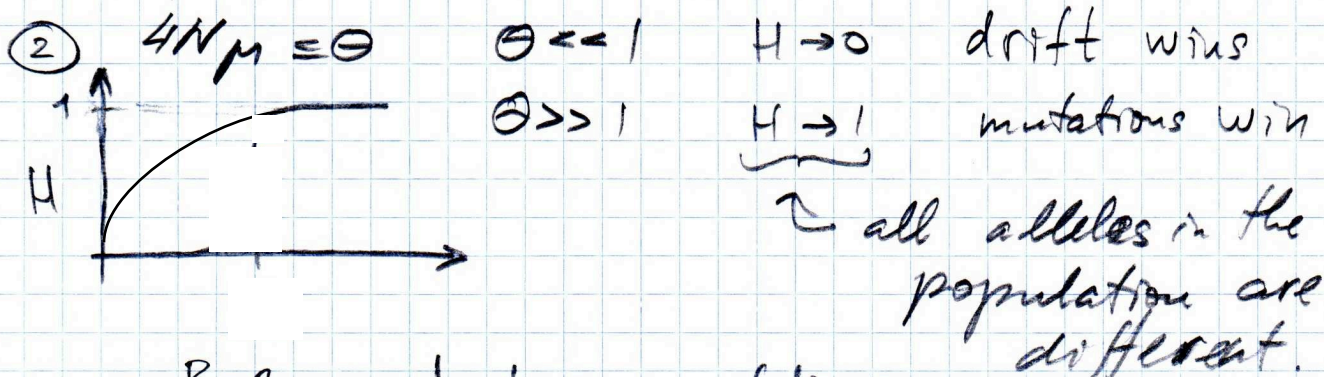
$$\frac{H_t}{1 - H_t} = 4N\mu$$

$$H_t = \frac{4N\mu}{1 + 4N\mu} ; G = \frac{1}{1 + 4N\mu}$$

$\uparrow$  heterozyg.                       $\uparrow$  homozyg.

①  $\Delta H = -\frac{H_t}{2N} + 2\mu(1 - H_t)$

see above  $\leftarrow$  drift (decreases  $H$ )      mutation (increases  $H$ )  
doesn't depend on  $N$



$\Rightarrow$  Conclusion:

Balance between mutations and drift leads to a constant (steady state) diversity of the population

## Selection and drifts Halden's equation

Selection (frequency independent selection)

$A_1 A_1$	$A_1 A_2$	$A_2 A_2$	
$w_{11}$	$w_{12}$	$w_{22}$	← fitness
$p^2$	$2pq$	$q^2$	← freq. before selection
$\frac{w_{11}}{\bar{w}} p^2$	$\frac{w_{12}}{\bar{w}} 2pq$	$\frac{w_{22}}{\bar{w}} q^2$	← after selection
↑	$1 - hs$	$1 - s$	← relative fitness

Where  $\bar{w} = p^2 w_{11} + 2p(1-p) w_{12} + (1-p)^2 w_{22}$  ← mean fitness of the population

$\frac{w_{11}}{\bar{w}}$        $\frac{w_{12}}{\bar{w}}$        $\frac{w_{22}}{\bar{w}}$  ← relative fitness

s - selection coefficient (strength of selection),  $s > 0$   
 h determines the type of selection

\* Types of selection

- h=0 A1 dominant
- h=1 A2 dominant
- $0 < h < 1$  incomplete dominance -> directional selection
- $h < 0$  overdominance -> balancing selection
- $h > 1$  underdominance -> disruptive selection

\* Special type of balancing selection -  
 - frequency-dependent selection.

\* Let's calculate the effect of selection on allele freq. p  
 (in one generation)

$$p_{t+1} = p^2 \frac{w_{11}}{\bar{w}} + \frac{2(1-p)p}{2} \frac{w_{12}}{\bar{w}}$$

$$p' = p^2 \frac{w_{11}}{\bar{w}} + 2 \frac{(1-p)p}{2} \frac{w_{12}}{\bar{w}}$$

$$\bar{w} = p^2 w_{11} + 2p(1-p) w_{12} + w_{22} (1-p)^2$$

$$\frac{d\bar{w}}{dp} = 2pw_{11} + 2(1-2p)w_{12} - 2w_{22}(1-p)$$

$$\Delta p = \frac{1}{\bar{w}} [p^2 w_{11} + (1-p)p w_{12} - p[p^2 w_{11}] - 2p^2(1-p)w_{12} - p(1-p)^2 w_{22}]$$

$$= \frac{1}{\bar{w}} [w_{11}(1-p)p^2 + w_{12}(1-p)p(1-2p) - w_{22}p(1-p)^2]$$

$$= \frac{p(1-p)}{2\bar{w}} [2w_{11}p + 2w_{12}(1-2p) - 2w_{22}(1-p)]$$

$$\Delta p = \frac{p(1-p)}{2\bar{w}} \frac{d\bar{w}}{dp} \quad \text{Wright Equation}$$

$$= \frac{p(1-p)}{2} \frac{d \log \bar{w}}{dp}$$

1. Direction of evolution is determined by  $\frac{d \log \bar{w}}{dp} \leftrightarrow$  adaptive landscape

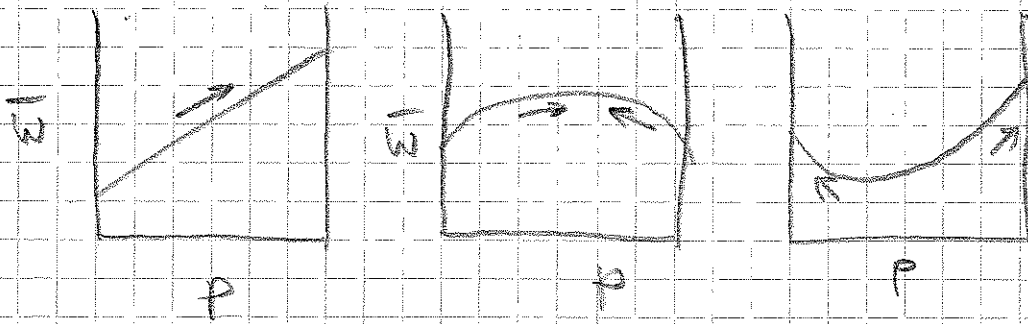
$$\frac{d \log \bar{w}}{dp} > 0 \Rightarrow \Delta p > 0$$

2. Stable equilibria are points where

$$\frac{d \log \bar{w}}{dp} = 0, \text{ i.e.}$$

local maxima of  $\bar{w}$





3. Evolution maximizes mean fitness of the population

\* Selection + drift

Probability of fixation

Selection  $\langle \Delta p \rangle = \frac{p(1-p)}{2\bar{w}} \cdot \frac{d\bar{w}}{dp}$  assuming  $h = 1/2$

$\bar{w} = p^2 + 2p(1-p)(1-hs) + (1-p)^2(1-s) \approx 1 + ps - s = 1 - qs$

$\langle \Delta p \rangle = \frac{p(1-p)}{2(1+ps-s)} \cdot s \approx \frac{p(1-p)}{2} s$  assuming  $s \ll 1$

$\pi(p) \langle \Delta p \rangle + \frac{1}{2} \pi''(p) \langle \Delta p^2 \rangle = 0$

$\pi(p)$   
||  
 $M(p)$

$\frac{1}{2} \pi''(p) \langle \Delta p^2 \rangle$   
||  
 $V(p) = \text{Var}(p) + \langle \Delta p \rangle^2$

$\frac{1}{2} \pi''(p) V(p) + M \pi'(p) = 0$

$\frac{\pi''}{\pi'} = -\frac{2M}{V}$

$M(p) \equiv \langle \Delta p \rangle = \frac{p(1-p)s}{2}$

$V(p) = \frac{p(1-p)}{2N} + \frac{s^2}{4}$

$-\log \pi' = 2 \int_0^p \frac{M(x')}{V(x')} dx' + C$

$= 2 \int_0^p Ns dx + C = 2Nsp + C$

$\pi'(p) = C_1 e^{-2Nsp} \Rightarrow \pi(p) = C_1 e^{-2Nsp} + C_2$

$\Rightarrow \pi(0) = C_1 + C_2 = 0$   
 $C_2 = -C_1$

$\pi(1) = C_1 e^{-2Ns} - C_1 = C_1 (e^{-2Ns} - 1) = 1$   
 $C_1 = \frac{1}{e^{-2Ns} - 1}$

boundary  
 $\pi(0) = 0$   
 $\pi(1) = 1$

# Haldein's equation

$s > 0 \Leftrightarrow A_2$  disadvantageous mutation

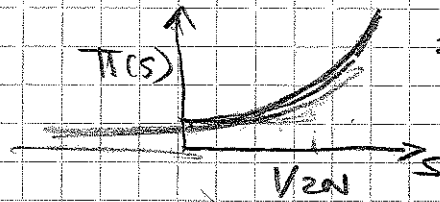
$$\pi(\bar{p}) = \frac{1 - e^{-2Ns p}}{1 - e^{-2Ns}}$$

1) New mutations

$$p = \frac{1}{2N}$$

$$\pi\left(\frac{1}{2N}\right) = \frac{1 - e^{-s}}{1 - e^{-2Ns}} \approx \frac{s}{1 - e^{-2Ns}} \approx s \text{ for } 2Ns \gg 1$$

$$\frac{s}{2Ns} = \frac{1}{2N} \text{ for } 2Ns < 1$$



2) For  $h \neq \frac{1}{2}$

$$\pi\left(\frac{1}{2N}\right) \approx 2(1-h)s$$

selective advantage of heterozygote

depends on the fitness of heterozygote

3) Deletions  $\pi_1 = \pi$

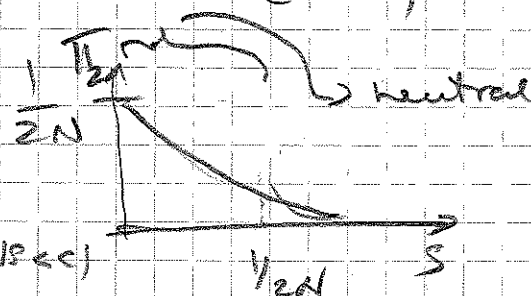
$$\pi_2 = 1 - \pi_1(1-q) = \frac{e^{-2Ns q} - 1}{e^{-2Ns} - 1} = \frac{1 - e^{-2Ns(1-q)}}{1 - e^{-2Ns}} = \frac{1 + e^{-2Ns q}}{e^{-2Ns} - 1}$$

New mutation

$$q = \frac{1}{2N}$$

$$\pi_2 = \frac{s}{e^{-2Ns} - 1}$$

$$\approx \frac{1}{2N} \text{ for } 2Ns \ll 1$$



$$\frac{1}{2} = \frac{1}{2} \frac{2N_{MS}}{2N_{MS} - 1}$$

rate of fixation

rate of a a. subst = rate of neutral subst.

8.08

Evolution, lecture 2

1. Selection (Wright's equation)

$A_1$	$A_2$	
$p$	$1-p$	frequency
$1$	$1-s$	fitness

in the next generation  $0 < s < 1 \leftarrow$  small disadvantage

$p \cdot \frac{1}{\bar{w}}$      $(1-p) \cdot \frac{1-s}{\bar{w}}$     frequency

$p \frac{1}{\bar{w}} + \frac{(1-p)(1-s)}{\bar{w}} = 1$      $\left\{ \begin{array}{l} \bar{w} - \text{normalization} \\ \bar{w} - \text{mean fitness of} \\ \text{the population} \end{array} \right.$

$\bar{w} = p + (1-p)(1-s)$   
 $= p + 1 - p - s + ps = 1 - s + ps$

$\Delta p = p \cdot \frac{1}{\bar{w}} - p = p \frac{1-\bar{w}}{\bar{w}} = p \frac{1-1+s-ps}{\bar{w}}$

$\Delta p = p(1-p) \cdot \frac{s}{\bar{w}}$

notice  $\frac{d\bar{w}}{dp} = s$

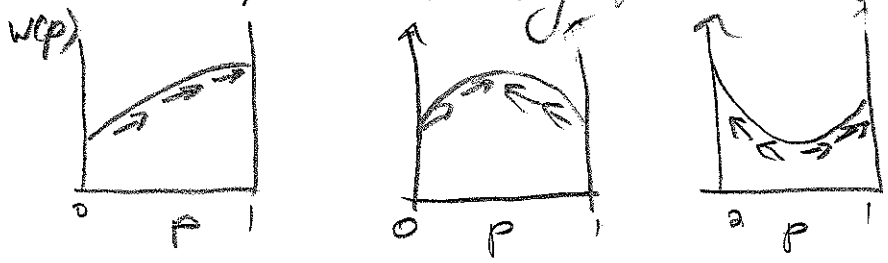
$\Rightarrow \Delta p = p(1-p) \frac{d\bar{w}}{dp} \cdot \frac{1}{\bar{w}} = p(1-p) \frac{d \log \bar{w}}{dp}$

Wright's equation

Conclusions:

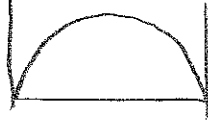
- 1) Allele frequency changes to increase fitness of the population  
 $(\Delta p \sim \frac{d \log \bar{w}}{dp})$

$\bar{w}(p)$  arbitrary function of  $p$



2) The rate of change (rate of evolution) is higher for more diverse populations

$p(1-p) \leftarrow$  variance of the population



$$\Delta p \sim p(1-p)$$

## 2. Selection + drift (Haldane's equation)

Probability of fixation  $\pi(p)$

Backward Kolmogorov equation

$$\langle \Delta p \rangle \pi'(p) + \frac{\langle \Delta p^2 \rangle}{2} \pi''(p) = 0$$

Drift  $\langle \Delta p^2 \rangle = \frac{p(1-p)}{2N}$        $\langle \Delta p \rangle = 0$

Selection  $\langle \Delta p \rangle = p(1-p)s$       (assuming  $w=1$ )

$$p(1-p)s \pi'(p) + \frac{p(1-p)}{2N} \pi''(p) = 0$$

Boundary

$$\begin{aligned} \pi(0) &= 0 \\ \pi(1) &= 1 \end{aligned}$$

$$\pi(p) = \frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}}$$

# \* Selection and drift (Haldane's equation).

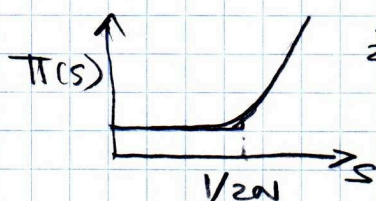
$$\pi = \frac{1 - e^{-2Nsp}}{1 - e^{-2Ns}}$$

$s > 0 \Leftrightarrow A_2$  disadvantageous mutation

1) New mutation (new  $A_1$ )

$$p = \frac{1}{2N}$$

$$\pi\left(\frac{1}{2N}\right) = \frac{1 - e^{-s}}{1 - e^{-2Ns}} \approx \frac{s}{1 - e^{-2Ns}} \approx s \text{ for } 2Ns \gg 1$$



$$\frac{s}{2Ns} = \frac{1}{2N} \text{ for } 2Ns < 1$$

2) For  $h \neq \frac{1}{2}$

$$\pi\left(\frac{1}{2N}\right) \approx 2(1-h)s$$

selective advantage of heterozygote

depends on the fitness of heterozygote

3) Deleterious allele  $A_2$

$$\bar{\pi}_1 = \pi$$

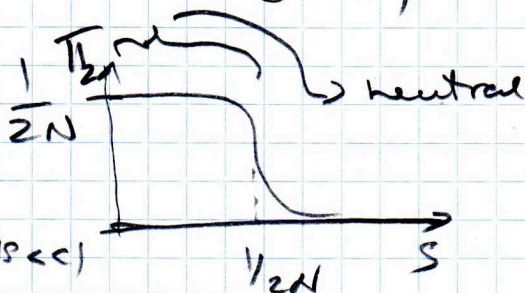
$$\pi_2 = 1 - \bar{\pi}_1(1-q) = \frac{e^{2Nsq} - 1}{e^{2Ns} - 1} = \frac{1 - e^{-2Ns} - 1 + e^{-2Ns(1-q)}}{1 - e^{-2Ns}} =$$

New mutation (new  $A_2$ )

$$q = \frac{1}{2N}$$

$$\pi_2 = \frac{s}{e^{2Ns} - 1}$$

$$\approx \frac{1}{2N} \text{ for } 2Ns \ll 1$$



$$k = \frac{\pi(1/2) N \mu}{e^{2N\mu} - 1} = \frac{2N\mu s}{e^{2N\mu} - 1}$$

↑ rate of fixation

rate of a. a. subst < rate of neutral subst.

## Effective population size

Recall

$$H_{t+1} = H_t \left(1 - \frac{1}{2N}\right)$$

if  $N_t$ , then

$$H_{t+1} = H_t \left(1 - \frac{1}{2N_t}\right)$$

$$H_t = H_0 \prod_{i=0}^{t-1} \left(1 - \frac{1}{2N_i}\right)$$

H - heterozygosity  
(fraction of different nucleotides)

compare to  $H_t = H_0 \left(1 - \frac{1}{2N_e}\right)^t$

↑  
effective pop. size

Then

$$\left(1 - \frac{1}{2N_e}\right)^t = \prod_{i=0}^{t-1} \left(1 - \frac{1}{2N_i}\right)$$

$$\prod_{i=1}^n (1+x_i) = e^{\sum \log(1+x_i)}$$

$x \ll 1$

using  $\log(1+x) \approx x - \frac{x^2}{2} \approx x$   
we obtain for  $x \approx 0.1$

$$= e^{\sum x_i} \approx 1 + \sum x_i$$

$$e^{-t/2N_e} = e^{-\sum_{i=0}^{t-1} \frac{1}{2N_i}}$$

$$N_e = \left( \frac{1}{t} \sum_{i=0}^{t-1} \frac{1}{N_i} \right)^{-1} \leftarrow \text{harmonic mean}$$