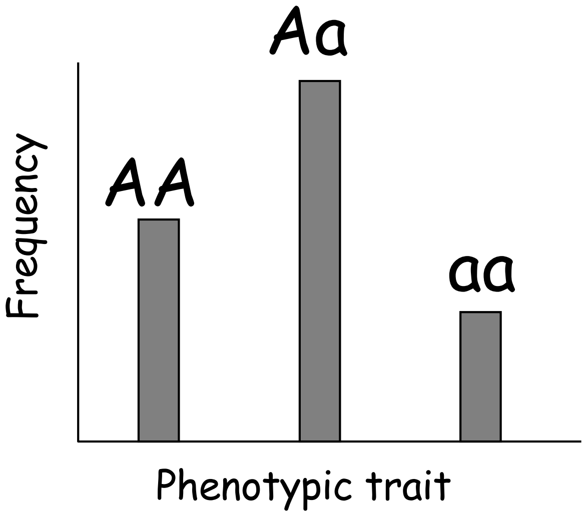
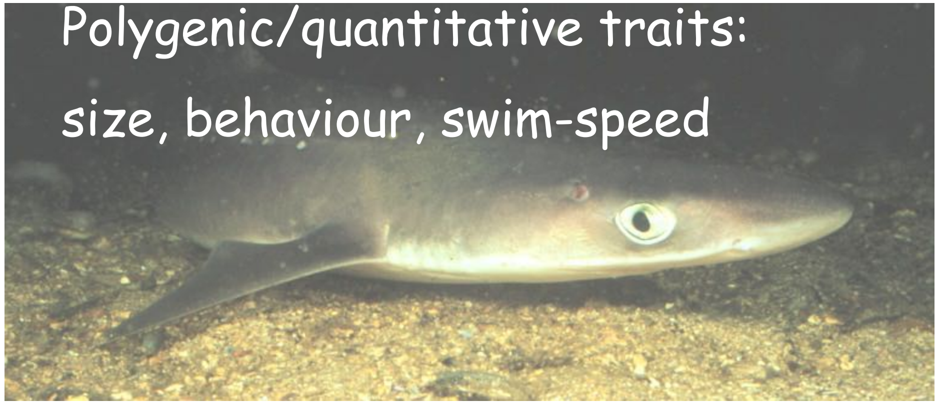
Population Genetics

Brief overview of basic concepts and processes
Kerstin Johannesson

Single and polygenic traits

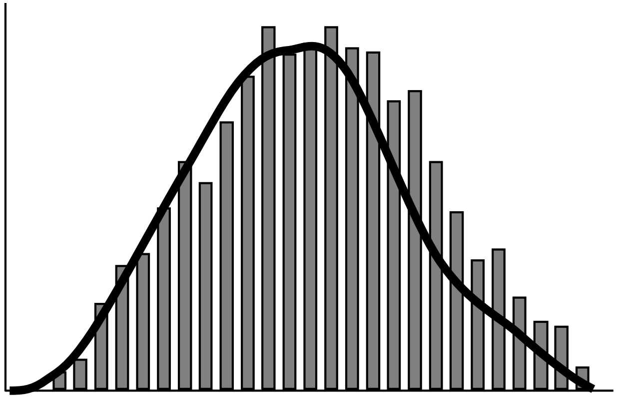
Few-loci traits:
colour, sex-determination
inherited diseases





Polygenic/quantitative traits:

size, behaviour, swim-speed



How genetic variation arises

1. **Mutation**

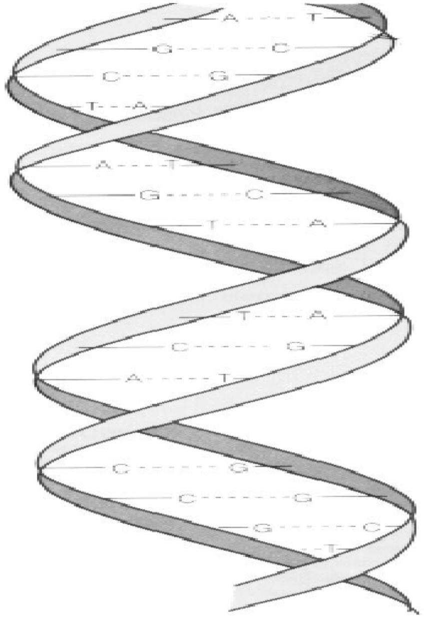
a. Deletions or insertions of nucleotides

protein	1×10^{-6}
microsatellites	$0.7\text{-}10 \times 10^{-5}$
SNPs	$1\text{-}10 \times 10^{-9}$

1 ind. out of 1.000.000 carries a new mutation in a protein

b. Chromosomal rearrangements
(inversions, duplications, fusions, fissions…)

- Results in big or no problems




Dog whelk

The fate of a new mutation:

Initial frequency = $1/2N_a$

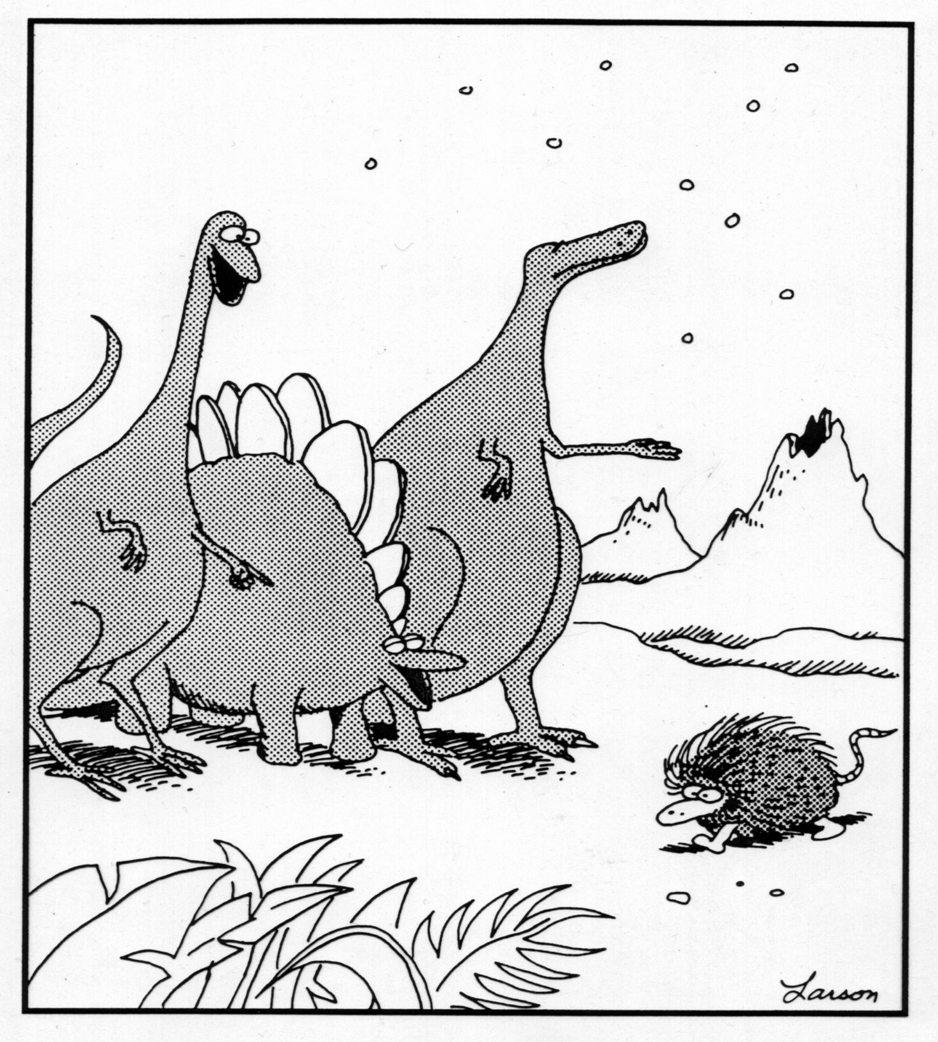
carried as a heterozygote by only one individual of a population

Positive for fitness:

A_1A_1	$1+s$
A_1A_2	$1+s/2$
A_2A_1	1

Probability of fixation* = $2s\ (N_e/N_a)$, if $2s$ is small and N_e is large

*Fixation = frequency of 1.00



The fate of a new mutation:

Initial frequency = $1/2N_a$

Negative for fitness:

A_1A_1	$1-s$
A_1A_2	$1-s/2$
A_2A_2	1

Mutation will disappear if $2N_e s \gg 1$ (relatively large population)
but it will take $2(N_e/N_a)[\ln(2N_a/2N_e s)+1-0.58]$ to get rid of it

For $N_e=N_a=10\,000$ and $s=1\%$ disadvantage, ≈9.6 generations
For $N_e=N_a=10\,000$ and $s=10\%$ disadvantage, ≈5 generations

Neutral for fitness:

A*A*	1
A*A	1
AA	1

Probability of fixation = $1/2N_a$
Probability of loss = $1-1/2N_a$

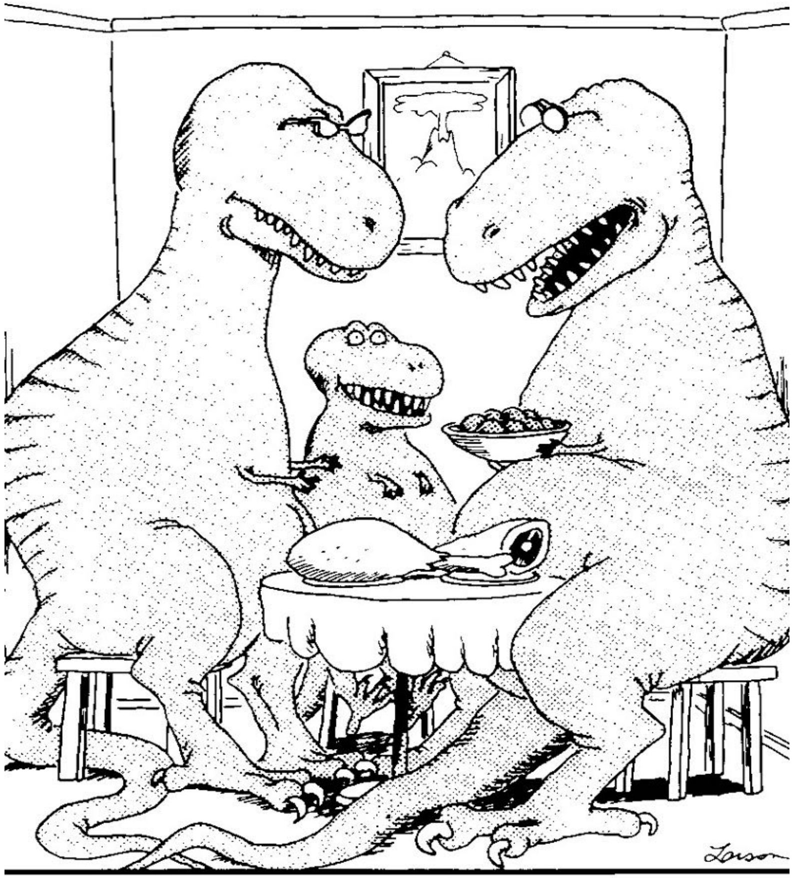
Time to fixation = $4N_e$ generations

For $N_e=N_a=10\,000$, 40 000 generations

Time for disapperance = $2(N_e/N_a)\ \ln(2N_a)$

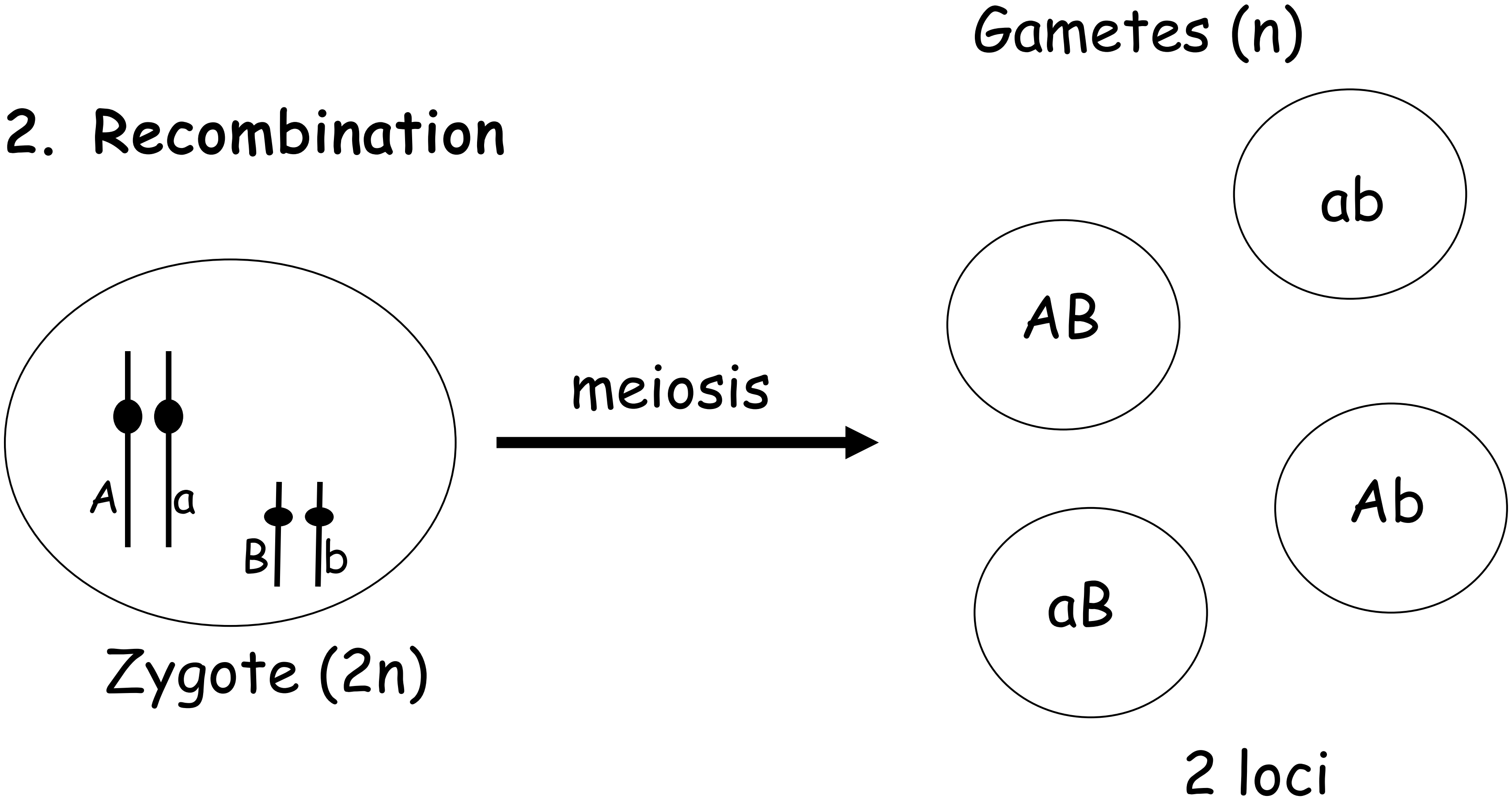
For $N_e=N_a=10\,000$, ≈ 20 generations

Positive, negative or neutral?



"Hey! I'm *trying* to pass the potatoes! ... Remember, my forearms are just as useless as yours!"

2. Recombination



Gametes (n)

meiosis

Zygote (2n)

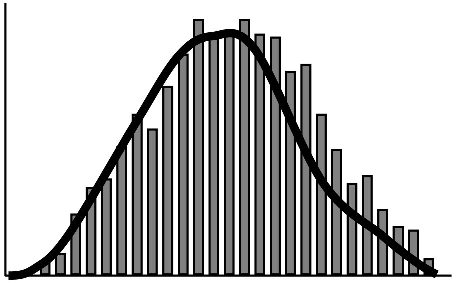
ab

AB

Ab

aB

2 loci



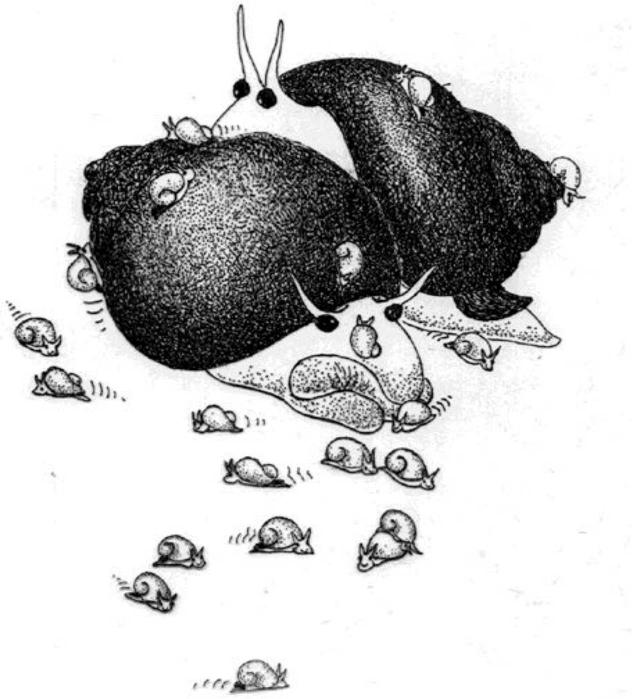
Extended to polygenic traits - infinite # of combinants

Hardy-Weinberg equilibrium

(the "null hypothesis" of population genetics)

Populations maintain stable allele frequencies from generation to generation if:

1. Diploid organisms
2. Sexual reproduction
3. Random mating
4. Nonoverlapping generations
5. No selection
6. No migration
7. No mutation
8. Infinite population size



$(AA)=p^2$ p=freq of allele A

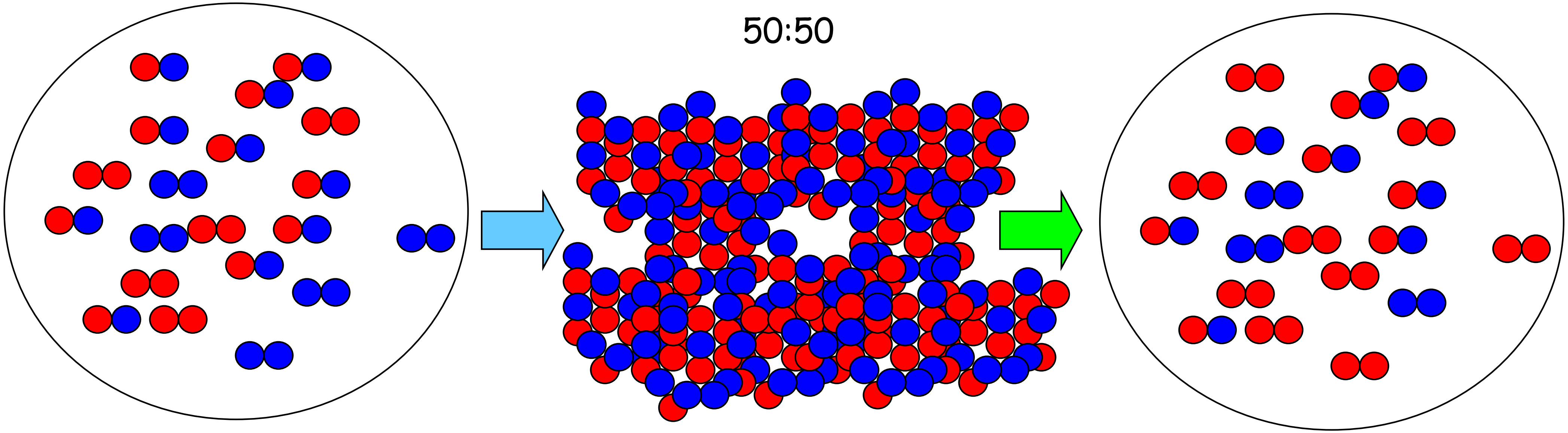
$(Aa)=2pq$ q=freq of allele a

$(aa)=q^2$

With finite population size → genetic drift

$N_e = 20$
$p = q = 0.5$

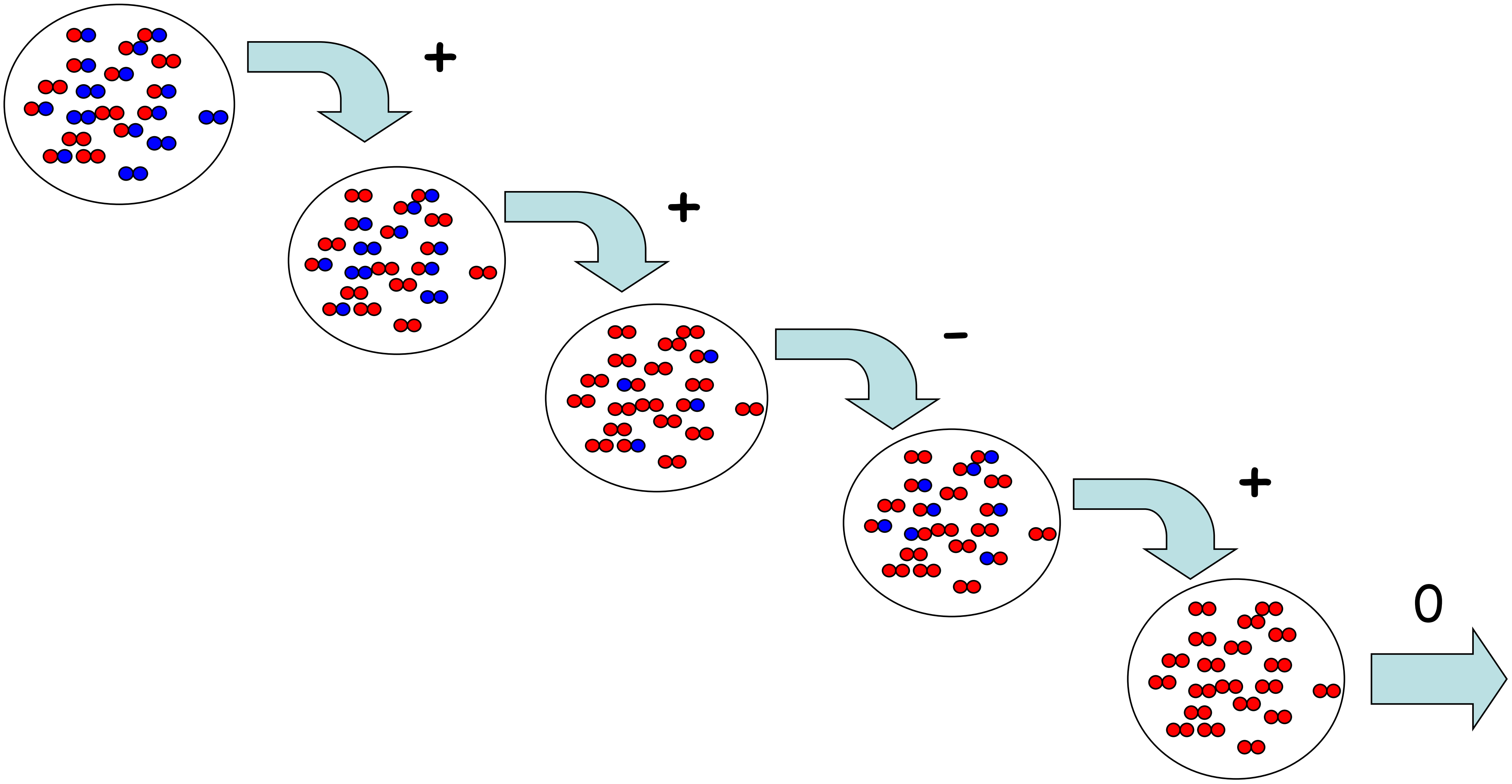
$N_e = 20$
$p = 0.7$ och $q = 0.3$



An infinite number of gametes are produced

New adults are sampled *at random* among the gametes

Drift removes genetic variation



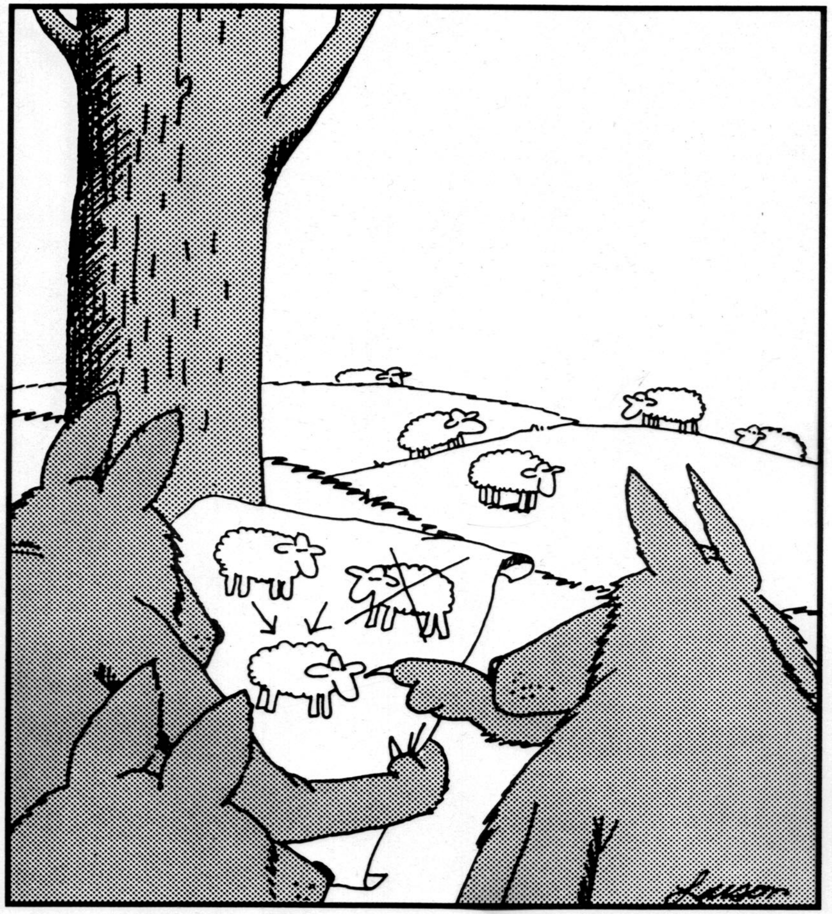
Random genetic drift is strongest if

a) populations are **small**

and/or

b) alleles are **neutral**

Selection most often violates HW equilibrium



Natural selection at work

Large populations:

Allele frequencies will change according to genotype fitnesses

Genotype	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Fitness	w_{11}	w_{12}	w_{22}

Mean fitness:

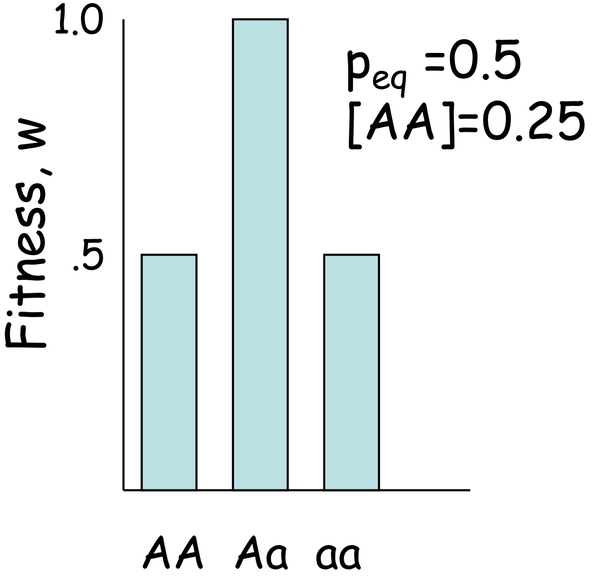
$w = p^2 w_{11} + 2pq\, w_{12} + q^2 w_{22}$

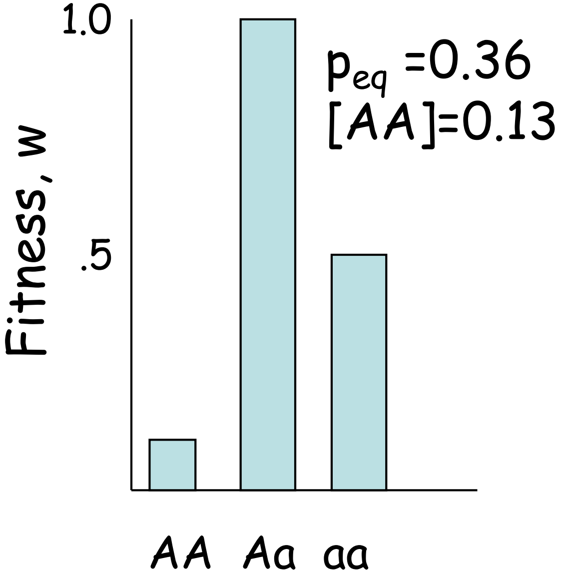
Change in *p* per generation:

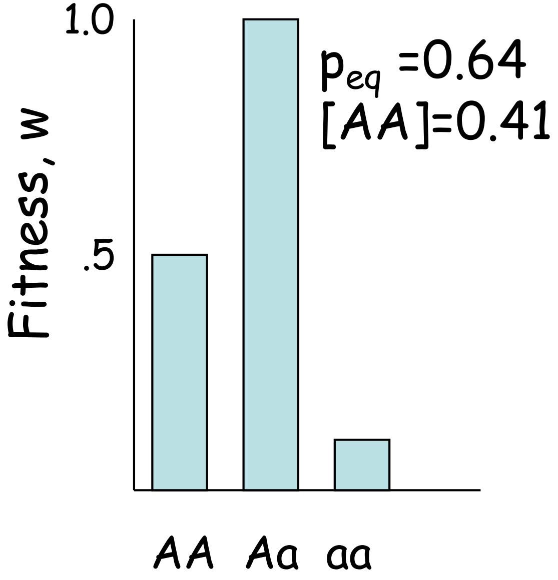
$\Delta p = pq[p(w_{11} - w_{12}) + q(w_{12} - w_{22})] / w$

Constant selection generates equilibrium frequencies

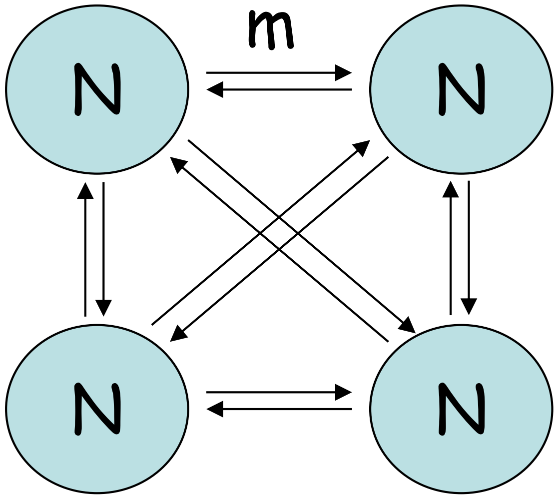
$w_{11} > w_{12} > w_{22}$	**A** will be fixed and **a** lost
$w_{11} < w_{12} < w_{22}$	**a** will be fixed and **A** lost


p_{eq} =0.5
[AA]=0.25


p_{eq} =0.36
[AA]=0.13

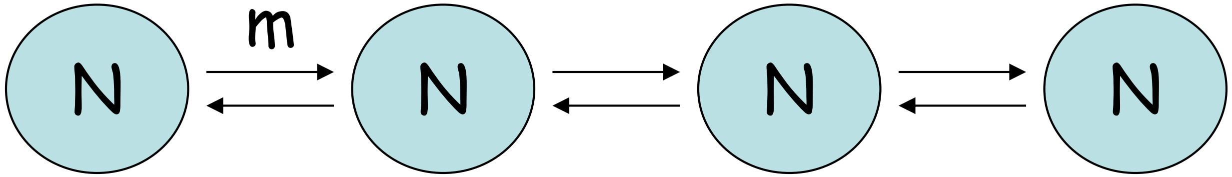

p_{eq} =0.64
[AA]=0.41

Gene flow



Island model

$N = (N_e)$
m = proportion of N migrating each generation
Nm = number of migrants



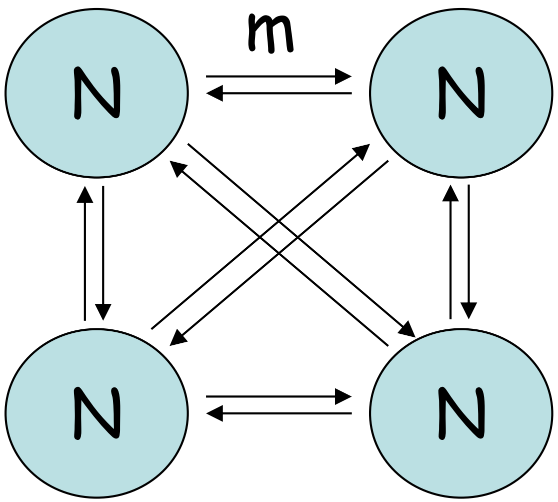
Stepping-stone model

Differentiation between populations (F_{ST})

F_{ST} = Degree of genetic differentiation between populations

$F_{ST} = 1$, max differentiation	Fixed for different alleles in a SNP locus
$F_{ST} = 0$, no differentiation	Same allele frequencies in a SNP locus

Gene flow, mutation and drift in an island model (no selection)

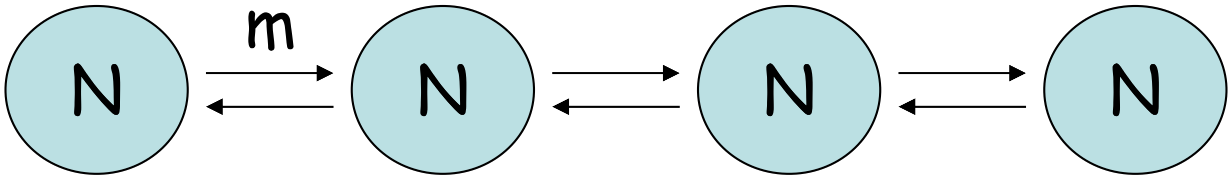


Equilibrium between
drift (loss of variation),
migration and mutation (adding variation)

For $\mu = 10^{-6}$, $m = 0.1$, $N = 50$

Island model:
$F_{ST} = 1/[4N(m+\mu) +1]$

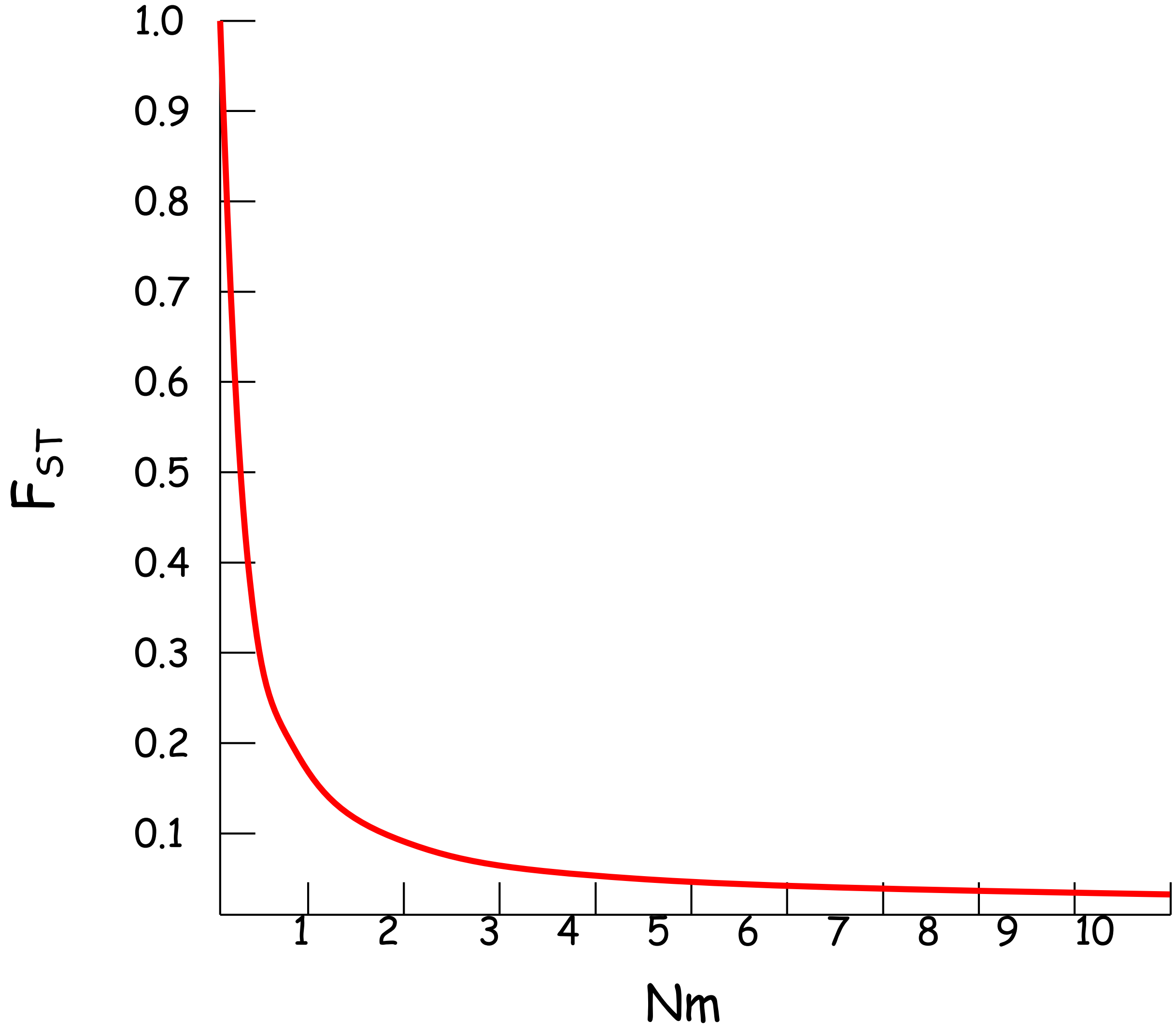
0.33

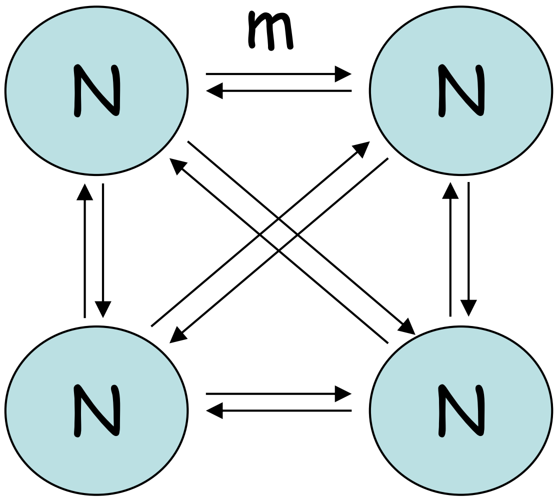


Stepping-stone model:
$F_{ST} = 1/[(4N\sqrt{2m\mu}) +1]$

0.97

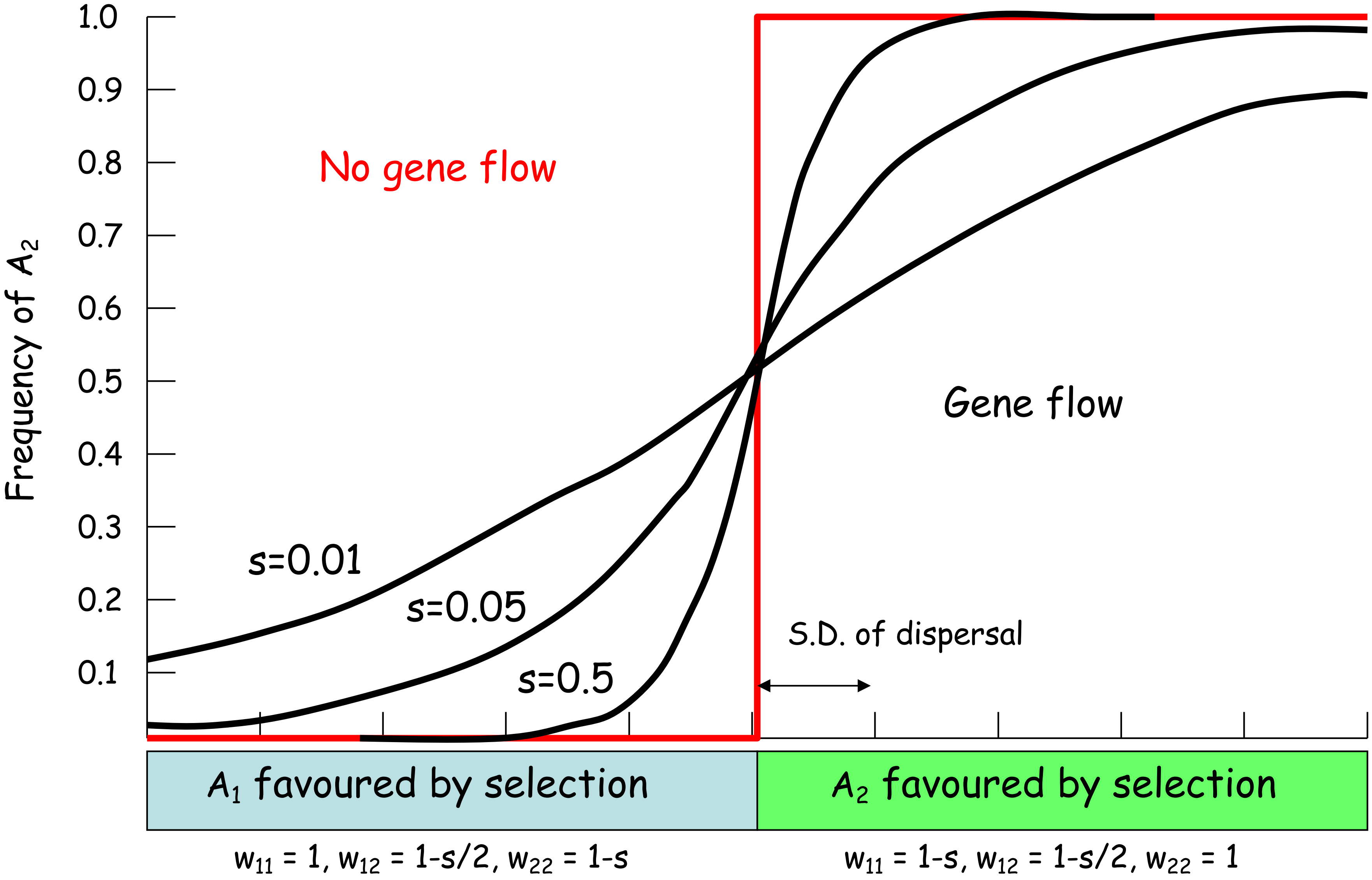
Gene flow and drift - island model



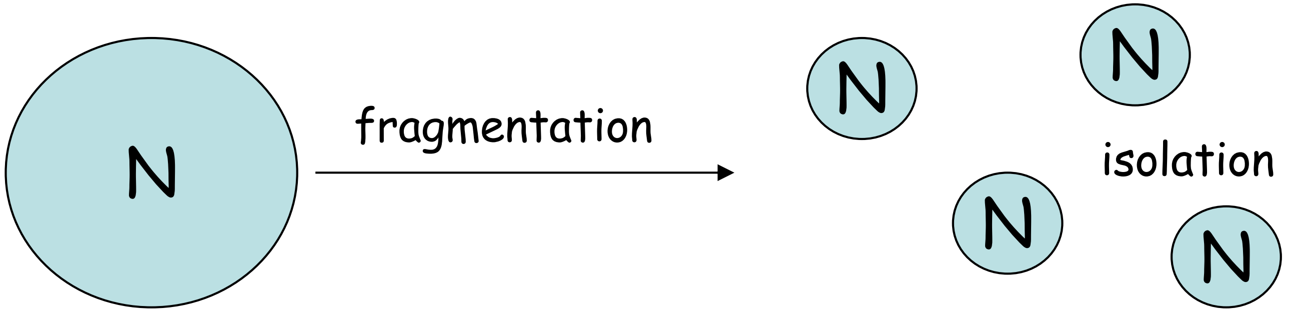


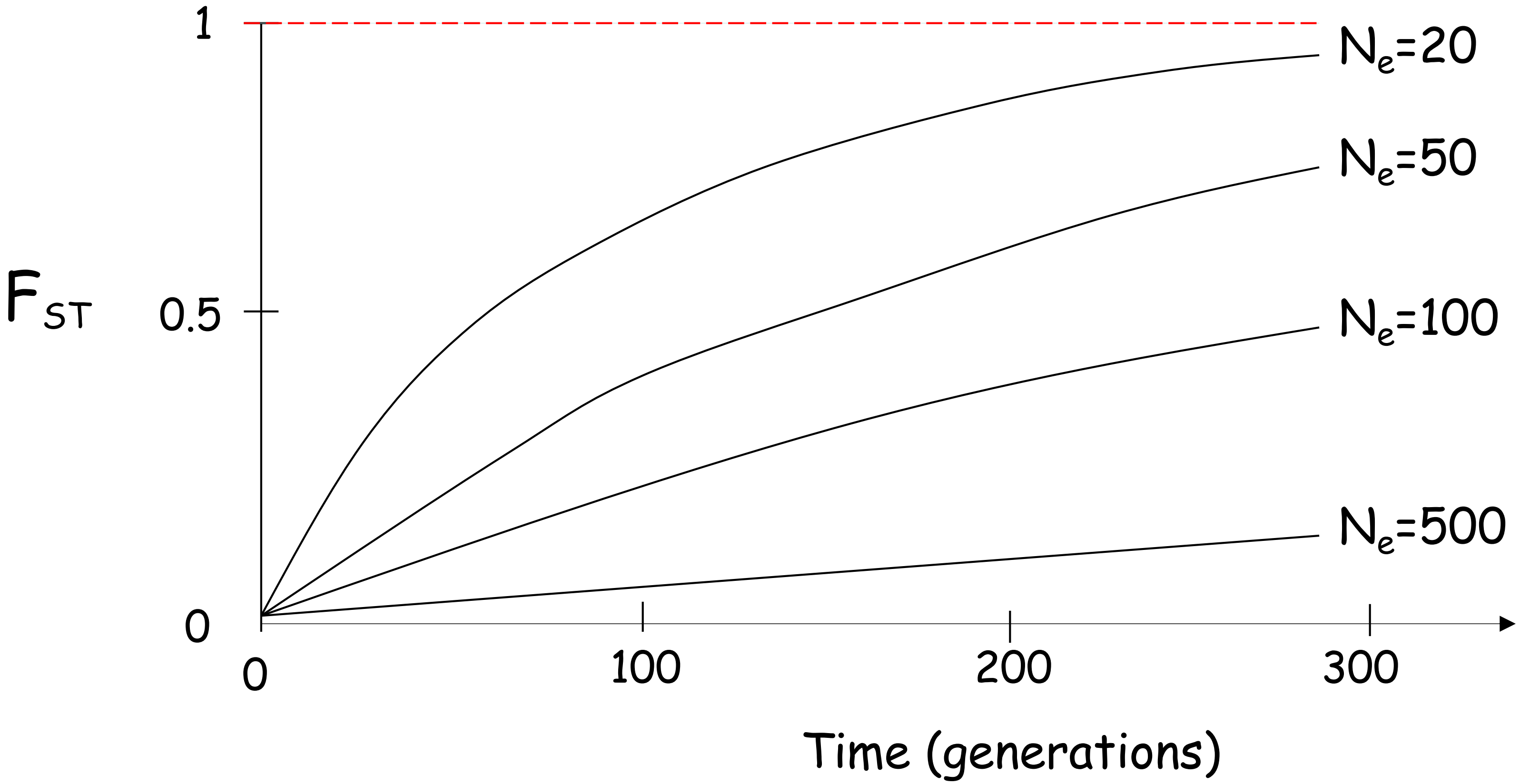
Island model:
$F_{ST} = 1/(4Nm + 1)$

Gene flow and selection over an ecotone

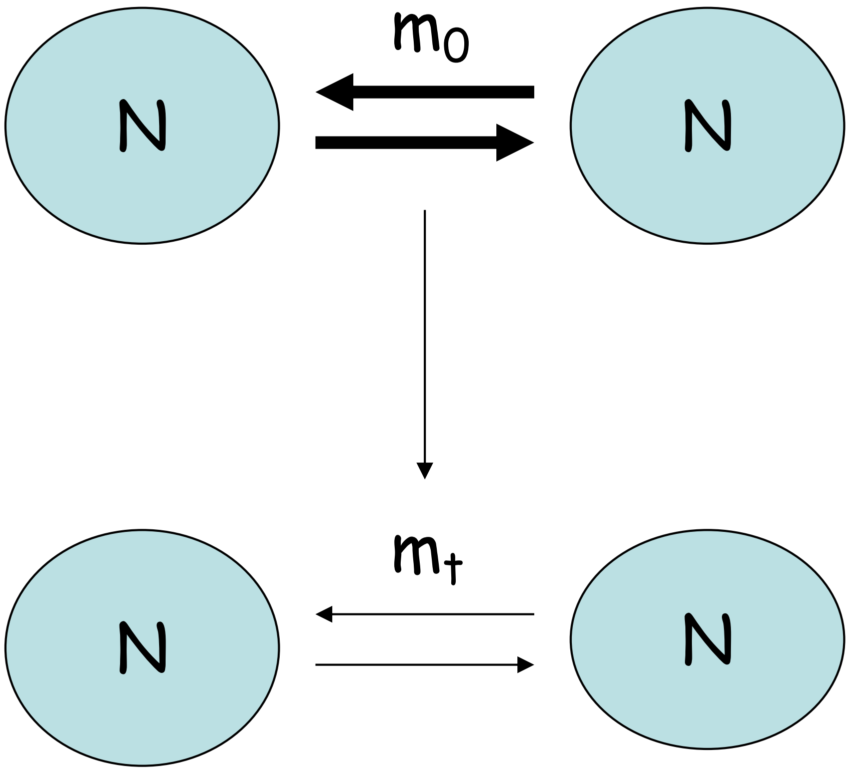


Population fragmentation and changes in F_{ST}





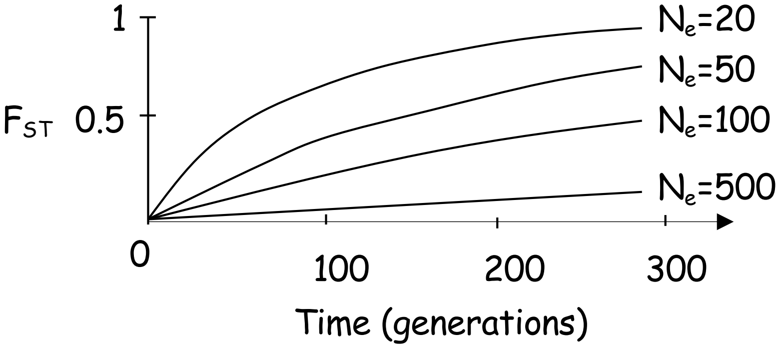
Non-equilibria is likely to be common



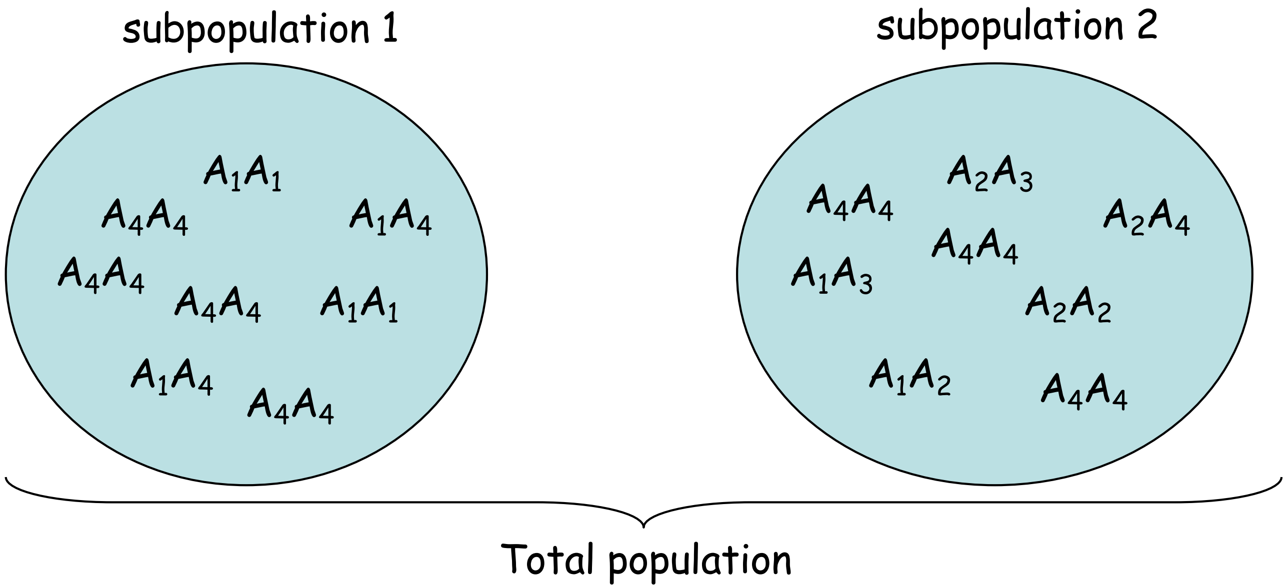
N = 1000
$m_0 = 0.1$
$F_{ST} = 0.0025$

N = 1000
$m_0 = 0.001$
$F_{ST} = 0.2$

$$F_{ST} = 1/(4Nm + 1)$$



More about F-statistics



F_{IS}, inbreeding coefficient (deviation from random mating) within a subpopulation

F_{IT}, inbreeding coefficient (deviation from random mating) between/among subpopulations

F_{ST}, differentiation between/among subpopulations

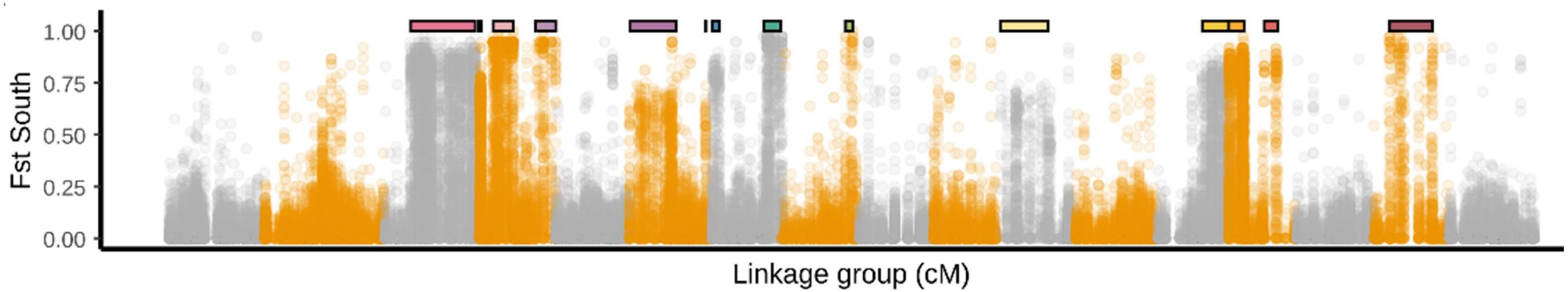
$$F_{IS} = 1-(H_{obs}/H_{exp})$$

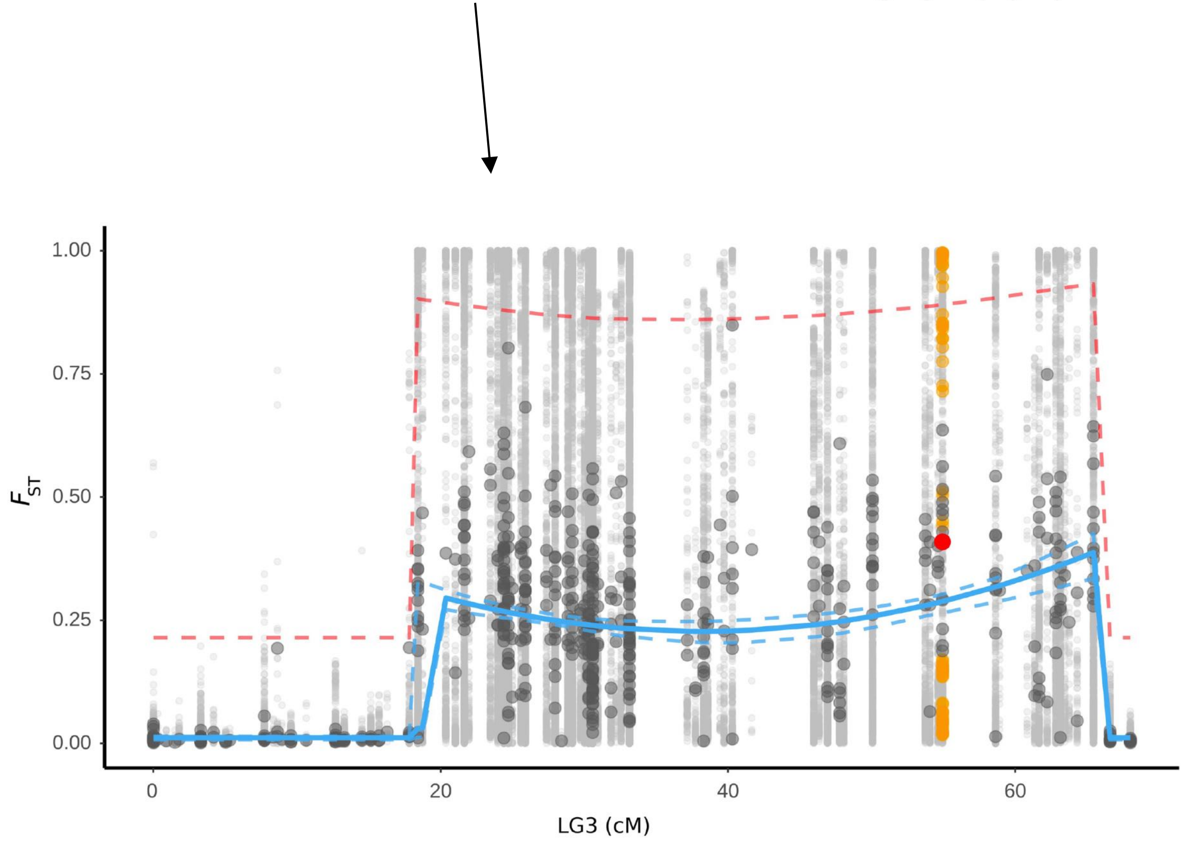
$$F_{ST} = 1-(\text{average } H_{\text{exp of all subpops}}/\text{total } H_{exp})$$

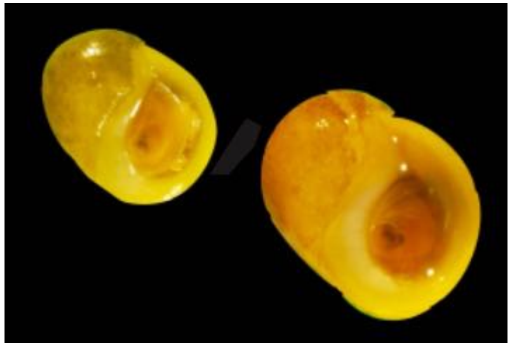
$$F_{IT} = 1-(H_{obs}/\text{total } H_{exp})$$

$$\boxed{F_{ST} = (F_{IT} - F_{IS})/(1-F_{IS})}$$

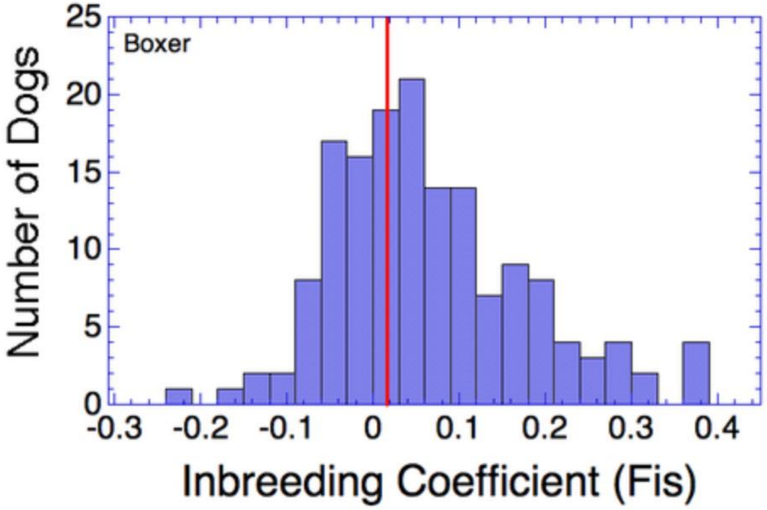
F_{ST} – still one of the most used indexes of genetic differentiation



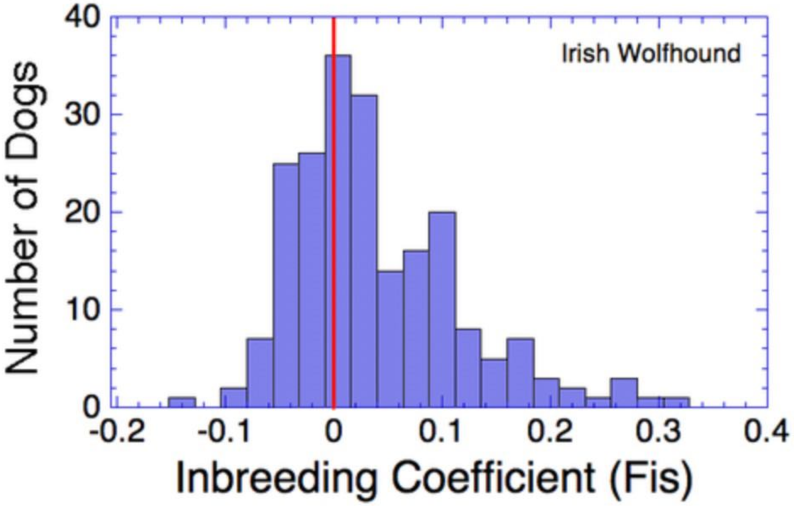




F_{IS} – is a classical index of inbreeding…



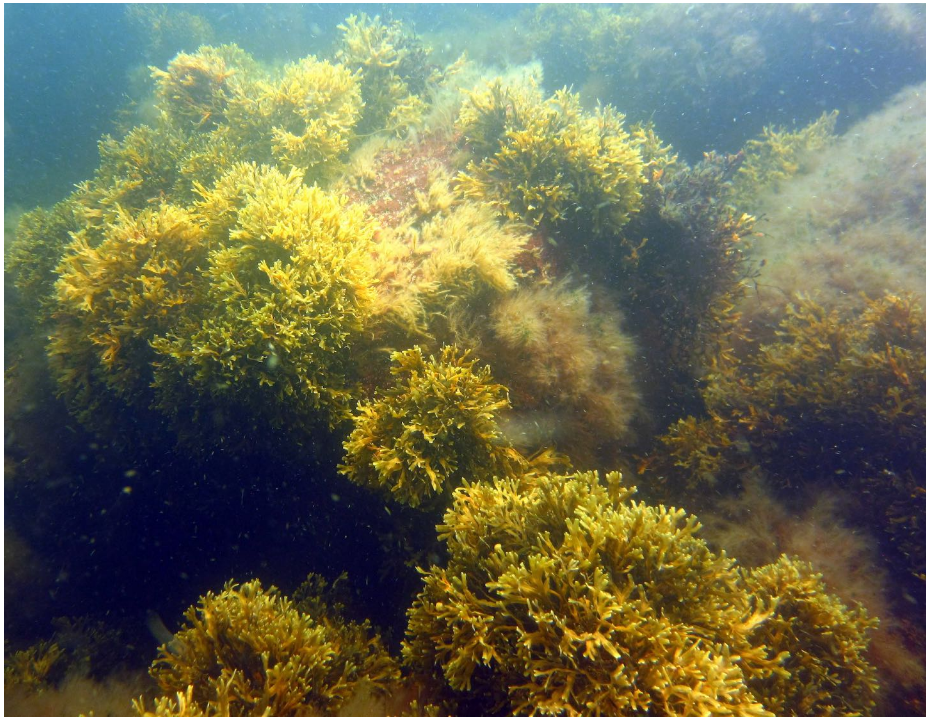
$F_{IS} = 1 - (H_{obs}/H_{exp})$



positive values → $H_{obs} < H_{exp}$

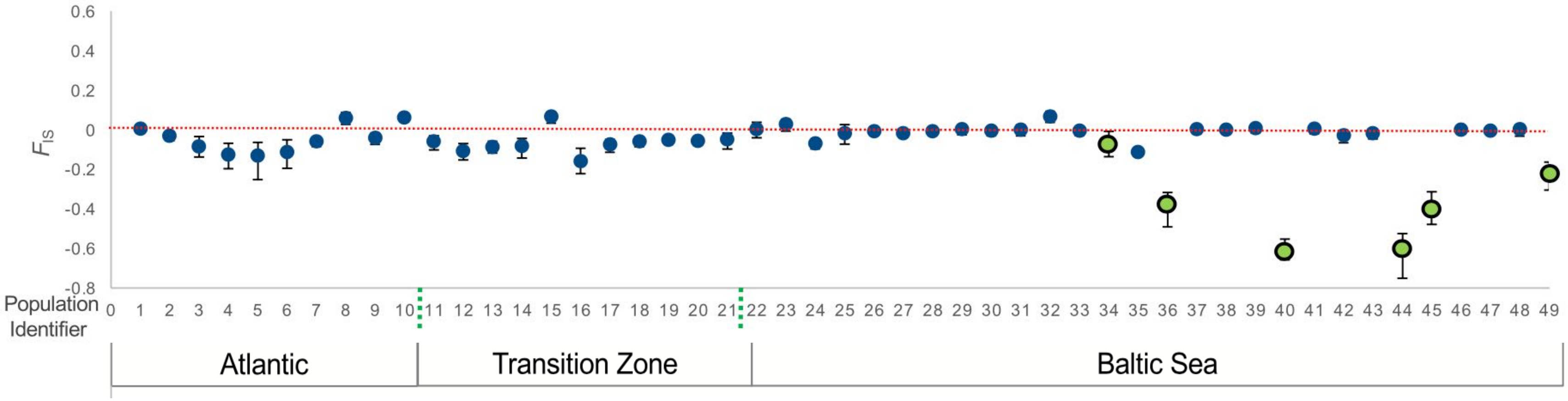
Source: https://www.instituteofcaninebiology.org/

F_{IS} and may also be used to estimate cloning

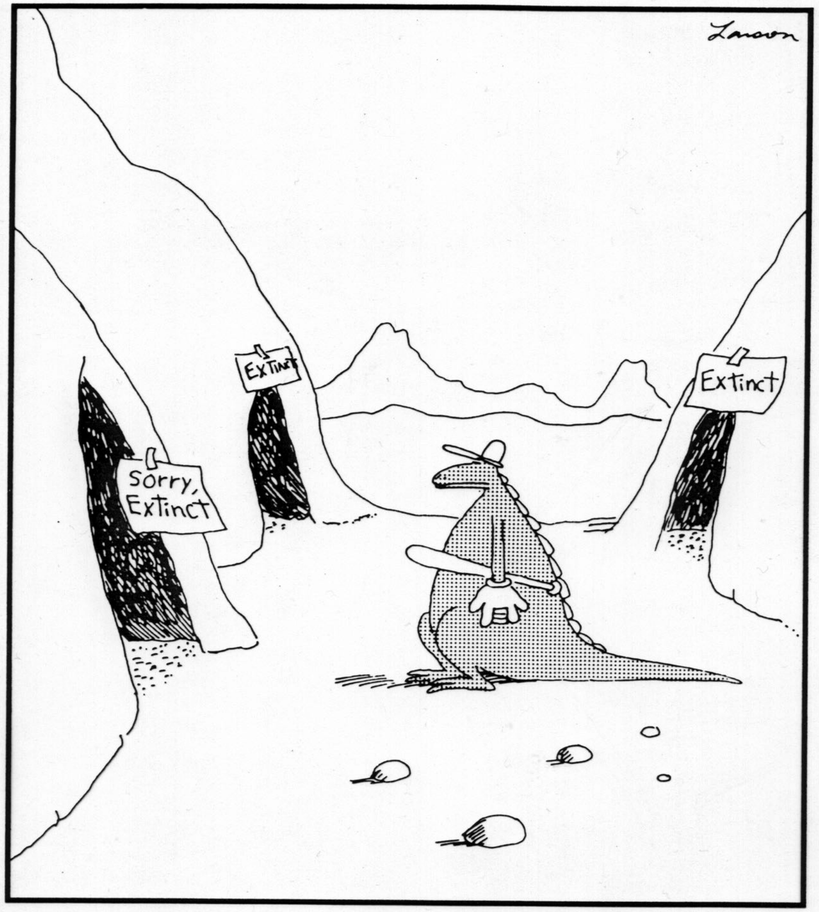


$F_{IS} = 1 - (H_{obs}/H_{exp})$

negative values → $H_{obs} > H_{exp}$



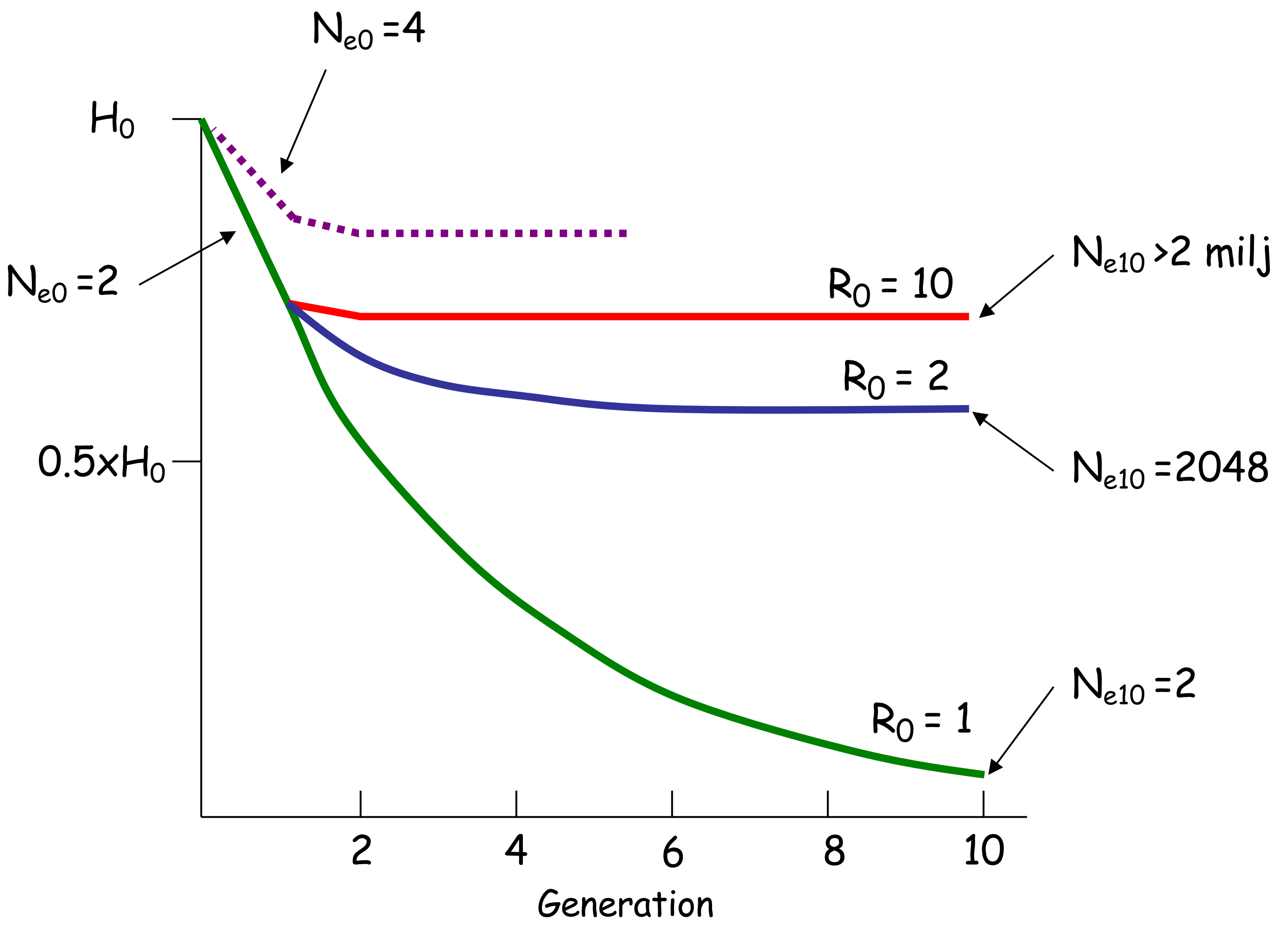
Genetic effects of population bottlenecks



Suddenly, Bobby felt very alone in the world.

Genetic variation is lost in relation to effective population size (N_e) and (population increase) R_0

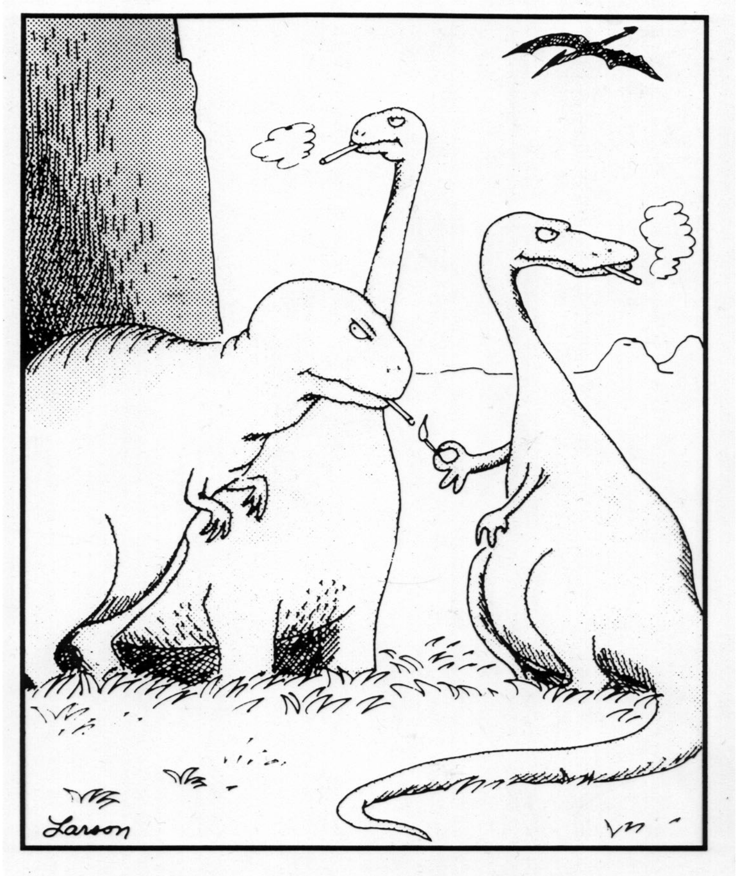
1/(2N) of the heterozygosity is lost per generation



Lost genetic variation is restored by mutation and gene flow (if present)

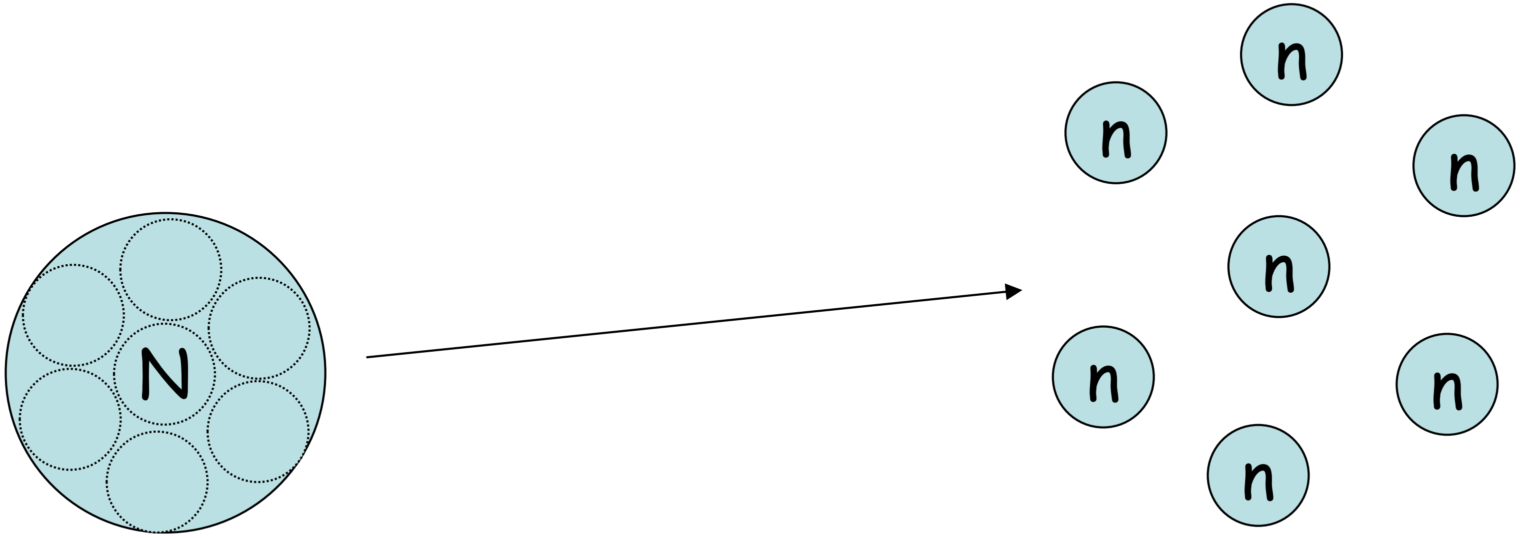
If only mutations - restoration will take 100,000s of generations for specific SNPs (see mutation rates),
but will be more rapid for quantitative traits (e.g., 100-1000 generations)

Demographic inferences of DNA sequencing data can be used to trace historic bottlenecks



The real reason dinosaurs became extinct

The Wahlund effect



$[A_1 A_1] = p^2$
$[A_1 A_2] = 2pq$
$[A_2 A_2] = q^2$

$[A_1 A_1] = p^2 + \sigma_p^2$
$[A_1 A_2] = 2pq - 2\sigma_p^2$
$[A_2 A_2] = q^2 + \sigma_p^2$

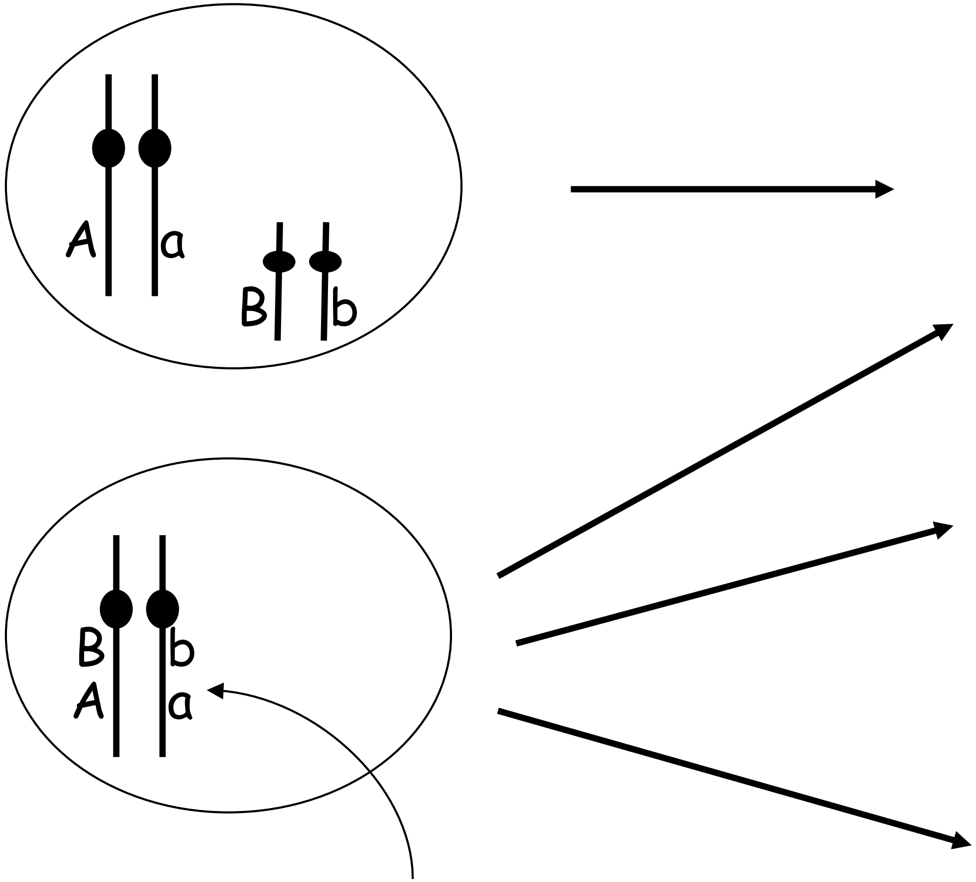
Fragmentation creates a deficiency of heterozygotes
- a Wahlund effect

recombined gametes

linked gametes

Genetic linkage

r, recombination fraction
D, disequilibrium
$D=p_{AB}p_{ab}-p_{Ab}p_{aB}$
$D_{max}=p_A q_b$ or $p_a q_B$



crossing-over
breaks linkage

AB-25%
Ab-25%
aB-25%
ab-25%

free recombination
r =0.5
D=0.0

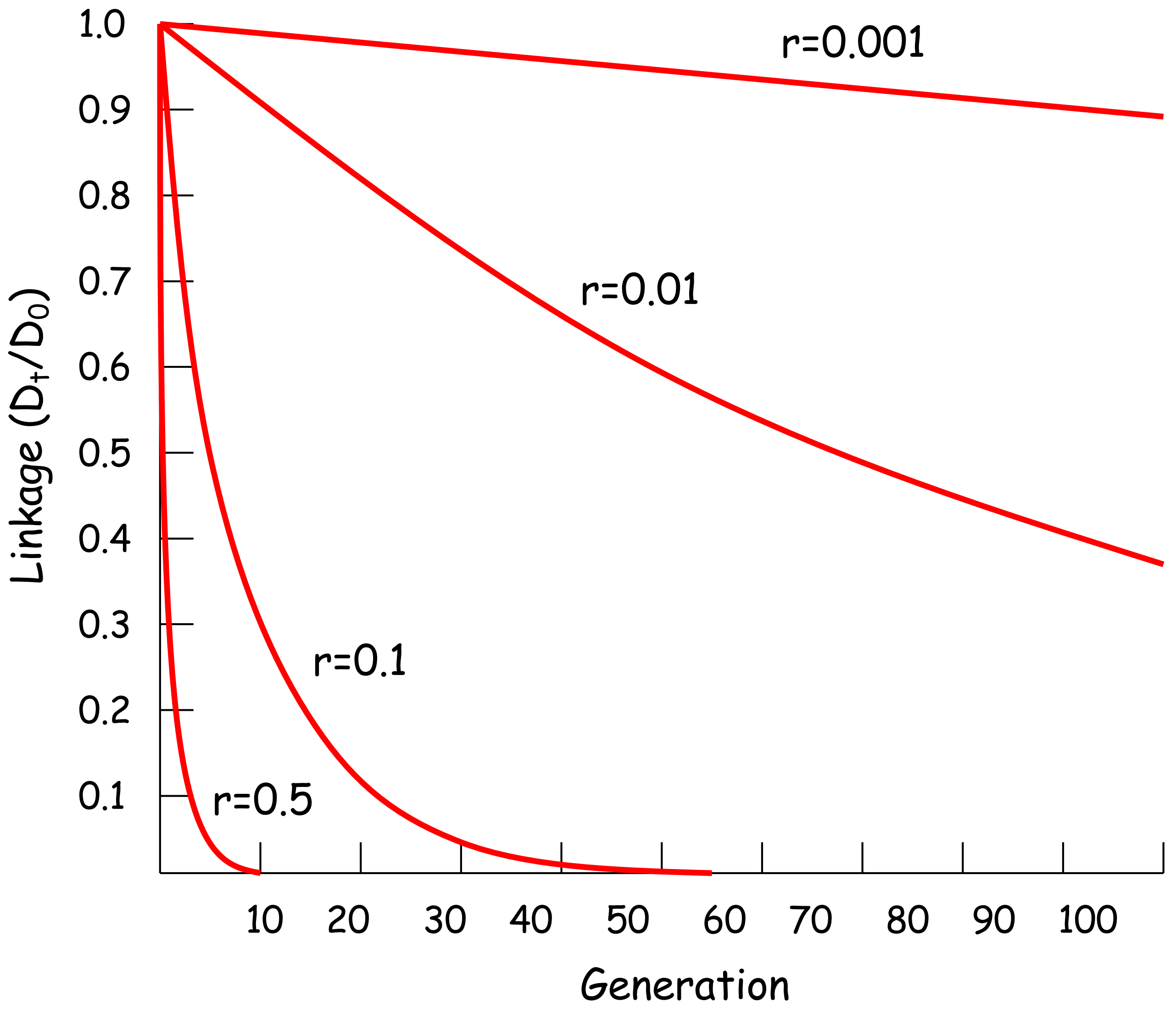
AB-40%
Ab-10%
aB-10%
ab-40%

linkage disequilibrium
r =20/100=0.2
D=0.15

AB-50%
Ab
aB
ab-50%

absolute linkage
r =0
D =0.25
$D_{max}=0.5\times0.5=0.25$

Decay of linkage over time



Only genes very
close on the chromosome
will remained linked for
longer periods of time,
and genes inside inversions

Overdominance can explain stable polymorphisms

Heterozygote favoured by selection

AA	Aa	aa
1	1+s*	0

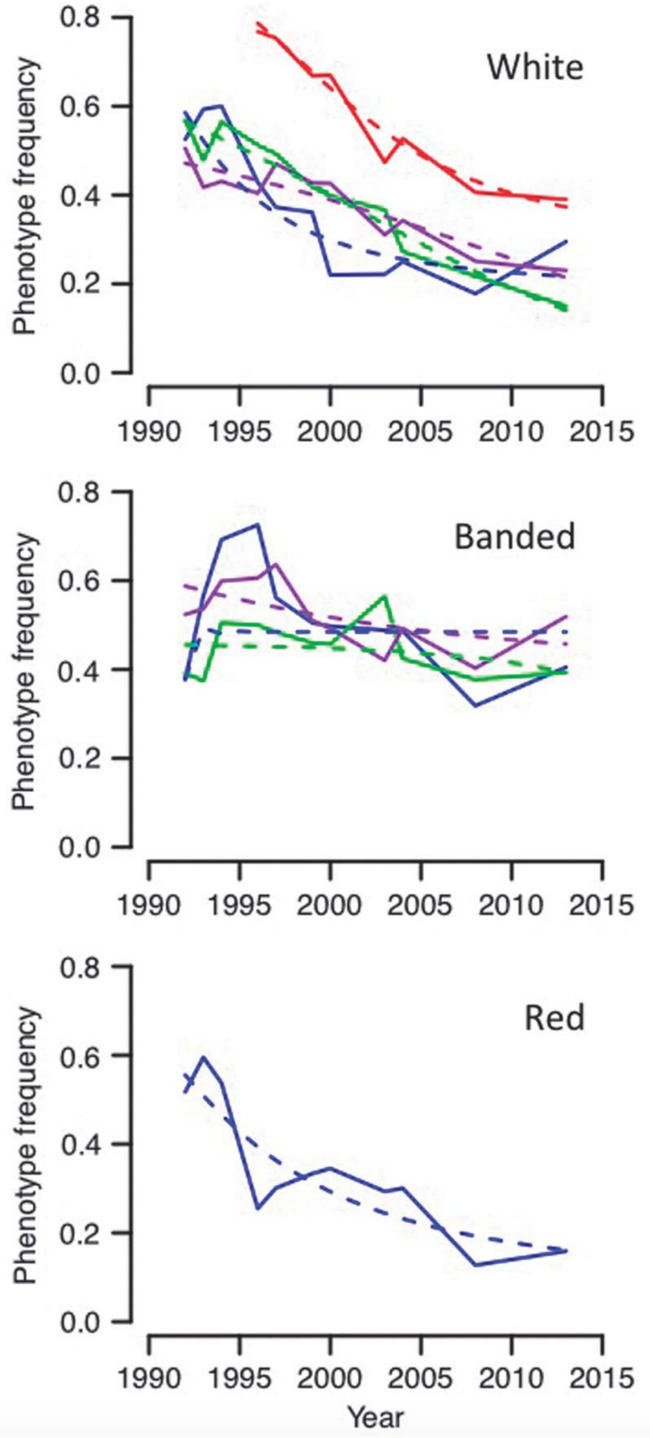
Even under extreme conditions, aa will remain in the population.
This will generate a "genetic load"

Negative frequency dependent selection can explain stable polymorphisms

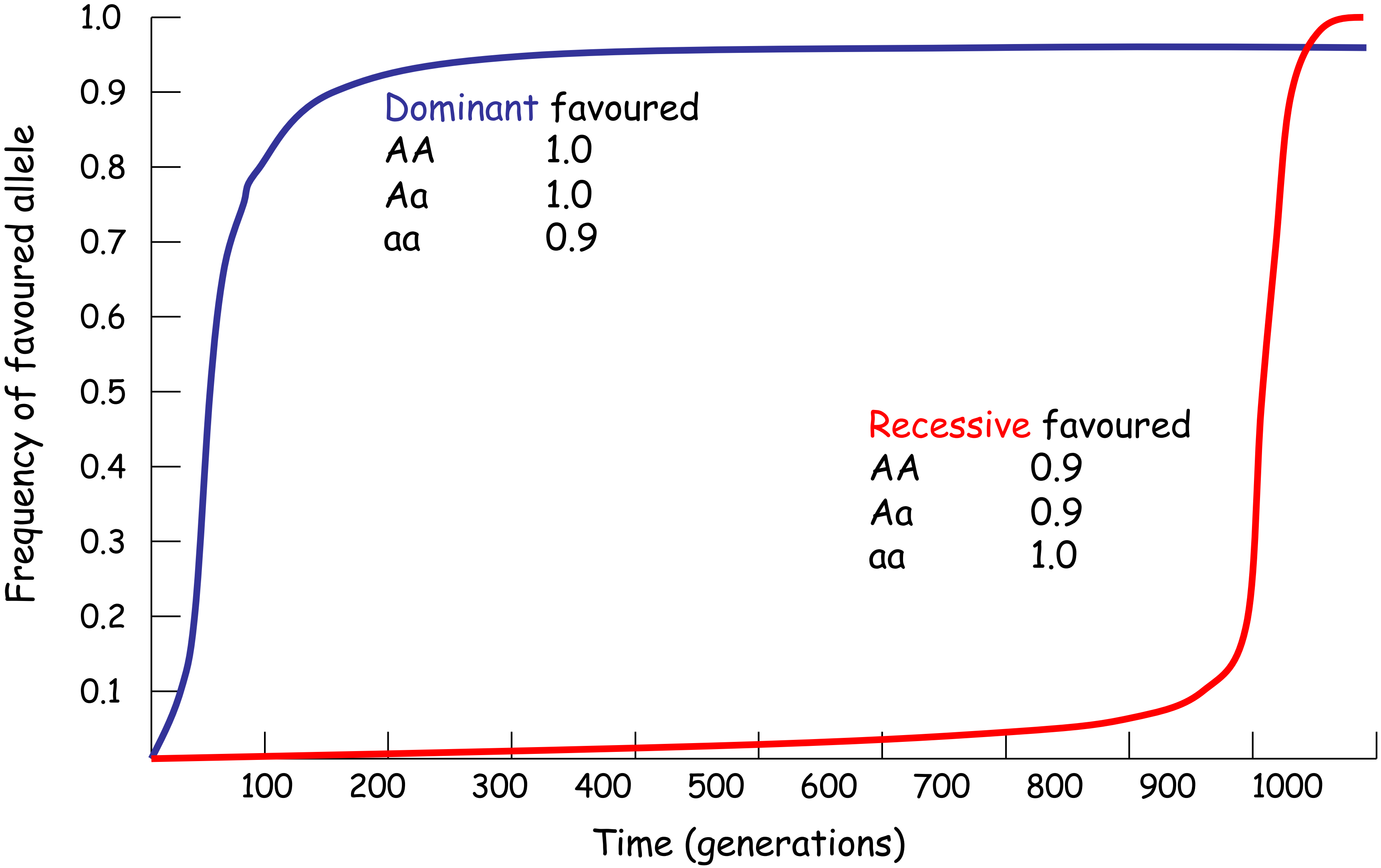
Genotype favoured by selection while rare

AA	Aa	aa
1	1	>1 while rare

aa will increase up to equilibrium when no longer favoured due to increased frequency



Dominant or recessive makes a difference



Dominant favoured

AA	1.0
Aa	1.0
aa	0.9

Recessive favoured

AA	0.9
Aa	0.9
aa	1.0