BIOL 434/509: Population genetics

Professor

Dr. Michael Whitlock
Zoology Department
Biodiversity 216
whitlock@zoology.ubc.ca

Office Hours: 1:30-2:30 Mondays and after class most days

course web-page: http://www.zoology.ubc.ca/~whitlock/bio434/

What is Population Genetics?

The genetical study of the process of evolution

(The study of the change of allele frequencies, genotype frequencies, and phenotype frequencies)

Population Genetics is…

- About microevolution (evolution within species)
- Strongly dependent on mathematical models (which have been more successful than most areas of mathematical biology)
- A relatively young science (most important discoveries are from after 1930)

Factors causing genotype frequency changes

- Selection
- Mutation
- Random Drift
- Migration
- Recombination
- Non-random Mating

What's the most important factor in evolution?

SELECTION

Natural selection causes evolution if

(1) There is variation in fitness (selection)

(2) That variation can be passed from one generation to the next (inheritance)

This is the central insight of Darwin.

Review of Hardy-Weinberg

Allele Frequency

The proportion of all alleles in all individuals in the group in question which are of a particular type. (often referred to as "gene frequency")

e.g. 40 individuals which are AA
47 individuals which are Aa
13 individuals which are aa

	GENOTYPE			
	AA	Aa	aa	Total
# of individuals	40	47	13	100
# of A alleles	80	47	0	127
# of a alleles	0	47	26	73
Total # of alleles				200

Allele frequency of A = 127/200 = 0.635

p_A=0.635

p_a = 73/200 = 0.365 = 1- p_A

Genotype Frequency

The proportion of individuals in a group with a particular genotype.

(Genotype can refer to one locus, two loci, or the whole genome, depending on the context)

40 AA 47 Aa 13 aa = 100 Total individuals

p_{AA} = 40/100 = 0.4

p_{Aa} = 47/100 = 0.47

p_{aa} = 13/100 = 0.13

Hardy-Weinberg Equilibrium

How to predict genotype frequencies from allele frequencies

- A useful null model
- Actually predicts genotype frequencies quite well, quite often

Assumptions

(1) Organism is diploid
(2) Reproduction is sexual
(3) Generations are non-overlapping
(4) Mating occurs at random
(5) Population size is very large
(6) Migration is zero
(7) Mutation is zero
(8) Natural selection does not affect the gene in question

1-locus, 2-alleles

$p_{AA} + p_{Aa} + p_{aa} = 1$

Frequency of A allele?

$$p_A = \frac{2p_{AA} + p_{Aa}}{2}$$

$$p_a = \frac{2p_{aa} + p_{Aa}}{2}$$

Mating Table – Random union of gametes

Assume that all of the H-W conditions are met, and that gametes find each other at random.

Combination	Freq.	Offspring Frequencies		
		AA	Aa	aa
A x A	p_A^2	1		
A x a	$p_A p_a$		1	
a x A	$p_a p_A$		1	
a x a	p_a^2			1
		p_A^2	$2\ p_A p_a$	p_a^2

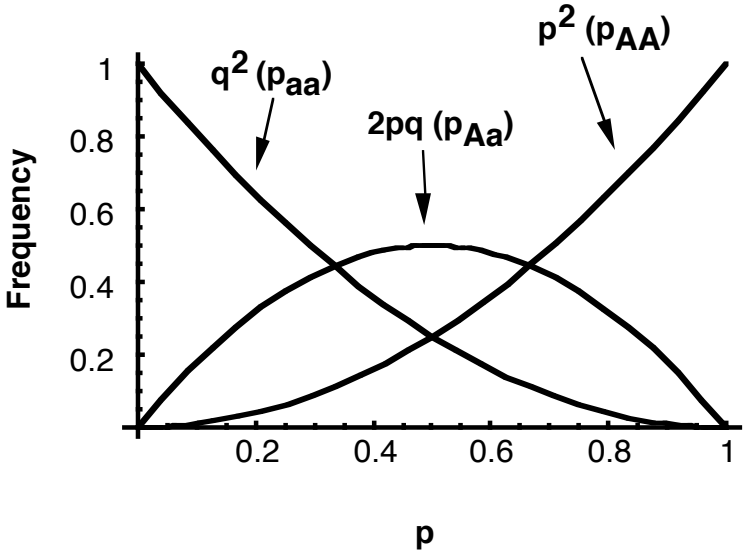
These are the Hardy-Weinberg frequencies.

	A (frequency p)	a (frequency q=1–p)
A	AA ($p_{AA}=p^2$)	Aa (pq)
a	Aa (pq)	aa (q^2)

What if individuals (instead of gametes) pair at random?

Pair	Frequency	Offspring Frequency		
		AA	Aa	aa
AA x AA	p_{AA}^2	1		
AA x Aa	$2p_{AA}p_{Aa}$	1/2	1/2	
AA x aa	$2p_{AA}p_{aa}$		1	
Aa x Aa	p_{Aa}^2	1/4	1/2	1/4
Aa x aa	$2p_{Aa}p_{aa}$		1/2	1/2
aa x aa	p_{aa}^2			1
		p_A^2	$2p_Ap_a$	p_a^2

Note that we also get the Hardy-Weinberg frequencies, but *we didn't specify that the parents were in Hardy-Weinberg*.



Hardy-Weinberg Principle

With the assumptions listed before, the frequencies of AA, Aa, and aa are p^2, 2pq, and q^2.

Note: The allele frequency has not changed: $p^2 + pq = p$

(Early Mendelians believed that dominant alleles would sweep through populations without selection -- WRONG!)

H-W proportions are *necessary, but not sufficient* to demonstrate that all of the assumptions are true.

Example: Scarlet Tiger Moth (*Panaxia dominula*)

FIGURE 9.1 Variation due to two alleles at a single locus in the moth *Panaxia dominula*. Top: the most common genotype, A_1A_1; middle: the heterozygote, A_1A_2; bottom: the rare homozygote, A_2A_2. In A_2A_2, the central white spot on the forewing is reduced or absent and the amount of black on the hindwing is less than in A_1A_1. The heterozygote is intermediate. (After Ford 1971.)

1469 white spotted (AA)
138 medium number of spots (AA')
5 few spots (A'A')
1612

$$p = \frac{1469}{1612} + \frac{1}{2}\left(\frac{138}{1612}\right) = 0.954$$

$q = 0.046$

Expected Frequency	x N	Observed
$p^2 = (0.954)^2 = 0.9101$	1467	1469
$2pq = 2\ (0.954)(0.046)$	142	138
$q^2 = (0.046)^2 = 0.0021$	3	5

H-W with more than 2 alleles

n alleles: $A_1, A_2, A_3, A_4, \ldots A_n$

frequencies: $p_1, p_2, p_3, p_4, \ldots p_n$

$$\sum_{i=1}^{n} p_i = 1$$

Frequency of A_iA_i homozygote is p_i^2
Frequency of A_iA_j heterozygote is $2p_ip_j$

Heterozygosity: $H=1-\Sigma p_i^2$

H-W for an X-linked locus

Allele frequency: $p = \frac{p_m}{3} + \frac{2p_f}{3}$
where p_m is the allele frequency in males and p_f is the allele frequency of females.

Allele frequency change over generations:

$$p'_m = p_f$$

$$p'_f = \frac{p_m + p_f}{2}$$

So the Hardy-Weinberg equilibrium is NOT reached in one generation for X-linked loci, if $p_m \neq p_f$

Genotype frequencies for 2 loci

Let p_A be the frequency of allele A (so $(1-p_A)$ is the frequency of a) at the A locus

and p_B be the frequency of allele B (so $(1-p_B)$ is the frequency of b) at the B locus

[Note: p_A, p_B are used differently in different contexts!]

At Hardy-Weinberg equilibrium, the frequency of a 2-locus genotype is the product of the frequencies of the two 1-locus genotypes it contains.

i.e. freq(Aabb) = $2p_A(1-p_A) \times (1-p_B)^2$

BUT, with two loci, this equilibrium is not reached immediately (unlike the one locus case)

If recombination occurs between the 2 loci at rate *r*, then the frequency of the AB gamete in the next generation is given by:

$$P'_{AB} = (1-r)P_{AB} + r\ p_A\ p_B$$

The $(1-r)P_{AB}$ term comes from the non-recombinants, and the $r\ p_A\ p_B$ term comes from the recombinants.

So if $P_{AB} \neq p_A\ p_B$, then the system is not in equilibrium.

This is called *linkage disequilibrium*. This term is a misnomer because it does not require physical linkage and it may have a non-zero value at equilibrium, when selection. migration or another process acts to create associations between alleles.

$D = P_{AB}\ P_{ab} - P_{Ab}\ P_{aB}$

$$P_{AB} = p_A p_B + D$$
$$P_{Ab} = p_A p_b - D$$
$$P_{aB} = p_a p_B - D$$
$$P_{ab} = p_a p_b + D$$

$D_t = (1-r)^t D_0$, where D_t is the linkage disequilibrium at time.

Probability Review

Distributions

A *distribution function* describes the probability of any given outcome for some process

There are continuous distributions (e.g. normal, χ^2, Γ) and discrete distributions (e.g. binomial and Poisson).

Binomial Distribution

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

p: probability of a particular event
n: number of trials
P[x]: probability that the event occurs exactly x times in N trials.

Remember, $\binom{n}{x} = \frac{n!}{(n-x)!x!}$

For example, the number of alleles of type A in one generation, taken as a sample from the alleles of the previous generation.

Poisson Distribution

$$P[x] = \frac{e^{-\mu}\mu^x}{x!}$$

where e is the base of the natural log, and μ is the expected number of successes per unit time.

x! is "x factorial" = x (x-1) (x-2) (x-3) …3 x 2 x 1

Discrete, like the binomial, but the number of trials is undefined

e.g., the number of offspring per mother
the number of flowers per m^2

Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean of the distribution and σ^2 is the variance.

Continuous
Very common in nature, because of the Central Limit Theorem

e.g., distribution of heights or weights, sampling error

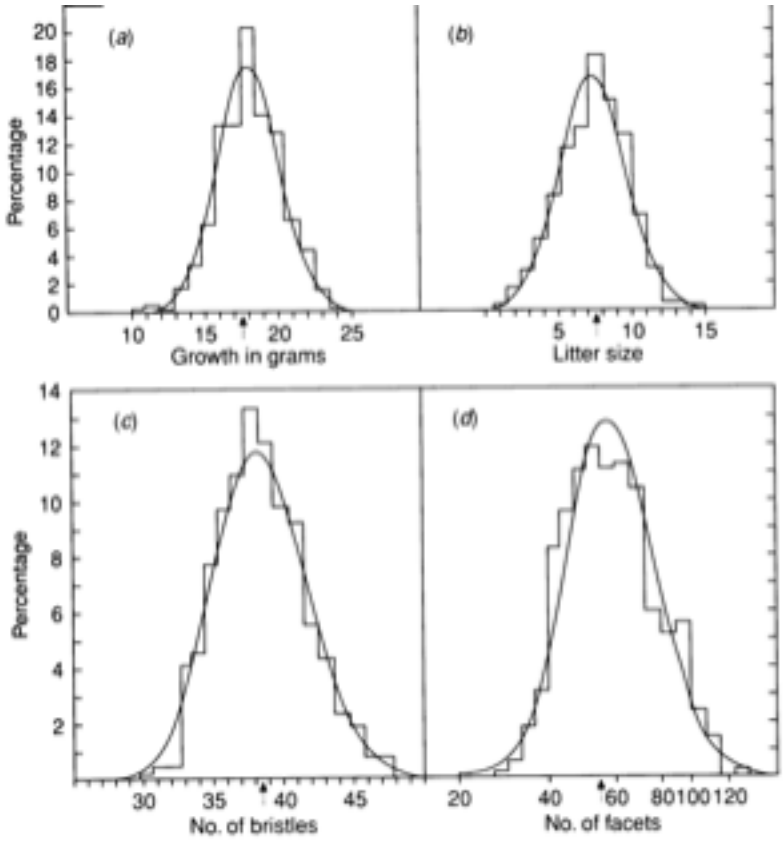


Fig. 6.2. Frequency distributions of four metric characters, with normal curves superimposed. The means are indicated by arrows. The characters are as follows, the number of observations on which each histogram is based being given in brackets:

(*a*) Mouse (♂♂): growth from 3 to 6 weeks of age. (380)

(*b*) Mouse: litter size (number of live young in 1st litters). (689)

(*c*) *Drosophila melanogaster* (♀♀): number of bristles on ventral surface of 4th and 5th abdominal segments, together. (900)

(*d*) *Drosophila melanogaster* (♀♀): number of facets in the eye of the mutant "Bar". (488)

(a), (b), and (c) are from original data; (d) is from data of Zeleny (1922).

Expected Values, Means, Variances

The expected value is the sum (or integral) of all possible values of quantity in question times their probabilities.

Represented by E[y]

$E[x] = \sum_{all\ x} x\ f(x)$ for discrete distributions

$= \int_{-\infty}^{\infty} x\ f(x)\ dx$ for continuous distributions

The mean of x is the expected value of x.

μ or $\bar{x} = \sum_{all\ x} x\ f(x)$ or $\int_{-\infty}^{\infty} x\ f(x)\ dx$

The mean measures the central tendency of a distribution.

The variance is the expected value of $(x-\mu)^2$

The variance measures the spread of the distribution.

V or σ^2 or $s^2 = E\left[(x-\mu)^2\right] = \sum_{all\ x}(x-\mu)^2\ f(x)$ or $\int_{-\infty}^{\infty}(x-\mu)^2\ f(x)\ dx$

Mean and Variance of basic distributions

	Mean	Variance
Binomial	pN	Np(1-p)
Poisson	μ	μ
Normal	μ	σ^2

Useful facts about means and variances

	Mean	Variance
X + Y	E[X] + E[Y]	V[X] +V[Y] + 2Cov[X,Y]
-X	- E[X]	V[X]
a X	a E[X]	a^2 V[X]
X + a	E[X] + a	V[X]

Cov[X, Y] is the covariance of X and Y

$Cov[X,Y] = E\left[(x-\mu_x)(y-\mu_y)\right]$

Conditional Probability

The probability that X is true, *given that Y is true*, is written as P[X | Y].

e.g. The probability that a 20 year-old is in college = 40%
The probability that a 60 year-old is in college = 0.5%

P[an individual is in college | that individual is 20 years old] = 0.4.

We can find the probability of an event, if we know the probability of that event under each of possible scenarios and the probability of those scenarios:

$$P[X] = \sum_{all\ Y} P[X \mid Y] P[Y]$$

e.g. Hardy- Weinberg: What is the probability that an allele drawn from an individual is type A?

P[A | Individual is of genotype AA] = 1
P[A | Individual is of genotype Aa] = 1/2
P[A | Individual is of genotype aa] = 0

So

P[A] = P[A |AA] P[AA] + P[A |Aa] P[Aa] + P[A |aa] P[aa]
= (1) p^2 + (1/2) 2pq + (0) q^2
=p^2+ pq = p(p + q) = p

P(X and Y) = P(X) P(Y |X)

If X and Y are *independent*, then P(X and Y) =P(X) P(Y)

(Relate this to linkage disequilibrium.)

Random Genetic Drift

Most populations are small enough that, by chance, sampling will result in a different allele frequency from one generation to the next.

This is called "*Genetic Drift*".

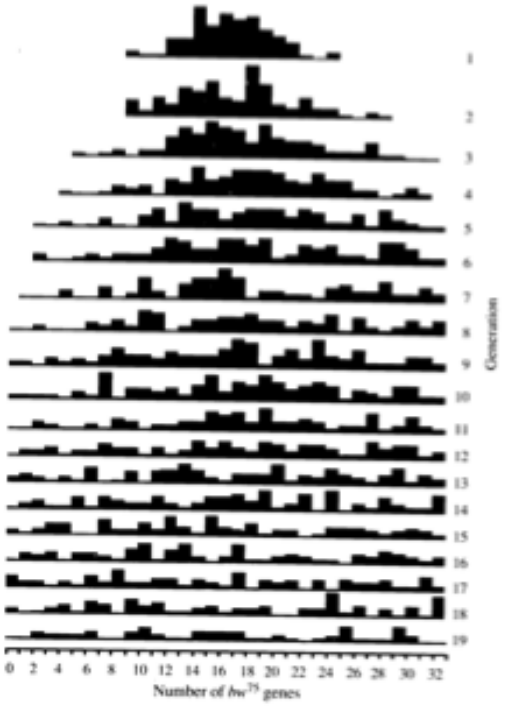


Fig. 3.1. Distributions of gene frequencies in 19 consecutive generations among 105 lines of *Drosophila melanogaster*, each of 16 individuals. The gene frequencies refer to two alleles at the 'brown' locus (bw^{75} and bw), with initial frequencies of 0.5. The height of each black column shows the number of lines having the gene frequency shown on the scale below, previously fixed lines being excluded. (*After Buri, 1956.*)

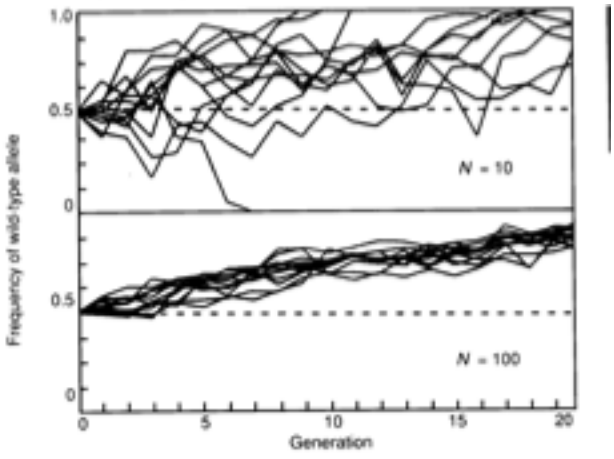
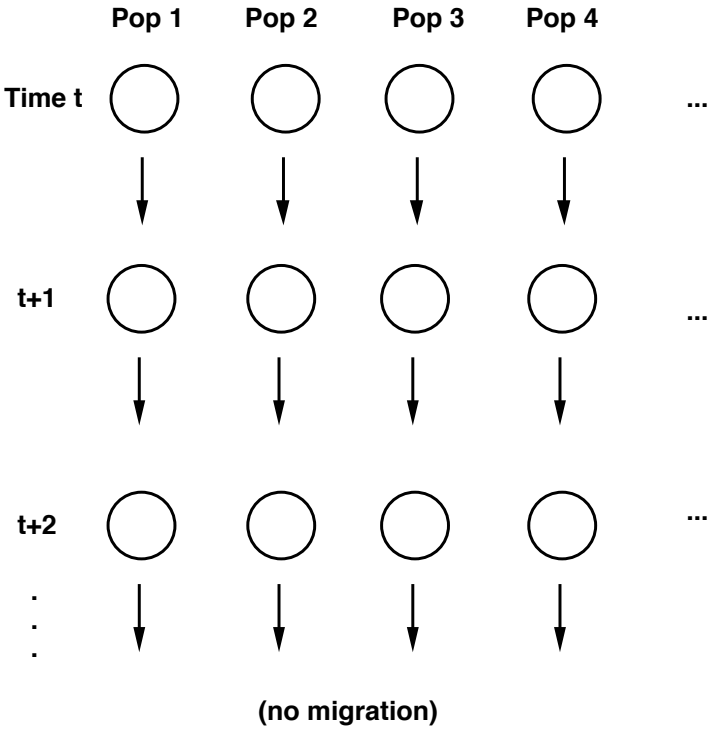


Fig. 3.2. Random drift of a colour gene ('black') in *Tribolium*. Heterozygotes were recognizable, so the gene frequencies were estimated exactly by counting. The figure shows the results with two population sizes, $N = 10$ and $N = 100$. There were 12 lines with each population size. Natural selection favoured the wild-type allele and led to an overall increase in its frequency, random drift causing variation of the lines around the mean, more marked in the smaller than in the larger populations. (*After Rich, Bell, and Wilson, 1979.*)

The two founders of population genetics, Sewall Wright and Ronald Fisher disagreed most strongly about the importance of genetic drift.

What happens in an individual population is unpredictable, but we can describe the distribution of allele frequencies among replicate populations.



(no migration)

Sampling

Start with an *ideal population* (= all individuals have an equal probability of being the parent of an individual from the next generation). (We know that these are unreasonable assumptions. We'll fix this later with the idea of an effective population size.)

With N individuals (or $2N$ alleles)

With an ideal population, and two different alleles, the number of copies of a particular allele in the *next* generation follows a binomial distribution

$$P(x) = \binom{2N}{x} p^x (1-p)^{2N-x}$$

Digression into the binomial distribution

General	Population Genetics for diploids
n = # of trials	2N = # of alleles
X = number of "successes"	X = # of A alleles
p = probability of success	p = allele frequency
E[X] = np	E[X] = 2Np
V[X]=np(1-p)	V[X]=2Np(1-p)

p' = X/(2N), so…

$$E[p'] = p \quad \left(= E\left[\frac{X}{2N}\right] = \frac{E[X]}{2N} = \frac{2Np}{2N} = p \right)$$

$$V[p'] = \frac{p(1-p)}{2N} \quad \left(= V\left[\frac{X}{2N}\right] = \frac{V[X]}{(2N)^2} = \frac{2Np(1-p)}{(2N)^2} = \frac{p(1-p)}{2N} \right)$$

$$\boxed{E[p'] = p \qquad V[p'] = \frac{pq}{2N}}$$

So with pure drift…

- The expected value of the allele frequency doesn't change.
- The amount of drift is inversely proportional to population size.

(For haploids, there are *N* alleles in the population, so $\boxed{E[p'] = p \qquad V[p'] = \frac{pq}{N}}$)

Drift over time

Random genetic drift can continue until one allele is *fixed* (i.e. reaches a frequency of 1) or *lost* (reaches a frequency of 0).

Without selection, mutation or migration, eventually every allele will be either fixed or lost.

The probability of eventual fixation of an allele affected only by drift = *p* (its allele frequency)

{Think about why.}

4 ways to model drift

1. Wright Fisher Model (Transition matrix)

With drift, there is some probability that the allele frequency in the next generation could be any value. The binomial distribution gives the probability of drawing a certain number of alleles of a given type from the pool of available parental alleles. The probability of drawing *j* alleles of a particular type out of 2*N*, given that there are *i* alleles of that type in the parental generation (let's call this T_{ij}) is:

$$T_{ij} = \binom{2N}{j}\left(\frac{i}{2N}\right)^{j}\left(1-\frac{i}{2N}\right)^{2N-j}$$

Remember that *p* is *i*/2*N* (and $q = 1 - p$), so T_{ij} is really a function of *p*.

So if we want to calculate the distribution of allele frequencies after one generation of drift, we can use T_{ij}. This T_{ij} is called a *transition probability* -- it is the probability of transitioning from *i* copies of the allele to *j* copies in one generation.

$$\Pr\left[j \text{ copies after one generation}\right] = \sum_{i=0}^{2N} T_{ij} \Pr[i \text{ copies now}]$$

What if we wanted to know the probability of changing from *i* copies in one generation to *k* copies in *two* generations? Then we use conditional probability. The probability of having a certain number of copies in two generations depends on what happens in the next generation. So

$$\begin{aligned} &P[k \text{ copies in } 2 \text{ generations}] \\ &= \sum_{j=0}^{2N} P[k \mid j \text{ copies in } 1 \text{ generation}] P[j \text{ copies in one generation}] \\ &= \sum_{j=0}^{2N}\sum_{i=0}^{2N} T_{jk}\left(T_{ij}\right)\Pr[i \text{ copies at start}] \end{aligned}$$

This could be continued for as many generations as desired.

Fortunately, this can be eased by noticing that the transition probabilities make up a matrix $\boldsymbol{T}$, which can be manipulated by matrix algebra to find results.

The Wright Fisher model has the advantages of (1) being exact, and (2) allowing the inclusion of other evolutionary forces like selection, mutation and migration, by

changing the form of **T**. It is an example of a *Markov chain*, which means that a whole range of mathematical techniques can be applied to population genetic problems.

2. Diffusion models (Continuous approximation of the Wright-Fisher model)

A continuous approximation to the Wright-Fisher model, largely due to Motoo Kimura.

See Figures 3.6, 3.7

Time to fixation (assuming that the allele starts at frequency *p* and ultimately fixes) :

$$\bar{t}_1(p) = -4N\left(\frac{1-p}{p}\right)\ln[1-p]$$

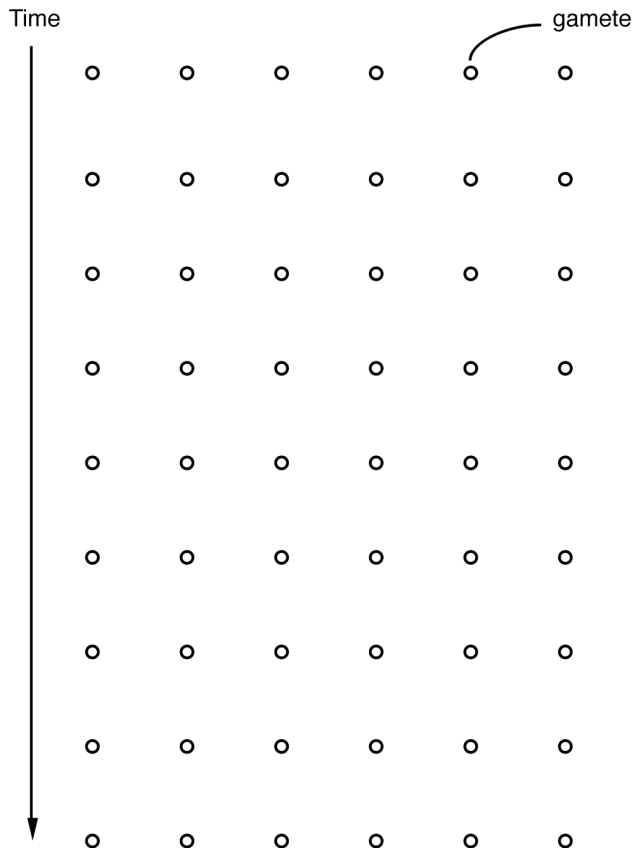
(If the allele starts as a single copy, this is approximately 4N generations.)

Time to loss (assuming that the allele starts at frequency *p* and ultimately is lost) :

$$\bar{t}_0(p) = -4N\left(\frac{p}{1-p}\right)\ln[p]$$

(If the allele starts as a single copy, this is approximately 2 ln[2N] generations.)

3. Changes in Identity


Time
gamete

Let *F* be the probability that two alleles drawn at random are the same. (This is called the probability of identity in state.)

$$F' = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F$$

The first term is the probability that two alleles descend from the same allele in the previous generation.

The second term is the probability that they didn't, times the probability (F) that they were identical anyway.

This is an example of using conditional probability:

P[2 alleles are the same] =

P[2 alleles are the same | they came from the same parent allele] x P[they came from the same parent allele] +
P[2 alleles are the same | they came from different parent alleles] x P[they came from different parent alleles]

= (1) (1/(2N)) + F (1-1/(2N))

More generally

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

$$1 - F_t = 1 - \frac{1}{2N} - \left(1 - \frac{1}{2N}\right) F_{t-1}$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)\left(1 - F_{t-1}\right)$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)\left[\left(1 - \frac{1}{2N}\right)\left(1 - F_{t-2}\right)\right]$$

$$1 - F_t = \left(1 - \frac{1}{2N}\right)^t \left(1 - F_0\right)$$

$$F_t = 1 - \left(1 - \frac{1}{2N}\right)^t \left(1 - F_0\right)$$

(For haploids, $2N$ in the previous equation would be N.)

Genetic Variance among individuals

If the population is polymorphic at a locus (the allele frequency is not 0 or 1), then there will be variance among individuals in that population in the number of copies of an allele that they have.

With random mating, the variance in number of copies of an allele among individuals is given by the binomial distribution, with N=2 (for diploids) and p = the allele frequency. Thus this variance is $2pq$.

Later we will use talk about the genetic variance in terms of the phenotypic effects of these alleles, as well. We will also learn what happens to the genetic variance when we relax the assumption of random mating.

Identity by descent

Similar equations also describe the probability of identity by descent, which we'll call f here. We say that two alleles are identical by descent if they share a common ancestor within a set period of generations. For example, we might calculate the probability of two alleles sharing an ancestor between an arbitrary time 0 and time t. Identity by descent behaves like identity in state except that $f_0 = 0$ by definition.

Relationship between drift, genetic variance, and heterozygosity

Note that the heterozygosity at time t (H_t) is

$$H_t = (1 - f_t)H_0$$

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

With Hardy-Weinberg frequencies, heterozygosity is $2pq(1-f)$.

Genetic variance among individuals is proportional to $(1-f)2pq$.

The variance among populations is $2fpq$.

So, as drift progresses,

- Heterozygosity decreases (inversely proportional to f)
- Genetic variance within populations decreases (inversely proportional to f)
- Genetic variance among populations increases (directly proportional to f)

4. The coalescent

With coalescence, we look at the genealogy of alleles going *backwards* in time. We start with a sample of alleles taken at one point in time, and then calculated the possible patterns of ancestry of those alleles.

The process is called "coalescence" because we watch the histories of independent alleles coalesce into the same historical path. That is because if two alleles ever share the same ancestor alleles, then all their history prior to that point in time is completely shared. So their histories "coalesce."

The coalescent approach focuses much more on *samples* of alleles, rather than on entire populations, typically. As such it is designed and used to make inference about populations from a sample of individual taken at one point in time. The coalescent approach can be used to generate ideas about the process of evolution, but it is much more often used as a statistical tool for making inference about the history of a population.

We'll do the coalescent with haploids. The diploid case is the same, but replace (N) with (2N) in each equation.

Start by thinking of two alleles independently sampled from the population. The probability that they had a common ancestor one generation ago (i.e., that they coalesce t = 1 generations ago) is $\frac{1}{N}$, because this is the probability that the second allele had the same parent allele as the first allele. There is a (1-1/N) probability that they did not coalesce in that generation.

The probability that the two alleles coalesce t = 2 generations ago is 1/N (1-1/N). For a pair of alleles to coalesce two generations ago, they had to *not* coalesce one generation ago (probability 1-1/N) and then *do* coalesce two generations ago (probability 1/N). Extending this, the probability that two alleles have their most recent common ancestor t generations ago is $\frac{1}{N}\left(1-\frac{1}{N}\right)^{t-1}$. This is the geometric probability distribution, and its mean and variance are known: Mean = N and variance = N^2. So on average it takes N generations for two randomly chosen alleles to coalesce, and there is a large variance around that expectation.

Let's look at larger samples, with n alleles in the sample. (We'll assume that n is much less than N, for most purposes.) The probability that there was a coalescent event in the previous generation is most easily calculated from the one minus the probability of a coalescent event, in this case. The probability that all n alleles had different parent alleles is given by :

$$P[n] = \prod_{i=1}^{n-1}\left(1-\frac{i}{N}\right)$$

because the second allele has probability (1-1/N) of having a different parent from the first allele, the third allele has probability of (1-2/N) of having a different parent from the first two alleles, etc.

That probability can be approximated (assuming that $n << N$) by

$$P[n] \cong 1 - \frac{\binom{n}{2}}{N} \cong \exp\left[-\frac{\binom{n}{2}}{N}\right]$$

Following similar logic as above, the probability that the last coalescent event happened t generations ago when there are n alleles in the sample is given by a geometric distribution:

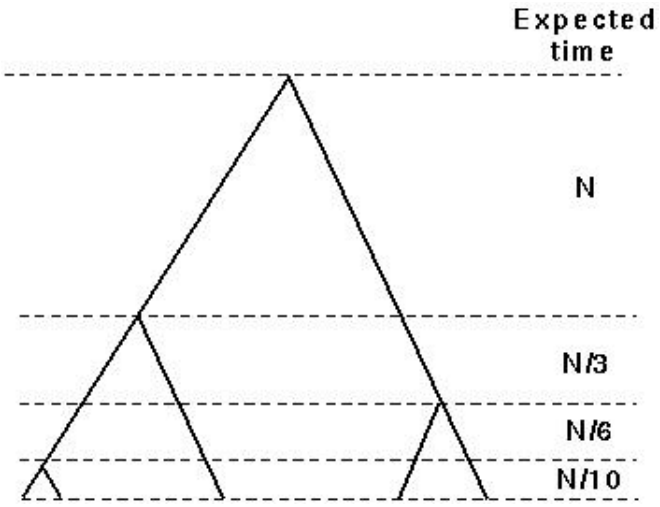
$$\Pr[\textit{last coalescent event from n alleles happened t generations ago}] = (1 - P[n])P[n]^{t-1}$$

The mean of this distribution is approximately $\frac{N}{\binom{n}{2}}$, which is much smaller with larger n.

Coalescent events are more common with more alleles, because there are more chances for alleles to have common ancestors. Therefore the most recent event in a larger sample is on average much more recent than when the sample is small. Coalescence also happens faster in small populations (when N is small).

You can find the time to coalescence for all n alleles (in other words until they all share the same common ancestor) by adding the time is take to go from n to $n-1$ alleles, to the time from n-1 to n -2, etc., all the way down to where there is only one ancestral allele left.

Expected time to coalescence of all n alleles = $\sum_{i=2}^{n} \frac{N}{\binom{i}{2}} = 2N\frac{n-1}{n}$. (Note that this reduces to N for n=2.)



The effective population size

The ideal population assumed before makes many simplifying assumptions, but the results are pretty straightforward.

Can we relax these assumptions?

The *effective population size* of a population is the size of an ideal population which acts the same as the real population in question.

The ideal population

The ideal population assumes:

(1) No selection
(2) Random mating
(3) Random chance of each offspring having a particular parent

Assumption (3) is well-approximated by a Poisson distribution of reproductive success by the parents.

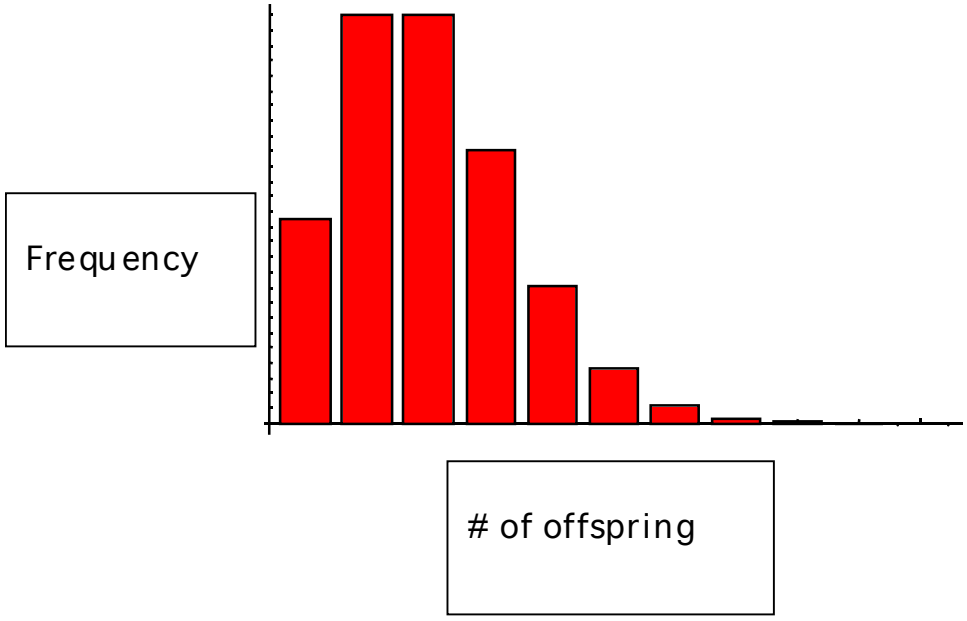
N_e (as we abbreviate the effective size) can deal with assumptions 2 + 3. Later we will learn how to combine the effects of drift and selection.

What is an ideal population like?

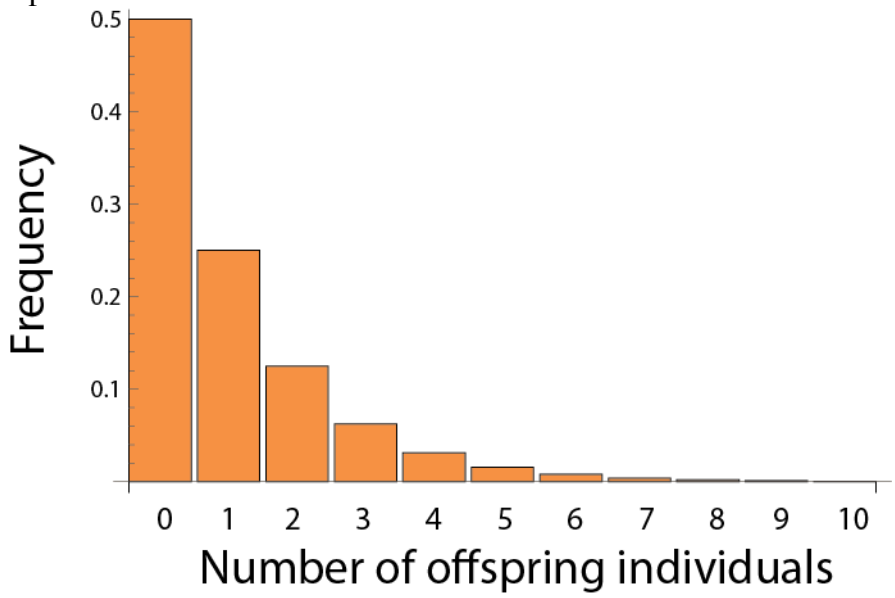
(Remember - each parent has an equal and independent chance of being the parent of each descendent allele.)

This is approximated by a Poisson distribution of reproductive success, because the Poisson describes the distribution that results when "successes" occur with equal probability and independently in each "block".

(Reproductive success = # of offspring per parent, or per parental allele.)



A more realistic distribution of reproductive success would have a higher variance in reproductive success:



With this distribution, there is a much higher chance of an allele leaving no copes of itself, as well as a much higher chance of an allele leaving a lot of copies in the next generation. This the change in allele frequency from one generation to the next would be on average more pronounced.

Kinds of effective size

- Variance effective size: predicts changes in genetic variance
- Inbreeding effective size: predicts changes in heterozygosity
- Eigenvalue effective size: predicts changes in allele frequencies

In general the different N_e's are *very* similar (as you would expect, since genetic variance, heterozygosity, and allele frequencies are so related).

Factors which control effective size

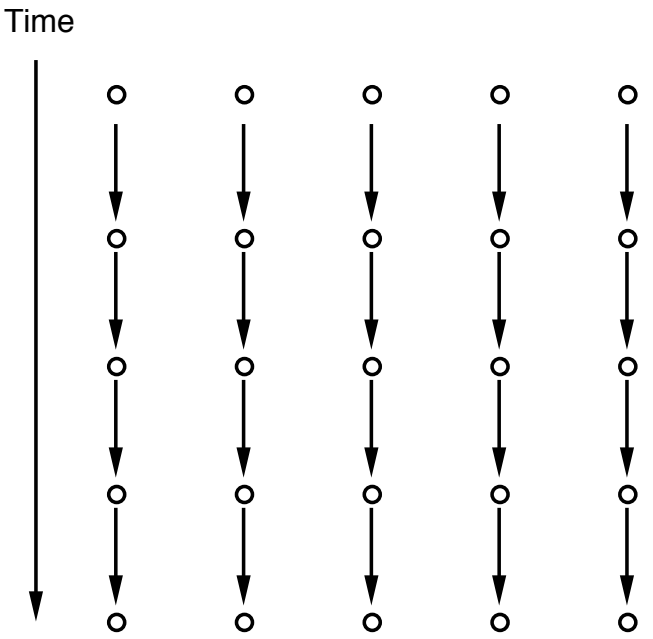
- Variance in reproductive success
- Correlations in reproductive success (including non-random mating)

Why does variance in reproductive success matter?

Variance in reproductive success implies that some individuals get more offspring and some get less, therefore there is a higher probability that offspring have the same parents.

Extreme example: No variance in reproductive success (Haploids)

All alleles have exactly one offspring allele:



$$F_t = F_{t-1}$$

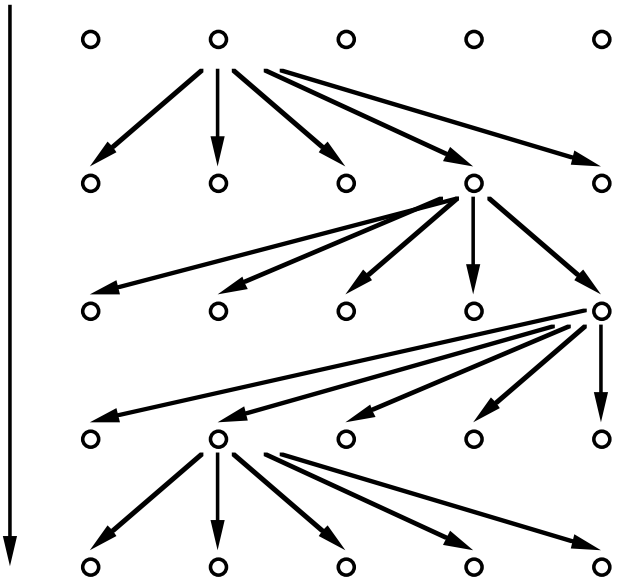
$$so\ if\ F_t = \left(1 - \frac{1}{N_e}\right) F_{t-1} + \frac{1}{N_e}$$

$$N_e = \infty$$

Extreme example: Maximum variance in reproductive success (Haploids)

If all offspring descend from one parent, then the variance in reproductive success is maximized:

Time



$F_t = 1$

$$so\ with\ F_t = \left(1 - \frac{1}{N_e}\right) F_{t-1} + \frac{1}{N_e}$$

$N_e = 1$

Haploids

The N_e for haploid populations is simple:

$$N_e = \frac{N}{V}$$

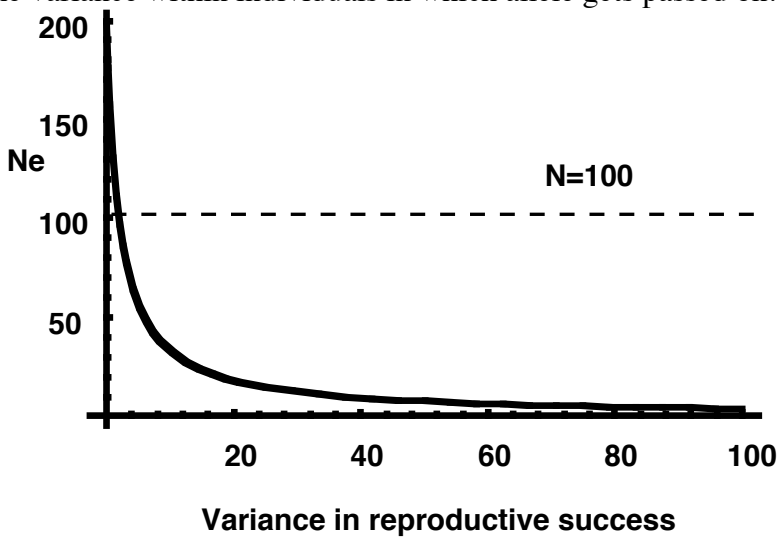
High Variance in reproductive success means low N_e

For diploids:

$$N_e = \frac{4N - 2}{V + 2}$$

where V is the variance in reproductive success among diploid individuals.

The reason for the more complicated formulae with diploids is that there is an extra source of variation in the success of gametes over and above the variance among individuals: the variance within individuals in which allele gets passed on.

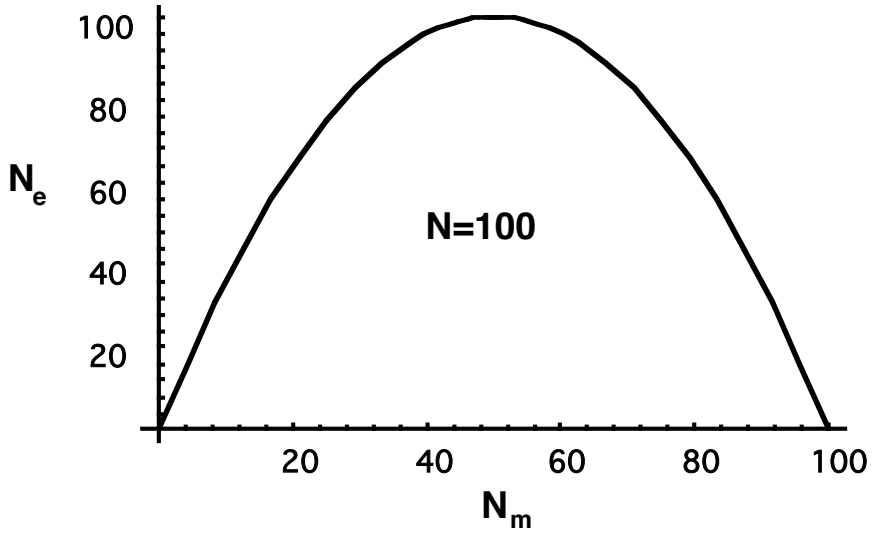


(Note that the variance among individuals in reproductive success with a Poisson distribution is 2 in a steady-state population, so that $N_e = N-1/2$.)
Note also that N_e can also be as great as 2N-1, if there is no variance in reproductive success. This fact is often used in animal and plant breeding to slow the loss of genetic material.

Unequal sex ratios

A special case of the variance in reproductive success comes from unequal sex ratios.

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$



Can be derived from

$$\frac{1}{2N_e} = \frac{1}{4}\left(\frac{1}{2N_m}\right) + \frac{1}{4}\left(\frac{1}{2N_f}\right) + \frac{1}{2}(0)$$

Implications of the relationship between V and Ne

Almost all natural populations have more variance in reproductive success than expected by random

Therefore, N_e is almost always lower than N.

Table 7. Summary of effective population sizes (N_e), census sizes (N), and N_e/N ratios in captive populations of *Drosophila*.

Maintenance	*N*	*N_e*	*N_e/N*	*Method*
Drosophila melanogaster				
Cages	5000	185	0.037	Heterozygosity
		253	0.051	V_A
		190–1252	0.038–0.250	Lethal allelism
Cage	3500	15	0.004	Heterozygosity
20 bottles	1000	16	0.016	Heterozygosity
Cages	3500	272–999,∞	0.078–0.285,∞	Lethal allelism
Cages	10,000	∞	∞	Lethal allelism
	1000	256	0.256	
Drosophila pseudoobscura				
Cages	500	18	0.036	Heterozygosity
Cages	4365	52	0.012	Heterozygosity

References
1. *Current paper.*
2. *Malpica and Briscoe (1981).*
3. *Lopez-Fanjul and Torroja (1982).*
4. *Prout (1954).*
5. *Derived from McDonald and Ayala (1974).*
6. *Derived from Powell and Wistrand (1978).*

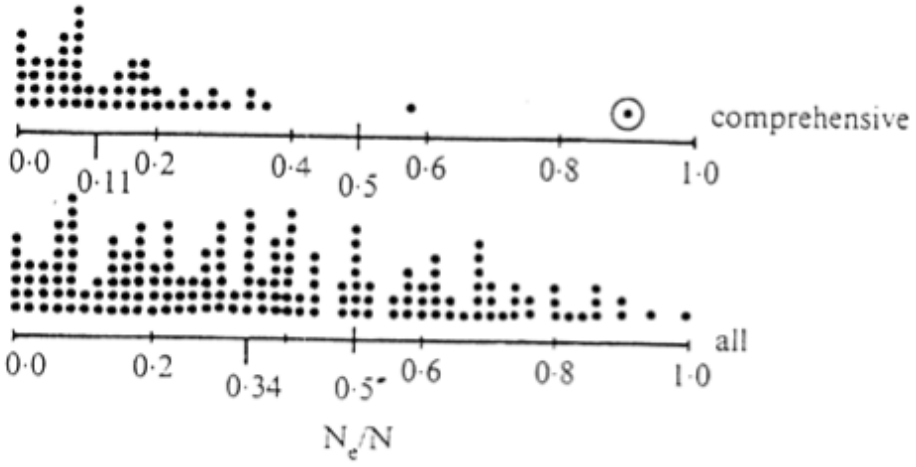


Fig. 1. Distributions of estimates of effective/actual population size (N_e/N) ratios. Comprehensive estimates (that include the effects of fluctuation in population size. variance in family size and unequal sex-ratio) are above and all estimates below. The circled outlier is for a pair mated rainbow trout (*Oncorhynchus mykiss*) population. Means of estimates are indicated below vertical lines.

N_e can be kept artificially higher than N

(For diploids, as high as 2N.)

By equalizing reproductive success

(Animal and plant breeders do this to preserve genetic variation.)

Selection causes variance in reproductive success.

So strong selection can cause genetic drift!

Genetic variance in reproductive success causes correlations across generations in reproductive success, which further decrease N_e.

This is called genetic hitchhiking.

For many (if not most species) N_e is reduced, because the number of breeding males is much less than the number of breeding females.

Unequal sex ratios = lower N_e.

Correlations in reproductive success

Inbreeding cause alleles in the same individual to be correlated

(i.e. more similar than average --more likely to be identical by descent)

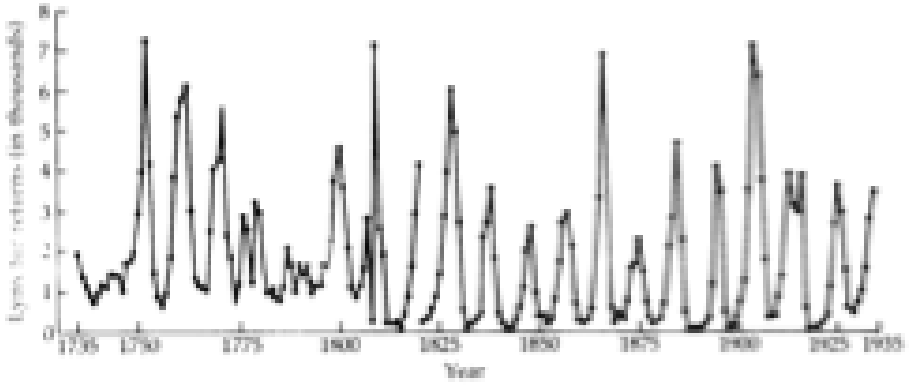
This correlation magnifies drift.

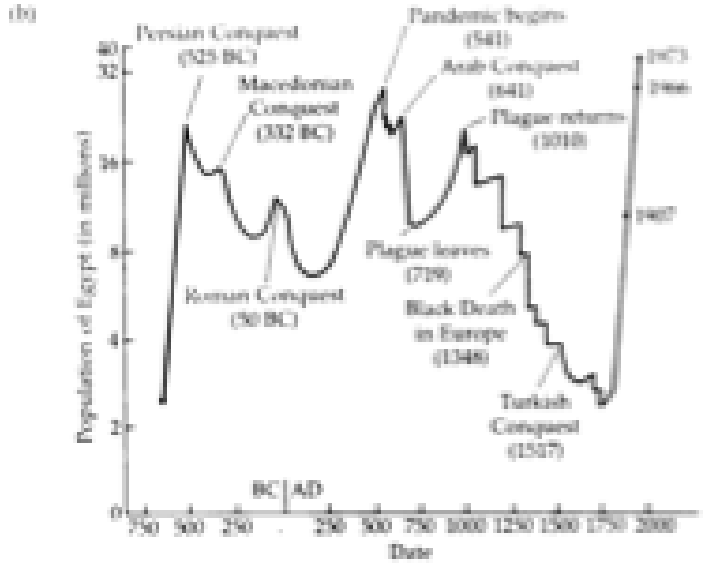
Imagine the case of complete selfing

- As we have seen, with selfing heterozygosity quickly goes to zero - almost all individuals carry two identical alleles
- With selfing, sampling is effectively among N individuals instead of 2N alleles. N_e is reduced in proportion to the amount of inbreeding.

Variable population size

Population size in natural populations does not remain constant





$$1 - F_t = \left(1 - \frac{1}{2N_{t-1}}\right)\left(1 - \frac{1}{2N_{t-2}}\right)\left(1 - \frac{1}{2N_{t-3}}\right)...\left(1 - F_0\right)$$

We want to find N_e so that

$$1 - F_t = \left(1 - \frac{1}{2N_e}\right)^t \left(1 - F_0\right)$$

N_e with population size fluctuations is approximately the harmonic mean of N over time:

$$\tilde{N} = \frac{1}{\frac{1}{t}\sum_i \frac{1}{N_i}}$$

The harmonic mean is very sensitive to small values.

($N_e << \overline{N}$) if N is variable

Example

$N_1 = 1{,}000{,}000{,}000$

$N_2 = 2$

$N_3 = 1{,}000{,}000{,}000$

$$\overline{N} = \frac{2{,}000{,}000{,}002}{3} = 666{,}666{,}667.3$$

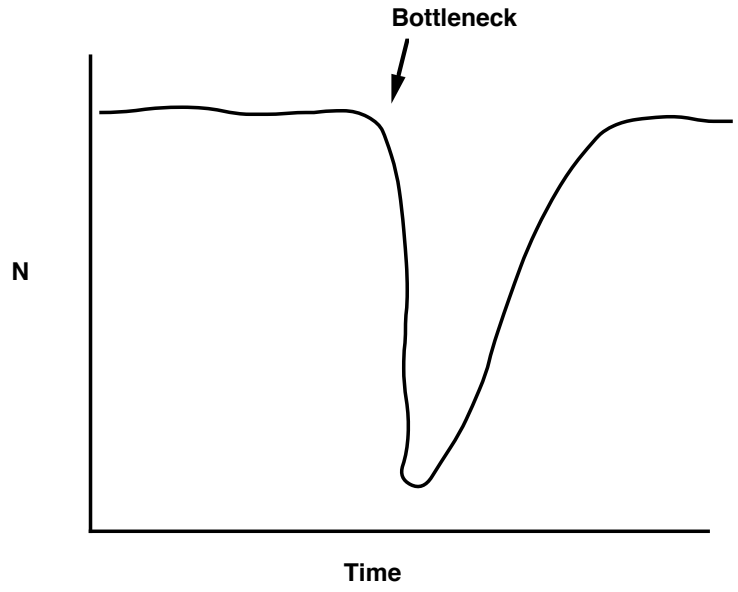
$$\tilde{N} = 5.999999976 \cong 6$$

$$\tilde{N} << \overline{N}$$

Population bottlenecks

A population bottleneck is a severe temporary reduction in population size.

They cause strong effects on N_e and therefore on genetic variance.



Implications

Since N_e << N (because V>2 and $\tilde{N} << \overline{N}$) the effective populations size is usually much less than the census size.

Homozygosity (and therefore genetic variance) is strongly correlated with population size in nature:

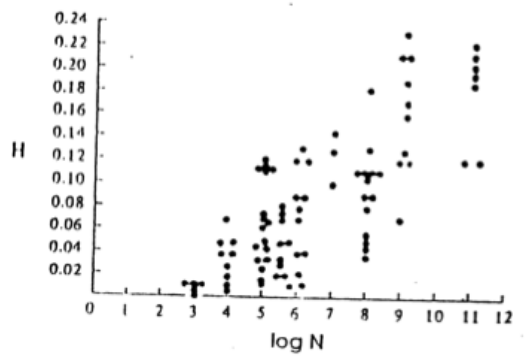


Figure 2. Correlation between heterozygosity (H) and logarithm of populations size (log N) for animal species, as given by Soulé (1976).

Examples

A population of *Drosophila* is kept at a constant population size and starts with 4% of its individuals heterozygous, on average, for any given locus. Every generation 4 males and 20 females are transferred to a new container. What is the expected heterozygosity after 10 generations?

If the variance in reproductive success among diploid individuals is 6, and there are 40 individuals per population per generation, what should the genetic variance within populations be after 30 generations (relative to the starting genetic variance)?

$$N_e = \frac{4N - 2}{V + 2} = \frac{4(40) - 2}{6 + 2} = 19.75$$

$$\begin{aligned} V_t &= \left(1 - \frac{1}{2N_e}\right)^t V_0 \\ &= \left(1 - \frac{1}{2(19.75)}\right)^{30} V_0 \\ &= 0.463\, V_0 \end{aligned}$$

Mutation

Mutation is the ultimate source of genetic variation and therefore of evolution.

Types of mutations

Nucleotide mutations

same as "point mutation"

A → G

G → A } **Transistions**

C → T

T → C

A → T **T → G**

T → A **G → T** } **Transversions**

A → C **C → G**

C → A **G → C**

Transitions are more common than transversions.

Estimated mutation rates in humans (Nachman and Crowell 2003, Genetics):

Mutation type	Mutation rate
Transition at CpG	1.6×10^{-7}
Transversion at CpG	4.4×10^{-8}
Transition at non-CpG	1.2×10^{-8}
Transversion at non-CpG	5.5×10^{-9}
All nucleotide substitutions	2.3×10^{-8}
Length mutations	2.3×10^{-9}
All mutations	2.5×10^{-8}

As estimated by substitutions in pseudogenes.

The genetic code is redundant, so different codons can code for the same amino acid.

Mutations which result in a different codon for the same amino acid are called synonymous mutations or silent mutations.

Mutations which affect the amino acid sequence are called nonsynonymous mutations.

Deletions and Insertions

Deletions are loss of genetic material

…GAACTG → ...GACTG…

Insertions are a gain of genetic material

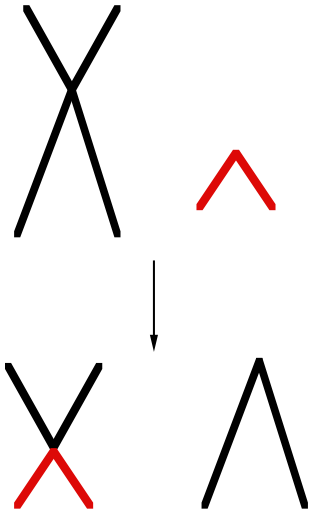
…GAACTG → ...GAATCTG…

Both insertions and deletions can result in frameshift mutations.

Chromosomal mutations

Inversions
1 2 3 4 5 6 7 8 → 1 2 6 5 4 3 7 8

Translocations



Mutational causes and rates

Causes of mutations

Mutations are caused by a variety of factors: chemical mutagens, radiation (x-rays, gamma rays, UV), and transposable elements

Mutation rates

Mutation rates per generation

Per base pair	$\sim 10^{-8} - 10^{-9}$
Per gene	$\sim 10^{-6} - 10^{-5}$
Per genome	~0.02 - 1

BUT -- These are highly variable from gene to gene, individual to individual, species to species

Modeling mutation

Equilibrium frequency with mutation - the bi-allelic case

Let μ be the mutation rate from A → a, and ν be the mutation rate from a → A

(μ is "mu" and ν is "nu")

p_t = the frequency of A at time t

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

(1-μ) is the probability that an A allele remains an A allele; ν is the probability that an a allele mutates to become an A allele.

$\hat{p} \equiv$ equilibrium frequency of A

At equilibrium, $p_t = p_{t-1} (= \hat{p})$

so…

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

becomes

$$\hat{p} = (1-\mu)\hat{p} + \nu(1-\hat{p})$$

$$\hat{p} = (1-\mu-\nu)\hat{p} + \nu$$

$$\hat{p}(\mu+\nu) = \nu$$

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

And it can be shown that

$$p_t = \hat{p} + (p_0 - \hat{p})(1-\mu-\nu)^t$$

Because μ and ν are usually *very* small, reaching equilibrium can take a very long time (tens of thousands of generations).

Infinite alleles models

Simplification of reality -- assumes that each gene has infinitely many possible alleles

(300 amino acid protein → 900 nucleotides → $4^{900} = 10^{542}$ possible sequences)

What do we expect heterozygosity to be?

Homozygosity = 1-H = $\Sigma p_i^2 = F$

Identity in state

$$F' = \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)F\right](1-\mu)^2$$

Are two randomly chosen alleles the same before mutation?

Are these two alleles still the same after the possibility of mutation?

Mutation-drift Balance

Mutation brings in new alleles; drift on average removes alleles

The increasing effects of mutation and the decreasing effects of drift reach a *mutation-drift balance* for the amount of heterozygosity and genetic variance.

The equilibrium value of F

At equilibrium $F_t = F_{t-1} \equiv \hat{F}$

So

$$\hat{F} = \left[\frac{1}{2N} + \left(1 - \frac{1}{2N}\right)\hat{F}\right](1-\mu)^2$$

solves to approximately

$$\hat{F} = \frac{1}{4N\mu + 1}$$

(this is valid for small values of μ).

The heterozygosity at the equilibrium between drift and mutation is

$$\hat{H} = 1 - \hat{F} = \frac{4N\mu}{4N\mu + 1}$$

Neutral Theory

The idea that many mutations have little or no effect

Due to Motoo Kimura.

For neutral alleles, the substitution rate equals the mutation rate.

Why?

The substitution rate can be found by multiplying the number of new mutations per locus per population per generation times the probability that a single new mutation fixes.

For a new mutation, the allele frequency is $p = 1/2N$.

For a neutral allele, the probability of fixation is p.

# of new mutations/locus/population/generation = $2N\mu$.

New mutation fixed per generation = $2N\ \mu\ (1/2N) = \mu$.

Measuring mutation rates

Possible ways:

- Directly observe changes from one genotype to another
- Assume selective neutrality - measure substitution rates
- Accumulate mutations over many generations to magnify the effects

Mutation accumulation experiments

A way to measure mutational effects on fitness or some other quantitative character

(both the rate of mutation and the effects of those mutations)

Therefore it is important to eliminate selection (to keep and measure the deleterious mutations as well).

One way is to use *balancer chromosomes*.

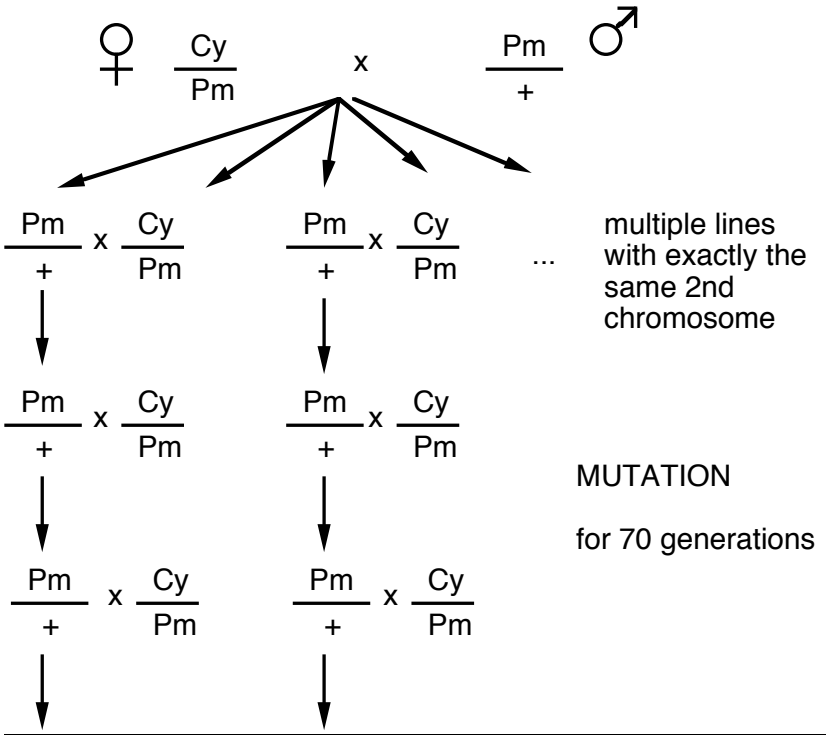
- Contain inversions, which suppress recombination
- Have dominant alleles for some easily seen phenotype
- Have recessive lethal allele (to kill homozygotes of the balancer)

These balancer chromosomes were developed to more easily maintain sick chromosomes in culture

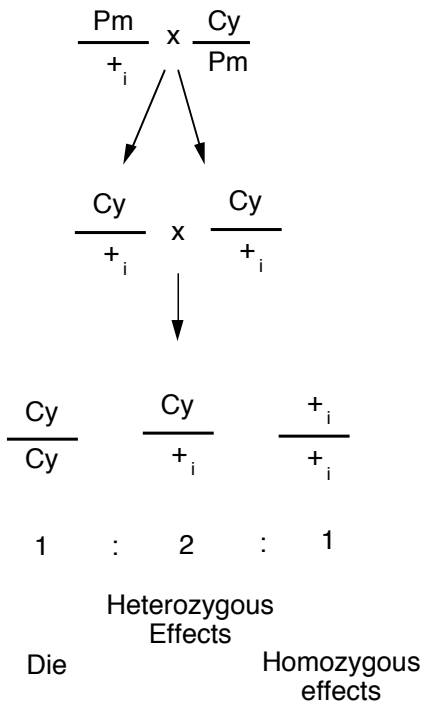
Cy (*Curly*) is a balancer for the 2nd chromosome in *Drosophila melanogaster*.

Pm (Plum) is a dominant mutation in *D. melanogaster*, also on the 2nd chromosome.

(Male *Drosophila* don't have recombination.)



Next, test the relative fitness of each mutation accumulation line:



The results:

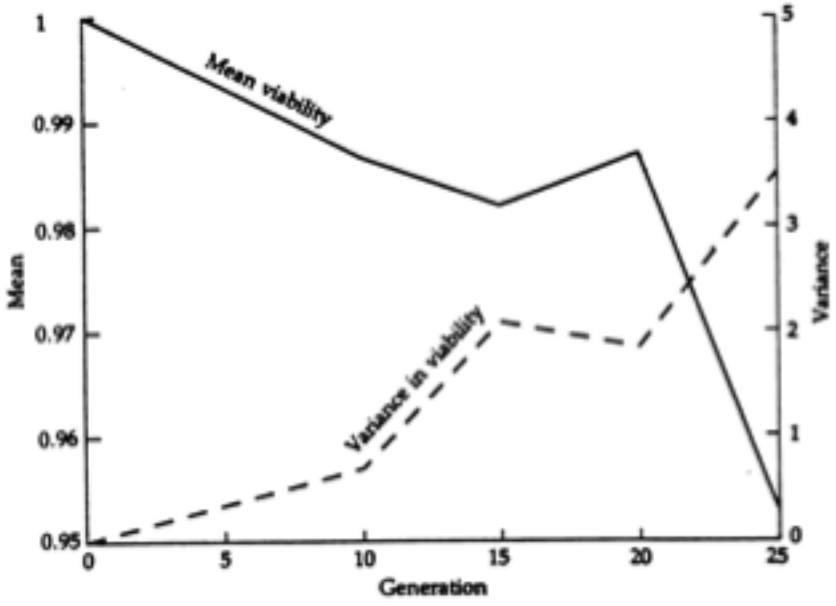


Figure 11. Results of Mukai's (1964) mutation accumulation experiment with *Drosophila*. The mean viability of the 104 lines decreased over time as they accumulated deleterious mutations, and the variance among lines increased. (From Mukai 1964.)

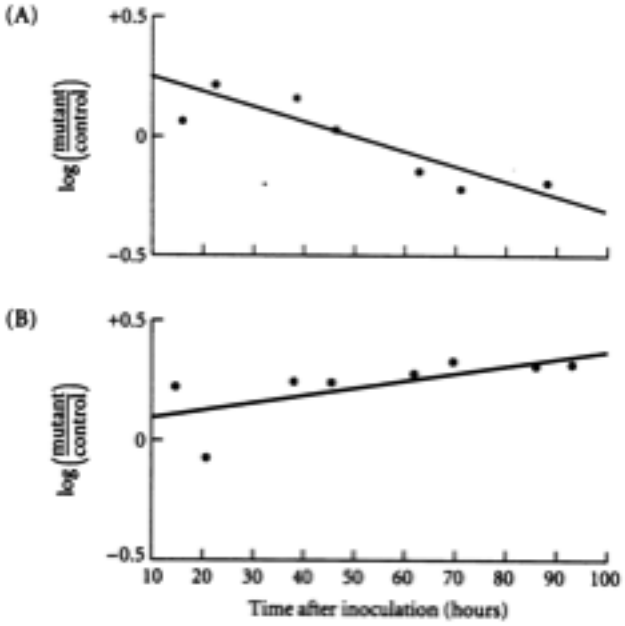
Selection

Natural Selection

The central idea of modern evolutionary biology, and Darwin's great contribution

- There is almost always *reproductive excess* (more offspring are produced than can survive and reproduce)
- Organisms vary in their ability to survive and reproduce (i.e. in their *fitness*)
- Genotypes with higher fitness leave more offspring, therefore these genotypes increase in frequency

FIGURE 12.8 Natural selection on mutations in the β-galactosidase gene of *Escherichia coli* in laboratory populations maintained on lactose. In each case, a strain bearing a mutation competed with a control strain bearing the wild-type allele. Populations were initiated with equal numbers of cells of each genotype, i.e., with log (mutant/control) initially equal to zero. Without selection, no change in the log ratio would be expected. (A) Mutation TD10.3 decreased in frequency, showing a selective disadvantage. (B) Mutation TD10.4 increased in frequency, demonstrating its selective (adaptive) advantage. (After Dean et al. 1986.)



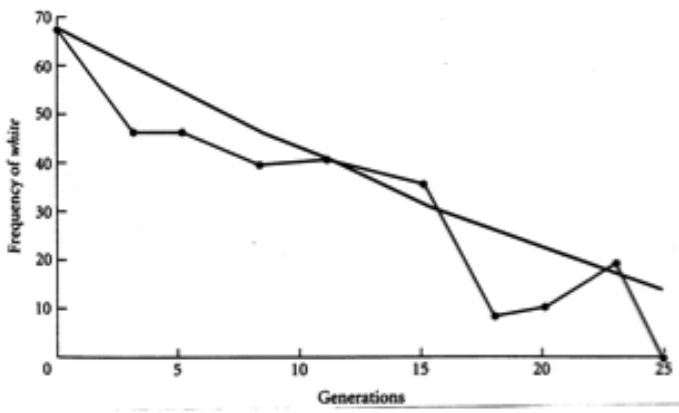
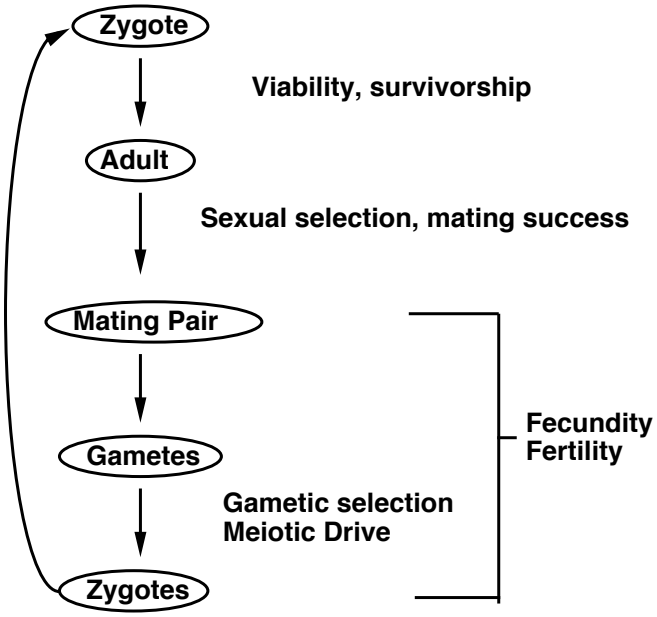


FIGURE 12.11 Elimination of the mutant allele *white* from a laboratory population of *Drosophila melanogaster*. The black line shows the decrease expected as a result of the lower mating success of white-eyed than wild-type males. (After Wallace 1968.)

Fitness and Fitness Components



Relative Fitness

Fitness of a genotype, standardized to a reference genotype

Genotype	Absolute Fitness	Relative Fitness *
AA	1.4	1.4/1.4 = 1
Aa	1.2	1.2/1.4 = 0.86
aa	0.8	0.8/1.4 = 0.57

* In this case, the fitnesses were standardized relative to the fitness of AA.

Measuring Selection

Measuring the fitness of different genotypes can be very difficult

- Each fitness component can be important, so all must be accounted for
- Must eliminate random effects
- Fitness can be genetically complex: dominance effects, interactions between loci, etc.
- Statistical problems: 1% fitness difference can be very important, but very difficult to reliably measure
- Fitness effects can be very environmentally dependent, and therefore vary over time and space

Simple Models of selection

Haploid selection

1 gene, 2 alleles

Let A and B be the numbers of the 2 alleles with growth rates a and b, so

$A_t = (1+a)^t A_0$

$B_t = (1+b)^t B_0$

$$\frac{A_t}{B_t} = \frac{(1+a)^t}{(1+b)^t}\frac{A_0}{B_0} = w^t \frac{A_0}{B_0}$$

where $w = \frac{(1+a)}{(1+b)}$ is the relative fitness of A.

Selection happens if a≠b.

If a > b , then the relative frequency of A will increase through time.

$$p_t = \frac{A_t}{A_t + B_t} = \frac{p_t N_t}{p_t N_t + (1 - p_t) N_t}$$

$$p_t = \frac{w\ p_{t-1}}{w\ p_{t-1} + (1 - p_{t-1})}$$

Proof:

$$p_t = \frac{A_t}{A_t + B_t} = \frac{(1 + a) A_{t-1}}{(1 + a) A_{t-1} + (1 + b) B_{t-1}}$$

$$p_t = \frac{w\ A_{t-1}}{w\ A_{t-1} + B_{t-1}}$$

Divide the top and bottom by ($A_{t-1} + B_{t-1}$):

$$p_t = \frac{w\ p_{t-1}}{w\ p_{t-1} + (1 - p_{t-1})}$$

So population size cancels out in modeling the effects of selection (when the relative fitness is not a function of population size)

Change in allele frequency (haploids)

$$\Delta p = p_t - p_{t-1} = \frac{w\ p}{w\ p + q} - p$$

$$= \frac{pq(w-1)}{w\ p + q}$$

The Δp notation refers to " the change in p"

Diploid selection

Let w_{11}, w_{12}, and w_{22} be the relative fitnesses of AA, Aa, and aa.

Life cycle:

Random mating →

Hardy-Weinberg frequencies

$$p^2 : 2pq : q^2$$

→ Differential survivorship

$$w_{11}\, p^2 : w_{12}\, 2pq : w_{22}\, q^2$$

→Standardize to sum to 1

$$(w_{11} / \overline{w})\, p^2 : (w_{12} / \overline{w})\, 2pq : (w_{22} / \overline{w})\, q^2$$

where $\overline{w}$ is the mean fitness of the genotypes:

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

The new allele frequencies are then found from these new genotype frequencies:

$$p' = p'_{11} + \frac{p'_{12}}{2}$$

$$= \frac{p^2 w_{11} + pq w_{12}}{\overline{w}}$$

$$q' = \frac{q^2 w_{22} + pq w_{12}}{\overline{w}}$$

We can also write this as $p' = p\left(\frac{p w_{11} + q w_{12}}{\overline{w}}\right)$

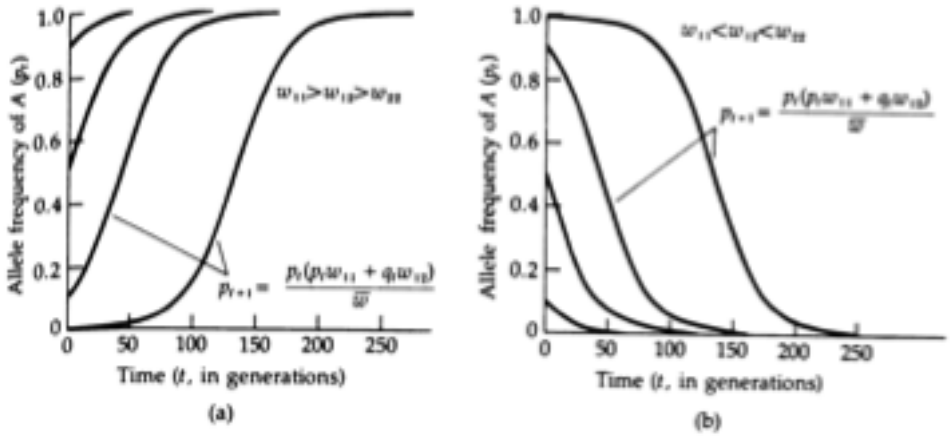


Figure 2. Changes of allele frequency against time when (a) allele *A* is favored, and (b) allele *a* is favored. (a) Fitnesses assumed are $w_{11} = 1$, $w_{12} = 0.9500$, $w_{22} = 0.9025$. (b) Fitnesses assumed are $w_{11} = 0.9025$, $w_{12} = 0.9500$, and $w_{22} = 1$. In both cases $w_{12} = \sqrt{w_{11}w_{22}}$, which allows an explicit expression for p_t to be obtained.

Additive effects

Let's look at the case with no dominance (i.e. the heterozygote is exactly intermediate to the homozygotes in fitness)

$w_{11} = 1+2s$, $w_{12} = 1+s$ and $w_{22} = 1$

We refer to s as the *selection coefficient*.

$$p' = \frac{p^2(1+2s) + pq(1+s)}{p^2(1+2s) + 2pq(1+s) + q^2(1)}$$

$$= \frac{p(1+s+sp)}{1+2ps}$$

$$\Delta p = p' - p = \frac{p(1+s+sp)}{1+2sp} - p = \frac{spq}{1+2ps} \approx spq$$

Notice that Δp is a function of two aspects of biology: the strength of selection and the amount of genetic variation.

Response to Selection

depends on the strength of selection and the genetic variance,

so the rate of change of allele frequency increases with

- greater selective differences
- intermediate p

Equilibria

Let the equilibrium allele frequency under selection be $\hat{p}$

With no mutation or migration bringing in new alleles, p = 0 and p = 1 are always possible equilibria

If $w_{11} < w_{12} < w_{22}$ then p→0

If $w_{11} > w_{12} > w_{22}$ then p→1

If $w_{11} < w_{12} > w_{22}$ then there will be a stable intermediate allele frequency (*overdominance*)

If $w_{11} > w_{12} < w_{22}$ then p→1 or p→0, depending on the initial allele frequency (*underdominance*)

Nature of equilibria: Unstable v. stable

Stable equilibrium: system moves to this point from nearby

Unstable equilibrium: system doesn't change if exactly at this point, but moves away with a small perturbation.

e.g. If $w_{11} < w_{12} < w_{22}$ then p=0 is a stable equilibrium point, but p=1 is an unstable equilibrium point.

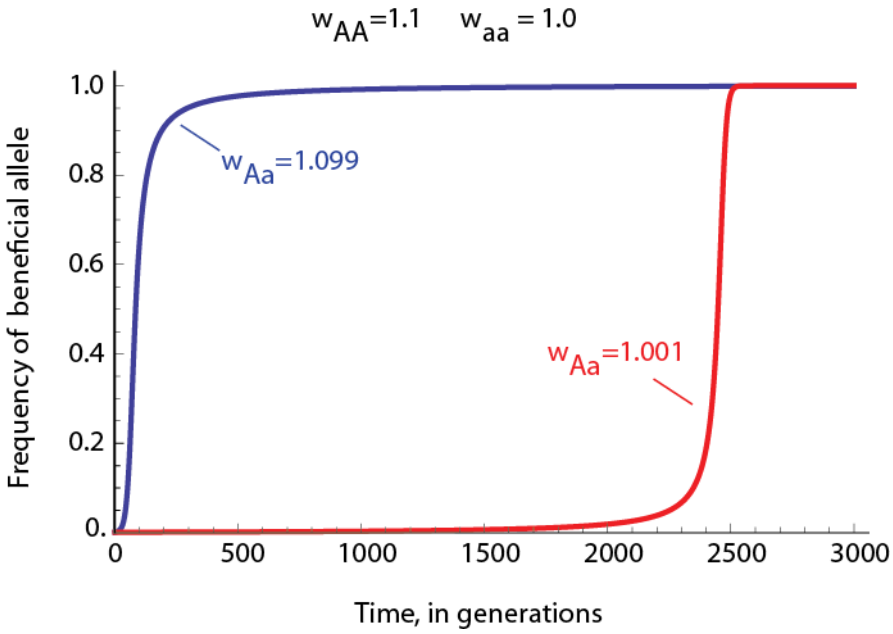
Dominance

AA	Aa	aa	
1+s	1+hs	1	Relative fitnesses

If h = 1/2, then these alleles are co-dominant with respect to fitness

If h <1/2, then A is recessive (or partially recessive)

if h >1/2 then A is dominant (or partially dominant)

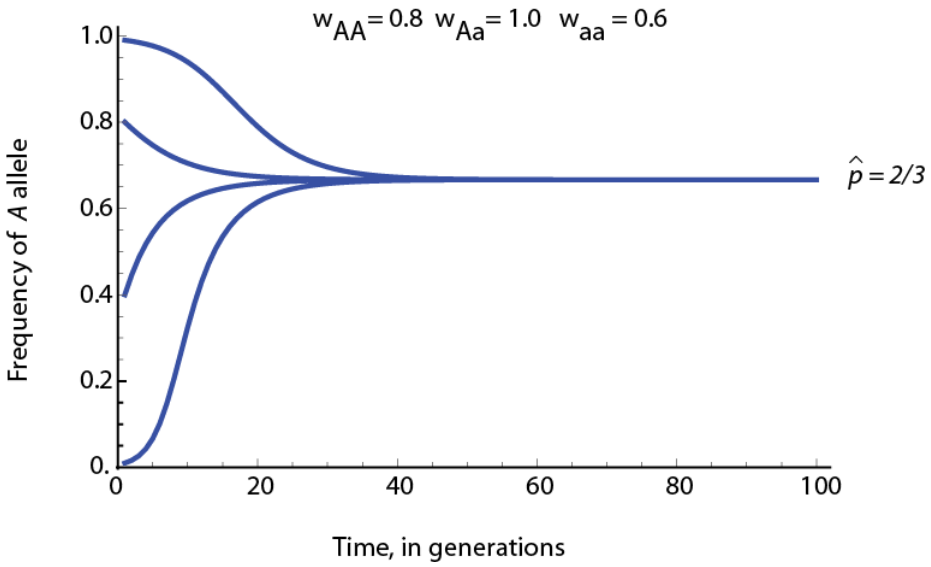


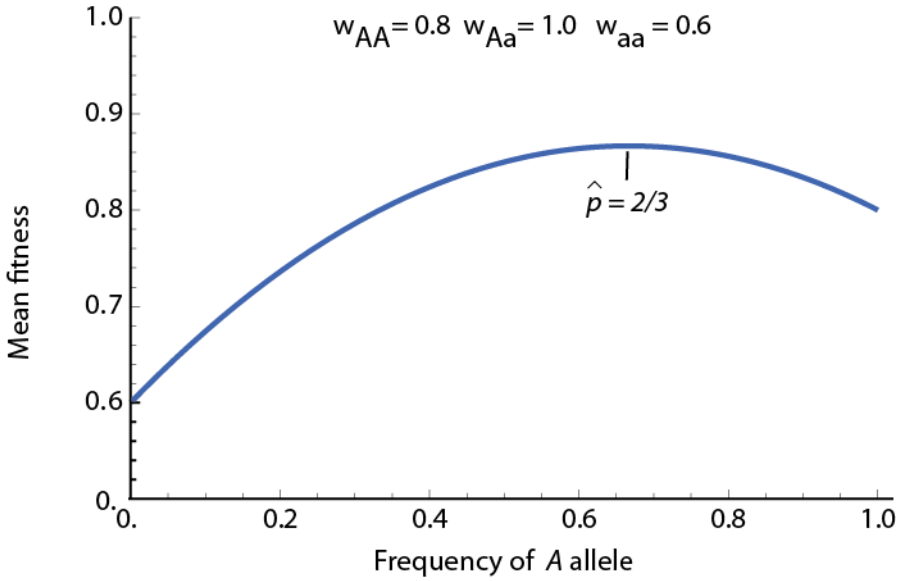
Overdominance

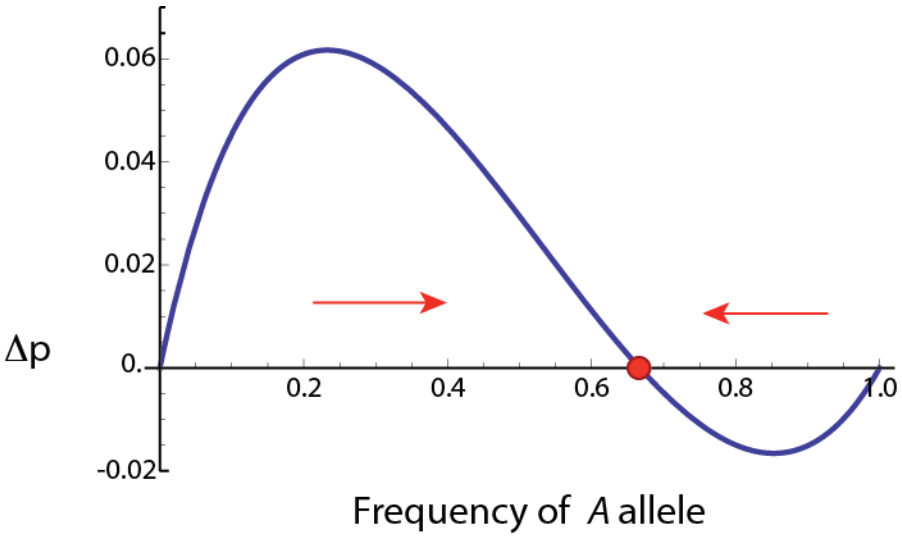
Heterozygote is the most fit genotype

AA	Aa	aa	
1-s	1	1-t	Relative fitnesses

$$\hat{p} = \frac{t}{s+t}$$


$w_{AA} = 0.8$ $w_{Aa} = 1.0$ $w_{aa} = 0.6$
1.0
0.8
0.6
0.4
0.2
0.
Frequency of A allele
$\hat{p} = 2/3$
0
20
40
60
80
100
Time, in generations


$w_{AA} = 0.8$ $w_{Aa} = 1.0$ $w_{aa} = 0.6$
1.0
0.9
0.8
0.7
0.6
0.
Mean fitness
$\hat{p} = 2/3$
0.
0.2
0.4
0.6
0.8
1.0
Frequency of A allele



Overdominance can maintain genetic variation ---

but relatively few examples are known.

The classic example is β hemoglobin in humans:

Two of the alleles of the β hemoglobin locus are A and S.

			Fitness *
AA	"normal"	sensitive to malaria	0.89
AS	heterozygote	resistant to malaria, slight sickling of red blood cells	1
SS	sickle cell	sickle cell anemia	0.2

- These are the fitnesses as measured in West Africa, in a malarial area

Multiple alleles and marginal fitness

n alleles with frequencies $p_1, p_2, p_3, \ldots p_n$

$\Sigma p_i = 1$

Use the marginal fitnesses of the alleles

$$w_i = \sum_j w_{ij} p_j$$

which is the average fitness of an individual with allele i.

$$p_i' = \frac{p_i w_i}{\overline{w}}$$

Table 4. Observations of hemoglobin *A*, *S*, and *C* genotypes and Hardy-Weinberg expectations for 72 West African populations.

	GENOTYPE						
	AA	*SS*	*CC*	*AS*	*AC*	*SC*	Total
Observed count	25,374	67	108	5482	1737	130	32,898
Expected count	25,615.5	306.87	74.69	4967.2	1768.6	165.01	32,898
Fitness[a]	0.991	0.218	1.446	1.104	0.982	0.788	—
Standardized fitnesses and their standard errors[b]	0.89 ± 0.03	0.20 ± 0.11	1.31 ± 0.29	1	0.89 ± 0.035	0.70 ± 0.07	—

[a] Fitnesses were calculated as the ratio of the observed to the expected counts.
[b] Fitnesses standardized by taking $\overline{AS} = 1$.
(From Cavalli-Sforza and Bodmer 1971.)

Fisher's Fundamental Theorem

The rate of increase of mean fitness is equal to the additive genetic variance for fitness

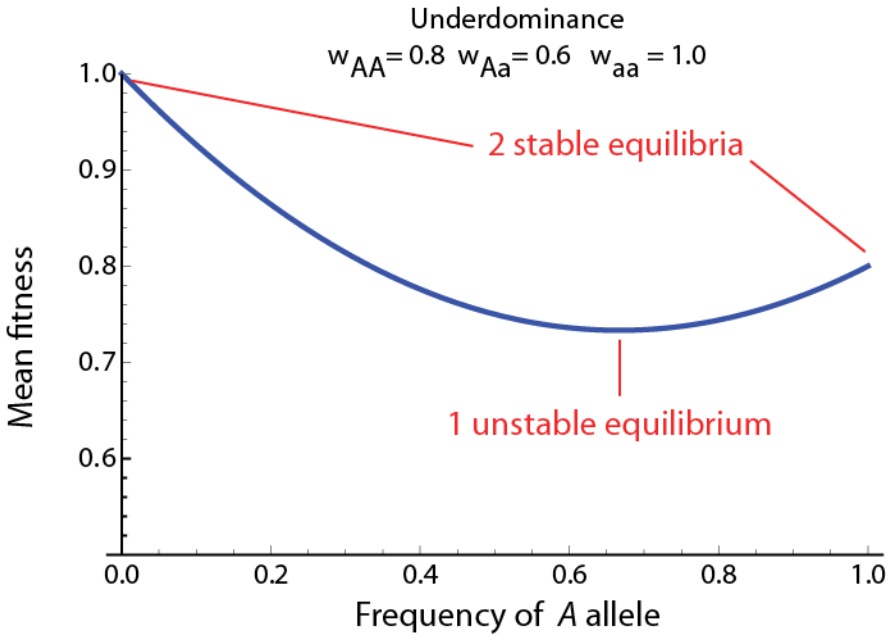
$$\Delta\overline{w} = V_A$$

The overall variance in fitness can be partitioned into genetic variance and environmental variance: V_A and V_E

For fitness, $V_E >> V_A$

The major implication of the fundamental theorem is that mean fitness increases by selection

(so we can find equilibria by finding the maximum $\overline{W}$.)


Underdominance
w_{AA}= 0.8 w_{Aa}= 0.6 w_{aa} = 1.0
2 stable equilibria
1 unstable equilibrium
Mean fitness
1.0
0.9
0.8
0.7
0.6
0.0
0.2
0.4
0.6
0.8
1.0
Frequency of A allele

Assumptions of the fundamental theorem

- No migration, mutation, or drift
- Genes interact additively
- Fitness effects are measured by the marginal effects of the alleles

$w_A = p\, w_{AA} + q\, w_{Aa}$

Sex specific selection

Selection on alleles can vary between sexes

Different ranks of fitness in males and females can act to maintain polymorphism.

Frequency - dependent selection

Fitness of an allele depends upon the frequency of that allele

e.g. coiling in snails or butterfly mimicry patterns

Frequency dependence can be positive or negative

Positive Frequency dependence -

Allele becomes more favored as it becomes more common

Negative Frequency dependence -

Allele becomes less fit as it becomes more common

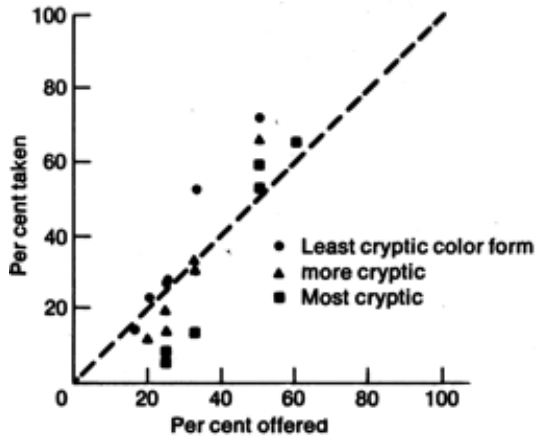
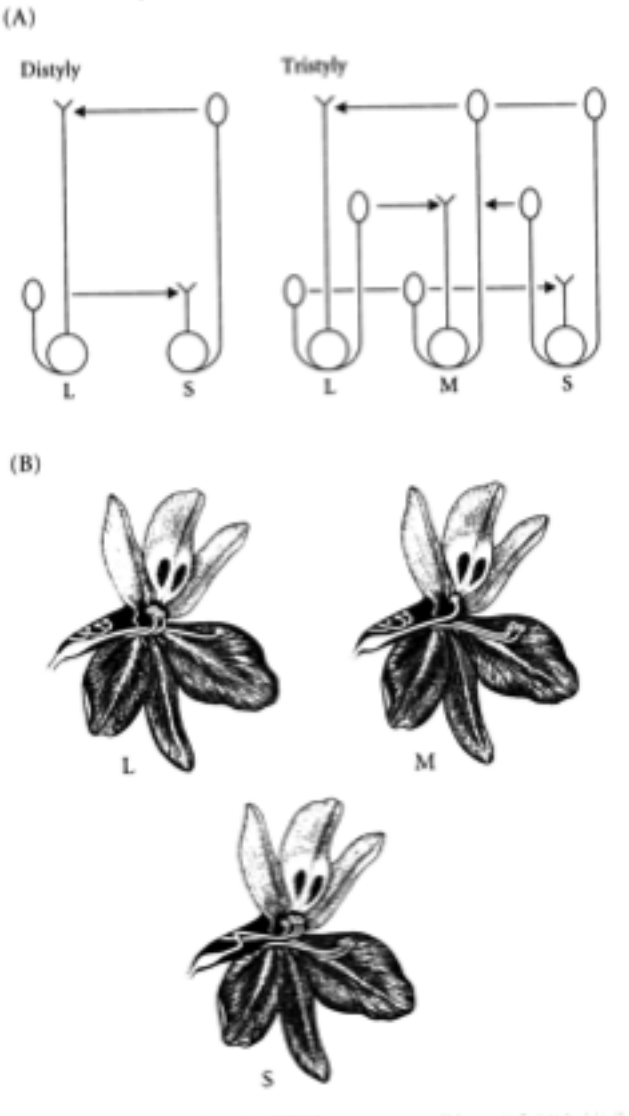


FIGURE 9
An example of frequency-dependent selection: predation by a fish on three color forms of the corixid bug *Sigara distincta*. Each suffers disproportionately higher predation (percentage taken) when it is the more common form. Compare with Figure 6 in Chapter 2. (After Clarke 1962, based on data of Popham 1942)

Heterostyly:

FIGURE 21.16 (A) A diagram of flower morphs in distylous and tristylous species. Ovals represent anthers; forks, stigmas. Compatible pollinations are shown by arrows; other pollinations usually produce little or no seed due to incompatibility. L, M, and S refer to long-, mid-, and short-styled morphs. (B) The long-, mid-, and short-styled morphs of the tristylous species *Eichhornia paniculata*. (A after Barrett 1992b; B courtesy of S. C. H. Barrett; drawings by E. Campolin.)



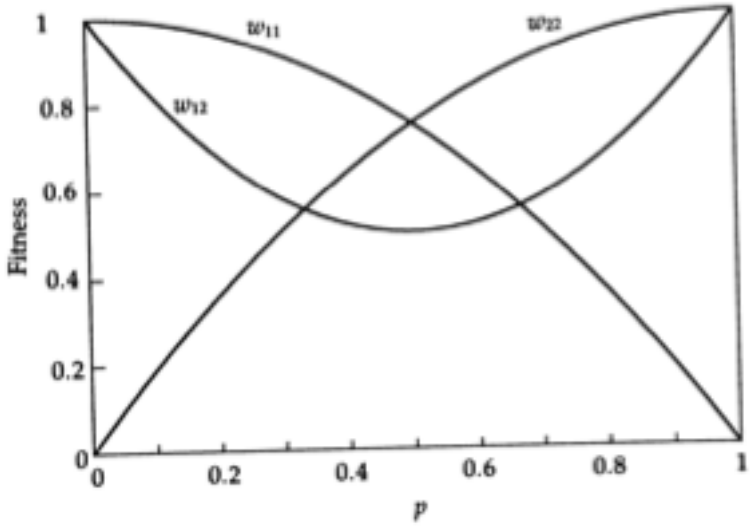


Figure 22. The model of frequency-dependent selection allows three genotypes to have fitnesses that vary with frequency, as plotted in the graph. Note that when the *A* allele is rare, the *AA* genotype has the highest fitness. In this example, the fitnesses are $w_{11} = 1 - p^2$, $w_{12} = 1 - 2pq$, and $w_{22} = 1 - q^2$.

Density - dependent selection

Relative fitnesses depend on population size

e.g. Selection on reproductive strategy
Feeding strategies

Sexual Selection

Differential mating success

2 types

- Intrasexual competition
  (Usually male-male competition)
  e.g. stags fighting over mates

- Choice
  (Usually female choice -- females choosing male mates)
  e.g. female peahens choosing peacocks with longer tails

Female choice in male traits can lead to exaggerated secondary sexual characteristics:
e.g. tail length, waddle size, plumage coloration, louder calls, etc.

Traits which increase mating success can evolve, even if they have some deleterious effects on survivorship

e.g. Forked fungus beetles

Males with longer horns on their back have higher mating success, but lower survivorship

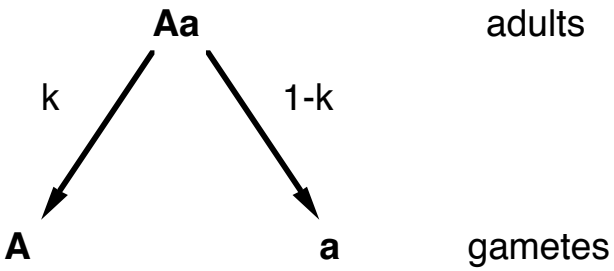
Fecundity selection

Relative fitnesses of alleles depend on producing differential numbers of offspring

Models of fecundity selection are not generally solvable and are much more difficult

Meiotic drive

Non-Mendelian production of gametes by heterozygotes



$p' = p^2 + 2kpq$

If k≠1/2, then meiotic drive is occurring.

Meiotic drive has been observed many times:

e.g. t alleles in mice
Segregation distorter (SD) in flies

Gametic selection

- differential success of gametes

e.g. pollen survival

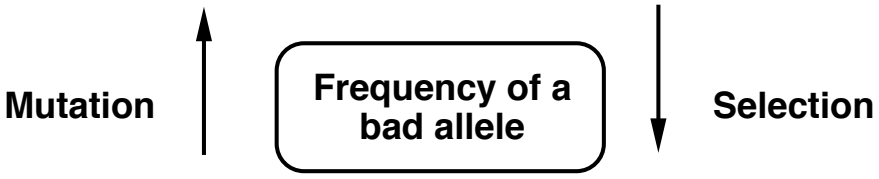
sperm competition (including sperm displacement, differential swimming rate, etc.)

Differential selection on different fitness components

e.g. Meiotic drive can be opposed by natural selection
Selection can be different in gametes and diploid phases
Viability selection can be different from fertility selection

Mutation-selection Balance

Most mutations are deleterious and decrease in frequency by selection



What is the equilibrium allele frequency?

AA	Aa	aa	
1	1-hs	1-s	Relative fitnesses

e.g. Huntington's Disease

$w_{11} = 1$, $w_{12} = 0.81$ and $w_{22} = 0$,

so s= 1 and h = 0.19.

a = deleterious allele

q = frequency of a
μ = mutation rate A → a

$$p' = \left(\frac{p^2 w_{11} + pq w_{12}}{\overline{w}} \right)(1 - \mu)$$

The first term is the change due to selection, and the second is the change due to mutation.

To find $\hat{p}$

$$\hat{p} = \frac{\hat{p}\left(\hat{p} w_{11} + \hat{q} w_{12}\right)(1 - \mu)}{\overline{w}}$$

$$\hat{q} hs\left(1 + \mu - 2\hat{q}\right) + \hat{q}^2 s = \mu$$

So..

If h=0 (complete recessive)

$$\hat{q}^2 s = \mu$$

$$\hat{q} = \sqrt{\frac{\mu}{s}}$$

If h>0 then $\hat{q}^2$ will be small, so

$$\hat{q} \cong \frac{\mu}{hs}$$

So for Huntington's disease, infer the mutation rate if $\hat{q} = 0.00005$.

$$\hat{q} \cong \frac{\mu}{hs}$$

$$0.00005 \cong \frac{\mu}{(1)(0{,}19)}$$

$$\mu \cong 9.5x10^{-6}$$

Mutation Load

The reduction in the mean fitness of a population due to deleterious mutations

$$\bar{w}(no\ mutations) \equiv 1$$

$$\bar{w}(with\ mutations) = 1 - q^2 s \quad (for\ h = 0)$$

$$\left[q^2 = \left(\sqrt{\frac{\mu}{s}} \right)^2 = \frac{\mu}{s} \right]$$

so

$$\bar{w}(h = 0) = 1 - \left(\frac{\mu}{s} \right) s = 1 - \mu$$

$$Load = \bar{w}(no\ mutations) - \bar{w}(mutations)$$

$$= \mu$$

Note that this load is independent of s.

(What's the load for mutations with h>0? 2μ)

In general the load is the number of mutations that come in per individual (2μ) divided by the number of mutations killed by an average selection event. This number of mutations removed per selective death is 2 for homozygous effects and 1 for heterozygous effects.

Drift and Selection

Genetic drift can obscure the deterministic process of selection. (In finite populations, the most fit allele does not always go to its deterministic equilibrium)

In small populations, drift can allow a deleterious allele to fix in a population. The reduction in fitness this causes is called *drift load*.

Table 2. The universal genetic code.

Codon	Amino acid	Codon	Amino acid	Codon	Amino acid	Codon	Amino acid
UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

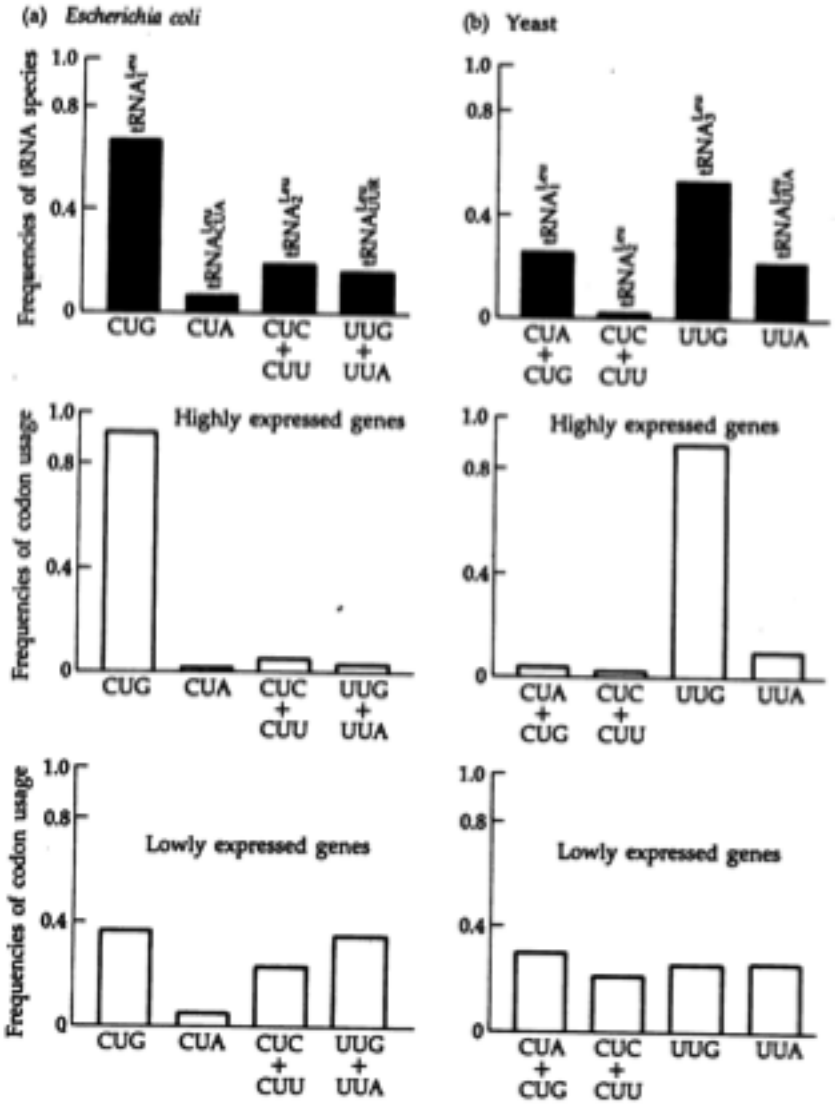


Figure 8. Diagram illustrating the relationship between the relative frequency of codon usage for leucine (open bars) and the relative abundance of the corresponding cognate tRNA species (solid bars) in (a) *Escherichia coli* and (b) *Sacharomyces cerevisiae*. The plus signs (e.g., between codons CUC and CUU for *E. coli*) indicate that each of these pairs of codons is recognized by a single tRNA species (e.g., $tRNA_2^{Leu}$ for CUC and CUU in *E. coli*).

Fixation of beneficial alleles

Even in large (or infinite) populations, a new mutation is present in only a few copies and is therefore likely to be lost by chance

The probability of fixation of a new beneficial allele is only 2s, where s is the increase in fitness of the heterozygotes.

In a non-ideal population, this probability becomes 2s N_e / N.
Note that by "new beneficial allele" we mean an allele that is present as a signle copy, i.e. at frequency p = 1/(2N).

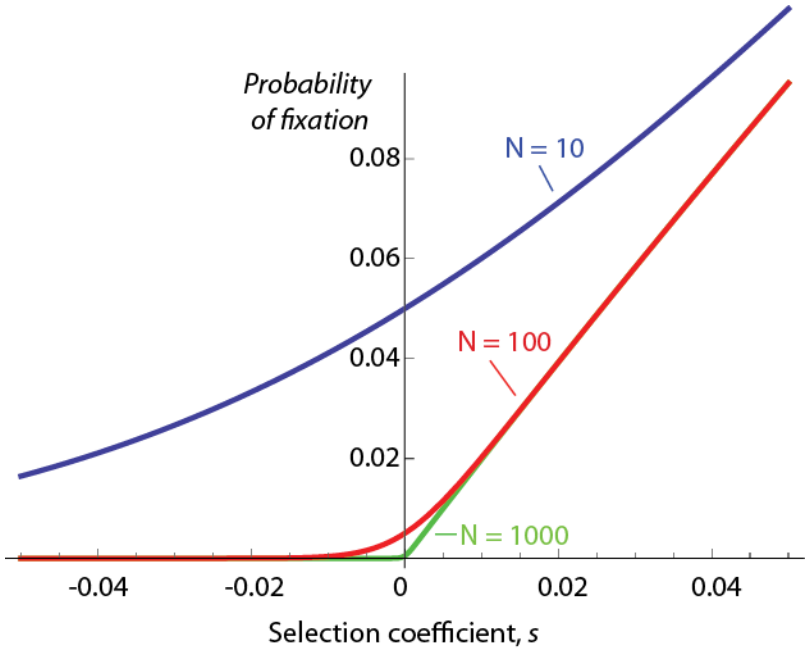
Fixation more generally

Probability of fixation of a new mutation (present initially as a single copy):

$$u \cong \frac{1-\exp\left[-2sN_e / N\right]}{1-\exp\left[-4sN_e\right]}$$

with fitnesses 1:1+s:1+2s.

Even when the allele is deleterious (i.e. s<0) there is some probability of fixation by chance.



Selection at multiple loci

If the fitness effects of two loci interact *multiplicatively*, then the 2-locus dynamics are completely described by the dynamics of each locus separately

i.e. $w_{AaBB} = w_{Aa}\, w_{BB}$ etc.

Epistasis

If the genes interact in some other way, we say that there is epistasis, and the changes at one locus affect what happens at the other
→

This tends to cause linkage disequilibrium (with an excess of the most fit genotypes), which is in turn broken down by recombination

So the fundamental theorem need not apply (because the genes are interacting in complex ways). Mean fitness *need not* be maximized.

BUT -- it turns out that if selection is weak and recombination is strong, then the fundamental theorem is very close to the right answer.

Example: *Moraba scurra*, a grasshopper from Australia

polymorphic for two inversions on two different chromosomes.

Fitnesses

	BB	BB'	B'B'
AA	0.79	1.00	0.83
AA'	0.67	1.006	0.90
A'A'	0.66	0.66	1.07

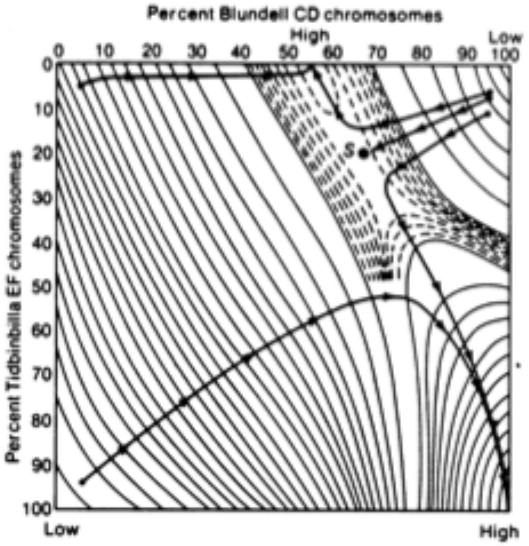


FIGURE 1

The adaptive landscape for a population of the Australian grasshopper *Moraba scurra* that is polymorphic for both the *EF* and *CD* chromosomes. Based on genotype frequencies in the field, viabilities were calculated for each genotype, and from these the theoretical fitness $\bar{w}$ of populations of each possible chromosomal constitution was calculated. Compositions of equal $\bar{w}$ are indicated by contour lines; the dashed lines indicate finer distinctions of $\bar{w}$ than the solid lines. There are two peaks (high) and a saddle point (*S*). The trajectories are theoretical changes in genetic composition a population would follow from five initial states. (After Lewontin and White 1960)

Detecting selection on proteins

There are many tests of selection at the protein level. One very commonly used approach is the McDonald-Kreitman test, which asks whether there are more coding changes between species than you would expect based on the variation within species. The McDonald-Kreitman test compares synonymous and non-synonymous differences within and among species, looking for an excess of coding changes between species.

The null model of the McDonald-Kreitman test, with no selection causing divergence between the species:

	Within species (Polymorphism)	Among species (Divergence)
Synonymous	$4 N_e \mu_S$	$2 t (\mu_S)$
Non-synonymous	$4 N_e \mu_N f$	$2 t (\mu_N) f$

where

N_e is the effective population size
μ_S is the mutation rate to synonymous codons
μ_N is the mutation rate to non-synonymous codons

f is the fraction of mutations which are not deleterious (the method assumes that deleterious alleles are efficiently removed from the population and are not observed.)

t is the number of generations since the common ancestor of the two species (so that there have been $2t$ generation s of evolution separating them, because both species have been evolving separately for t generations.)

These calculations assume that the polymorphism of a particular type is predicted by $4 N_e \mu$, and that the substitution rate per generation of neutral mutations is the mutation rate per individual, as shown by Kimura. (Hence the divergence is proportional to the number of generations in which a substitution could have occurred($2t$) times the substitution rate per generation (μ).)

Note that the rows and columns of this table can be constructed by the product of a term for the column ($4N_e$ or $4t$) and a term for the row (μ_S or $\mu_N f$). This means that to test this null model, we can use a contingency analysis; we can count the number of differences between individuals within or between species, and separate those into synonymous or non-synonymous differences. A contingency test allows a statistical test of whether this null model fits.

For example, McDonald and Kreitman (1991) looked at the Alcohol dehydrogenase gene in *Drosophila melanogaster*. Looking at the DNA sequence in the coding region of this gene (and comparing it to the sequence form other Drosophila species), they found:

	Within species (Polymorphism)	Among species (Divergence)
Synonymous	42	17
Non-synonymous	2	7

This shows a statistical significant excess of non-synonymous changes between species ($P = 0.006$), and so evidence of adaptive evolution at this locus.

Adam Eyre-Walker and his colleagues have taken this one step further. Comparing synonymous and non-synonymous changes across many genes in many species, they have estimated the fraction of amino acid changes between species that are due to positive selection. Their estimates range from about 0% in some plants to 10% in humans to up to 50% in rodents or flies.

Inbreeding

Inbreeding is the mating of relatives - a form of non-random mating

Inbreeding alone does not change allele frequencies, but inbreeding *does* change genotype frequencies.

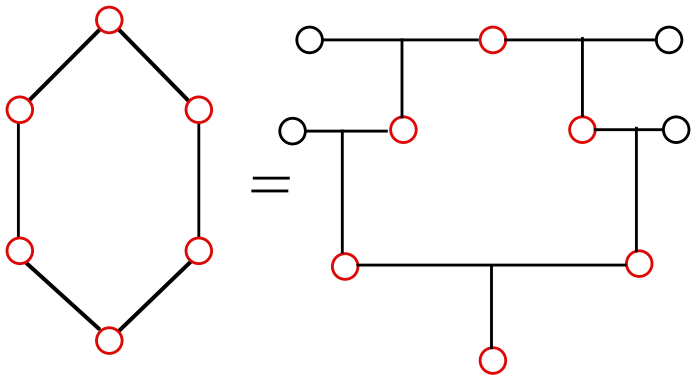
Inbreeding can affect allele frequencies, by changing how selection operates

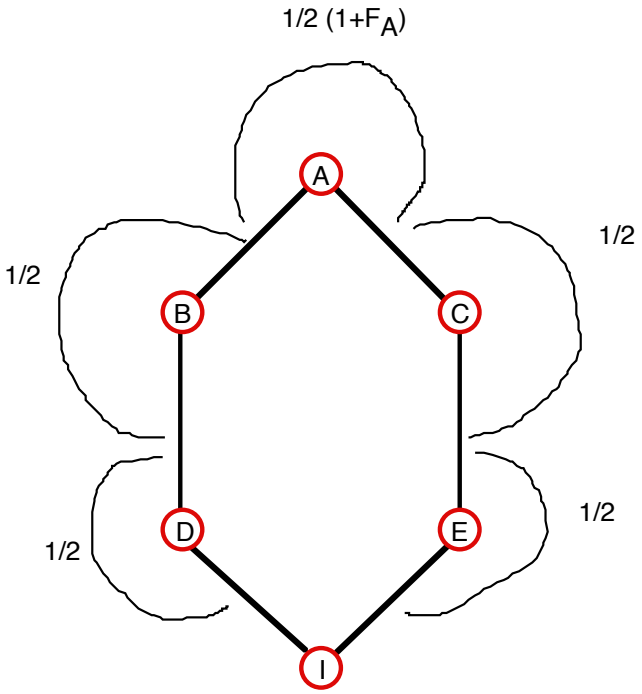
f: Inbreeding coefficient: the probability that 2 alleles in the same individual are identical by descent.

Calculation of F from pedigrees

Inbreeding—and the probability of identity by descent of two alleles in an individual—comes from cases when the two alleles in an individual may have the same ancestor.

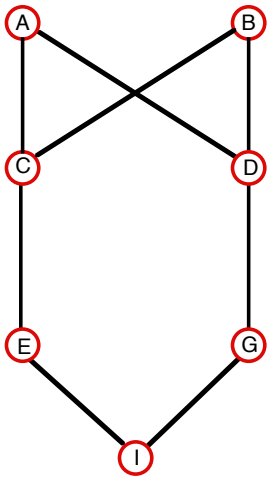
For example, in the following pedigree, the individual at the bottom descends has two great-grandparents which are the same individual.

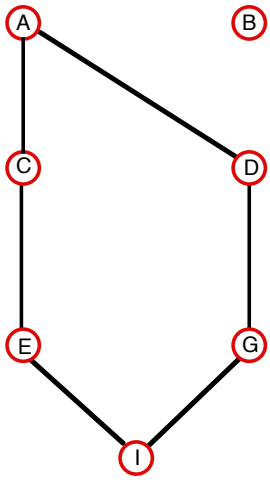


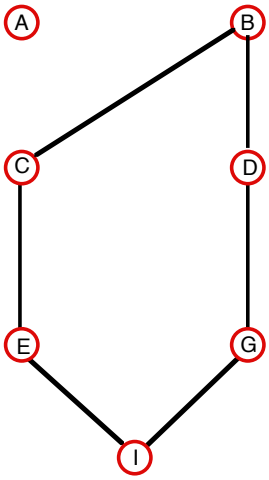


The probability that 2 alleles in individual I are identical by descent is $(1/2)^5(1+f_A)$.

There can be multiple paths through a pedigree: f is the sum of the probabilities from all of these paths







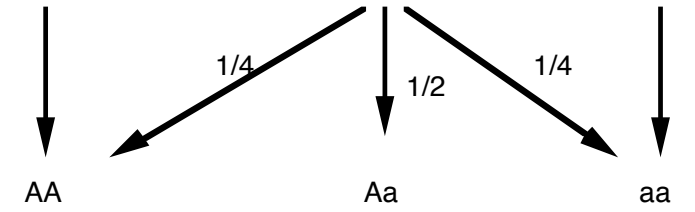
$f_I = (1/2)^5(1+f_A) + (1/2)^5(1+f_B)$

Genotype frequency changes with inbreeding

Main result: Excess Homozygotes

Remember the example of self-fertilization:

Genotype	AA	Aa	aa
Genotype frequencies	x	y	z
Self-fertilization	AA x AA	Aa x Aa	aa x aa
	AA	Aa	aa
Genotype frequencies next generation	x+y/4	y/2	z+y/4



We can measure inbreeding by the reduction of heterozygosity

$$f = \frac{(H_0 - H)}{H_0},$$

where H_0 is the heterozygosity of a randomly mating population with the same allele frequencies.

For the bi-allelic case, $H_0 = 2\ pq$, so $H = 2pq(1-f)$.

Genotype frequencies with inbreeding

AA	Aa	aa

$p^2 + fpq$	2pq(1-f)	$q^2 + fpq$
$pF +(1-f)p^2$	2pq(1-f)	$qf+(1-f)q^2$

i.e. 1/2 the "missing" heterozygotes become AA and 1/2 become aa, which leave the allele frequency unchanged:

$p = p^2 + pqf + pq(1-f) = p^2 + pq = p(p + q) = p$

Relationship between inbreeding and drift

As a population is divided into separate groups, each of size N, mating becomes non-random (i.e. mating only occurs within groups)

Therefore the probability that an individual has 2 alleles which are identical by descent goes up as $f' = 1/2N + (1-1/2N)f$

Inbreeding and fitness

Inbred individuals usually have lower fitness than outbred individuals.

This reduction in fitness with inbreeding is called *inbreeding depression*.

$$\delta = \frac{w_{outbred} - w_{inbred}}{w_{outbred}}$$

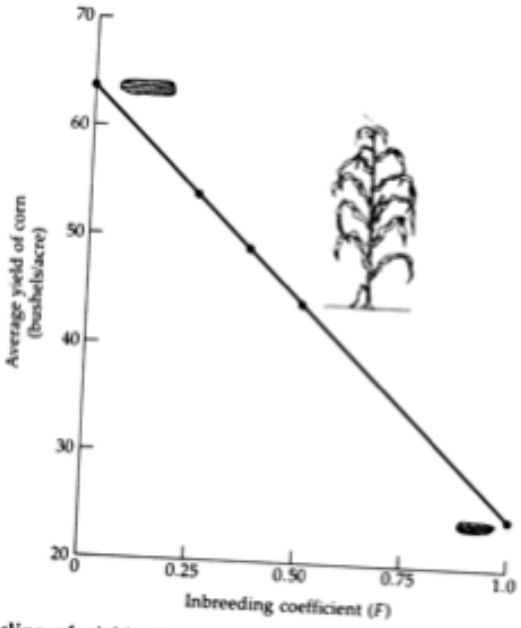


Figure 12. Decline of yield with inbreeding in corn (inbreeding depression). A linear inbreeding depression, as observed here, is expected to result from detrimental recessive alleles that are rendered homozygous by inbreeding. (Data from Neal 1935.)

Why is there inbreeding depression?

2 possible reasons:

(1) deleterious recessive alleles
(2) overdominance

In both cases, heterozygotes are more fit than the average of the homozygotes, and inbreeding increases homozygosity

Inbreeding depression by deleterious recessives

Genotype	AA	Aa	aa
Frequency in outbreds	p^2	2pq	q^2
Frequency in inbreds	p^2 + fpq	2pq(1-f)	q^2+ fpq
Fitness	1	1	1-s

$$\overline{w}_{outbred} = p^2 + 2pq + q^2(1-s)$$
$$= 1 - sq^2$$

$$\bar{w}_{inbred} = p^2 + fpq + 2pq(1-f) + (q^2 + fpq)(1-s)$$
$$= 1 - s(q^2 + fpq) < \bar{w}_{outbred}$$

The effects of inbreeding on selection

Inbreeding alone does not affect allele frequency; but it can influence the outcome of selection

Why?

Because an excess number of homozygotes are produced, and selection on homozygotes can be different from that on heterozygotes

e.g. Selfing

A fraction S of a population selfs each generation, and 1-S outcross (mate at random):

Let x, y, and z be the frequencies of AA, Aa, and aa:

$$x' = \frac{\left\{(1-S)p^2 + S\left[x + \frac{y}{4}\right]\right\}w_{11}}{\bar{w}}$$

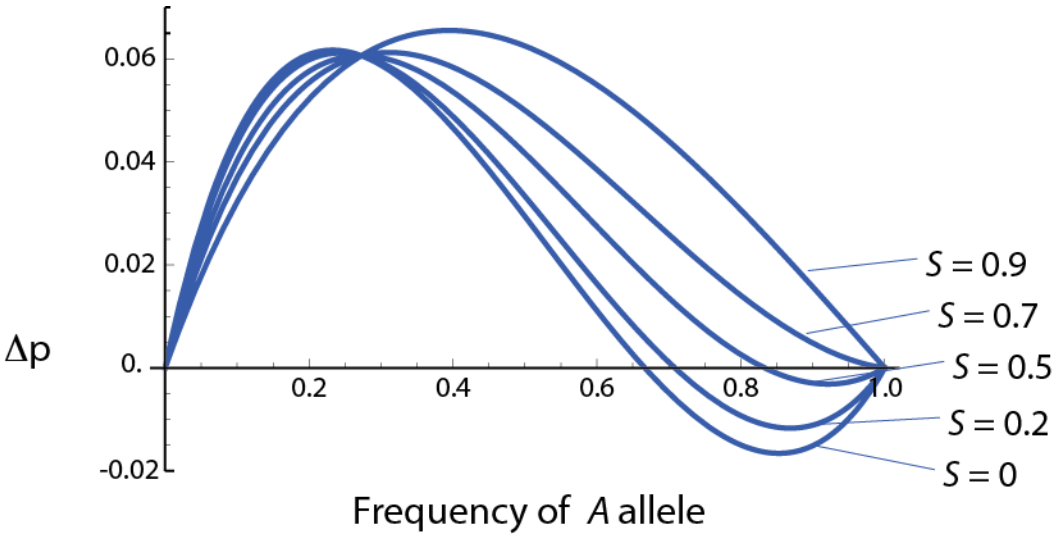
$$y' = \frac{\left\{(1-S)2pq + S\left[\frac{y}{2}\right]\right\}w_{12}}{\bar{w}}$$

$$z' = \frac{\left\{(1-S)q^2 + S\left[z + \frac{y}{4}\right]\right\}w_{22}}{\bar{w}}$$

where

$$\bar{w} = (1-S)\bar{w}_{outbred} + S\,\bar{w}_{inbred}$$

$$\bar{w}_{inbred} = w_{11}\left(x + \frac{y}{4}\right) + w_{12}\left(\frac{y}{2}\right) + w_{22}\left(z + \frac{y}{4}\right)$$



With sufficient selfing in a population, the internal equilibrium with overdominance moves towards more of the allele with the fitter homozygote, and the internal equilibrium can disappear. This is because the homozygote fitness matters more and more with selfing, because more alleles are expressed in homozygotes than expected with random mating.

Hybrid Vigor = Heterosis

Increased fitness of outcrossed individuals

e.g. hybrid corn increases yield 15-30%

Mixed Mating Model

Some individuals self S
Some outcross 1-S

With complete selfing: $f_{t+1} = 1/2\ (1+f_t)$

With partial selfing: $f_{t+1} = 1/2\ (1+f_t)\ S$

At equilibrium:

$$\hat{f} = S\left(\frac{1}{2}\right)\left(1+\hat{f}\right)$$

$$\hat{f} = \frac{S}{2-S}$$

which is used to estimate selfing rates.

Evolution of Inbreeding and inbreeding avoidance

Many adaptations have evolved to avoid inbreeding (and inbreeding depression)

e.g. tristyly, self-incompatibility alleles, sequential hermaphrodism, sex-biased dispersal

There are also many adaptations for selfing; e.g. cleistogamous (closed) flowers), small flowers, anthers close to stigma, etc.

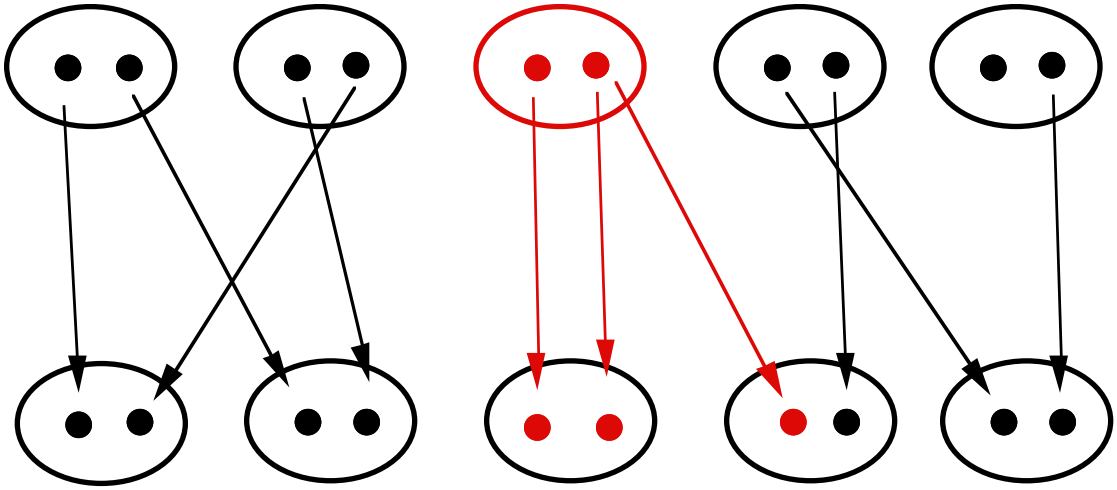
Evolution of selfing

There are obvious costs to selfing (inbreeding depression)

But also advantages:

- Reproductive assurance (don't need others to reproduce)
- Cost of outcrossing (producing offspring is expensive - why share with other genes?)

Individuals which self ***and*** act as males to other individuals get 3/2 as many alleles into the next generation, ***all else being equal***



Except for inbreeding depression, mutation which increase selfing rates would be more fit and therefore increase in frequency.

The fitness of an individual which selfs at rate r in a population that selfs on average $\bar{r}$:

$$w(r) = rw_s + \frac{1}{2}(1-r)w_O + \frac{1}{2}(1-\bar{r})w_O$$

where the three terms correspond to reproduction obtained through selfing, though outcrossed ovules and through pollinating other ovules, respectively.

w_S = fitness of selfed offspring, w_O = fitness of outbred offspring

When does Selfing increase fitness?

$$\frac{\partial w}{\partial r} > 0$$

$$\frac{\partial w}{\partial r} = w_s - \frac{w_O}{2}$$

So selfing increases fitness when $w_S > w_O/2$, in other words, when inbreeding depression is less than a half.

Inbreeding changes the frequency of deleterious mutations

For deleterious recessives, selfing exposes the homozygous alleles to selection much more often

$$\hat{q}_{selfing} \approx \frac{\mu}{s} < \hat{q}_{outcrossing} \approx \frac{\mu}{hs}$$

So inbreeding depression will lessen with selection → "*purging*"

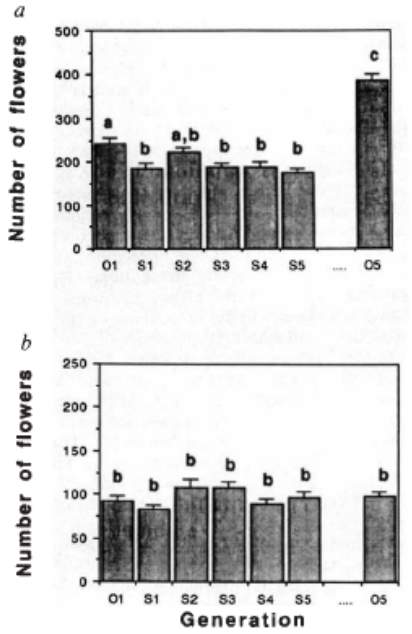


FIG. 1 Mean flower numbers of plants in various inbred and intercrossed generations. *a*, Naturally outbred population B11, NE Brazil. *b*, Naturally inbred population J13, Jamaica.

METHODS. Seed from open-pollinated families were randomly sampled from the two populations, B11 and J13. Lines were established in the glasshouse from each of 30 families from B11, and from each of 10 families from J13. One plant from each line was selfed and randomly outcrossed to other plants from the same population, to give the S_1 and O_1 generations, respectively. A small amount of S_1 seed was sown and a single randomly chosen plant was self-pollinated to give the S_2 generation. The remaining S_1 and O_1 seeds were stored dry at room temperature. The selfing was repeated for four successive generations with seeds from the S generations being stored. Plants from the lines were then selfed and outcrossed (within populations) to give the S_5 and O_5 generations. Stored seeds (maximum age 3 years) from all generations were then sown and the multi-generation plants grown in the glasshouse under uniform conditions. Fitness components were compared among generations in a randomized block design with all lines represented. Sample sizes were $n = 1007$ for population B11 (average 4.8 plants per line per generation), and $n = 378$ for J13 (average 5.4 plants per line per generation). Means and standard errors are shown, and the letters above the bars indicate statistically significant differences (5% level).

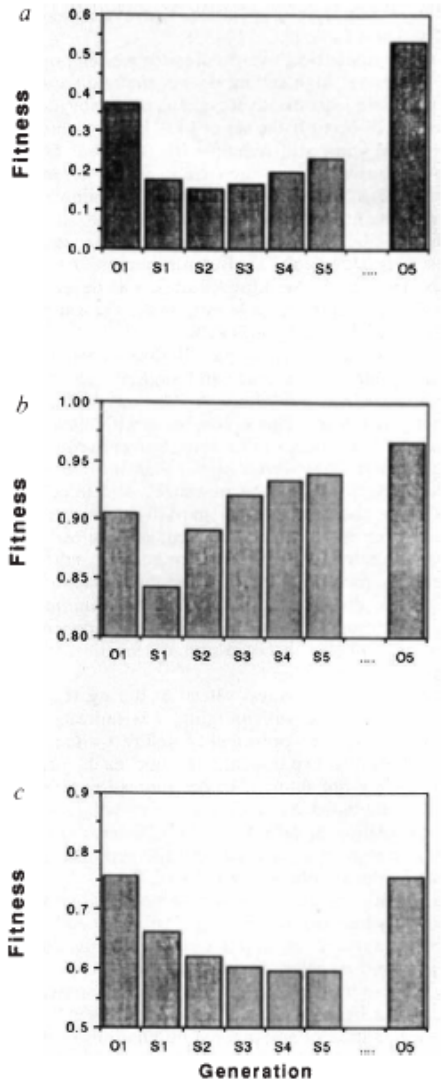


FIG. 2 Theoretical results for different selfed and intercrossed generations, assuming unlinked loci. *a*, Mutational model: selection coefficient $s = 0.2$; dominance coefficient $h = 0.2$; mutation rate per diploid genome $U = 1.0$. *b*, Mutational model: $s = 0.9$; $h = 0.2$; $U = 0.1$. *c*, Overdominance model: selection coefficients $s_1 = 0.1$ and $s_2 = 0.2$ at four loci.

METHODS. The computer calculations for the mutational model were obtained using the infinite population size model of Kondrashov[24] programmed as described[5]. In this model, the fitness of a homozygote for a mutant allele at one locus was denoted by $1 - s$, where s is the selection coefficient, and the fitness of heterozygotes by $1 - hs$, where h is the dominance coefficient. For the overdominance model, a five-locus deterministic program was used, with symmetric selection coefficients s_1 and s_2 (with the biologically implausible assumption of symmetrical overdominance, the mean fitness declines with inbreeding, even though variation is maintained with both alleles at frequencies of 0.5, and the fitness on intercrossing inbred lines will not exceed the original outbred value). To combine the fitness effects of different loci, multiplicative fitnesses were assumed in both models. For each model populations were run until equilibrium was reached under almost complete outcrossing (outcrossing rate 0.99), and then the selfing rate was altered to a high value (outcrossing rate 0.01). The time-course of changes in the population was then followed. Each generation, the fitnesses of inbred and outbred progeny produced in the population were calculated.

Therefore repeated inbreeding can create the conditions for selfing to evolve -- But this says that you have to have selfing to evolve selfing!

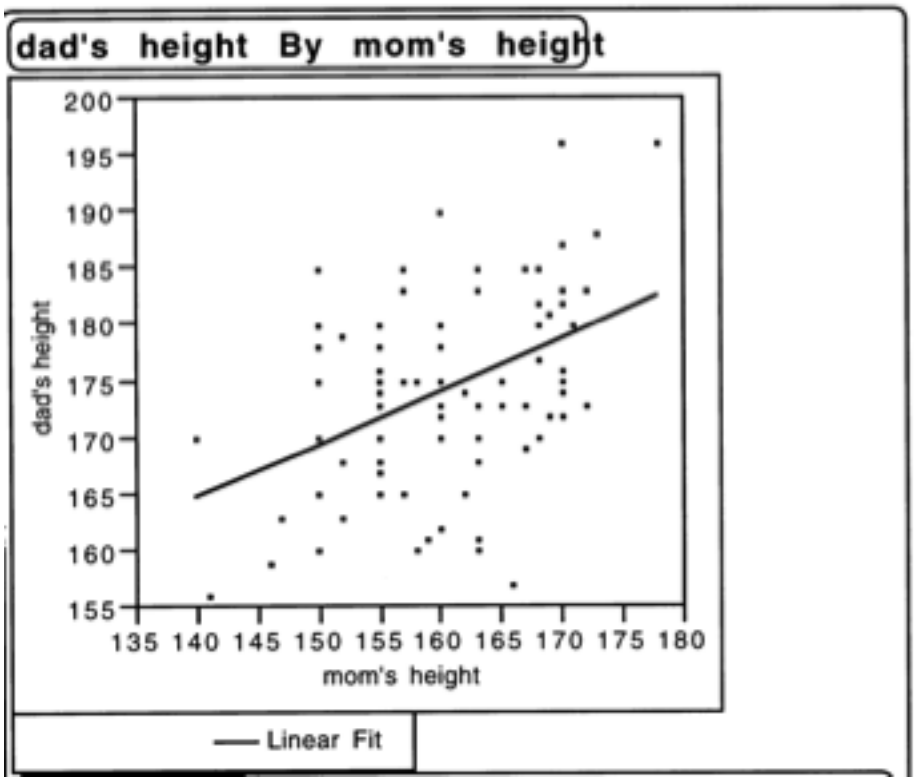
Other factors can allow for, or select for selfing

- Isolation → A colonizing individual may have no one else to mate with (most weeds are selfers)
- Other inbreeding → small population size can also result in a reduction in inbreeding depression

Assortative mating

Positive assortative mating → "like mates with like"→ positive correlation of mates

e.g. flowering time, human height, IQ



Negative assortative mating → negative correlation of parents

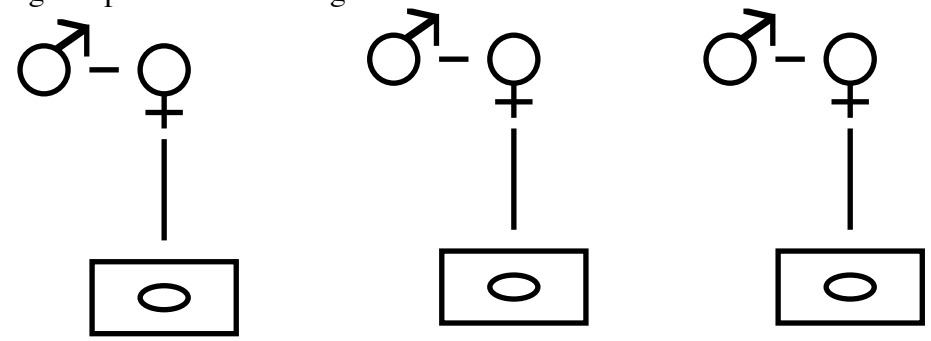
e.g. heterostyly, self-incompatibility alleles

Positive assortative mating tends to increase homozygosity

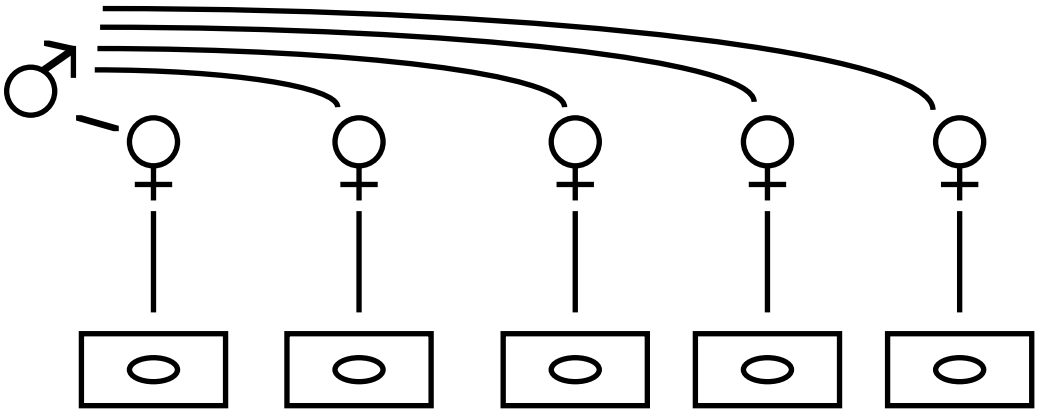
Negative assortative mating tends to increase heterozygosity, and usually acts as negative frequency dependence

Sex Ratio Evolution

A population of nearly all females with just enough males to allow fertilization has the highest possible intrinsic growth rate.



versus



So -- Why so many males?

Because

- All offspring have one male parent and one female parent
- If males are rare, than males will have a high fitness relative to females (males would contribute more of the genes of the next generation
- Same in reverse, if females are rare

Fisher showed that this causes an equal sex ratio

Subsequently it has been shown that this result holds regardless of the mechanism of sex determination (e.g. chromosomal sex determination, genic sex determination, environmental sex determination)

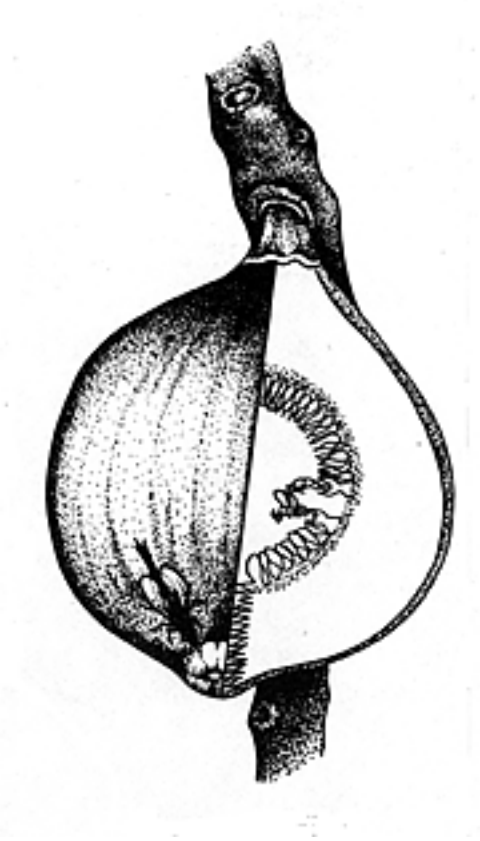
More precisely, mothers will equalize investment in reproductive male and female offspring
(which resolves the apparent contradictions in the Hymenoptera)

Exceptions:

If mating occurs among close relatives before mixing with the general population, this selects for female-biased sex ratios

- This is called local mate competition

e.g. fig wasps, *Nasonia* parasitic wasps



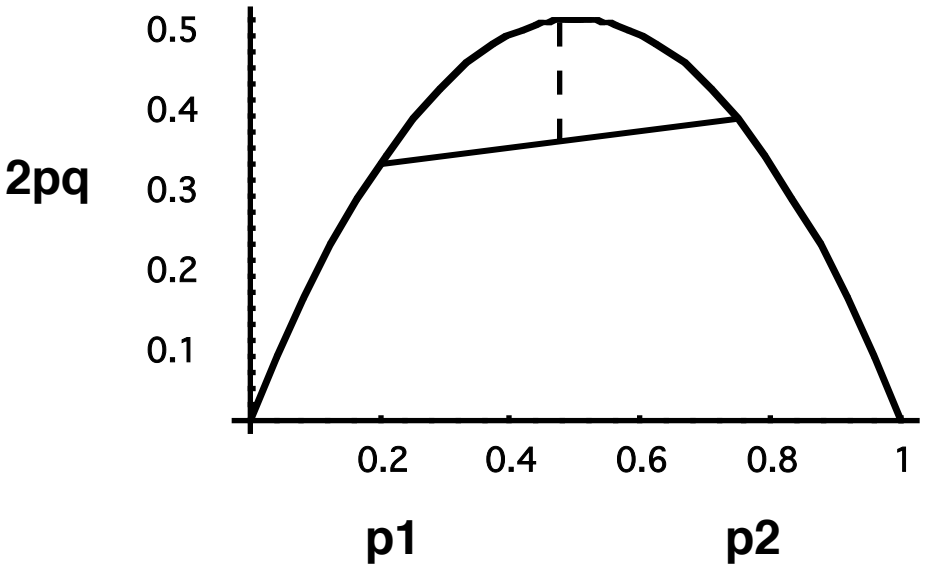
Population Structure

Migration and Population Structure

deme - semi-isolated sub-population

Wahlund Effect

The number of homozygotes in 2 separate populations is equal to or greater than the number in a randomly mating mixture of these 2 populations



F in subdivided populations

	AA	Aa	aa
Population 1	p_1^2	$2\ p_1\ q_1$	q_1^2
Population 2	p_2^2	$2\ p_2\ q_2$	q_2^2
Pooled Populations	$(p_1^2 + p_2^2)/2$	$p_1\ q_1 + p_2\ q_2$	$(q_1^2 + q_2^2)/2$
Pooled (in terms of F)	$\bar{p}^2(1-f)+\bar{p}f$	$2\bar{p}\bar{q}(1-f)$	$\bar{q}^2(1-f)+\bar{q}f$

This F we call F_{ST}.

$$\frac{p_1^2 + p_2^2}{2} = \bar{p}^2(1 - F_{ST}) + \bar{p}F_{ST}$$

$$E[p^2] = \bar{p}^2 + F_{ST}(\bar{p} - \bar{p}^2)$$

$$E[p^2] - \bar{p}^2 = F_{ST}\bar{p}\bar{q}$$

$$Var[p] = F_{ST}\bar{p}\bar{q}$$

$$F_{ST} = \frac{Var(p)}{\bar{p}(1-\bar{p})}$$

Figure 2. Map of Scandinavia showing the Skaggerak, the Kattegat, and the Baltic and the arbitrary subdivision into three areas used by Sick (1965) in a study of cod.

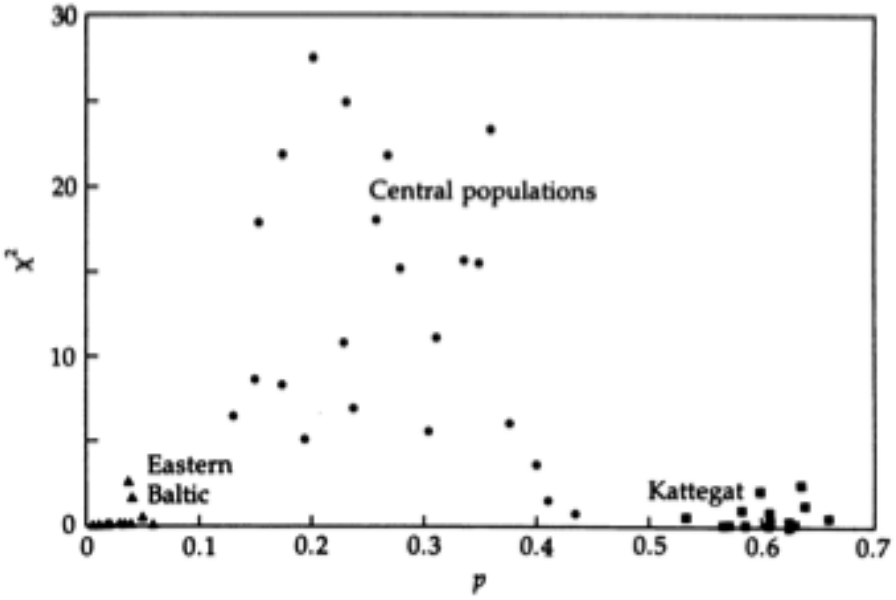


Figure 3. Chi-square values for the departure from panmixia in samples of cod taken from the Eastern Baltic (where $p < 0.1$), from central populations (where $p \approx 0.3$), and from the Kattegat (where $p > 0.5$). Note that the central populations differ greatly from Hardy-Weinberg proportions, suggesting that the sample is a mixture (there is a deficit of heterozygotes in the central populations).

2-locus Wahlund Effect

Differences in allele frequencies at 2 loci in multiple populations generates linkage disequilibrium

Wright's F-statistics

Population structure leads to increased homozygosity

Inbreeding within subpopulations also increases homozygosity

Wright's F-statistics can describe the effects of both

Definitions:

H_I : heterozygosity of an individual in a subpopulation (averaged over subpopulations)

H_S : Expected heterozygosity of an individual in a randomly mating subpopulation with the same allele frequencies (averaged over subpopulations) $= 2\, p_i\, q_i$

H_T: Expected heterozygosity of an individual in a randomly mating total population (given the average allele frequency of the total population) $= 2\bar{p}\bar{q}$

F_{IS}

Inbreeding coefficient *within* subpopulations

$$F_{IS} = \frac{\overline{H}_S - H_I}{\overline{H}_S}$$

F_{IS} → inbreeding coefficient of individuals within subpopulations

F_{ST}

Measure of the non-random mating *among* subpopulations

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T}$$

F_{ST} → inbreeding among subpopulations (S) within the total population (T)

G_{ST} is another measure of the genetic differences among subpopulations, but which allows for multiple alleles

F_{IT}

Overall inbreeding coefficient

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

$$\left(1 - F_{IT}\right) = \left(1 - F_{ST}\right)\left(1 - F_{IS}\right)$$

For outcrossing organisms, F_{IS} is usually small (<0.01)

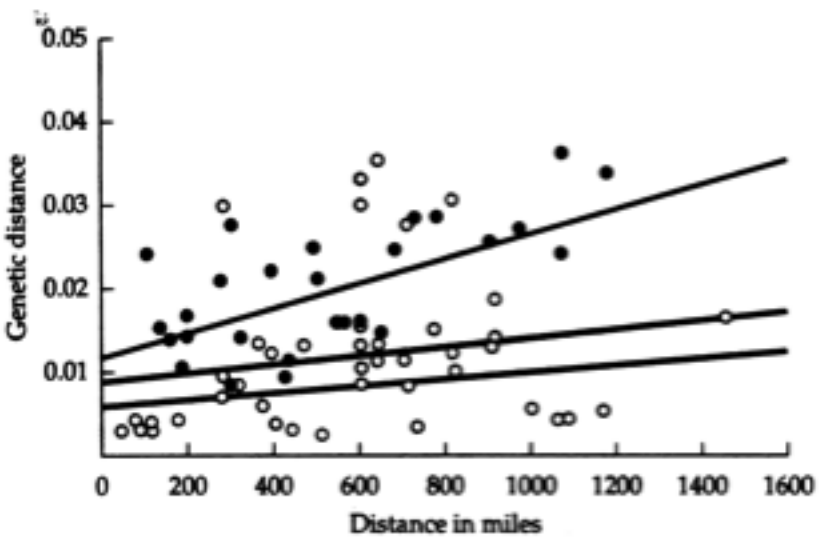
F_{ST} is variable among species (0 → 0.4)

Nei's D

(Note - this D does not mean linkage disequilibrium)

Another measure of the genetic differentiation of 2 populations

"Genetic distance"



ure 11. Regression of Nei's genetic distance on geographic distance among pop-tions of *Drosophila melanogaster* (solid dots; black graph) and *D. pseudoobscura* en circles; gray graphs). The top graph of *D. pseudoobscura* includes all popu-ons; the lower trace excludes the population at Nelson Ranch, Colorado (top en points). (From Singh and Rhomberg 1987b.)

Migration and Selection

Consider the simple situation where there is one-way migration from a mainland to an island. The mainland is fixed for an allele A, which is less fit on the island than another allele a.

What is the allele frequency on the island?

$$\Delta q_{mig} = -mq$$

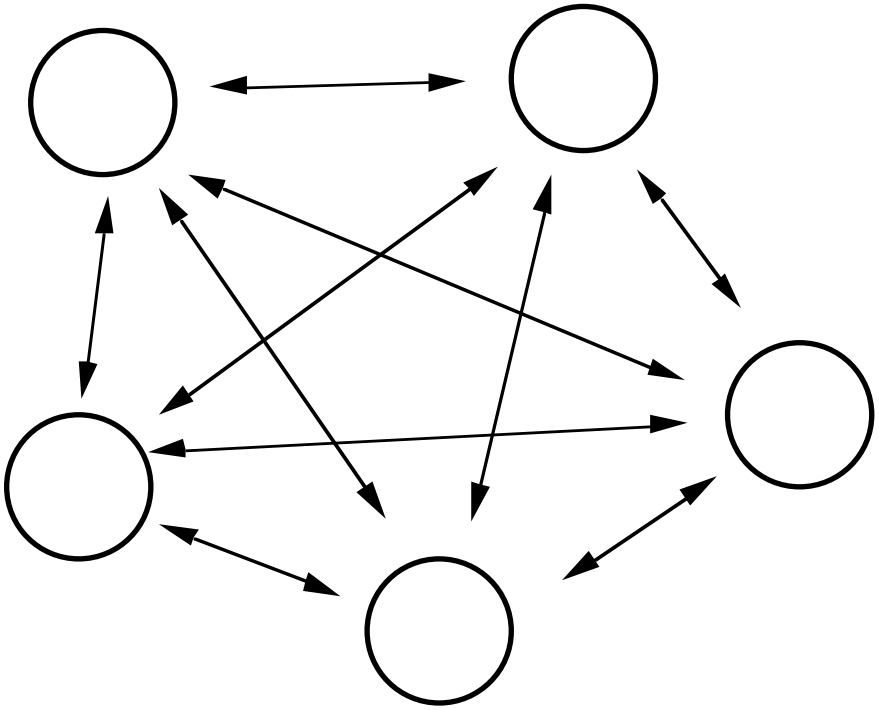
$$\Delta q_{sel} = \frac{pqw_{12} + q^2 w_{22}}{\overline{w}} - q$$

There will be an equilibrium when $\Delta q_{mig} + \Delta q_{sel} = 0$

a will be able to invade the island if s > m, where s is the selective benefit on the island of Aa over AA

Migration and Drift

Migration -drift balance; the island model



Each deme has N individuals and a proportion m of the individuals are migrants.

Assumptions:

- No selection
- No mutation
- All populations are the same size: N
- All populations contribute m of their individuals to the migrant pool
- All populations have m of their individuals arriving as migrants per generation
- Migration is random with respect to distance

$$F'_{ST} = \frac{1}{2N} + (1-m)^2\left(1-\frac{1}{2N}\right)F_{ST}$$

$$F_{ST} = \left(\frac{1}{2N} + (1-m)^2 \left(1 - \frac{1}{2N}\right) F_{ST} \right)$$

$$F_{ST} \left(1 - (1-m)^2 \left(1 - \frac{1}{2N}\right) \right) = \frac{1}{2N}$$

$$F_{ST} = \frac{1}{2N\left(1-(1-m)^2\right)\left(1-\frac{1}{2N}\right)} \cong \frac{1}{4Nm+1}$$

F_{ST} is often related to the number of migrants per generation by the formula

$$F_{ST} = \frac{1}{4Nm+1}$$

Migration is a potent way of reducing the genetic variance among populations.

e.g. if $Nm = 1$ (one migrant per generation) then $F_{ST} = 0.2$, which is not as much inbreeding as one generation of sib-mating.

So in principle, we can estimate something about the rate of dispersal:

$$Nm = \frac{1/F_{ST} - 1}{4}$$

Table 5. Estimates of Nm and $\hat{F}_{ST}$.

Species	Type of organism	Estimated Nm	Estimated $\hat{F}_{ST}$
Stephanomeria exigua	Annual plant	1.4	0.152
Mytilus edulis	Mollusc	42.0	0.006
Drosophila willistoni	Insect	9.9	0.025
Drosophila pseudoobscura	Insect	1.0	0.200
Chanos chanos	Fish	4.2	0.056
Hyla regilla	Frog	1.4	0.152
Plethodon ouachitae	Salamander	2.1	0.106
Plethodon cinereus	Salamander	0.22	0.532
Plethodon dorsalis	Salamander	0.10	0.714
Batrachoseps pacifica ssp. 1	Salamander	0.64	0.281
Batrachoseps pacifica ssp. 2	Salamander	0.20	0.556
Batrachoseps campi	Salamander	0.16	0.610
Lacerta melisellensis	Lizard	1.9	0.116
Peromyscus californicus	Mouse	2.2	0.102
Peromyscus polionotus	Mouse	0.31	0.446
Thomomys bottae	Gopher	0.86	0.225

(Data from Slatkin 1985a.)

We can estimate F_{ST} from genetic data:

$$F_{ST} = \frac{Var(p)}{\bar{p}(1-\bar{p})}$$

or

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T}$$

We can use data from multiple individuals from many populations to estimate distribution of allele frequencies., e.g. protein electrophoresis, DNA (such as microsatellites)

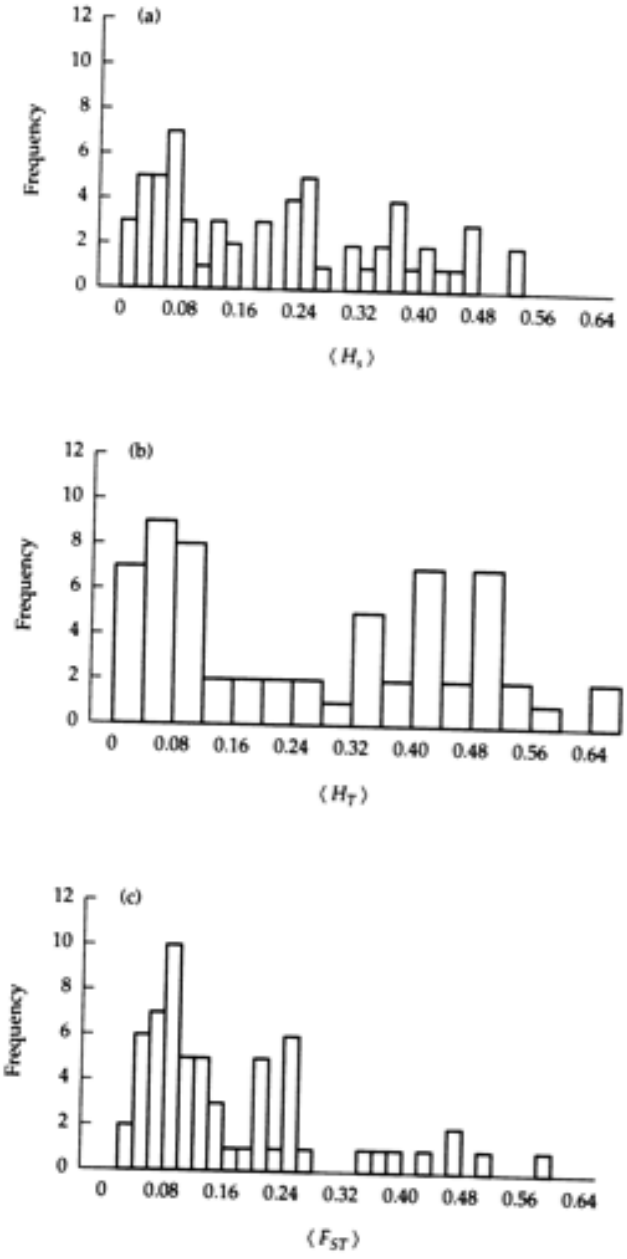


Figure 20. Distribution of (a) single locus heterozygosity (H_S), (b) total genic diversity (H_T), and (c) fixation index (F_{ST}) at polymorphic loci in geographic populations of *Drosophila melanogaster*. (From Singh and Rhomberg 1987b.)

BUT..... there are two problems:

- The real world is not like the island model
- Even the island model is not always like the island model: statistical problems

The real world is not like the Island model-- take the assumptions one by one:

- All populations are created equal, with N individuals and equal contributions to the migrant pool -
  - ⇒ Population sizes are extremely variable, both in space and time
  - ⇒ Populations are variable in their contributions to the migrant pool (e.g. sources and sinks)
  - ⇒ populations vary through time in migration rates
- There is NO spatial structure: in effect all populations are equally close to all other populations (no isolation by distance)
  - ⇒ Dispersal is almost always distance related; there is isolation by distance
  - ⇒ Dispersal is also often affected by other factors: rivers, roads, mountains, etc.
- Everything is at equilibrium, nothing is changing.
  - ⇒ Populations often go extinct, and new ones form by colonization
  - ⇒ History matters -- often the circumstances which determine the current population structure are the conditions of the past, which may have changed
  - ⇒ There may be migration in from outside the study system, changing allele frequencies over time
- No selection
  - ⇒ There's ALWAYS selection
- no mutation
  - ⇒ Mutation can be at very fast rates, for example in microsatellites

The island model is not always like the island model

- For mitochondrial markers (or others inherited uniparentally) $F_{ST} = 1/(2Nm+1)$
- The statistical properties of F_{ST} are not well worked out, but they're ugly - see the figures

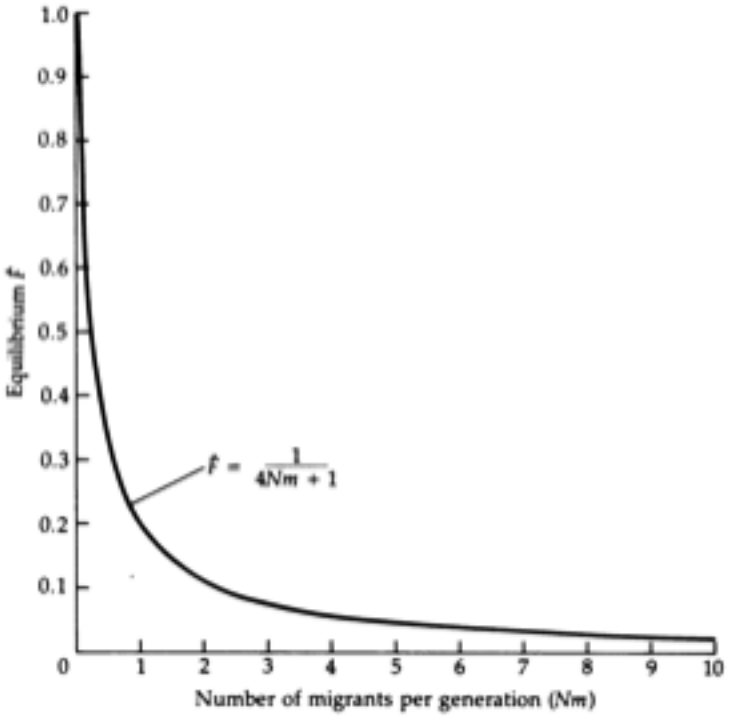


Figure 15. Decrease of equilibrium fixation index ($\hat{F}$) against number of migrants per generation (Nm). Note that only a few migrants are necessary to reduce $\hat{F}$, and thus population differentiation, to very small levels.

- Dispersal rates for genetic purposes are often quite different than what is needed for ecological studies
  - ⇒ Genetic dispersal only counts if the migrants reproduce effectively
  - ⇒ Genetic dispersal only counts if the reproduction of migrant individuals is equal to resident individuals (i.e., migrants have to move before their reproductive life starts)
  - ⇒ Selection can over-amplify migrant genetic contributions
- Problems of scale : Genetic analysis only tells you about migration at the geographical scale at which the samples are drawn from.

$$\psi(p) = C \quad p^{4N\mu+4Nm\bar{p}-1} q^{4N\nu+4Nm\bar{q}-1} \overline{W}^{2N}$$

This is Wright's distribution.

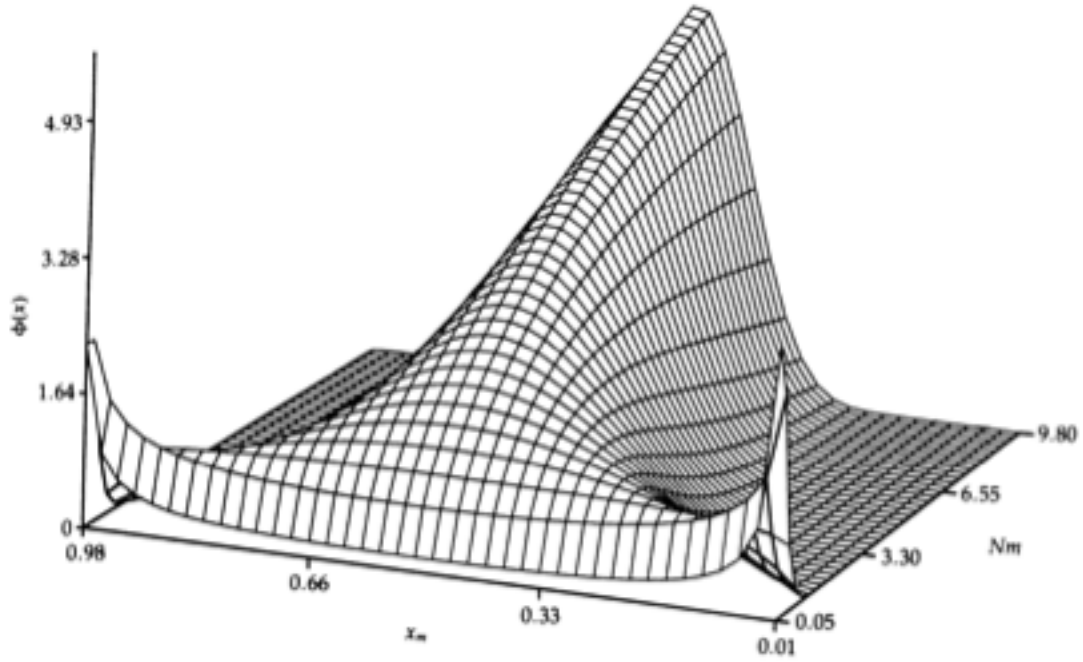
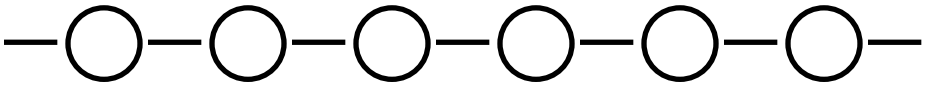


Figure 16. Wright's island model. The solution to the diffusion approximation (Equation 6.26) gives $\phi(x)$, the proportion of populations having allele frequency x, for a range of levels of migration Nm. With very low levels of migration, most populations are near fixation, but with high levels of migration (where migrants have an allele frequency $x_m = 0.5$), the populations tend to have intermediate allele frequencies.

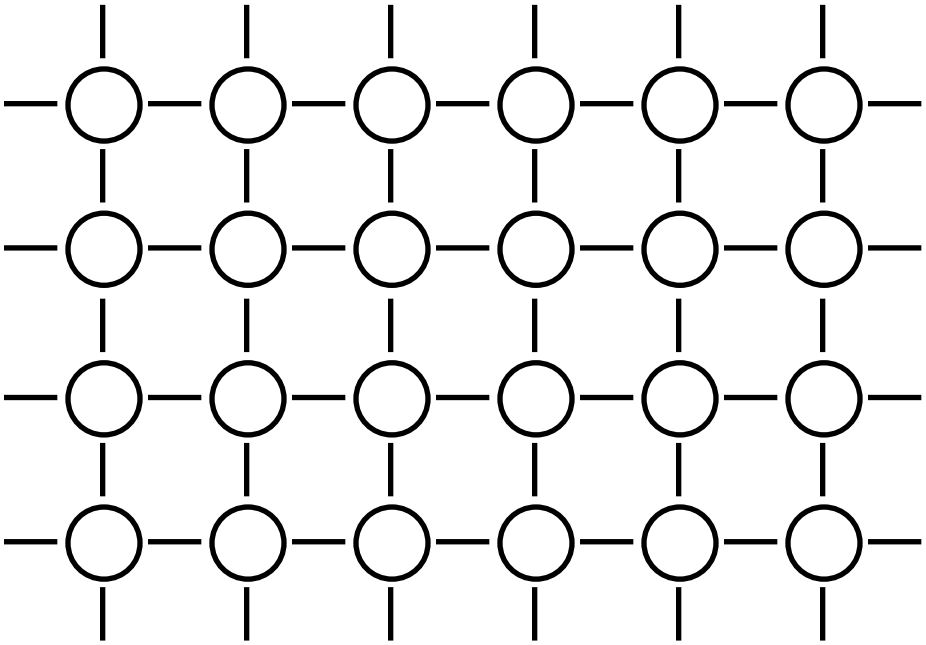
Alternate Models of Population Structure

Stepping stone models -- migrants come from adjacent demes

1-dimensional



2-dimensional



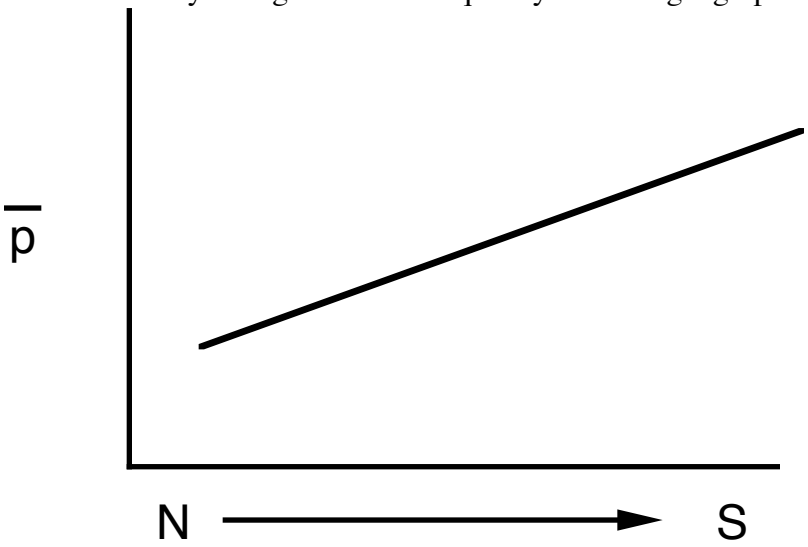
Local adaptation

Environments are heterogeneous at many spatial scales

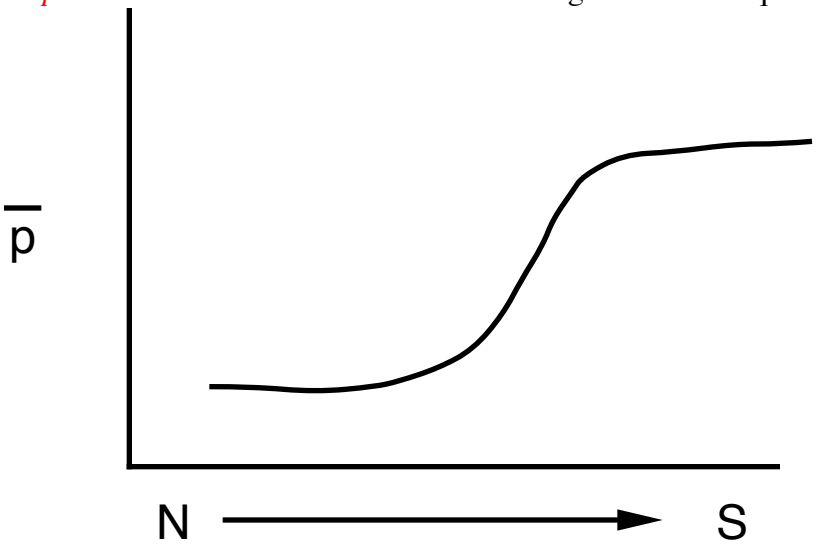
One patch of habitat will be different from any other patch

Populations may adapt to local conditions

Cline - a steady change in allele frequency across a geographic region



Step cline - a cline with a more sudden change in allele frequency



Trifolium

FIGURE 13

Frequency of the cyanide-producing form in populations of white clover (*Trifolium repens*), represented by the black section of each circle. The cyanogenic form is more common in warmer regions. Thin lines are January isotherms. (Modified from Jones 1973, after Daday 1954)

Cyanide production reduces herbivory, but when cells freeze and burst, this releases cyanide to within the clover itself.

Agrostis tenuis - bent grass

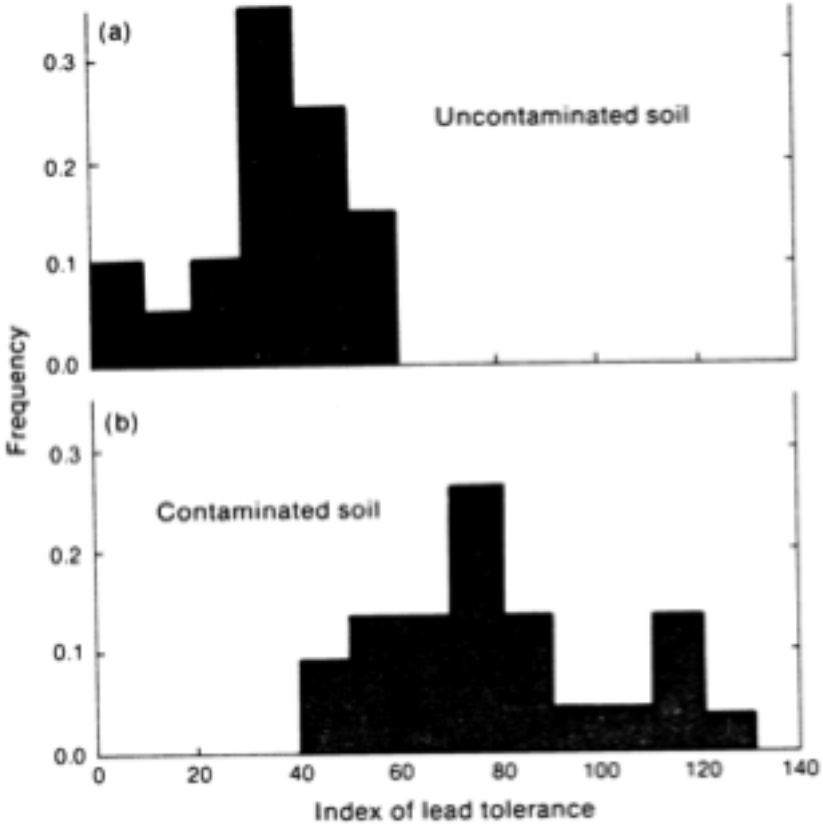


Figure 16.3. The frequency of individuals with different levels of lead tolerance in an *Agrostis tenuis* population sampled from contaminated and uncontaminated soil (after Bradshaw et al., 1965).

These plants were taken from seeds on plants only 7 m apart, in mine tailings and just outside

Strong selection
Agrostis selfs, so reduced recombination with other types

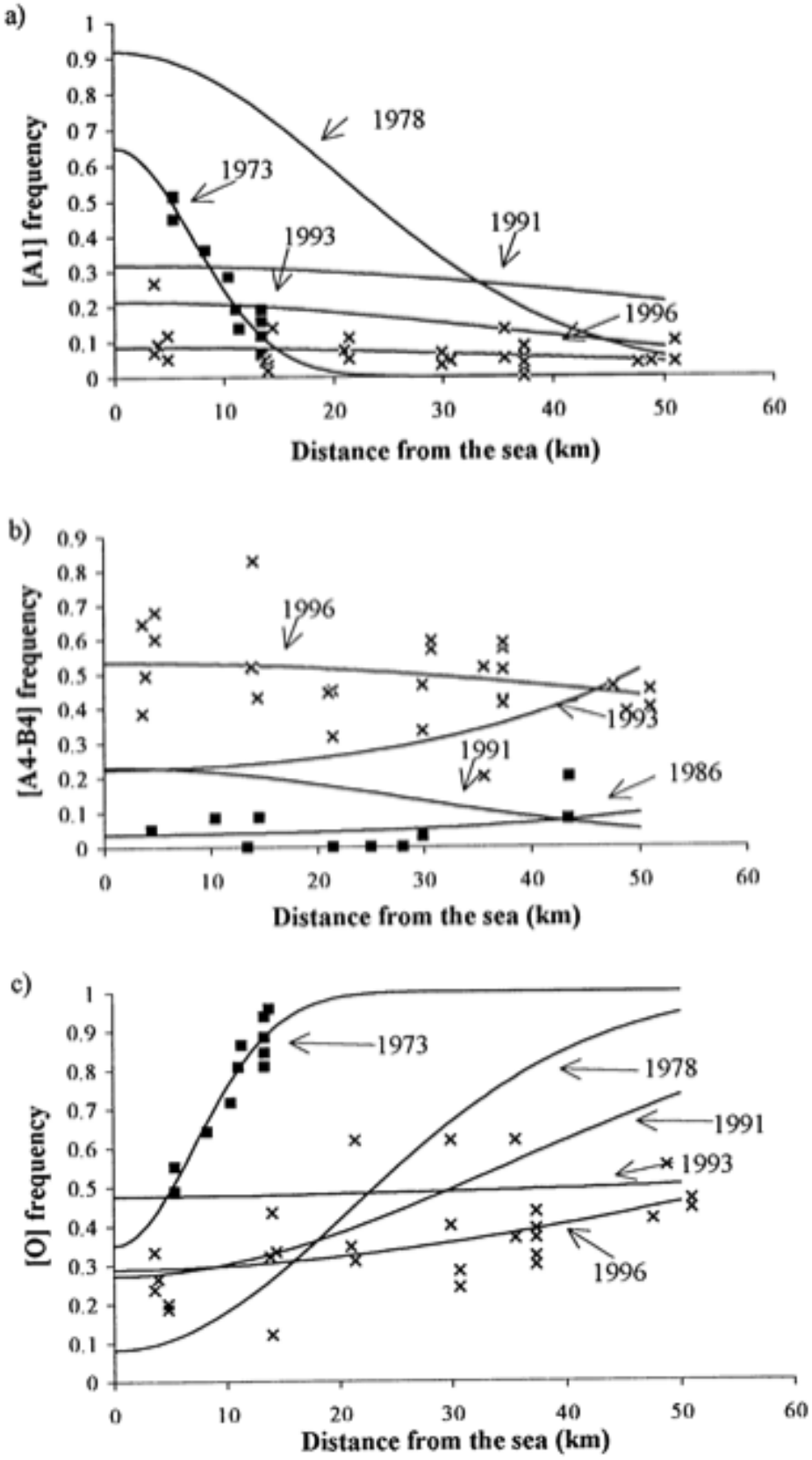
Culex pipiens

Organophosphate pesticides applied near south coast of France beginning in 1972

In 1973, a resistant allele of esterase appeared which could metabolize the pesticide

In the 1980's, the broad scale application of the pesticide stopped

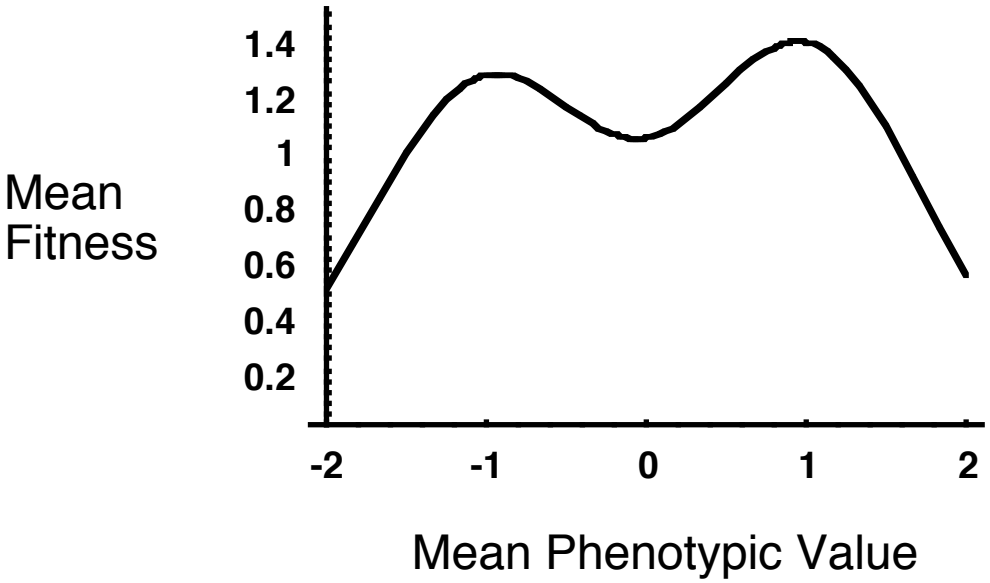
In 1987, an allele of acetocholineesterase appeared which was even more resistant.


a)
[A1] frequency
1978
1973
1993
1991
1996
Distance from the sea (km)
b)
[A4-B4] frequency
1996
1993
1991
1986
Distance from the sea (km)
c)
[O] frequency
1973
1978
1991
1993
1996
Distance from the sea (km)

Evolution on complex landscapes

Adaptive landscape - graph of the function which relates mean fitness of a population to its genotype or phenotype frequencies

Q: If mean fitness always increases as a result of selection, AND there is a valley between two peaks, how can a population evolve from one local fitness maximum to another?



Possible answers:

(1) Mean fitness function changes as a result of environmental changes

(2) Changes in other loci may change fitness interactions

(3) Wright's shifting balance process

Viability of two-locus genotypes of *Drosophila melanogaster* in two experimental environments[a]

(A) Ethanol Present

		Adh Genotype		
		SS	*SF*	*FF*
α-*Gpdh* Genotype	*SS*	0.596	1.288	0.932
	SF	0.964	1.000	0.836
	FF	0.909	0.968	0.864

(B) Ethanol Absent

		Adh Genotype		
		SS	*SF*	*FF*
α-*Gpdh* Genotype	*SS*	0.992	1.059	0.863
	SF	1.080	1.000	0.935
	FF	0.765	1.164	0.750

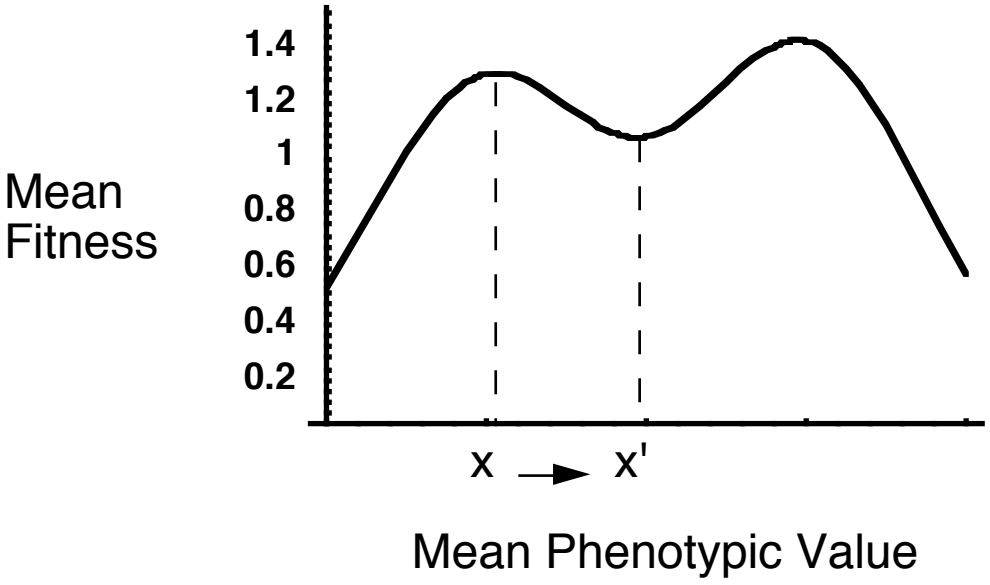
(Modified from Cavener and Clegg 1981)
[a]The viability of the double heterozygote, taken as a standard, has been set at 1.000 in each environment.

Wright's Shifting Balance theory

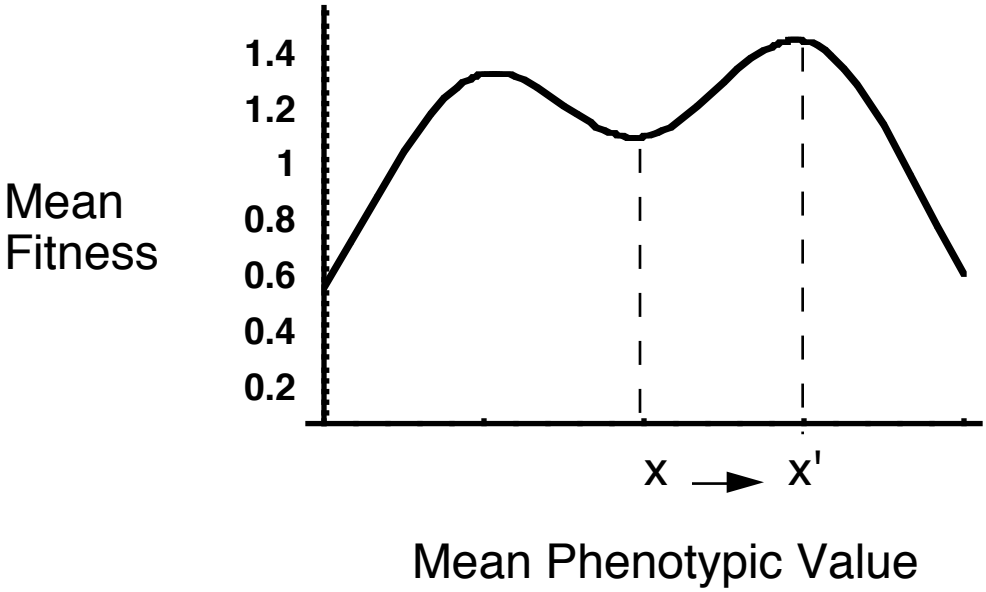
suggests that drift in local populations can allow a shift from one adaptive peak to another.

Three phases:

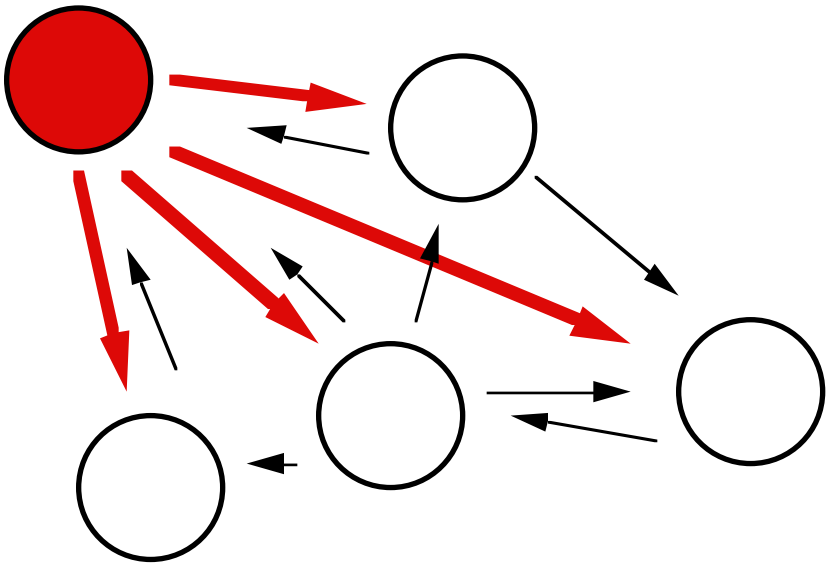
Phase I: Drift



Phase II: Selection within demes



Phase III: Interdemic selection



Problems with the SBT

- Sufficient drift probably occurs very rarely
- Response to interdemic selection is weak

Kin Selection

Any event which increases the frequency of an allele due to the properties of that allele is selection. If the effect of this selection is to increase the fitness of relatives, this is called *kin selection*.

e.g. worker ants, parental care

r = coefficient or relationship of 2 individuals

(= 2F/(1+F) where F is the inbreeding coefficient of offspring of those two individuals)

A trait will evolve by kin selection if

$c < r\,b$

c: cost to its own fitness

b: benefit to relative's fitness

Group selection

Kin selection is a special case of *group selection*

Groups of organisms may have differential survivorship (i.e. different extinction rates)

and/or

differential fertility (i.e. different emigration rates)

If groups are genetically differentiated, this allows for gene frequency change as a result of the properties of groups.

Broad scale group selection is probably a small factor in allele frequency change

Because F_{ST}'s are relatively small, change in allele frequency due to group selection is likely much smaller than the change due to individual selection

Introduction to Quantitative Genetics

What's a "quantitative character"?

A quantitative character is a trait which exhibits continuous variation.

Why study continuous variation, when we know the genetic basis of traits is discrete?

Even with a relatively small number of genes involved, the variation for a character will often appear continuous, due to measurement error and, more importantly, environmental effects.

Who cares?

- Most characters are continuously distributed.
- We do not fully understand the genetic basis of most traits, therefore we must describe them statistically.
- Even with better understanding of the genetic basis of a trait, a statistical description of that trait can be extremely useful.
- Quantitative genetic understanding has allowed substantial gains in agricultural output.
- QG concepts has allowed new understanding of evolutionary processes.

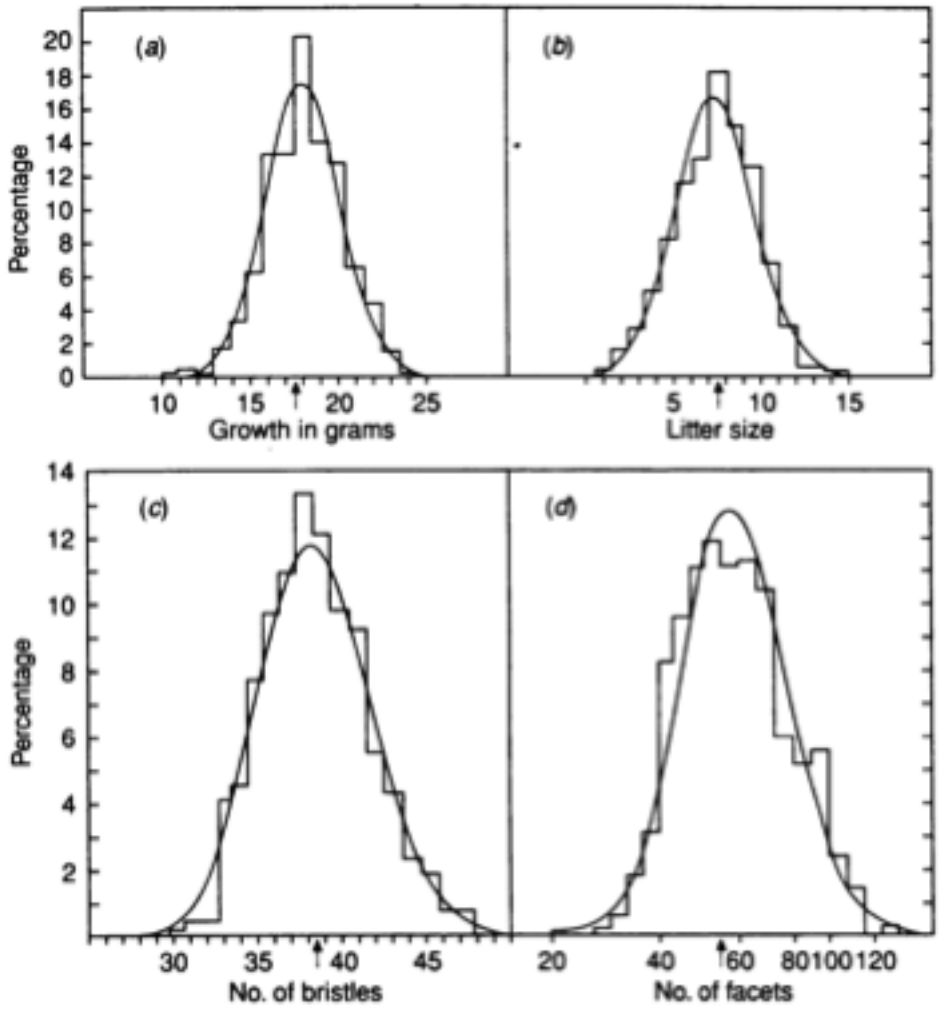


Fig. 6.2. Frequency distributions of four metric characters, with normal curves superimposed. The means are indicated by arrows. The characters are as follows, the number of observations on which each histogram is based being given in brackets:

(*a*) Mouse (♂♂): growth from 3 to 6 weeks of age. (380)

(*b*) Mouse: litter size (number of live young in 1st litters). (689)

(*c*) *Drosophila melanogaster* (♀♀): number of bristles on ventral surface of 4th and 5th abdominal segments, together. (900)

(*d*) *Drosophila melanogaster* (♀♀): number of facets in the eye of the mutant "Bar". (488)

(a), (b), and (c) are from original data: (d) is from data of Zeleny (1922).

Central Limit Theorem and the Normal approximation to the Poisson distribution

The CLT states that the sum of a large number of independent variables is distributed approximately by a normal distribution.

The Poisson distribution with large enough mean is closely approximated by the normal distribution.

Resemblance between relatives is in part a function of sharing alleles

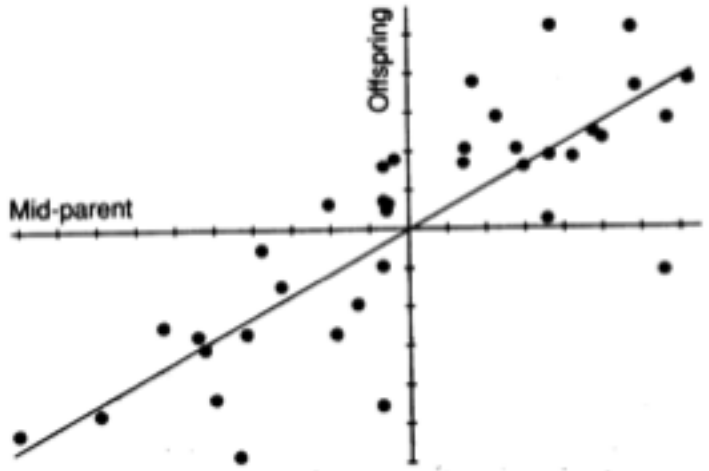


Fig. 10.1. Regression of offspring on mid-parent for wing length in *Drosophila*, as explained in Example 10.2. Mid-parent values are shown along the horizontal axis, and mean value of offspring along the vertical axis. (*Drawn from data kindly supplied by Dr E. C. R. Reeve.*)

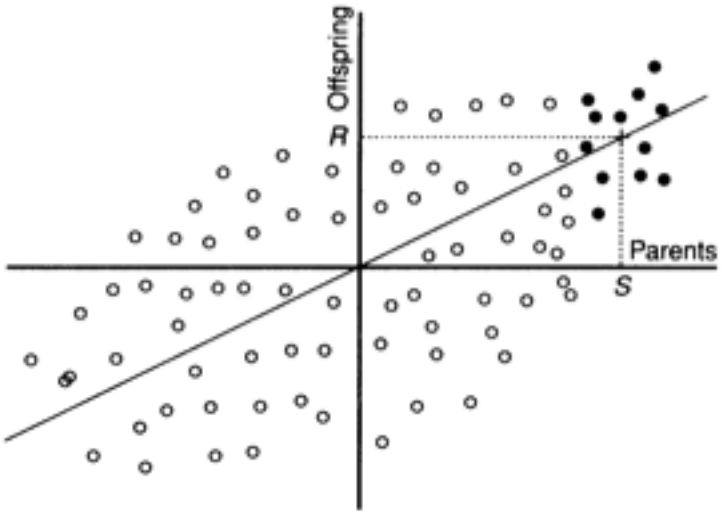


Fig. 11.1. Diagrammatic representation of the mean values of progeny plotted against the mid-parent values, to illustrate the response to selection, as explained in the text.

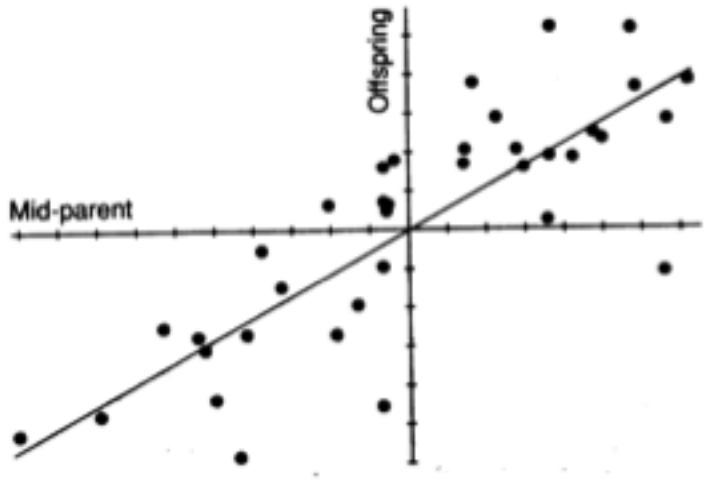


Fig. 10.1. Regression of offspring on mid-parent for wing length in *Drosophila*, as explained in Example 10.2. Mid-parent values are shown along the horizontal axis, and mean value of offspring along the vertical axis. (*Drawn from data kindly supplied by Dr E. C. R. Reeve.*)

Response to Selection and the correlation of relatives

The response to selection is predicted by the heritability and the selection differential:

$$\mathbf{R = h^2 S.}$$

Truncation Selection

Truncation selection is simply taking all of the individuals above a certain threshold value.

Table 10.1 Approximate values of the heritability of various characters in various animal species. The estimates are rounded to the nearest 5 per cent; their standard errors range from about 2 per cent to about 10 per cent.

	h^2(%)	Ref.
Man		
Stature	65	(1)
Serum immunoglobulin (IgG) level	45	(2)
Cattle		
Body weight (adult)	65	(3)
Butterfat, %	40	(4)
Milk-yield	35	(4)
Pigs		
Back-fat thickness	70	(5)
Efficiency of food conversion	50	(5)
Weight gain per day	40	(5)
Litter size	5	(6)
Poultry		
Body weight (at 32 wks)	55	(7)
Egg weight (at 32 wks)	50	(7)
Egg production (to 72 wks)	10	(7)
Mice		
Tail length (at 6 wks)	40	(8)
Body weight (at 6 wks)	35	(8)
Litter size (1st litters)	20	(9)
Drosophila melanogaster		
Abdominal bristle number	50	(10)
Body size	40	(11)
Ovary size	30	(12)
Egg production	20	(11)

Definitions: regression and correlations

The *correlation* of X and Y is defined as the covariance of X and Y divided by the standard deviations of X and Y. (This is the square root of the coefficient of determination.)

$$\rho_{XY} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

The *regression coefficient* of Y on X is the covariance of X and Y divided by the variance of X. This is the slope of the regression of Y on X.

$$b_{XY} = \frac{Cov[X,Y]}{Var[X]}$$

Mid-parent offspring regression

Define the mean of the two parents as the *mid-parent*.

$$b_{O\bar{P}} = \frac{Cov[O,\bar{P}]}{Var[\bar{P}]} = \frac{V_A}{V_P}$$

Heritability

The heritability is one of the most important quantitative genetic properties. It predicts the response to selection, and expresses the reliability of the phenotype in determining the breeding value.

Definitions

Heritability is represented by h^2. The narrow-sense heritability is defined as

$$h^2 = \frac{V_A}{V_P}$$

Variance is a property of the population; therefore the heritability is a function of the population, as well as of the character.

Variances must be non-negative. $V_A \leq V_P$. Therefore, the heritability must be between 0 and 1.

Predicting response to selection

Components of the phenotypic variance

Average Effect

The average effect is the mean deviation from the population mean of individuals which received that allele from on parent, when the other allele is chosen at random from the population.

Breeding Value

The value of an individual, as measured by the average value of its offspring, is called its *breeding value*.
This is twice the deviation of the offspring from the population mean (since the individual only contributes half of the alleles to its offspring).
This is also the sum of the average effects of the individual.
With random mating, the mean breeding value is zero.

Dominance

The genetic effects (G) can be further partitioned.
Ignoring interactions among loci, **G=A+D.**
In this equation, the **A** refers to the *additive effects*, which is the sum of the breeding values, and the **D** refers to dominance deviations.
These dominance deviations refer to the deviation of diploid genotypic values from the sum of the average effects of those genotypes, due to the interaction between alleles at the same locus.

Interaction

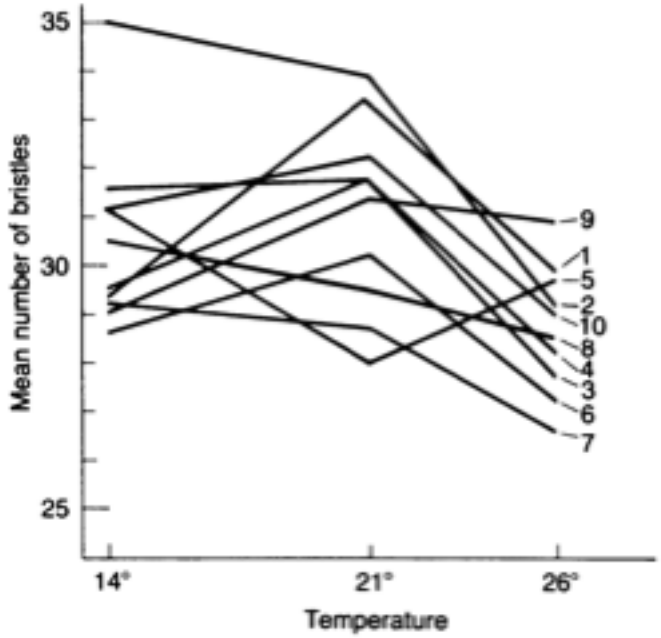
If different loci interact to form the phenotype, then there is *epistasis*, and this interaction affects the composition of phenotypes in the population.

Components of variance

$$V_P = V_G + V_E$$

$$= V_A + V_D + V_I + V_E$$

V_P= Phenotypic Variance
V_G= Genetic Variance
V_E= Environmental Variance
V_A= Additive genetic Variance
V_D= Dominance Variance
V_I= Epistatic Variance



Heritability: the proportion of variance which is determined by genetics.
Broad-sense heritability: V_G/V_P.
Narrow-sense heritability: V_A/V_P. *** This is the important one, generally.

Analysis of variance

The partitioning of variance follows an ANOVA kind of logic. In fact ANOVA's were invented for exactly this reason by R. A. Fisher.

Additive, dominance, epistatic variance

Having the additive variance be greater than zero does not imply that alleles interact additively. Even with complete dominance, or with complete epistasis, there will *usually* be additive variance.

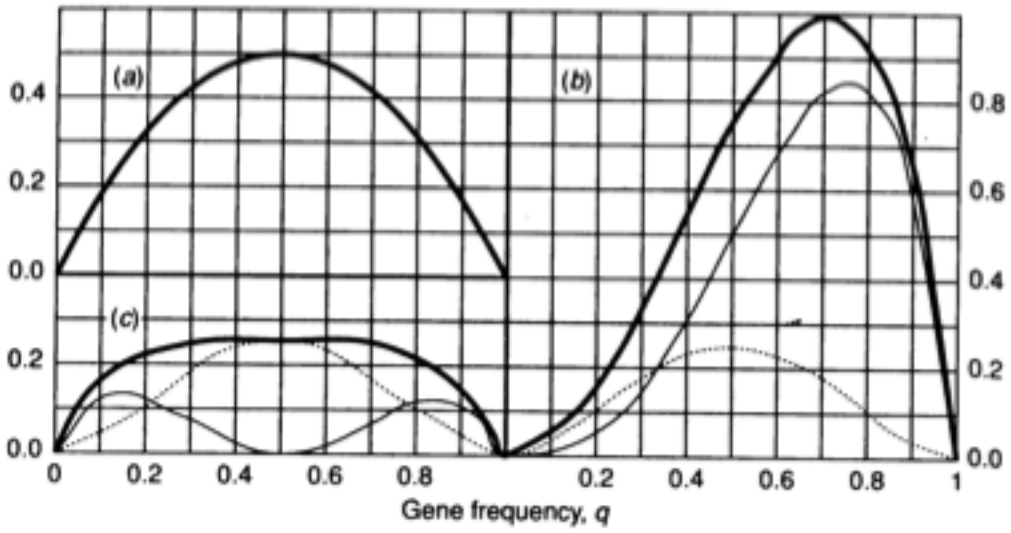


Fig. 8.1. Magnitude of the genetic components of variance arising from a single locus with two alleles, in relation to the gene frequency. Genotypic variance – thick lines; additive variance – thin lines; dominance variance – broken lines. The gene frequency, q, is that of the recessive allele. The degrees of dominance are: in (*a*) no dominance ($d = 0$); in (*b*) complete dominance ($d = a$); and in (*c*) 'pure' overdominance ($a = 0$). The figures on the vertical scale, showing the amount of variance, are to be multiplied by a^2 in graphs (*a*) and (*b*), and by d^2 in graph (*c*).

Correlations and interactions between G and E

G and E can be correlated (i.e. "good" offspring being provisioned more); more importantly, there can be interaction terms.

Estimating heritability by the resemblance among relatives

There are two ways of estimating heritability.
1. By examining the covariance among relatives.
2. By measuring the response to selection.

There are various problems with using correlations of relatives to estimating heritability:
1. There can be common environmental effects.
2. There can be maternal effects.
3. *All* of the covariances include epistatic variance terms; many also include dominance variance.
4. Measurement error.
5. Assortative mating can skew results.
6. Differences in the variance of the different sexes must be accounted for.

- **Consider the problems of each type of relatives**

Intuition: Relatives resemble one another. Why? Because they share genes and because they share environments.

Genetic Covariance among relatives

Parent-offspring

The regression of the mean value of offspring on the value in a parent:

$$b_{OP} = \frac{Cov[O,P]}{Var[P]} = \frac{1}{2}\frac{V_A}{V_P}$$

Notice that this is equal to half the heritability!
Define the mean of the two parents as the *mid-parent*. In this case, the covariance of offspring with the mid-parent is also 1/2 V_A. But the variance among mid-parents is equal to $V_P/2$, so the regression of offspring on mid-parent is equal to the heritability:

$$b_{O\bar{P}} = \frac{Cov[O,\bar{P}]}{Var[\bar{P}]} = \frac{V_A}{V_P}$$

Half-sibs

The covariance among half-sibs is

$$Cov\left[Half-sibs\right] = \frac{1}{4}V_A.$$

Full-sibs (Dominance rears its ugly head)

The covariance among full sibs turns out to be:

$$Cov\left[Full\ sibs\right] = \frac{1}{2}V_A + \frac{1}{4}V_D$$

This is because full sibs share not only alleles, but also can have the same *genotype* at a greater than random rate.
Think of it as this: full sibs have a 1/2 probability of sharing alleles, so the contribution of the additive variance (i.e. that due to single allelic effects) to the covariance is half the additive variance. Furthermore, they have a 1/4 chance of sharing the same genotype, therefore they "share" 1/4 of the dominance variance.

Identical Twins

Identical (or monozygotic) twins share all of their genes and genotype, so the genetic covariance among twins is equal to V_G.

Environmental covariance among relatives

Here's a problem: family members are likely to experience similar environments as well. Therefore there is likely to be some covariance among relatives which is due to environmental effects, not genetical effects. Hence we can naïvely overestimate heritability and V_A.

Common environmental effects

Common environmental effects are those environmental effects which different groups (such as families) experience in common. Represented by V_{Ec}.
These can be **huge**.
Examples: varying food quality, common density, cultural effects (i.e. wealth in humans)
These can completely skew genetic variance component estimations. The covariance among any class of relatives is equal to the genetic covariance plus the common environment variance. Thus if we measure VA in half sibs by multiplying the covariance times 4, we would also be multiplying the error introduced by V_{Ec} by a factor of 4 as well.
This is a huge problem for studies on humans, where these sorts of environmental covariances cannot be controlled for. Controlling for V_{Ec} in other experiments is a major source of expense, since it requires offspring to be raised separately.
Competition is a particularly bad source of common environmental effects; it can even cause negative covariances.

Maternal effects

Maternal effects are the covariances that result from the maternal environment. Mothers contribute not only genes to their offspring, but also cytoplasm, nutrients, and sometimes parental care. These factors can co-vary with the mother's *condition*, and therefore a source of environmental variation can result in pernicious covariance.
This can be controlled for partially by using half-sib designs, with the sire as the common parent, by cross-fostering, and by comparing maternal covariance to paternal covariance.

Assortative mating

Assortative mating, the mating of like with like, results in inflated estimates of heritability.
For example, the covariance of parents and offspring is inflated by a factor of $(1+r)$ where, confusingly, r is the correlation of phenotypic values of mates.
Intuition: If mates are correlated, then the offspring of those individuals will be not only correlated with parents because of the alleles they inherit from that parent, but also because the alleles they inherit from the other parent have a higher than average probability of being the same as well.

Response to Selection

The *response to selection* (i.e. the difference in the mean of one generation from the mean of the previous generation) is directly proportional to the heritability.
Selection on Quantitative Characters

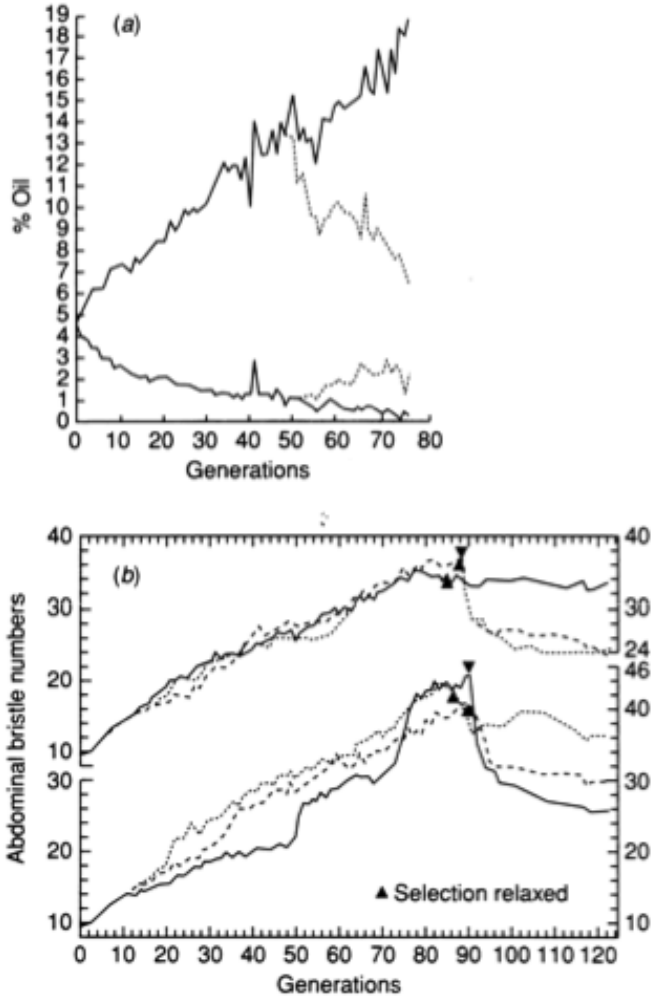
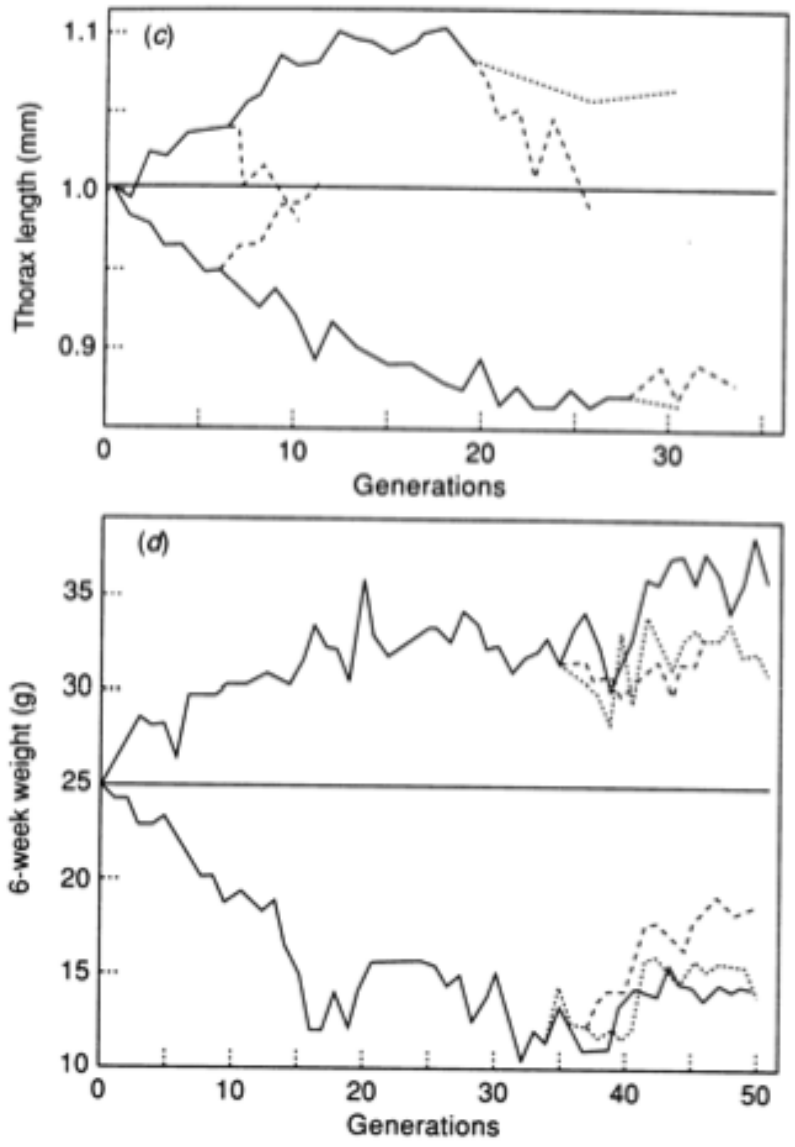


Fig. 12.3. Four experiments illustrating long-term responses.
(*a*) Two-way selection for oil-content of maize seeds. Broken lines are reversed selection. (*After Dudley, 1977.*)
(*b*) Six replicate lines of *Drosophila melanogaster* selected upwards for abdominal bristle number. Selection was suspended at the points marked. (*After Yoo, 1980a.*)



(*c*) *Drosophila melanogaster*, thorax length. (*After F. W. Robertson, 1955.*)
(*d*) Mouse, six-week body weight. (*Adapted from Roberts, 1966b.*)
Dashed lines are responses to selection in the reverse direction; dotted lines are responses t natural selection, with artificial selection suspended.
(*All figures redrawn from the above sources with permission of the authors and publishers.*)

Response to selection
The *response to selection* is the change in the population mean after selection. -> **R**
Selection Differential
The *selection differential* is the difference in the mean of the selected parents from the mean of the population as a whole. -> **S**

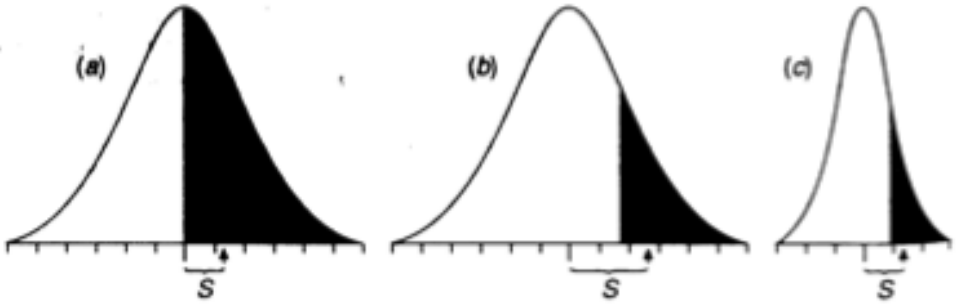


Fig. 11.2. Diagrams to show how the selection differential, S, depends on the proportion of the population selected, and on the variability of a normally distributed character. All the individuals in the stippled areas, beyond the points of truncation, are selected. The axes are marked in hypothetical units of measurement.
(*a*) 50 per cent selected; standard deviation 2 units: $S = 1.6$ units.
(*b*) 20 per cent selected; standard deviation 2 units: $S = 2.8$ units.
(*c*) 20 per cent selected; standard deviation 1 unit: $S = 1.4$ units.

Intensity of Selection
This is just a standardized version of the selection differential; i.e., the selection differential divided by the standard phenotypic deviation of the trait. -> **i**

Measuring the response to selection

Variability among generations

Selected lines vary substantially in their response to selection even from one generation to the next.
WHY?

Error

Measurement error can result in substantially variable results. Selection on behavioral traits, or other traits with low repeatability, result in noise among generations in means.

Drift

Selection experiments are normally done with relatively small population sizes, because of the resource costs of measuring individuals. Therefore, there is a lot of latitude for genetic drift.

Environmental variance

Environmental conditions vary through time: this is why it is essential to have *controls*.

Asymmetry of response

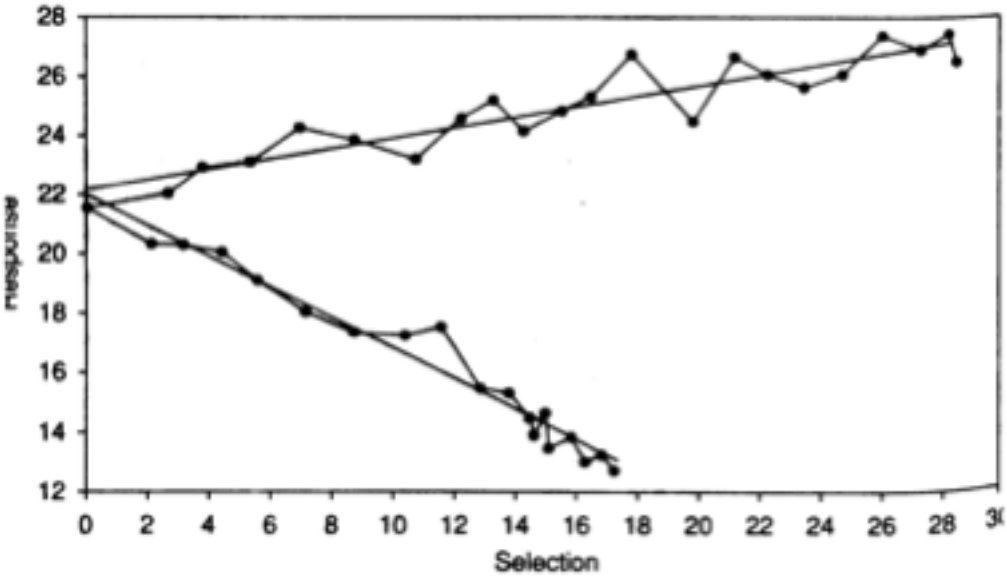


Fig. 11.5. Two-way selection for 6-week weight in mice. The generation means are plotted against the cumulated selection differentials, as explained in the text. The slopes of the regression lines fitted to the points measure the realized heritabilities, which were 0.175 for upward selection and 0.518 for downward selection. (*After Falconer, 1954.*)

Selection experiments are often done in both directions; i.e. up and down for the same trait. It is often the case that there is an asymmetry of the response to selection. In morphology, the lines for increased size often show more response than those for decreased size. There are several reasons for this asymmetry.

Drift

Drift can cause the cumulative response in one direction to be greater than the other. This often cannot explain a repeated bias in response in one direction across replicate lines, but must be rejected as a null hypothesis.

Natural selection

If there is stronger *natural* selection for the trait in one direction than the other, then natural selection will aid artificial selection n direction and hinder it in another.

Scale effects: the variance may change as a function of the mean

The genetic variance may scale as a function of the mean of the trait, and therefore the variance may be reduced in one direction, and therefore the response to selection.

Inbreeding depression

As lines become more selected in selection experiments, then the lines also typically become more inbred, for a variety of reasons. If there is *directional dominance* then this will result in a bias in response to selection.

"Genetical asymmetry": Gene frequencies divergent from p=0.5

If allele frequencies for genes which affect the trait are on average different from $p = q = 0.5$, then there will be an asymmetry of the amount to which selection can respond and how fast.

Special case: Genes of large effects

If much of the variance for a trait is due to a few important loci, then if the frequencies of even just those loci deviate from $p = 0.5$, then the asymmetry of response will also be seen

Non-linear interactions of genes and environments

Genotype-by-environment interactions can result in different effects of genes at different allele frequencies.

Long-term selection

Selection limits

Theoretical limits: the sum of the best alleles in the starting population

This depends on the number of loci contributing to the genetic variance.

Changes in phenotypic variance

Phenotypic variance often *increases* as a result of selection. This may be due to inbreeding effects on the environmental variance.

Mutation generates new variance

At equilibrium, the long-term response to selection allowed by mutation is $R = S\ 2N_e\ V_M/V_P$. This is because the variance coming into the population per generation is $2N_e\ V_M$, where V_M is the mutational variance per haplotype per generation. A typical value for V_M is approximately $10^{-3}\ V_E$.

What limits selection response?

Physical or logical constraints

The value in question cannot be greater or less than a certain physical limit, such as a proportion cannot be greater than 1 or less than 0.

Natural, or correlated, selection

Natural selection can limit the response to artificial selection.

Changes in the partitioning of variance (i.e. variance due to recessive alleles)

Dominance

When the unfavored recessives reach low frequencies, much of the variance is non-additive, and therefore not available for selection response.

Overdominance

In the (unlikely) event that the favored phenotype is produced by heterozygotes, then this also results in a lack of additive variance.

Exhaustion of genetic variance -- the effect of population size

Obviously, if selection "uses up" all of the genetic variance, then the response will slow. Larger populations allow more genetic variance.

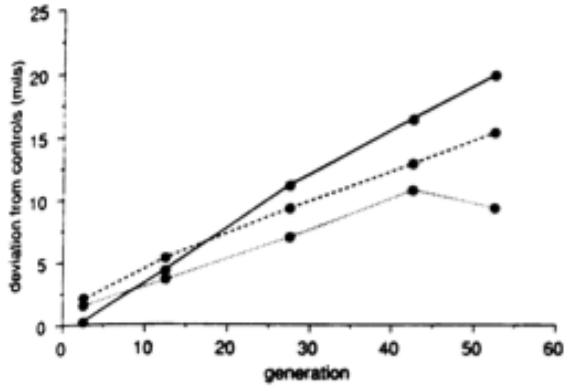


FIGURE 2.—Mean deviation from controls within treatments, averaged over successive periods of 5, 15, 15, 15 and 5 generations. Points are at midpoint of each period. Large lines (——); medium lines (- - - -); small lines (· · · · ·).

RESULTS

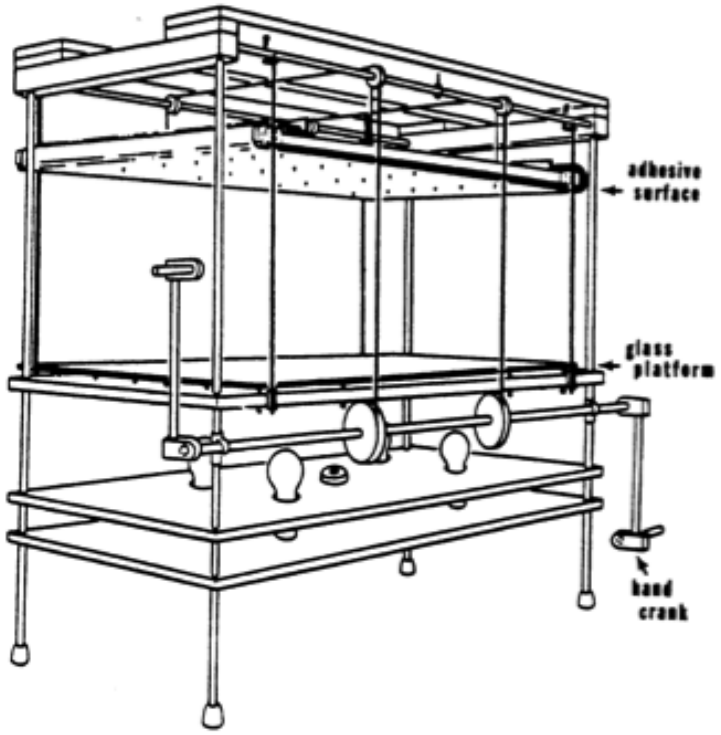


FIGURE 1.—Apparatus for mass selection on wing-tip height. See text for details.

clipped to the corners (not shown in figure). Detachable spacers of identical height stick to the adhesive surface in a 30-point grid. The feeler gauges and the spacers control the

Inbreeding and Quantitative Genetics

The importance of inbreeding in QG is two-fold: it affects the mean of traits (which results in inbreeding depression) and it affects the variance.

Changes in means

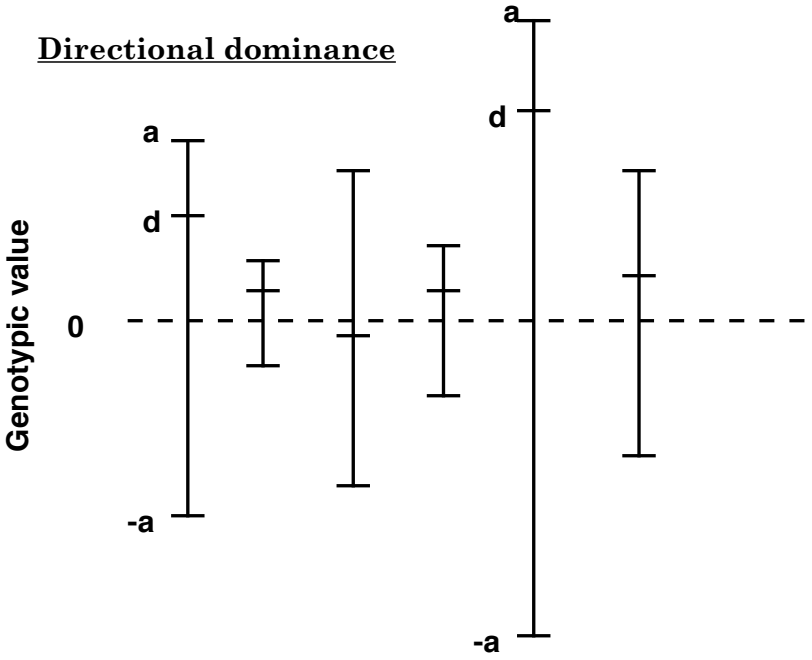
Inbreeding depression

Definitions

Inbreeding depression is the average reduction in fitness, or of a character, due to inbreeding.

Directional dominance

Directional dominance implies that the dominance of genes for a trait is on average biased in one direction.



If there is directional dominance, then a quantitative trait will on average change in mean as a result of inbreeding.

Changes in variance

Partitioning of variance

Inbreeding on additive characters results in greater genetic variance among populations and lower genetic variance within populations.
Additive genetic variance within populations, on average, decreases by F.

$$V_{A,w} = (1-F)V_A.$$

The additive genetic variance among populations increases in proportion to 2F:

$$V_{A,b} = 2FV_A.$$

The total additive genetic variance is therefore

$$\mathbf{V_{A,w} + V_{A,b} = (1+F)V_A}.$$

The heritability within lines is expected to decrease as a result of this reduction in V_A.

Changes in V_E

Sensitivity to environmental perturbations increases with inbreeding; in other words, there is inbreeding depression for *homeostasis*.

Correlated Characters

Characters are often correlated, that is, the phenotypic value of one character in an individual is correlated with the phenotypic value of another character on that individual. (The circumference of your head is about 1/3 of your height.)

These correlations can also be due to environmental effects or genetic effects. The genetic causes of correlation are *pleiotropy* (that genes affect more than one character) and *linkage disequilibrium*.

This need not be constant across genes: some genes can cause positive pleiotropy and others negative pleiotropy; the balance determines the genetic correlation of the two characters.

The sign of the genetic correlation need not be the same as the phenotypic correlation.

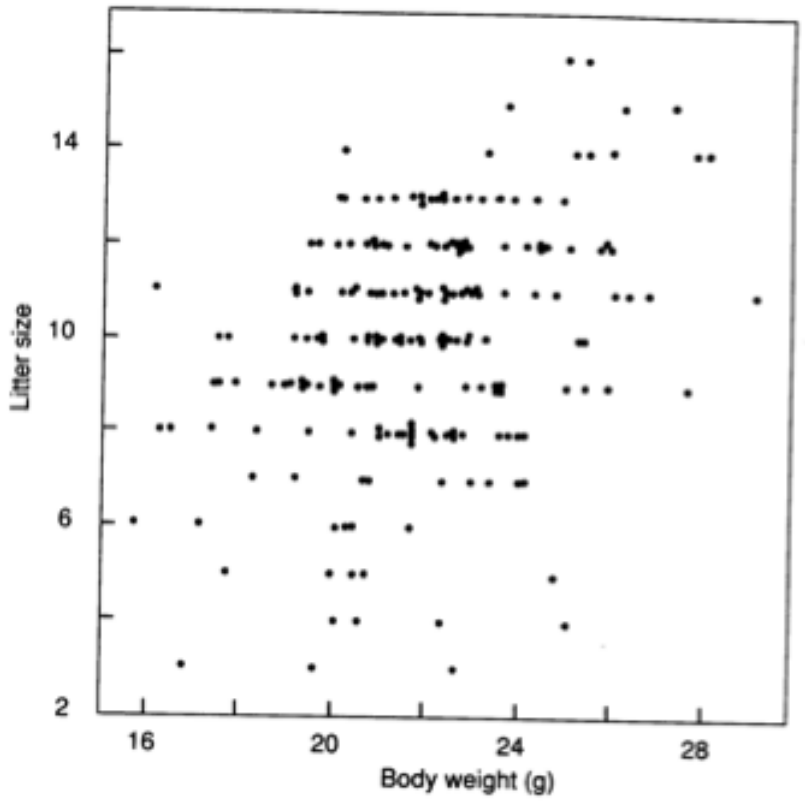


Fig. 6.3. Correlation between body weight and litter size in mice. Each point represents one individual female plotted according to its body weight at 6 weeks of age and the number of live young in its litter born some weeks later. The correlation among these 200 individuals is 0.3. (*Data kindly provided by Dr Ian Hastings.*)

Correlated response to selection

Selection on one trait will often result in response for another trait. This is due to genetic correlations.
Selection on one trait can cause an apparent selection differential at another trait, because of both the genetic and environmental correlations. This is a particularly huge problem when studying natural selection in natural conditions.

Indirect selection

Selecting for one character to get a response in another.

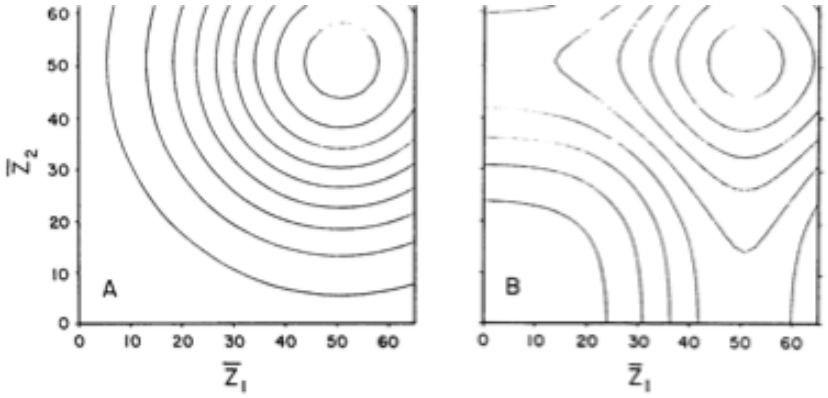


FIG. 1. Adaptive topographies for a character selected in two environments under either soft selection (A) or hard selection (B). The phenotypes of the trait in the two environments are considered to be two separate, but genetically correlated, character states with mean phenotypes $\bar{z}_1$ and $\bar{z}_2$. Contours represent levels of joint mean fitness at different combinations of mean phenotypes in the two environments. Contours are 0.1 units apart. Under soft selection, joint mean fitness is $\bar{W}_1^{q}\bar{W}_2^{(1-q)}$, while under hard selection, the joint mean fitness is $q\bar{W}_1 + (1-q)\bar{W}_2$, where $\bar{W}_i$ is the mean fitness in the ith environment, given by Equation (7). Parameters for both plots are $G_{11} = G_{22} = 10$, $P_{11} = P_{22} = 20$, $\omega_1^2 = \omega_2^2 = 200$, $q = 0.5$, and $\theta_1 = \theta_2 = 50$.

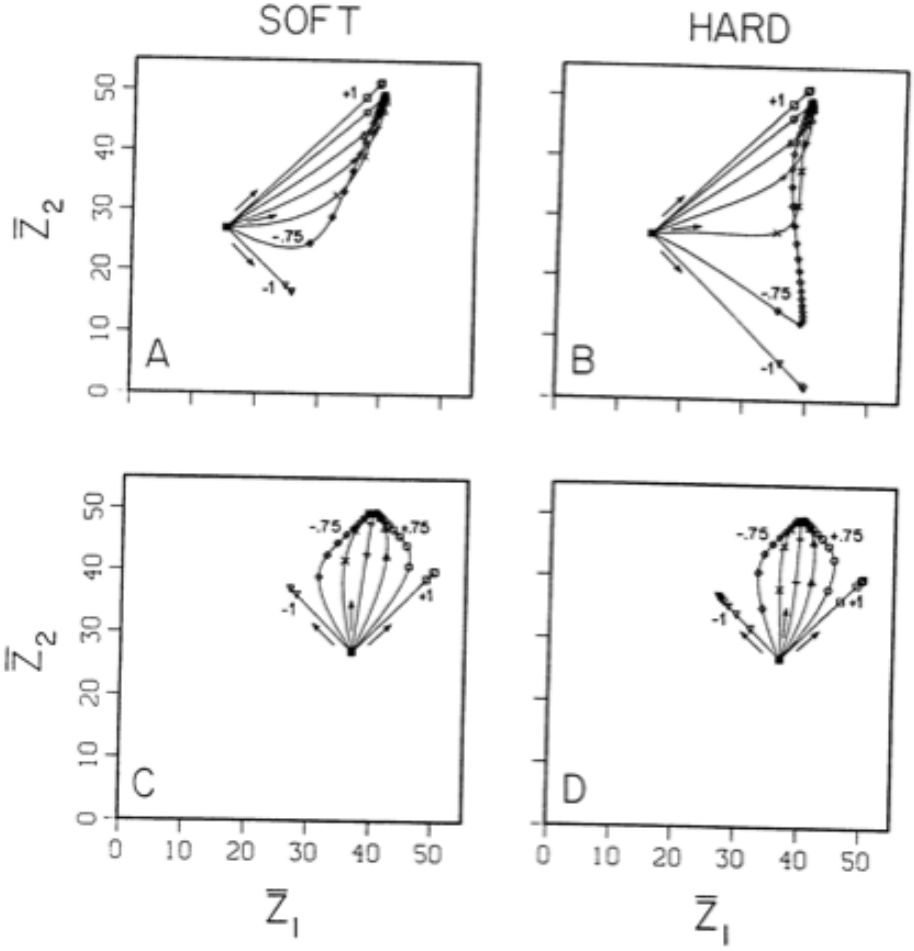


FIG. 3. Effects of genetic correlation across environments on the trajectory of the mean phenotype when environments are represented in unequal frequencies or when the population is initially well adapted to only one environment. Except as noted, parameters are the same as Figure 1; symbols are as in Figure 2. Optima are $\theta_1 = 40$ and $\theta_2 = 50$. Initial $\bar{z}_1 = \bar{z}_2 = 50$.

Fisher's Fundamental Theorem of Natural Selection

So the response to selection for fitness is given by

$$R_W = V_A[W]$$

Correlated Response to Natural Selection

By a similar derivation, the correlated response to selection on fitness for a character is given by

$$CR_Y = Cov_A[Y, W]$$

Evolutionary QG in a nutshell

Trait vectors and the G matrix

z is the vector of phenotypic values (i.e. {z1, z2, z3, ...} where each of these z's is the value of some character.

G is the genetic variance-covariance matrix:

$$\begin{bmatrix} V_A[X] & Cov_A[X,Y] \\ Cov_A[X,Y] & V_A[Y] \end{bmatrix}$$

The **P** matrix, similarly, is the phenotypic variance-covariance matrix.

The Response to Selection

The strength of selection on can be described for multivariate selection as well:

$$S = \bar{z}^{*} - \bar{z}$$

where the * indicates the value after selection.
Then there is a formula which describes the change in the mean from one generation to the next, analogous to $R=h^2 S$:

$$\Delta\bar{z} = G P^{-1} s$$

The value $\beta = \mathbf{P}^{-1}\,\mathbf{s}$ is called the *selection gradient*. It is equivalent to the vector of partial regression coefficients of relative fitness on character states.

Measuring the Strength of Selection

Two basic problems to all methods of measuring selection:
(1) There are always unmeasured characters, which may explain the true relationship of fitness and phenotype and have correlated effects on the characters studied.
(2) Fitness is ***hugely*** difficult to measure.

Stabilizing Selection

Definition

Selection for some intermediate value of the trait.

Effects

Stabilizing selection acts to reduce the phenotypic variance of the trait. It does this by increasing the *canalization* of the trait, as well as by reducing the genetic variance for the trait.
Canalization is the reduced sensitivity of an organism to noise from the environment.
Genetic variance can be reduced in two ways:

- By creating negative correlations of allelic effects: linkage disequilibrium.
- By changing gene frequencies to be closer to fixation.

Disruptive selection

Selection against intermediates

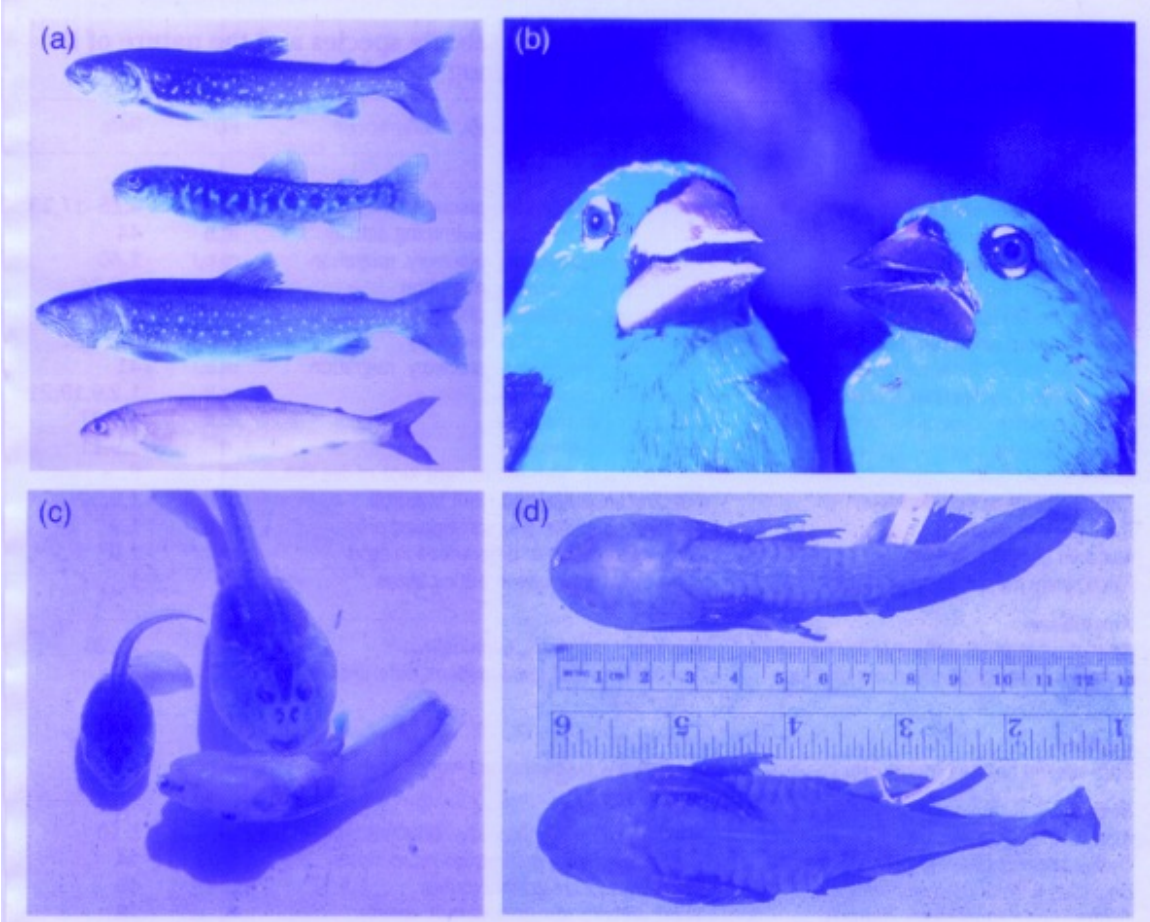


Fig. 1. (a) Four morphs of arctic charr (*Salvelinus alpinus*) from Thingvallavatn in Iceland (all individuals are adults). From top down: large benthivorous charr (33 cm long), small benthivorous charr (8 cm long), piscivorous charr (35 cm long) and planktivorous charr (19 cm long). *Photograph by S. Skúlason.* (b) Small- and large-billed male morphs of the African finch (*Pyrenestes ostrinus*). *Photograph by T. Smith.* (c) Larval morphs of spadefoot toad (*Scaphiosus multiplicatus*). *Photograph by D. Pfennig.* (d) Cannibal (top) and typical (bottom) larval morphs of the tiger salamander (*Ambystoma tigrinum*). *Photograph by J. Collins.*

Mutation

New variance comes from mutations at a rate approximately 10^{-2} to 10^{-4} V_E per generation.

V_M should be equal to $2nua^2$, where n is the number of genes which can mutate to have effect on the trait, u is the per locus mutation rate, and a is the effect of the mutation.

We can estimate nu from mutation accumulation experiments (like Mukai 1972). For viability, nu is in the range 0.1 to 1.
This means that either the mutation rate is extremely high, or there are many loci which mutate to affect viability. (Probably the latter is true.)

Maintenance of Genetic Variance

The effects of drift on neutral variation

The amount of genetic variation present when only mutation and drift are acting is $V_G=2N_eV_M$. Therefore the total variance is $2N_eV_M + V_e$. If we calculate the (broad-sense) heritability, and use 10^{-3} V_E for V_M, we find

$$h^2 = \frac{0.002N_e}{1 + 0.002N_e}$$

So we can see using this formula that the population size need not be large at all to allow high heritability to be maintained even with drift: