# INTRODUCTION TO GENETIC EPIDEMIOLOGY (1012GENEP1)

Prof. Dr. Dr. K. Van Steen

# BASIC POPULATION GENETICS

## 1 What is means and doesn't mean

### 1.a Introduction

Goals and aims

### 1.b The three domains of population genetics

Theoretical, empirical, experimental

## 2 How does evolution take place?

### 2.a Darwin

### 2.b Evolution

## 3 Distributions of genotypes in human populations

### 3.a Random mating

### 3.b Hardy-Weinberg equilibrium

### 3.c Allelic association

## 4 Natural selection revisited

## 5 Inbreeding

## 6 Fitness

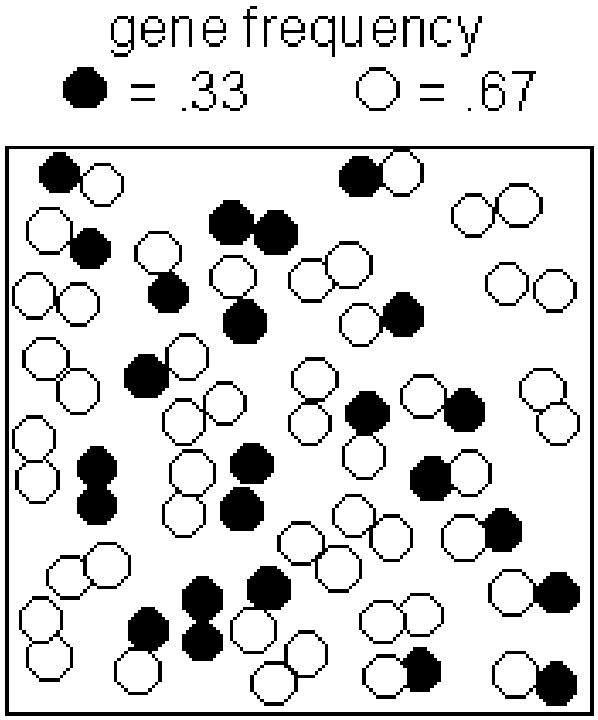
# 1 What it means and doesn't mean

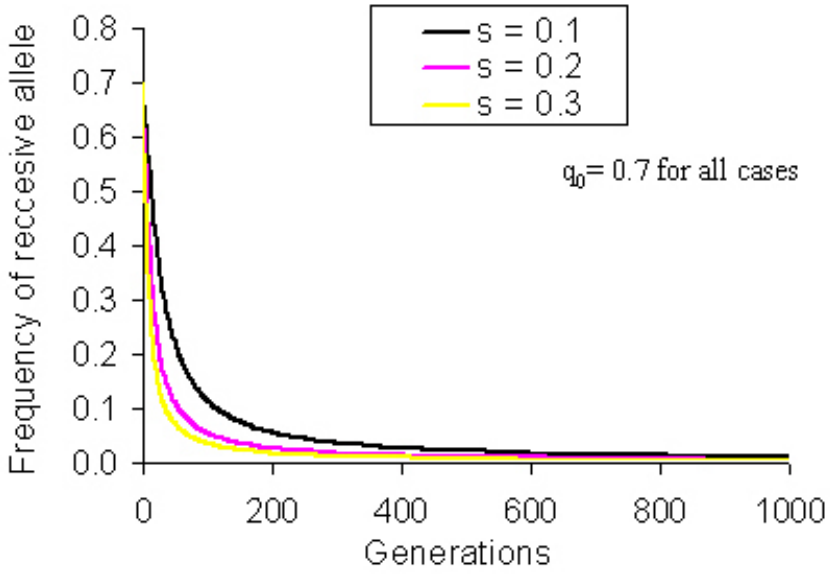
## Main references:

- Ritland K. Population genetics (course slides)

# 1.a Introduction

## Frequency at the heart of population genetics



## The essence of population genetics

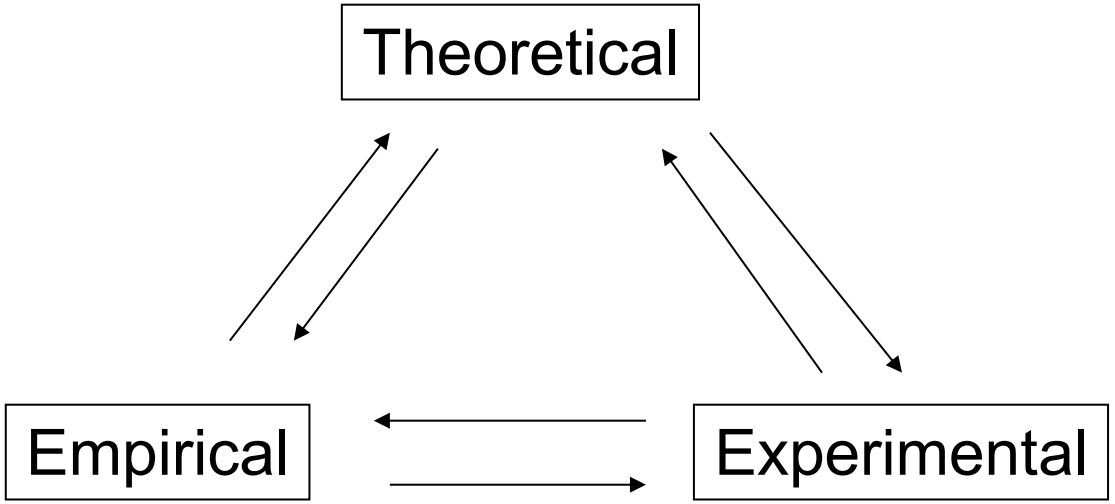
- A gene by itself is a constant entity (but perhaps harbors mutation)
- Alternative forms of a gene (allele) can exist at certain frequencies in a *population*
- These frequencies can change (via genetic drift, selection mutation)
- Resulting in *adaptation* and *evolution*

- There are two major facets:
  - Describing the pattern of genetic diversity
  - Investigating the processes that generate this diversity

# What is it useful for?

- Breeding
  - Plant and animal breeding
  - Pesticide and herbicide resistance
  - Effects of release of GM organisms
- People
  - Forensic analysis (DNA fingerprinting)
  - Identification of genes for complex human traits
  - Genetic counseling
  - Study of human evolution and human origin
- Ecology and evolution
  - Inferring evolutionary processes
  - Preservation of endangered species

# 1.b The three domains of population genetics



## Theoretical population genetics

- General theoretical models **predict** evolution of gene frequencies and other things
- Highly dependent upon **assumptions**
- May or may not be realistic
- Mathematically satisfying

VOL. 23, 1937 GENETICS: S. WRIGHT 307

## THE DISTRIBUTION OF GENE FREQUENCIES IN POPULATIONS

BY SEWALL WRIGHT

DEPARTMENT OF ZOÖLOGY, UNIVERSITY OF CHICAGO

Read before the Academy, April 26, 1937

The effects of the various evolutionary factors—mutation, cross breeding, selection and inbreeding—can be reduced to common terms by considering the rates of change which they tend to bring about in the relative frequencies of alleles.[1] In the absence of such factors, there is constancy of gene frequencies from the symmetry of the Mendelian mechanism.

The frequency ($q$) of a given gene changes at the rate $\Delta q = -uq$ per generation under recurrent mutation of the gene to alleles at the rate $u$. Mutation in the opposite direction at the average rate $v$ per generation changes the gene frequency at the rate, $\Delta q = v(1 - q)$.

If a certain gene has the frequency $q$ in a local population but $q_1$ in the species as a whole, exchange of the proportion $m$ of the local population with an equal number of random individuals from the whole species leads to change of gene frequencies in the former at the rate $\Delta q = -m(q - q_1)$. Cross breeding is, however, most likely to be with neighboring populations which differ but little in value of $q$. In this case the coefficient $m$ is only a small fraction of the actual amount of exchange. There may be other complications such as selective immigration or emigration, but the above simple form will suffice here to illustrate cross breeding or migration pressure.

The simplest kind of selection is that in which the heterozygote is exactly half way between the two homozygotes in the extent per individual to which it contributes to the next generation. The selective value of zygotes (relative to a certain standard) will be designated $w$ and the mean value for a population, $\bar{w}$.

| ZYGOTE | FREQUENCY | $w$ | |
|---|---|---|---|
| $AA$ | $(1 - q)^2$ | 1 | $\bar{w} = 1 - 2sq$ |
| $AA'$ | $2q(1 - q)$ | $1 - s$ | $\frac{d\bar{w}}{dq} = -2s$ |
| $A'A'$ | $q^2$ | $1 - 2s$ | |

$$\Delta q = \frac{(1 - s)q(1 - q) + (1 - 2s)q^2}{1 - 2sq} - q = \frac{-sq(1 - q)}{1 - 2sq} = \frac{q(1 - q)}{2\bar{w}} \frac{d\bar{w}}{dq}. \qquad (1)$$

In the more general case in which $\bar{w}$ is not related linearly to $q$, the momentary selective advantage ($-2s$) of replacing $A$ by $A'$ is still given

(Proceedings Nat. Acad. Sciences)

## EVOLUTION IN MENDELIAN POPULATIONS

SEWALL WRIGHT

*University of Chicago, Chicago, Illinois*

Received January 20, 1930

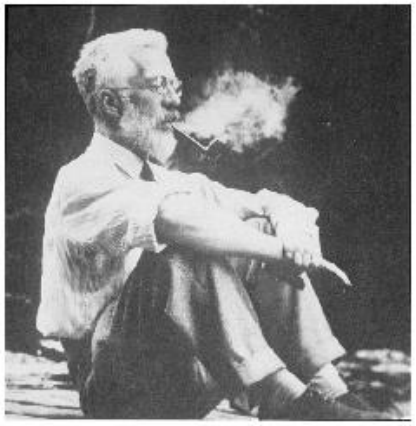
TABLE OF CONTENTS

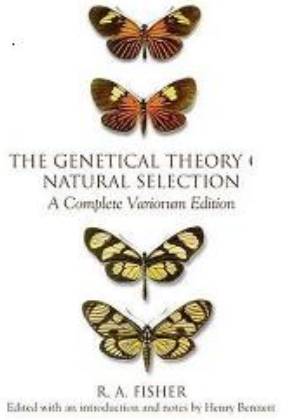
| | PAGE |
|---|---|
| Theories of evolution | 97 |
| Variation of gene frequency | 100 |
| Simple Mendelian equilibrium | 100 |
| Mutation pressure | 100 |
| Migration pressure | 100 |
| Selection pressure | 101 |
| Equilibrium under selection | 102 |
| Multiple allelomorphs | 104 |
| Random variation of gene frequency | 106 |
| Rate of decrease in heterozygosis | 107 |
| The population number | 110 |
| The distribution of gene frequencies and its immediate consequences | 111 |
| No mutation, migration or selection | 111 |
| Nonrecurrent mutation | 116 |
| Reversible recurrent mutation | 121 |
| Migration | 126 |
| Irreversible recurrent mutation | 128 |
| Selection | 129 |
| General formula | 133 |
| The distribution curves | 134 |
| Dominance ratio | 137 |
| The mean and variability of characters | 139 |
| The evolution of Mendelian systems | 142 |
| Classification of the factors of evolution | 142 |
| Lability as the condition for evolution | 147 |
| Control of evolution | 151 |
| Agreement with data of evolution | 153 |
| "Creative" and "emergent" evolution | 154 |
| SUMMARY | 155 |
| LITERATURE CITED | 158 |

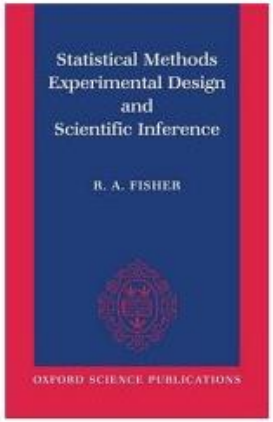
### THEORIES OF EVOLUTION

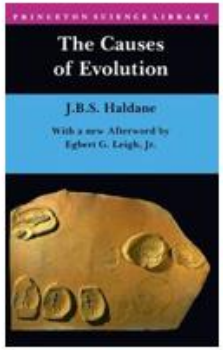
One of the major incentives in the pioneer studies of heredity and variation which led to modern genetics was the hope of obtaining a deeper insight into the evolutionary process. Following the rediscovery of the Mendelian mechanism, there came a feeling that the solution of problems of evolution and of the control of the process, in animal and plant breeding

GENETICS 16: 97 Mr 1931




THE GENETICAL THEORY
NATURAL SELECTION
A Complete Variorum Edition
R. A. FISHER
Edited with an introduction and notes by Henry Bennett


Statistical Methods
Experimental Design
and
Scientific Inference
R. A. FISHER
OXFORD SCIENCE PUBLICATIONS


The Causes
of Evolution
J.B.S. Haldane

# Empirical population genetics

- Apply statistical models to real data to infer underlying processes
- Again, adequate sampling is necessary to achieve statistical power
- Empirical population genetics is often emphasized due to the enormous volumes of genetic data that is out there.

## Experimental population genetics in practice

- Test **hypotheses** in population genetics using controlled experiments
- Design such that alternative outcomes possible, some of which can **reject** hypothesis
- Need **controls, replicates** and adequate **sample size**
- Usually restricted to **model organisms**
  - Drosophila, Neurospora, some crop plants

**Famous population geneticists in 1953**

Back row: Mather, da Cunha, Haldane, Dobzhansky, Waddington, Epling, Carson, Robertson, Falconer
Middle row (crouching): Ford, Wallace, ?, (large gap) Lerner, ?
Bottom row (sitting on ground): Mayr, Levine, Buzzati-Traverso, Fisher, Clausen, Pavan

The two unidentified ones may be (in order) Antonio Cordeiro and Renzo Scossiroli.
I have a complete list somewhere and will post it when I find it - Joe Felsenstein, 1998

# 2 How does evolution take place?

Main references:

- URLs:
  - http://www.biosci.ohio-state.edu/~pfuerst/course

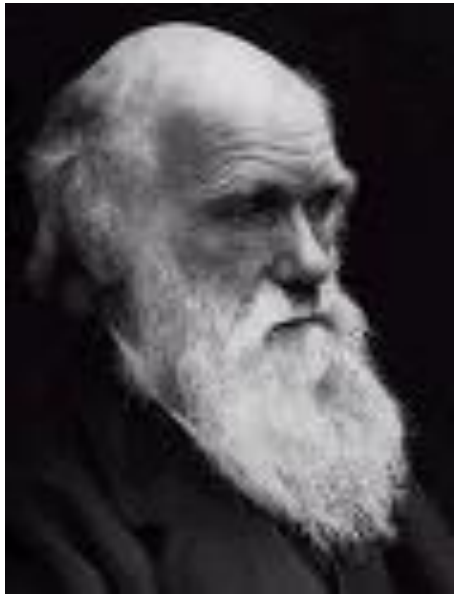
# 2.a Darwin's model in population genetics

## Introduction

- The modern evolutionary synthesis is a combination of Darwin's theory of the evolution of species by natural selection, Mendel's theory of genetics as the basis for biological inheritance, and mathematical population genetics.
- Put together by dozens of scientists throughout the 1930s and 1940s, Darwinian population genetics is our best model of the process that incrementally created all life on earth, evolution and natural selection.

## Charles Darwin (1803-1873)

- In the 19th century, a man called *Charles Darwin*, a biologist from England, set off on the ship HMS Beagle to investigate species in exotic places (e.g., Galapagos islands).
- After spending time on the islands, he soon developed a theory that would contradict the creation of man and imply that all species derived from common ancestors through a process called *natural selection*.
- Natural selection is considered to be the biggest factor resulting in the diversity of species and their genomes.



# Darwin's principles

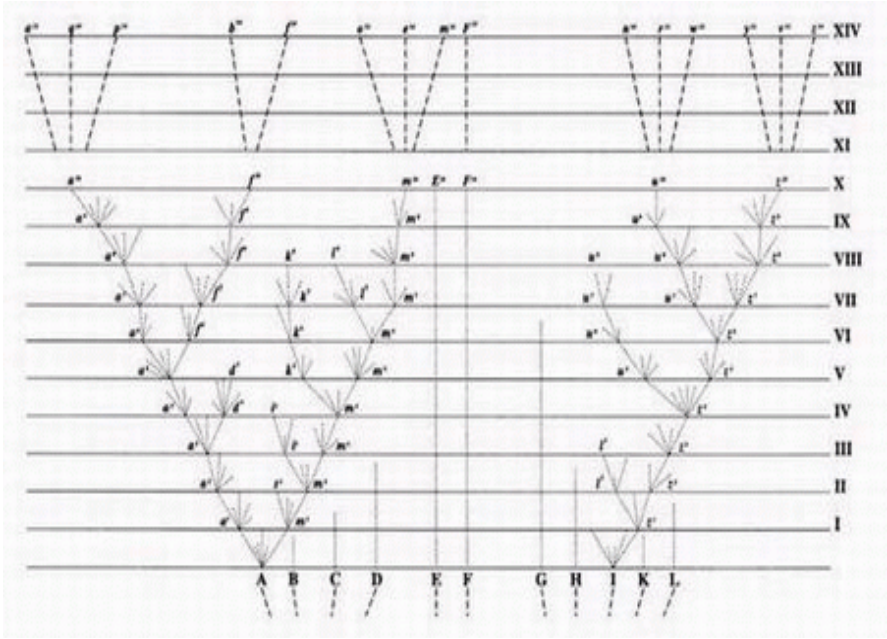
- One of the prime motives for all species is to reproduce and survive, passing on the genetic information of the species from *generation* to generation. When species do this they tend to produce more offspring than the *environment* can support.
- The lack of resources to nourish these individuals places pressure on the size of the species *population*, and the lack of *resources* means increased competition and as a consequence, some organisms will not survive.
- The organisms who die as a consequence of this competition were not totally random, Darwin found that those organisms more suited to their environment were more likely to survive.
- This resulted in the well known phrase *survival of the fittest*, where the organisms most suited to their environment had more chance of survival if the species falls upon hard times.
- Those organisms who are better suited to their environment exhibit desirable characteristics, which is a consequence of their genome being more suitable to begin with.

## Darwin's tree of life

- This 'weeding out' of less suited organisms and the reward of survival to those better suited led Darwin to deduce that organisms had evolved over time, where the most desirable characteristics of a species are favored and those organisms who exhibit them survive to pass their *genes* on.
- As a consequence of this, a changing environment would mean different characteristics would be favorable in a changing environment. Darwin believed that organisms had 'evolved' to suit their environments, and occupy an *ecological niche* where they would be best suited to their environment and therefore have the best chance of survival.

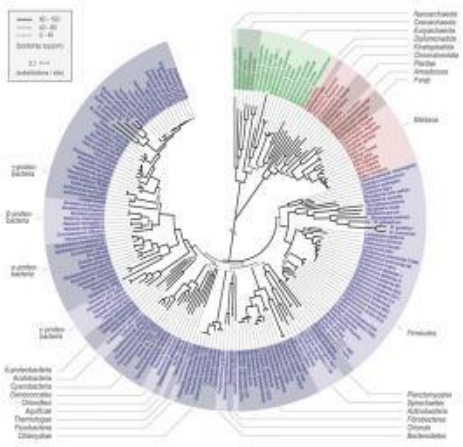
(http://www.biology-online.org/2/10_natural_selection.htm)

# Darwin's tree of life (continued)




The Tree of Life image that appeared in Darwin's *On the Origin of Species by Natural Selection*, 1859. It was the book's only illustration

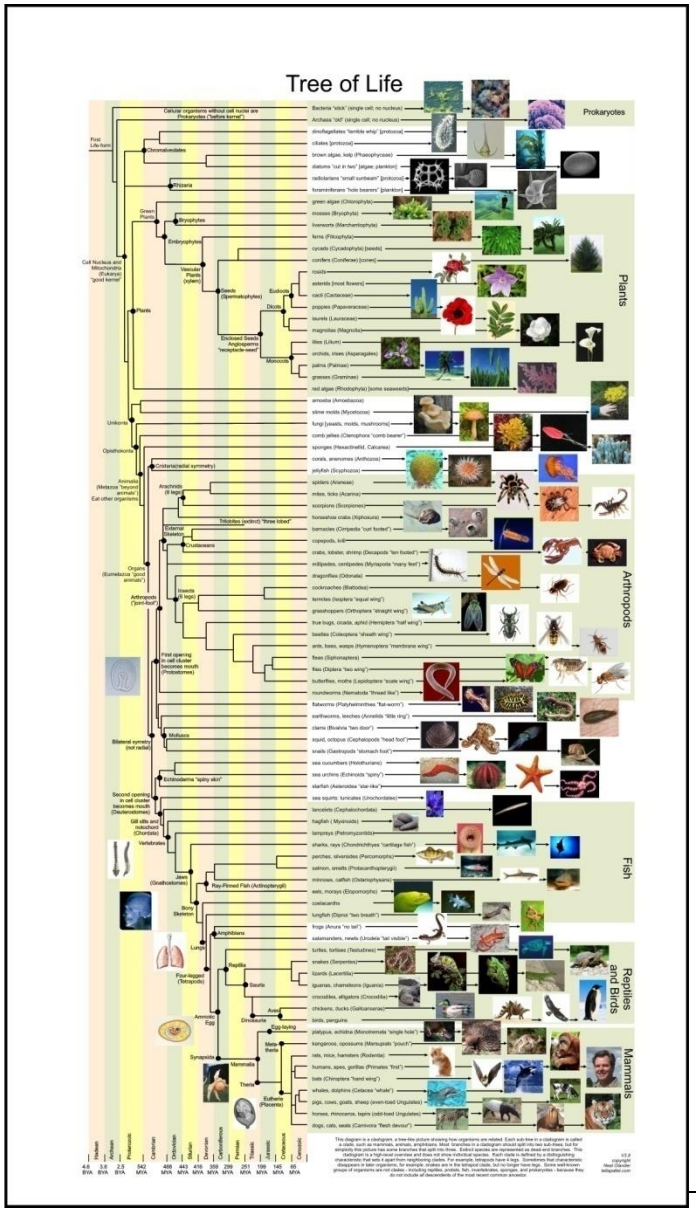
A group at the European Molecular Biology Laboratory (EMBL) in Heidelberg has developed a computational method that resolves many of the remaining open questions about evolution and has produced what is likely the most accurate tree of life ever:



# Modern tree of life

- A modern phylogenetic tree. Species are divided into *bacteria*, *archaea*, which are similar to bacteria but evolved differently, and *eucarya*, characterised by a complex cell structure
- A beautiful presentation can be downloaded from

http://tellapallet.com/tree_of_life.htm


Tree of Life
Prokaryotes
Plants
Arthropods
Fish
Reptiles and Birds
Mammals

# 2.b Evolution

## Evolution

- Evolution refers to the changes in gene (allele) frequencies in a population over time.
- Evolution takes place at the population, not species level. So populations, not species evolve

## Terminology

- A *population* refers to a group of interbreeding individuals of the same species sharing a common geographical area
- A *species* is a group of populations that have the potential to interbreed in nature and produce viable offspring
- *Gene pool* is the sum total of all the alleles within a population

## Four processes of evolution

- *Mutation*: changes in nucleotide sequences of DNA. Mutations provide new alleles, and therefore are the ultimate source of variation
- *Recombination*: reshuffling of the genetic material during meiosis
- *Natural selection*: differential reproduction (see later)
- *Reproductive isolation* (see later)

Mutation and recombination provide *natural variation*, the raw material for evolution.

# 3 Distributions of genotypes in human populations

## Main references:

- Ziegler A and König I. *A Statistical approach to genetic epidemiology*, 2006, Wiley.

  (Section 2.4)

- Clayton course notes on HWE (Bristol 2003)
- URLs:
  - Course notes on population genetics available from http://arnica.csustan.edu/boty1050

# 3.a Random mating

## The importance of random mating

- In chapter 2, we have considered the genetic inheritance processes underlying human reproduction, which basically allow us to describe the form of the conditional distribution of the genotype of a child $G_c$ given the parental genotypes $G_m$,$G_f$ :

$$P(G_c = g_c \mid G_m = g_m, G_f = g_f),$$

  since the genotype of the child is a stochastic quantity even if the parental genotypes are fixed.
- Suppose that we want to determine the children's genotype frequencies $P(G_c = g_c)$ using the above formula
- Note that in practice, the genotype frequencies among the children may be different from the genotype frequencies in the parental generation…

# The importance of random mating (continued)

- To compute the desired probabilities $P(G_c = g_c)$ we may treat the parental genotypes $G_m$,$G_f$ as unobserved, summing over all possible parental genotypes in the joint genotype distribution, i.e.

$$\begin{aligned} P(G_c = g_c) &= \sum_{g_m, g_f} P(G_c = g_c, G_m = g_m, G_f = g_f) \\ &= \sum_{g_m, g_f} P(G_c = g_c | G_m = g_m, G_f = g_f) P(G_m = g_m, G_f = g_f) \end{aligned}$$

  where the last identity is by definition of the conditional distribution
- The first factor is given by the inheritance laws

# The importance of random mating (continued)

- However, it is not possible to construct the joint distribution

$$P(G_m = g_m, G_f = g_f)$$

from the marginals $P(G_m = g_m)$, $P(G_f = g_f)$ without making additional assumptions, because we do not know the degree of dependence between the parental genotypes

- This problem is usually resolved by making the *random mating assumption*, which means that the parental genotypes at any locus are independent, i.e. for all $g_m$ and $g_f$:

$$P(G_m = g_m, G_f = g_f) = P(G_m = g_m)P(Gf = g_f)$$

- The random mating assumption will turn out to be the basis of many future modeling processes.

# 3.b Hardy-Weinberg equilibrium

## Background

- Early in the 20th century biologists believed that natural selection would eventually result in the dominant alleles driving out or eliminating the recessives. Therefore, over a period of time genetic variation would eventually be eliminated in a population
- Early in this century the geneticist Punnett was asked to explain the prevalence of blue eyes in humans despite the fact that it is recessive to brown. He couldn't do it so he asked a mathematician colleague named Hardy to explain it. A physician named Weinberg came up with a similar explanation, describing the genetics of non-evolving populations
- The Hardy-Weinberg law was born: the frequencies of alleles in a population will remain constant unless acted upon by outside agents or forces (see later).
  A non-evolving population is said to be in *Hardy-Weinberg equilibrium*

## Hardy-Weinberg conditions

- The Hardy-Weinberg principle sets up conditions which probably never occur in nature. One or more of mutation, migration, genetic drift, non-random mating or natural selection are probably always acting upon natural populations. This means that evolution is occurring in that population.
- The conditions for Hardy-Weinberg are:
    - Random mating (individuals mate independent of their genotype)
    - No selection (all genotypes leave, on average, the same number of offspring)
    - Large population size (genetic drift can be ignored)
    - Allele frequencies the same in both sexes
    - Autosomal loci

## Distorting factors to Hardy-Weinberg equilibrium causing evolution to occur

1. *Mutation* - by definition mutations change allele frequencies causing evolution

2. *Migration* - if new alleles are brought in by immigrants or old alleles are taken out by emigrants then the frequencies of alleles will change causing evolution

3. *Genetic drift* - random events due to small population size. Random events have little effect on large populations.

> E.g., consider a population of 1 million almond trees with a frequency of an allele r at 10%. If a severe ice storm wiped out half, leaving 500,000, it is very likely that the r allele would still be present in the population. However, suppose the initial population size of almond trees were 10 (with the same frequency of r at 10%). It is likely that the same ice storm could wipe the r allele out of the small population.

# Distorting factors to Hardy-Weinberg equilibrium causing evolution to occur

## 3. *Genetic drift (continued)*

a) Intense natural selection or a disaster can cause a *population bottleneck*, a severe reduction in population size which reduces the diversity of a population. The survivors have very little genetic variability and little chance to adapt if the environment changes.

By the 1890's the population of northern elephant seals was reduced to only 20 individuals by hunters. Even though the population has increased to over 30,000 there is no genetic variation in the 24 alleles sampled. A single allele has been fixed by genetic drift and the bottleneck effect. In contrast southern elephant seals have wide genetic variation since their numbers have never reduced by such hunting.

# Distorting factors to Hardy-Weinberg equilibrium causing evolution to occur

## 3. *Genetic drift (continued)*

b) Bottleneck effect, combined with inbreeding (see later), is an especially serious problem for many endangered species because great reductions in their numbers have reduced their genetic variability. This makes them especially vulnerable to changes in their environments and/or diseases. The Cheetah is a prime example.

c) Sometimes a population bottleneck or migration event can cause a *founder effect*. A founder effect occurs when a few individuals unrepresentative of the gene pool start a new population.

E.g., a recessive allele in homozygous condition causes Dwarfism. In Switzerland the condition occurs in 1 out of 1,000 individuals. Amongst the 12,000 Amish now living in Pennsylvania the condition occurs in 1 out of 14 individuals. All the Amish are descendants of 30 people who migrated from Switzerland in 1720. The 30 founder individuals carried a higher than normal percentage of genes for dwarfism.

## Distorting factors to Hardy-Weinberg equilibrium causing evolution to occur

4. *Nonrandom Mating* - for a population to be in Hardy-Weinberg equilibrium each individual in a population must have an equal chance of mating with any other individual in the population, i.e. mating must be random.

a) If mating is random then each allele has an equal chance of uniting with any other allele and the proportions in the population will remain the same. However in nature most mating is not random because most individuals choose their partner.

*Sexual selection* - nonrandom mating in which mates are selected on the basis of physical or behavioral characteristics.

## Distorting factors to Hardy-Weinberg equilibrium causing evolution to occur

5. *Natural Selection* - For a population to be in Hardy-Weinberg equilibrium there can be **no** natural selection. This means that all genotypes must be equal in reproductive success (see later for more details about natural selection).

But recall Darwin's reasoning:

- all species reproduce in excess of the numbers that can survive
- yet adult populations remain relatively constant
- therefore there must be a severe struggle for survival
- all species vary in many characteristics and some of the variants confer an advantage or disadvantage in the struggle for life
- the result is a natural selection favoring survival and reproduction of the more advantageous variants and elimination of the less advantageous variants

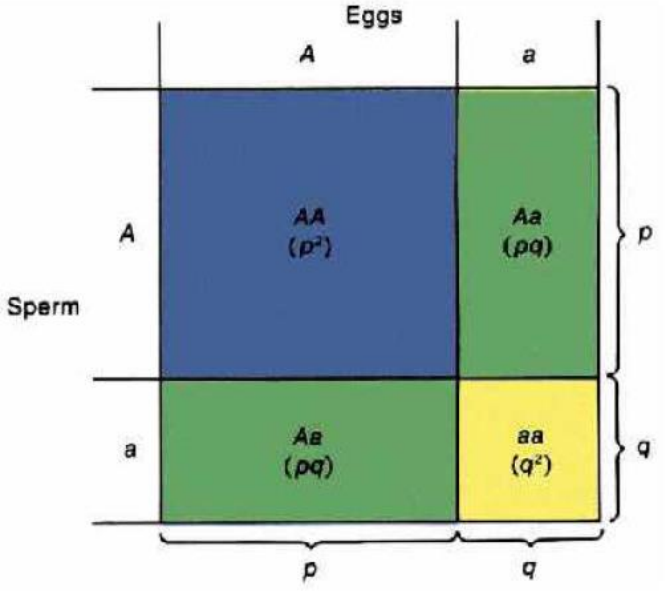
## Hardy-Weinberg equilibrium in formulae

- If alleles *i* and *j* have relative frequencies $p_i$ and $p_j$, then, under random mating, the genotype frequencies are

$$\Pr(i/j) = 2\pi_i\pi_j \qquad (i \neq j)$$
$$\Pr(i/i) = \pi_i^2$$

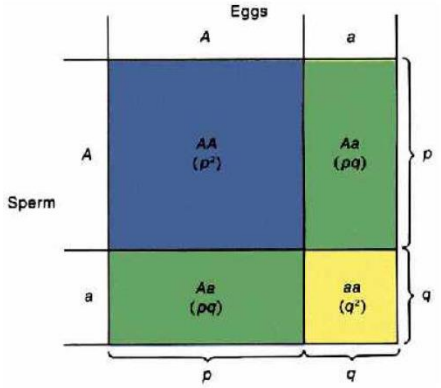
- The above is termed Hardy-Weinberg equilibrium
- Deviation from HWE may indicate population *stratification* and/or *admixture* or genotyping errors

(Note: this will become important when actually testing for genetic associations with a trait).



## When HWE is true, allele frequencies do not change from one generation to the next

- Consider allele frequencies *p* and *q* for alleles *A* and *a*, respectively, in generation zero, at a particular locus
- Then, assuming *A* and *a* are the only possible alleles, *p+q=1*
- Construct Punnett's square to obtain the genotype frequencies in



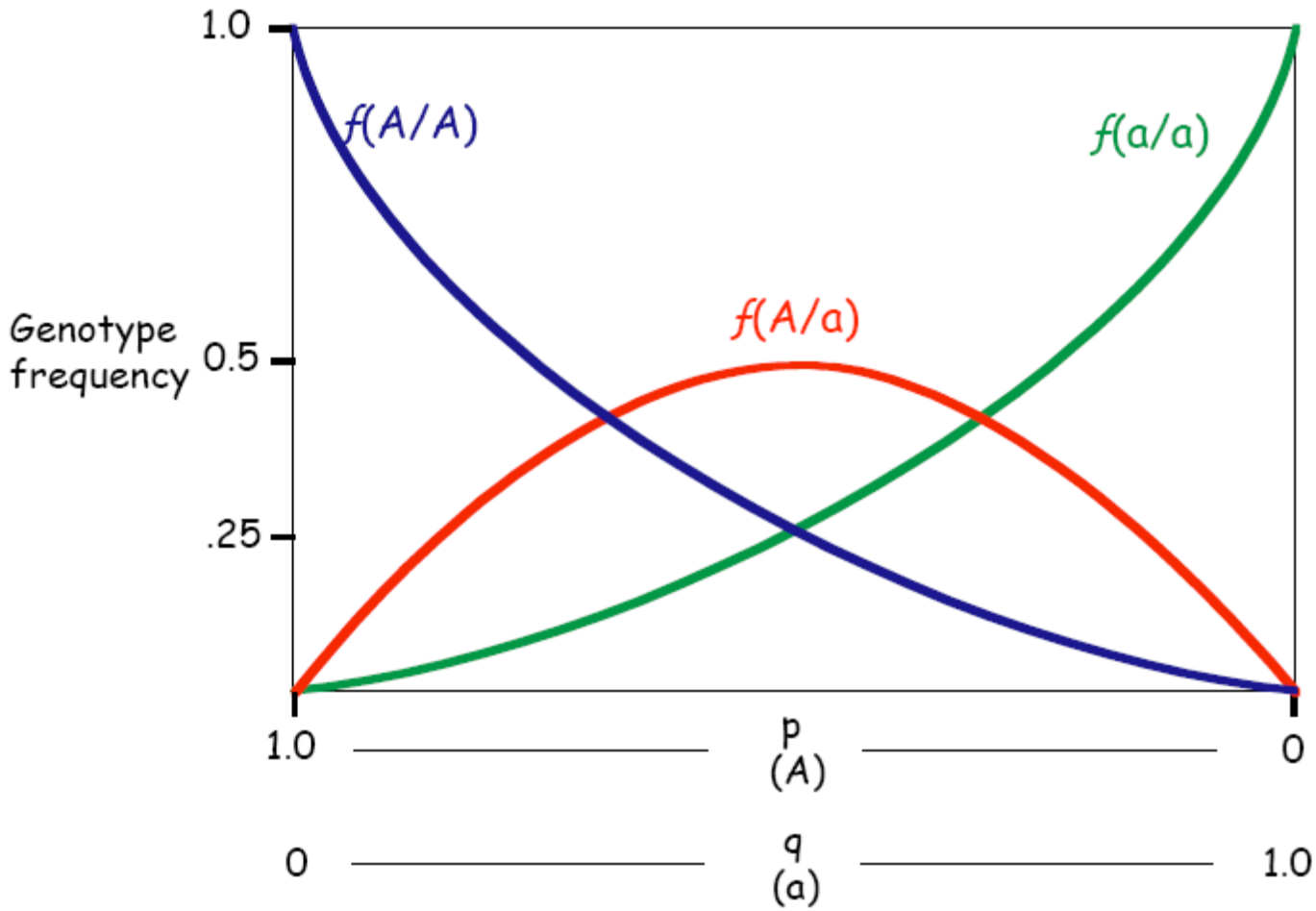

generation one, assuming random mating.

- The frequency f(A) of A in generation can be derived as follows:

$$f(A) = p^2 + \tfrac{1}{2}(2pq) = p(p+q)=p,$$

the frequency of allele A in generation zero

- A useful way to look at frequencies in HWE:


1.0
f(A/A)
f(a/a)
Genotype frequency
0.5
f(A/a)
.25
1.0
p
(A)
0
0
q
(a)
1.0

# 3.c Allelic association

## Setting

- Consider now two genetic loci, located on the same chromosome, with alleles A, a in locus 1 and alleles B, b in locus 2
- Thus, each chromosome in a homologous pair contains two alleles (one from each locus), which regarded jointly form a haplotype
- Four haplotypes are possible in the case of two diallelic loci: A-B, A-b, a-B and a-b.
- The haplotype on the ith chromosome in a pair may be regarded as a value of a random variable $H_i$, i = 1, 2.
- The respective probabilities $p(H_i = h)$ represent the *population haplotype frequencies.*

## Definition (continued)

- Suppose the genetic data on the same chromosome at the two loci are organized as follows:

| | | Locus 2 | | |
|---|---|---|---|---|
| | | B | b | |
| Locus 1 | A | | | $p(H_{1i}=A)$ |
| | a | | | $p(H_{1i}=a)$ |
| | | $p(H_{2i}=B)$ | $p(H_{2i}=b)$ | |

- Then the haplotype frequencies (e.g., $p(H_i=AB)$) can be computed via the corresponding allele frequencies $p(H_{1i}=A)$ and $p(H_{2i}=B)$
- To compute the "joint distribution" via the "marginal distributions", we need to know what the degree of dependence is between the constituting alleles

## Definition (continued)

- Denote $H_{1i}$ the random variable that refers to that component of $H_i$ at locus 1 ($H_{2i}$ is defined similary)
- The non-independence between the $H_{1i}$ and $H_{2i}$ in the given population is called allelic association, i.e. when

$$P(H_{1i} = x, H_{2i} = y) \neq P(H_{1i} = x)P(H_{2i} = y).$$

- Another example of non-independence:

> If the population initially consisted of individuals with AB haplotypes only and the two loci are close together, some new mutation could affect both loci at the same time resulting in a individual with an ab haplotype. Because the loci are close together, recombinations between them are rare, and the two alleles ab are always transmitted together to the children. This means, that if we know that the allele in the first locus is an a, we can also be pretty sure that the other allele is a b — this is means that the two alleles are dependent.

## Measure of allelic association

- Several measures of allelic association exist
- A popular one is determined by taken the difference between the joint probability $p(H_i=AB)$ and the product of the marginal probabilities $p(H_{1i}=A)$ and $p(H_{2i}=B)$
- This deviation from "statistical independence" is usually denoted by D (in particular for the previous example: $D_{AB}$)
- It can be shown that

$$D_{xy,t} = (1-\theta)^t D_{xy,0}$$

with $D_{xy,0}$ the measure of allelic association in the initial generation and $D_{xy,t}$ the one after *t* generations of random mating

## Measure of allelic association (continued)

- Thus, is the two loci are close ($\theta \approx 0$) it takes many generations to achieve *linkage equilibrium*.
- However, if the loci are far apart ($\theta \approx 1/2$), then $D_{xy,t}$ decreases to zero very quickly, since

$$D_{xy,t} \approx D_{xy,0}/2^t.$$

- The existence of allelic associations in human populations has some important implications for genetic analysis. First, it might be possible to indirectly predict the allele in one locus (e.g. an unobserved locus causing the disease) by observing the alleles in another nearby locus (e.g. a marker locus). This strategy is used in *association studies* to localize the unobserved disease loci. Second, many statistical techniques, simply assume that the marker loci are in linkage equilibrium.

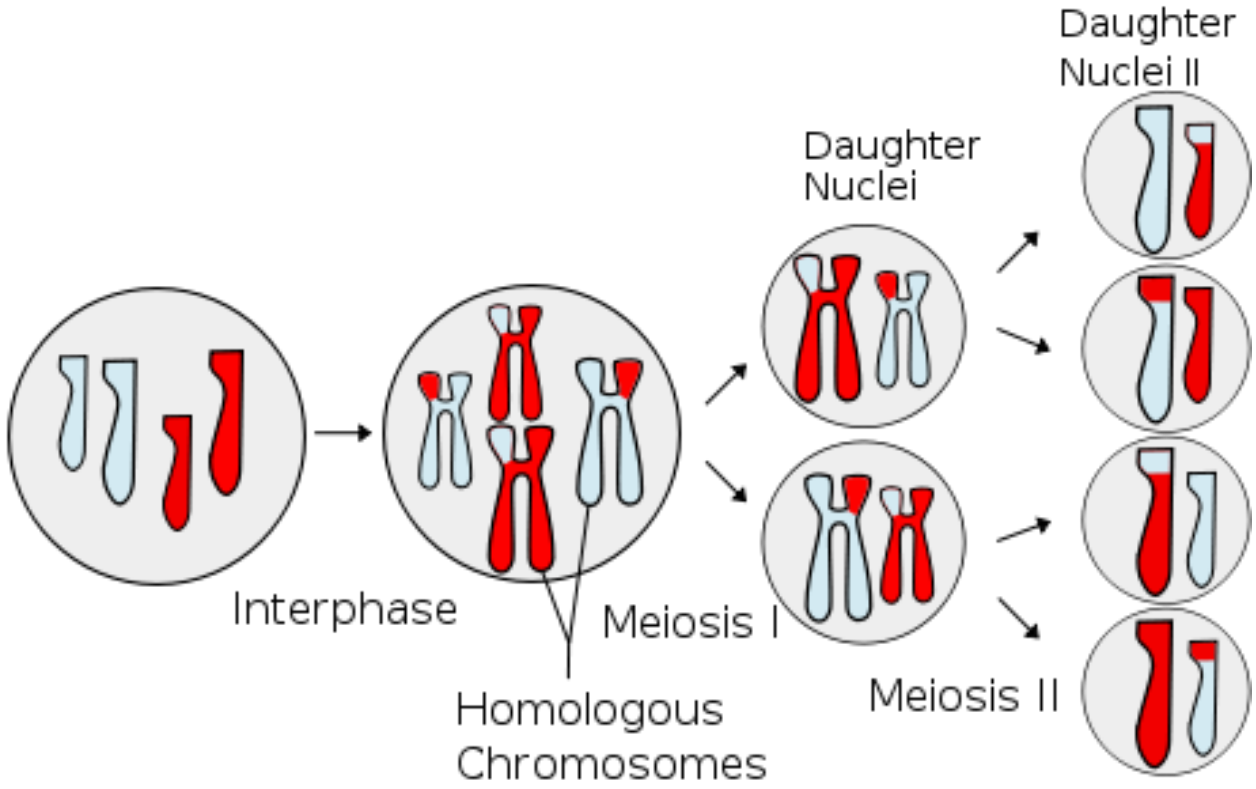
# Linkage disequilibrium and gamete frequencies

## *Two loci*

- When considering two (or more) loci, one must also account for the presence or linkage disequilibrium
- Under random mating, gametes combine at random.
- Hence, if the frequency of an A-B gamete is 0.4 and an a-b gamete is 0.1, then
  - freq(A-B/A-B) = $0.4^2$
  - freq(a-b/a-b) = $0.1^2$
  - freq(A-B/a-b) = $2\times0.4\times0.1$
- However, the frequencies of gametes in a population can change by recombination from generation to generation unless they are in linkage equilibrium (which is also called gametic-phase equilibrium; see later).

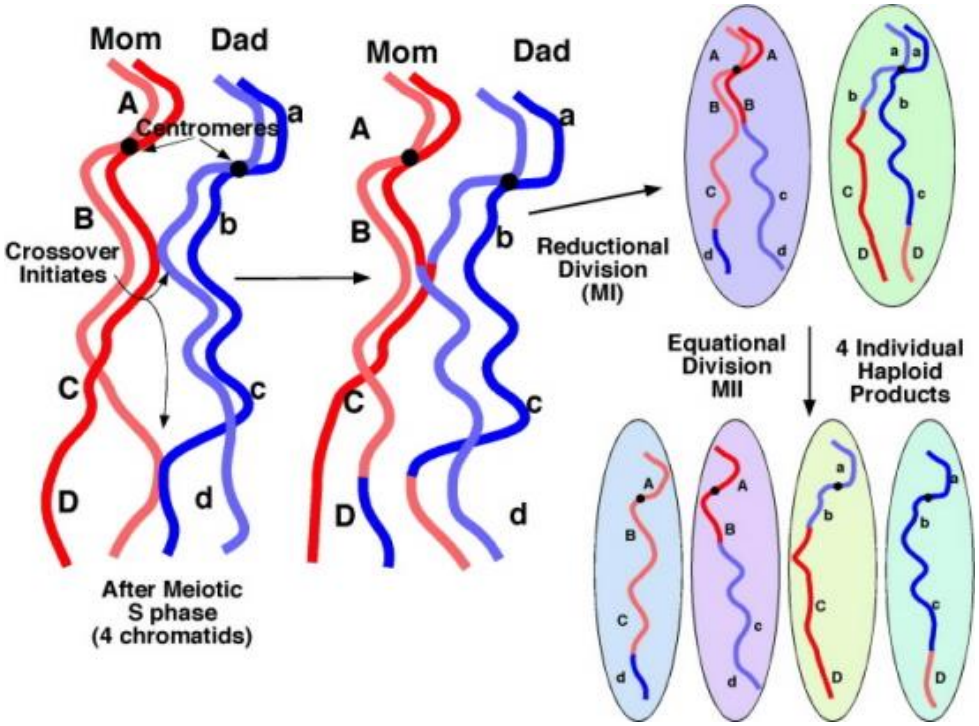
# Changing gamete frequencies

- Meiosis

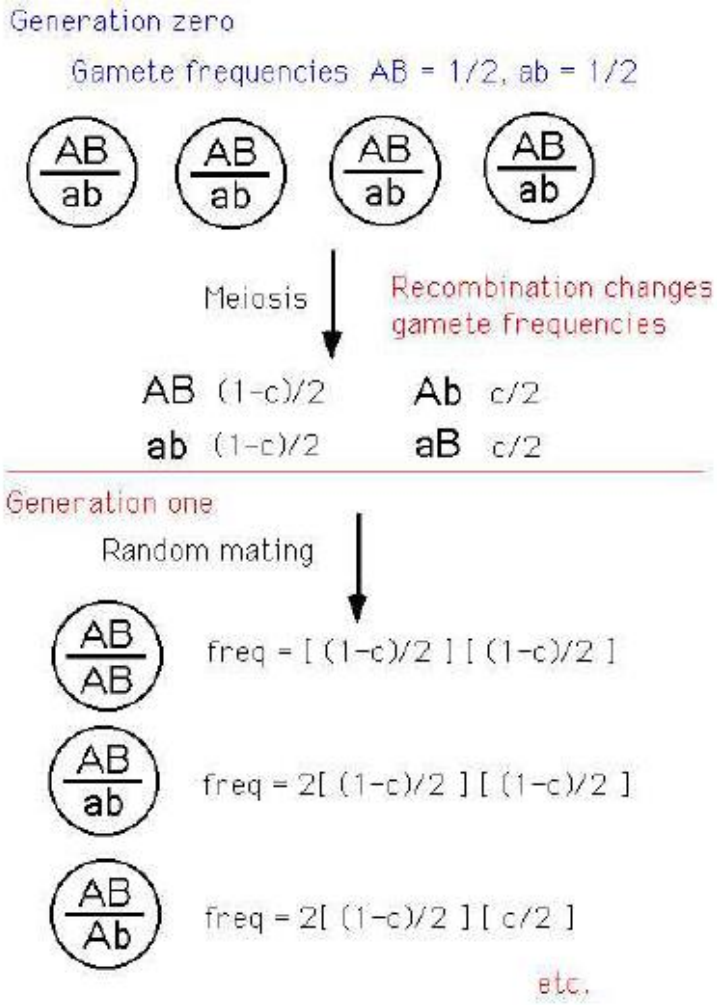


# Changing gamete frequencies

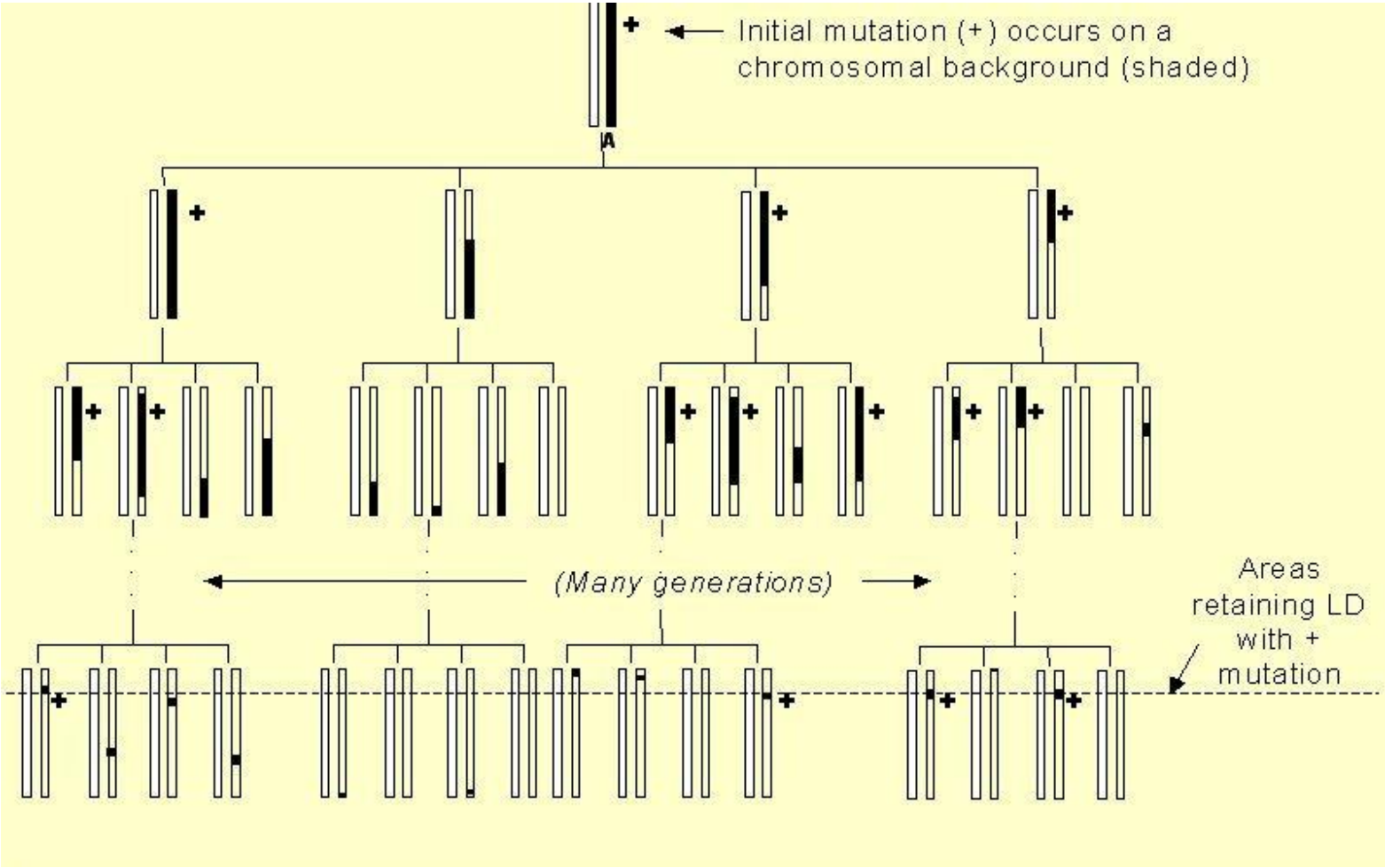
- Meiosis



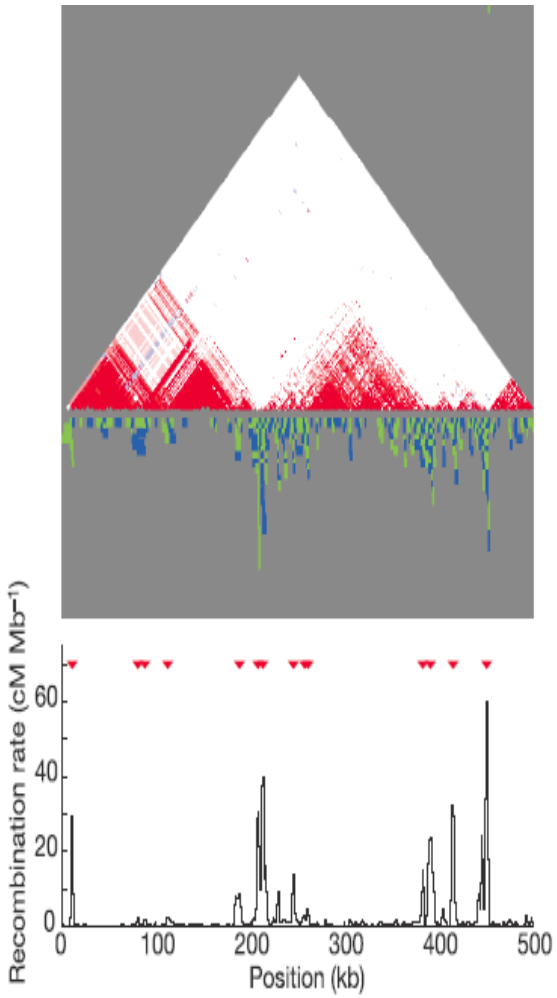
# Changing gamete frequencies – an example



# An example of linkage disequilibrium through generations



# Haplotype blocks



# 4 Natural selection revisited

## Main references:

- URLs:
  - Course notes on population genetics available from http://arnica.csustan.edu/boty1050

# 4.a Definition

## Natural selection

- *Natural selection* refers to differential reproduction. Organisms with more advantageous gene combinations secure more resources, allowing them to leave more progeny. It is a negative force, nature selects against, not for.
- Ultimately natural selection leads to **adaptation** - the accumulation of structural, physiological or behavioral traits that increase an organism's fitness.

# 4.b Three types of selection

## Stabilizing Selection

- Selection maintains an already well adapted condition by eliminating any marked deviations from it. As long as the environment remains unchanged the fittest organisms will also remain unchanged.

    - Human birth weight averages about seven pounds. Very light or very heavy babies have lower chances of survival. Fur color in mammals varies considerably but certain camouflage colors predominate in specific environments. Stabilizing selection accounts for "living fossils" - organisms that have remained seemingly unchanged for millions of years.

## Directional Selection

- favors one extreme form over others. Eventually it produces a change in the population. Directional selection occurs when an organism must adapt to changing conditions.

    - Industrial melanism in the peppered moth (Biston betularia) during the industrial revolution in England is one of the best document examples of directional selection.

        The moths fly by night and rest during the day on lichen covered tree trunks where they are preyed upon by birds. Prior to the industrial revolution most of the moths were light colored and well camouflaged. A few dark (melanistic) were occasionally noted.

        During the industrial revolution soot began to blacken the trees and also cause the death of the lichens. The light colored moths were no longer camouflaged so their numbers decreased quite rapidly. With the blackening of the trees the numbers of dark moths rapidly increased.

The frequency of the dark allele increased from less than 1% to over 98% in just 50 generations. Since the 1950's attempts to reduce industrial pollution in Britain have resulted in an increase in numbers of light form.

## Directional Selection (continued)

- Antibiotic resistance in bacteria is another example of directional selection. The overuse/misuse of antibiotics has resulted in many resistant strains.
- Pesticide resistance in insects is another common example of directional selection.

## Disruptive Selection

- Disruptive selection occurs when two or more character states are favored.

  - African butterflies (Pseudacraea eurytus) range from orange to blue. Both the orange and blue forms mimic (look like) other foul tasting species (models) so they are rarely eaten. Natural selection eliminates the intermediate forms because they don't look like the models.

# 5 Inbreeding

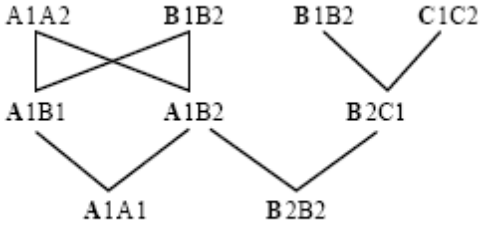
## Main references:

- Course notes GENE251/351 Lecture 8

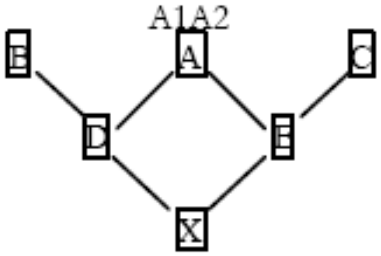
# 5.a Introduction

## Definition

- The *coefficient of inbreeding* (F) is the probability that two alleles at a randomly chosen locus are identical by descent (IBD)
- Hence, F ranges between 0 and 1



# F of an individual X: example 1



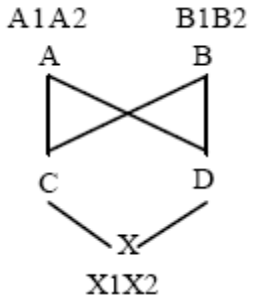

- The probability of 2 alleles at a randomly chosen locus being IBD can be computed via the probabilities

$$p_{A1A1} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

$$p_{A1A2} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

- Hence, $F_X = 2 \times 1/16 = 1/8$

# F of an individual X: example 2




- The probability of 2 alleles at a randomly chosen locus being IBD can be computed via the probabilities

$$p_{A1A1} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

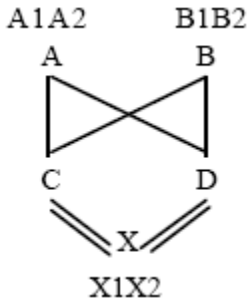
$$p_{A2A2} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

$$p_{B1B1} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

$$p_{B2B2} = \left(\frac{1}{2} x \frac{1}{2}\right) x \left(\frac{1}{2} x \frac{1}{2}\right) = \frac{1}{16}$$

- Hence, $F_X = 4 \times 1/16 = 1/4$

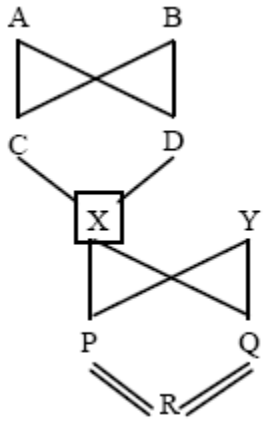
# F of an individual X: example 3




- The shortcut "loop" method determines for a loop (path through common ancestor) a contribution of $(½)^n$ to F, where n is the number of individuals in the loop, X excluded
- Hence, for more than one loop, determine each time $(½)^n$, and sum the contributions to F
- In the example, the loops are:
  - ADXC: $(½)^3$
  - BDXC: $(½)^3$

  leading to $F_X = (½)^3 + (½)^3 = ¼$

# F of an individual R: example 4

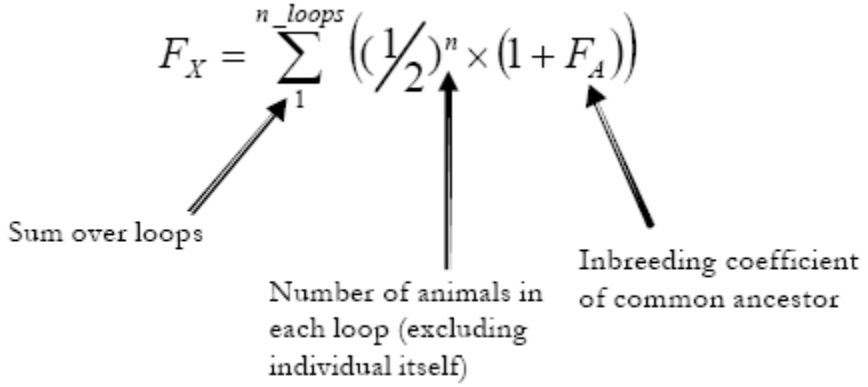



- When the common ancestor (X) is inbred, then a correction is needed
- Note that $F_X$=1/4 (example 3)
- The contribution from any loop with X must be increased for X itself being inbred
- The loops are
  - XQRP: $(½)^3(1+1/4)$
  - YQRP: $(½)^3$

  leading to $F_R$= 0.281

# General formula

- The formal equation for the “loop” method is:



# Change in genotype frequencies in response to inbreeding

- Inbreeding increases expression of recessive alleles:
  - Genotype frequencies for non-inbred: $p^2$, 2pq, $q^2$
  - Genotype frequencies for inbred: $p^2$+Fpq, 2pq-2Fpq, $q^2$+Fpq
- For instance, if q=0.02, then

| F | 0 | 0.125 | 0.25 | 0.50 |
|---|---|---|---|---|
| Prob. aa (recessive genotype) | 0.4 in 1000 | 2.9 in 1000 | 5.3 in 1000 | 10.2 in 1000 |

# Change in genotype frequencies in response to inbreeding (continued)

- For instance, if p=q=0.5

| Genotype | aa | Aa | AA |
|---|---|---|---|
| Frequency | $q^2+pqF$ | $2pq-2pqF$ | $p^2+pqF$ |
| At F=0 | 0.25 | 0.50 | 0.25 |
| At F=0.5 | 0.375 | 0.25 | 0.375 |
| At F=1.0 | 0.5 | 0 | 0.5 |

- Observe that the allele frequencies f(a)=q and f(A)=p do not change:
  - $f(a) = (q^2 + pqF) + ½ (2pq-2pqF) = q^2+pq=q(q+p)=q$
  - $f(A) = (p^2 + pqF) + ½ (2pq-2pqF) = p^2+pq=p(p+q)=p$

# 6 Fitness

## Main references:

- URLs:
  - Course notes on population genetics available from http://arnica.csustan.edu/boty1050

# 5.a Gentle introduction

## Meeting Darwin again

- Darwin marveled at the "perfection of structure" that made it possible for organisms to do whatever they needed to do to stay alive and produce offspring
- He called this perfection of structure *fitness*, by which he meant the combination of all traits that help organisms survive and reproduce in their environment
- Fitness is now measured as r*eproductive success*, i.e. the number of progeny left behind who carry on the parental genes. Those who fail to contribute to the next or succeeding generations are unfit.

## A measure of fitness

- In population genetics, we look for differences in the fitness of different genotypes at a particular locus (or set of loci).
  - For example, do AA individuals have (on average) more offspring than (say) aa individuals?
  - We denote the expected fitness of a particular genotype (say Aa) by $W_{Aa}$. If $W_{Aa} = 12$, Aa individuals leave (on average) 12 offspring.

- Selection on a single locus with two alleles

| Genotype | $AA$ | $Aa$ | $aa$ |
|---|---|---|---|
| Freq. before selection | $p^2$ | $2p(1-p)$ | $(1-p)^2$ |
| Fitness | $W_{AA}$ | $W_{Aa}$ | $W_{aa}$ |
| Freq. after selection | $p^2 \frac{W_{AA}}{\overline{W}}$ | $2p(1-p) \frac{W_{Aa}}{\overline{W}}$ | $(1-p)^2 \frac{W_{aa}}{\overline{W}}$ |

where $\overline{W} = p^2 W_{AA} + 2p(1-p) W_{Aa} + (1-p)^2 W_{Aa}$

The frequency of $A$ after selection, $p'$, is

$$p' = \text{freq}(AA \text{ after selection}) + \frac{1}{2}\text{freq}(Aa \text{ after selection})$$

$$= p^2 \frac{W_{AA}}{\overline{W}} + \frac{1}{2} 2p(1-p) \frac{W_{Aa}}{\overline{W}}$$

Thus,

$$\Delta p = p' - p = p\left(p \frac{W_{AA}}{\overline{W}} + (1-p) \frac{W_{Aa}}{\overline{W}} - 1\right)$$

## A measure of fitness (continued)

- In quantitative genetics, we look for differences in the fitness of different characters (phenotypes).
  - For example, do taller individuals have more offspring than shorter individuals.
  - We denote the expected fitness of a particular character value z by W(z). If z = height, then W(60) = 2.5 means that individuals who are 60 inches have, on average, 2.5 offspring.


**Figure 1.5**
*Differences in height in the same population:* heights of conscripts over 60 years ago. (From A. Blakeslee, *Journal of Heredity*, vol. 5, 1914.)