2004-2005 Molecular Medicine *M.Sc.*

Introduction to Population Genetics

Two Bengal tigers, one displaying the recessive white coat phenotype

Dr. A. Ruth Freeman
Department of Genetics
Trinity College Dublin.
Email: abigail.freeman@tcd.ie
www.gen.tcd.ie/molpopgen/ruth

Useful information

Forthcoming lectures

1) Genetic Variation and Introduction to Population Genetics(1) (2pm, Thurs 21st Oct)
2) Introduction to Population Genetics(2) (3pm, Thurs 21st Oct)
3) Genetics and Human Evolution 1 & 2 (next term)
4) Genetics and Human Evolution 3 & 4 (next term)

Recap- Why are we interested in Genetic Variation?

- 99.9 percent of your DNA is exactly the same as any other person's (except identical twins)
- Difference in the sequence of DNA among individuals is called genetic variation.
- Genetic variation explains some of the differences among people, such as eye color and blood group.
- Also plays a role in whether a person has a higher or lower disease risk. E.g. cystic fibrosis, sickle cell disease, Huntington's disease, Duchenne's muscular dystrophy etc.
- Genetic variation also exists that does not affect the phenotype → neutral variation

Terminology

- Allele – alternative form of a gene
- Locus – position of a gene on a chromosome
- Polymorphic – the presence of two or more forms of a gene (usually >1%)
- Mutation – an inherited genetic change caused by an alteration in the DNA sequence of an allele
- Dominant – an allele that masks the expression of another allele of the same gene
- Recessive – an allele whose effect is masked by the presence of a dominant allele of the same gene
- Genotype - genetic constitution of an individual
- Phenotype – the appearance and function of an organism as a result of its genotype and the environment

What is Population Genetics?

- The science of population genetics deals with Mendel's laws and other genetic principles as they affect entire populations of organisms. The organisms may be human beings, animals, plants, or microbes.
- Population genetics also includes the study of various forces that result in evolutionary changes in species through time and space.
- The principles of population genetics are basic to a broad evolutionary perspective on biology.
- Evolution provides a wealth of testable hypotheses for all other branches of biology. Many things in biology become comprehensible in the light of evolution: they result from shared ancestry among organisms, and they attest to the unity of life on earth.

Why is population genetics important in a medical context?

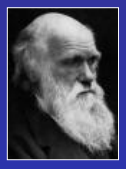
- Understanding the genetic basis of disease resistance and susceptibility in human populations.
- Population-associated inherited disorders - sickle cell anaemia, Tay-Sachs disease, cystic fibrosis etc…
- Genetic mapping and identification of genes for disease susceptibility in humans, including breast cancer, colon cancer, diabetes, schizophrenia etc….
- Genetic counselling of parents and other relatives of patients with hereditary diseases.

Why is population genetics important in a medical context?

- Forensic interpretation and statistical analysis of DNA evidence from criminal investigations.
- Design of studies to sample and preserve a record of genetic variation among human populations throughout the world. Understanding the evolutionary history of the human species and our close relatives.
- Understanding the evolution and population dynamics of human pathogens such as bacteria and viruses (e.g. the HIV - the causative agent of AIDS).

Population genetics is a result of the "synthesis" of Darwin's theory of evolution and Mendel's laws of heredity.

Darwin's theory of evolution through natural selection can be summarised in three principles (Origin of Species 1859):



- Principle of variation: Among individuals within any population, there is variation in morphology, physiology, and behaviour.
- Principle of heredity: Offspring resemble their parents more than they resemble unrelated individuals.
- Principal of selection: Some forms are more successful at surviving and reproducing than other forms in a given environment.

Darwin's problem – Blending inheritance

Darwin thought that particles called "gemmules" would move from the extremities where an organ was, back to the sex cells and, in modern terms, "reprogram" the sex cells with the new information.

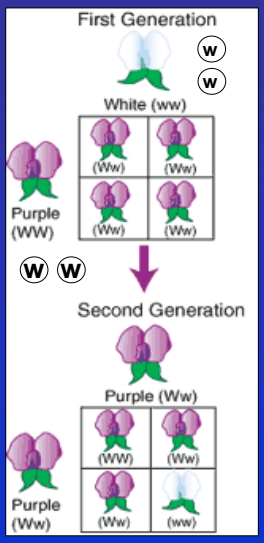
He thought that traits were "blended" (e.g. a tall and short producing offspring of intermediate height) during sexual reproduction.

However, this would gradually eliminate all variation, and would require the rate of mutation to be very high to produce new material on which natural selection could act.

NOTE THAT DARWIN LOOKED AT Quantitative/Continuous characteristics (e.g. height, speed, beak shape).

This usually indicates polygenic inheritance – where 2 or more genes affect a single phenotype.

Mendel's Laws (1865-9)



- First – Law of segregation

  "Contrasting forms of a character are controlled by pairs of unlike alleles that separate in equal numbers into different gametes as a result of meiosis"

Each individual carries 2 unblending copies of each gene....and each gamete receives only one copy of each gene chosen at random

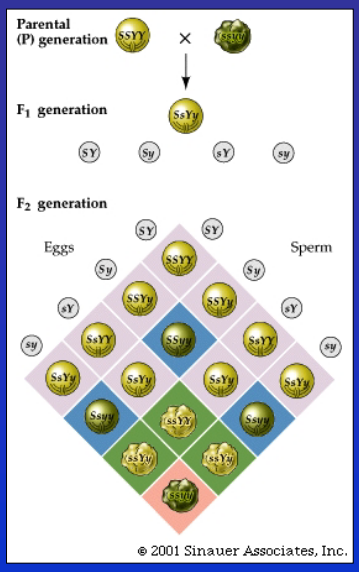
Mendel's Laws

•Second – Law of independent assortment

"2 or more unlike pairs of alleles segregate independently of each other as a result of meiosis, provided the genes concerned are unlinked."

Allele pairs separate independently during the formation of gametes. Therefore, traits are transmitted to offspring independently of one another.

NOTE THAT MENDEL WAS LOOKING AT DISCRETE TRAITS (single gene)



Biometricians vs. Mutationists

Darwin convinced most biologists that species evolve, but found it difficult to convince them that natural selection was the driving force – did not understand mechanism of inheritance, and how variation was transmitted through the population

Even though they were contemporaries, Mendel's contributions were not recognised.

Darwin's theories were based on continuous traits, while Mendel was looking at discrete traits.

Hugo De Vries – developed idea of mutation as main mechanism of evolution – Mutationism/Mendelism

Karl Pearson – defended Darwinian evolution and cumulative effect of small changes which passed from one generation to the next without being subject to Mendel's laws.

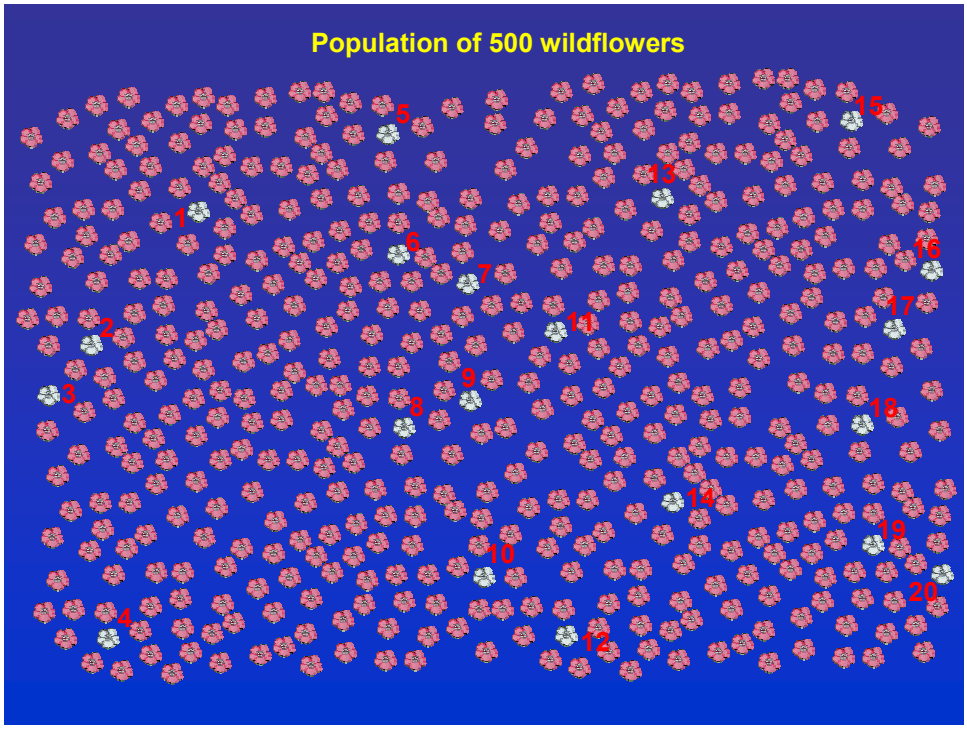
The simplest description of Mendelian variation is the frequency distribution of alleles (gene variants) and genotypes in a population or gene pool.

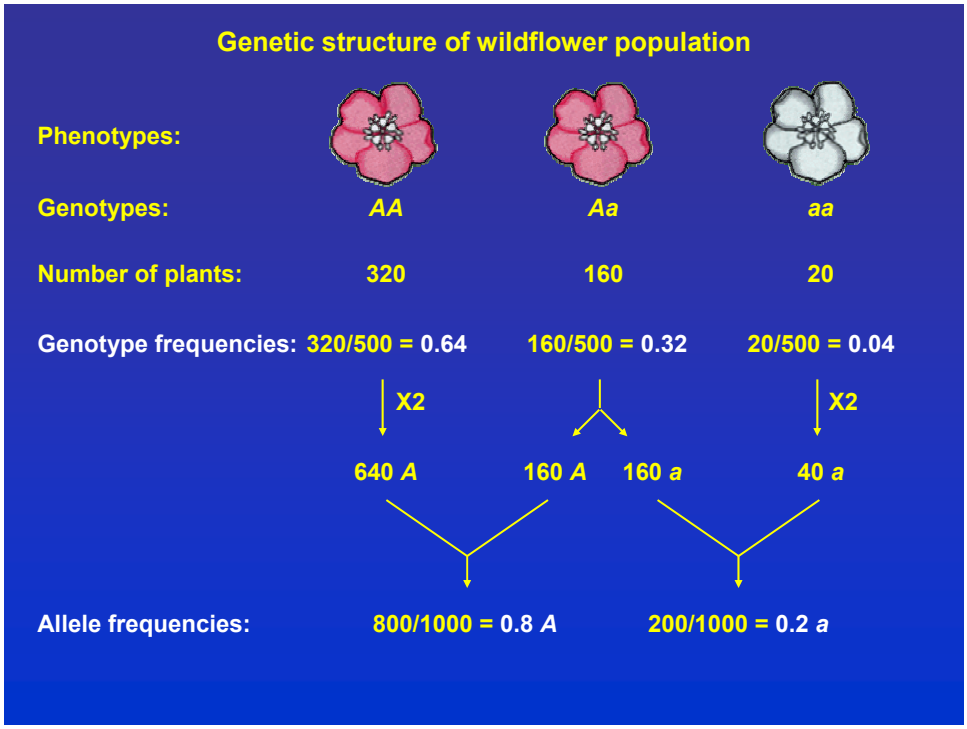
What exactly is a gene pool?

- Gene pool: all alleles at all gene loci in all individuals of the population at any one time.
- N.B. Diploid species ➔ each locus is represented twice in each individual.
- Can be either homozygous (same two alleles) or heterozygous (two different alleles) at an individual locus.
- If all members of a population - homozygous for the same allele ➔ allele is fixed in the gene pool.
- Usually, 2 or more alleles for a gene—each having a relative frequency (proportion) in the gene pool.

How do we calculate allele frequencies?

- A wildflower population with two varieties (pink and white) where an allele for pink flower (*A*) is completely dominant to an allele for white flower (*a*).
- Only 2 alleles at this genetic locus, *A* and *a*.
- Imagine a population of 500 plants with 20 white-flowered plants (must be homozygous *aa*).
- The other 480 plants have pink flowers (320 homozygous *AA* & 160 heterozygous *Aa*).
- N.B. Diploid ➔ 1000 copies of the gene for flower colour in the population.
- Dominant allele (*A*) = 320 x 2 = 640 for *AA* plants and 160 x 1 for *Aa* plants.
- Therefore: frequency of the *A* allele = 800/1000 = 0.8 = 80%
- Only 2 alleles ➔ frequency of the *a* allele = 0.2 = 20%.


Population of 500 wildflowers
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20


Genetic structure of wildflower population
Phenotypes:
Genotypes: AA Aa aa
Number of plants: 320 160 20
Genotype frequencies: 320/500 = 0.64 160/500 = 0.32 20/500 = 0.04
X2
X2
640 A
160 A
160 a
40 a
Allele frequencies: 800/1000 = 0.8 A 200/1000 = 0.2 a

The genetic structure of a population

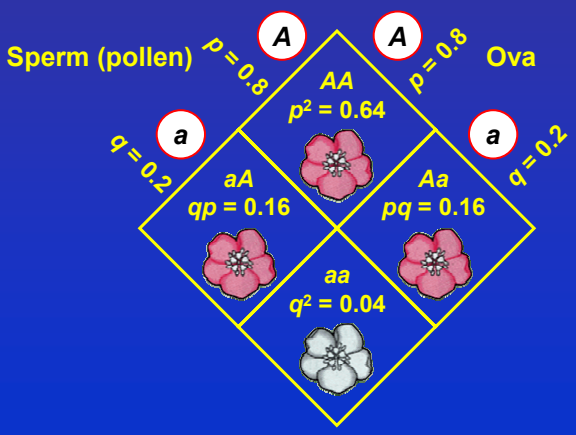
- N.B. *p* = frequency of *A* = 0.8 *q* = frequency of *a* = 0.2
- Population geneticists use the term genetic structure to refer to a population's frequency of alleles and genotypes.
- The frequencies of genotypes and alleles in a non-evolving population are described by the Hardy-Weinberg theorem.
- The mathematical formulation was first proposed independently in 1908 by G.H. Hardy a British mathematician and Wilhelm Weinberg, a German physician.

The Hardy-Weinberg principle or theorem

- The Hardy Weinberg theorem states: *the frequencies of alleles and genotypes in a population's gene pool remain constant over the generations unless acted upon by agents other than sexual recombination.*
- *i.e.* The sexual shuffling of alleles due to meiosis and random fertilisation has no effect on the overall genetic structure of a population.
- Wildflower population (frequency of *A* (*p*) = 0.8, frequency of *a* (*q*) = 0.2.
- How does the reshuffling and genetic randomisation during meiosis/fertilisation affect the allele frequencies in the next generation?
- Assume random mating: analogous to mixing all gametes in bag and drawing randomly.

Genetic structure of 2nd generation of wildflowers

N.B. $p = 0.8$; $q = 0.2$
Also: assuming completely random mating



The Hardy-Weinberg principle or theorem (cont.)

- 2nd generation:
  $p^2 = 0.64$ *AA* $2pq = 0.32$ *Aa* $q^2 = 0.04$ *aa* (genotype frequencies)
  Allele frequencies: $p = 0.8$ *A* $q = 0.2$ *a*
- The genetic structure of a non-evolving population remains constant over the generations (unlike blending inheritance).
- Sexual recombination alone will not alter the relative frequencies of alleles or genotypes.
- The flower-colour locus is a state of equilibrium ➔ Hardy-Weinberg Equilibrium (HWE).
- Simplest case: 2 alleles, one completely dominant.

The Hardy-Weinberg principle/theorem in general terms

- **N.B. general principle applies to 3 or more alleles and other dominance relationships.**
- **However, simplest case used to describe the theorem in general terms:**

2 alleles: $p + q = 1$

$p + q = 1$, then $p = 1 - q$ and $q = 1 - p$

Probability of *AA* genotype: p^2 (probability multiplication rule) = 0.64

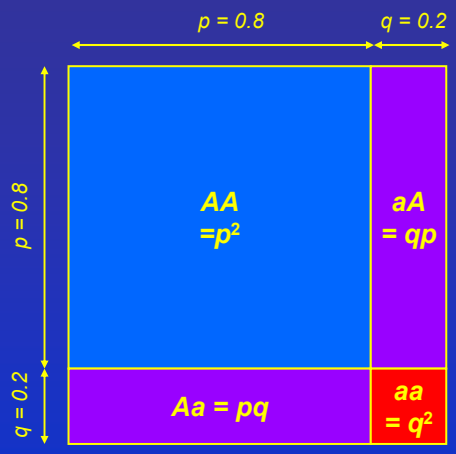
Probability of *aa* genotype: $q^2 = 0.04$

Two ways to produce *Aa* genotype: either *Aa*, $p \times q$ & *aA*, $q \times p$) ➔ $2pq = 0.32$

For 2 alleles, the summation over all genotypes is:

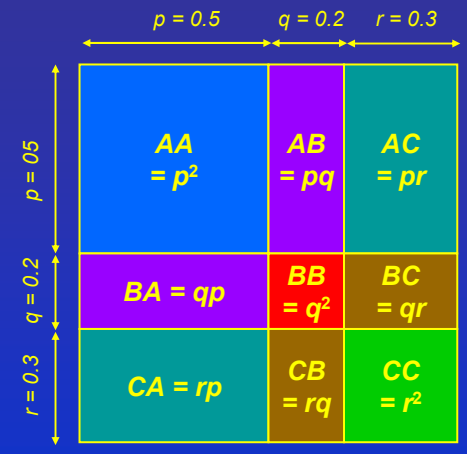
$$p^2 + 2pq + q^2 = 1$$ **The Hardy-Weinberg equation**

The Hardy-Weinberg equation from a geometrical perspective



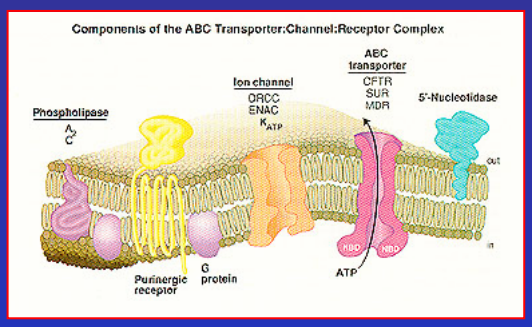
$$(p + q)^2 = p^2 + 2pq + q^2 = 1$$ **HWE**

The Hardy-Weinberg equation from a geometrical perspective: 3 alleles



$$(p + q + r)^2 = p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$$

An example: cystic fibrosis in Ireland



Ireland: highest incidence in the world: 1 in 1600 people are affected by this recessive genetic disorder.

1) What is the frequency of the defective CF gene?

2) What is the frequency of the normal functional CF gene?

3) What is the frequency of carriers (heterozygotes) of the CF gene?

Cystic fibrosis in Ireland

1 in 1600 people are affected by this recessive genetic disorder (Two versions of the gene - 2 alleles)

Assuming HWE:

1) q^2 = 1/1600 = 0.000625 = 0.06%

$q = \sqrt{0.000625} = 0.025 = 2.5\%$

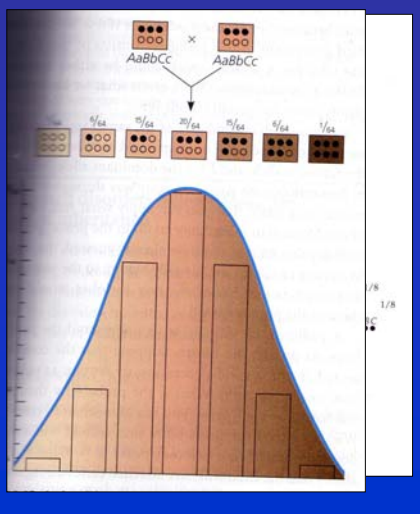
2) $p = 1 - q = 0.975 = 97.5\%$

3) Proportion of carriers (heterozygotes) = $2pq$ = 2 x 0.975 x0.025 = 0.0488 ~ 5% or 1/20 people.

Some early 20th century biologists (Biometricians) rejected Mendelism and argued that segregation and independent assortment could not explain the continuous variation that was predominantly observed in nature. They argued that the 3:1 ratio should be observed more frequently.

Quantatative characteristics usually indicate polygenic inheritance – where 2 or more genes affect a single phenotype in an additive way.

Example – at least three genes control skin pigmentation in humans.



Microevolution...and getting back to Darwin

- Microevolution is a generation-to-generation change in a population's allele or genotype frequencies.
- Hardy-Weinberg theorem describes a gene pool in equilibrium—a *non-evolving* population.
- If the frequencies of alleles or genotypes deviate from values expected from HWE, then the population is evolving.
- Therefore at the population level: Evolution is a generation-to-generation change in a population's frequencies of alleles or genotypes—a change in a population's genetic structure.
- Change in gene pool like this is evolution at the smallest scale: microevolution.

Causes of Microevolution

- NB. Microevolution is a departure from HWE. There are essentially five conditions that are required for HWE to be maintained in a population.

Features of a population in HWE

1) *Very large population size*: In a small population, genetic drift, which is chance fluctuations in the gene pool, can alter the frequencies of alleles.
2) *Isolation from other populations*: Gene flow, the transfer of alleles between populations due to the movement of individuals or gametes, can change gene pools.
3) *No mutations*: By changing one allele into another, mutations can alter the gene pool.
4) *Random mating*: If individuals select mates having certain heritable traits, then the random mixing of gametes required for HWE does not occur.
5) *No natural selection*: Differential survival and differential reproductive success alter a gene pool by favouring the transmission of some alleles at the expense of others.

The 5 conditions required to maintain Hardy-Weinberg equilibrium ➔ provide a framework for understanding the processes that cause microevolution.

Each is a departure from one of the conditions to maintain HWE.

1) Genetic drift

2) Gene flow

3) Mutation

4) Nonrandom mating

5) Natural selection

Only natural selection generally adapts a population to its environment.

The other four processes ➔ termed non-Darwinian - usually non-adaptive.

Reconciling Darwin and Mendel (1930-40)

R.A. Fisher and Sewall Wright in the 1930s showed that Darwinian selection flowed as a logical consequence of the results of the new genetics.

Under Mendel's laws, genetic variation can persist in a population for a very long time, which increases the likelihood that a new mutation will meet copies of itself and thus permit new traits to become fixed.

So the modern paradigm of evolutionary theory is that natural selection acts on genetically different individuals in populations, thus species change over time.

However, selection is not the only driving force in evolution. There are other forces such as genetic drift, gene flow, mutation and recombination.

The "Modern Synthesis" – 1940s

- **Population:** (defined as a localised group of individuals belonging to the same species) recognised as units of evolution
- **Natural selection:** central role as the mechanism of evolution.
- **Gradualism:** explained how large changes can evolve from many small changes accumulated over large periods of time.

How much genetic variation?

- Two schools of thought existed in the middle of the 20th century concerning the extent and nature of genetic variation.

CLASSICAL VIEW

- H.J. Muller.
- Genetic variation is limited.
- Natural selection removes deleterious mutations and beneficial mutations are spread rapidly through populations.
- High levels of polymorphism constituted a burden or genetic load.

BALANCE SCHOOL

- T. Dobzhansky.
- Genetic variation is extensive.
- Natural selection promotes genetic variation through providing a reservoir of adaptability to changing environment.

What is the origin of this variation?

- Most DNA substitutions do not alter the amino-acid sequence of proteins (silent or synonymous substitutions).
- Huge amount of genetic variation - could not be explained completely by natural selection ➨ Neutral Theory of Molecular Evolution (Kimura, 1968).
- Kimura suggested that alternative alleles in a population were generally selectively neutral or more properly - selectively equivalent. Allele frequencies in a population change primarily due to chance processes and genetic drift. *(Will deal with this later)*

What is the origin of this variation?

- Neutral variation is variation at the molecular level that does not necessarily influence the phenotype of an organism. Natural selection can still act on the phenotype of an organism.
- In 1969 King and Jukes expanded the neutral theory and published a famous paper in *Science* entitled:

  Non-Darwinian evolution: random fixation of selectively neutral mutations

- During the last 30 years the Neutral Theory has been borne out by numerous analyses and experiments.
- The Neutral Theory became the null hypothesis of molecular evolution - that is, the simplest way to interpret genetic variation at the molecular level.

Factors that disrupt Hardy-Weinberg and cause Microevolution

1) **Genetic drift**
2) **Gene flow**
3) **Mutation**
4) **Nonrandom mating**
5) **Natural selection**