

<http://www.ucl.ac.uk/~ucbhjow/bmsi/bmsi>

Genetics

introduction

Textbooks: There are many very well written human genetics textbooks on the market today. Which one you choose is very much a matter of individual preference, you must decide which author's style most suits you.

I particularly like:

- *Basic Human Genetics* by Mange and Mange, (2nd edition. Sinauer Associates Inc., 1999) (The first, 1994, edition is still a good book if last year's students try to sell it on.) The new edition includes a CDrom which has some helpful videos and animations.
- *Human Genetics, Concepts and Applications* 3rd edition by Lewis, (W.C. Brown, 1998).
- *Emery's Elements of Medical Genetics* by Mueller and Young (10th. edition, Churchill and Livingstone 1998) .
- *Medical Genetics* by Jorde, Carey, Bamshad and White (2nd edition Mosby, 1999)

The first two are entertainingly written for a predominantly American audience, the third is a somewhat drier British account and the final, American, book seems to be directly aimed at courses such as this.

A book which includes examples of almost all the commoner genetic and / or cytogenetic diseases is:

- *Essential Medical Genetics* by Connor and Ferguson-Smith (5th edition Blackwell Science Ltd. 1997)

A book which I have in the past recommended but which is a little difficult to read and is also now starting to show its age is:

- *Genetics in Medicine* by Thompson, McInnes and Willard (5th edition, Saunders 1992)

Click below for the individual lecture notes.

- [Lecture 1](#) Introduction and Mitosis
- [Lecture 2](#) Reproduction, meiosis and Mendel
- [Lecture 3](#) Apparent exceptions to Mendel's rules

- [Lecture 4](#) Mostly about sex
 - [Lecture 5](#) Tools of molecular genetics
 - [Lecture 6](#) Mutation
 - [Lecture 7](#) Cytogenetics
 - [Lecture 8](#) Genetic linkage
 - [Lecture 9](#) More on Linkage
 - [Lecture 10](#) Genes in populations. Cancer.
 -
-

Genetics Lecture 1

"There's a lot of it about!"

Scarcely a day seems to pass without some reference to genetics in the newspapers. Not since the furore stirred up by the publication of *The Origin of Species* has Genetics been so prominent in the headlines. Of all sciences which are likely to make changes in our lives in the near future outside the area of microelectronics, no science seems more likely than genetics to have profound effects. As doctors you should understand how genetics is of direct relevance to the lives of patients and you will be expected to deliver advice to patients worried or excited by newspaper stories (which are sometimes poorly written, more usually poorly edited, or based on overoptimism by the researchers concerned.)

In this series of lectures I will attempt to impart some basic principles of the subject (which have not changed since Mendel was rediscovered at the turn of the century) and to show how the techniques of "the new genetics" have revolutionised the subject.

Genetics in the news:

There are a number of headline catching themes, some of which are discussed below.

Forensic science

-

[DNA fingerprinting](#)

In 1985 sequences of DNA were discovered which were present at many sites in the genome and which varied in the numbers of copies present at any one location. Human fingerprints though composed of only a few basic elements, lines whorls and loops, are unique to any individual. In the same way, the pattern of variation of these simple DNA elements is sufficient to mark each human

individual uniquely (with the exception of identical twins). The added advantage is that any fragment of tissue from which DNA can be extracted, (and this can be as small as a dried drop of blood, or a single hair root) is enough to identify the human from which it originated. The latest high profile trial in which DNA evidence featured was the O.J.Simpson trial. However, in this case although DNA tests established beyond reasonable doubt the identity of various blood samples, enough questions remained as to how the blood came to be present to allow the defendant to be acquitted. This week we heard in the news of a murderer being convicted because of the DNA evidence obtained from minute bloodstains left on the clothes of his victim - herself a doctor at the Royal Free.

Because the DNA variation is inherited, following the rules of [Gregor Mendel](#) which we will discuss in lecture 2, it is also possible to use DNA fingerprinting to establish the relationships between individuals. One of the earliest such uses was in an immigration case to prove that a boy desiring entry to the UK was indeed the son of his mother who was resident here.

-

Facial features

"Forensic scientists will be able to predict a criminal's facial features from a hair root at the scene of the crime" - recent news story. Should we believe this headline? In fact, since the work is going on here at UCL, and since I and my family are some of the experimental guinea pigs, I feel fairly confident to report that, as so often, the position has been overstated by journalists in search of a good headline. Nevertheless, studies are underway to try to find measurable components of facial shapes which are controlled by the action of single genes. The type of feature which can be examined is sometimes subtle and can only be revealed by computer imaging techniques but can be much more obvious such as the prominent cleft in the chins of actors



Kirk Douglas and his son, Michael.

Once such features are found then the genes controlling their appearance can be sought using in the first instance genetic linkage analysis which features in lectures [8](#) and [9](#)

Agriculture

Working in the Galton Laboratory, the first laboratory set up specifically to study human inheritance, and doing research connected with the inheritance of human genetic disease

and with the human genome project, I am as guilty as any one of neglecting the tremendous importance of genetics to agriculture. You may feel that as tomorrow's doctors you too can safely neglect this area. However, as a figure of authority and with your scientific training you are bound to meet from time to time requests for information such as "Is it safe to eat genetically modified food?". In [lecture 5](#) we will consider the techniques of molecular genetics as applied to both human genetics and to agriculture so that even if you lack the detailed information to answer the above question, at least you will understand what processes were involved in creating a "genetically engineered" strain of plants or animals.

Some of the products which have made it to the supermarket shelves include:

- delayed ripening tomatoes
- freeze resistant strawberries
- fast growing fish

Last year [Tommy Archer](#) was acquitted in Borchester Crown Court for destroying a field full of genetically modified oilseed rape. Was he right or wrong to fear this crop? (Apologies to those of you who do not follow radio soap opera!) This year true life followed fiction as [Lord Peter Melchett](#) was found not guilty on similar charges.

Normal human variation

Human genetics is concerned with the causes and alleviation of disease. However, an important part of the subject is the study of normal human variation, if only to disprove that there is any such thing as a "true Aryan" genetic type. (That name incidentally only means the descendent of a person who spoke the original Aryan language). There are many traits which are the products of variation in the forms of a single gene present in different individuals. Some examples include:

- "[uncombable hair](#)" - wild tangly hair with a triangular cross section and a longitudinal groove
- "[wooly hair](#)" - hair which forms tight curls
- ear lobes - attached or unattached
- [ability to bend the first joint of the thumb back at a right angle](#)
- [ability to curl the tongue into a U section](#)
- [teeth present at birth](#) - King Louis XIV of France was born in this way, much to the distress of his wet nurses
- colour blindness - there are several forms but the most common, affecting about one male in twelve, is [deuteranopia](#), the inability to distinguish between red and green. It is caused by a missing gene which codes for one of the visual receptor pigments.

Genes can affect behaviour. In some cases we can begin to understand why because the gene in question is responsible for the production of a neurotransmitter or receptor. Other cases are so bizarre that we cannot even begin to guess what is the underlying cause.

Examples of "behavioural" genes include:

- [Tourette syndrome](#) - This inherited condition is very variable in its severity. The symptoms vary from no more than minor patterns of repetitive behaviour, motor ticks, to the most notorious symptom, an uncontrollable involuntary urge to utter swearwords and obscenities.
- "[The jumping Frenchmen of Maine](#)" - These curiously named people are the descendents of a French immigrant to America in the 18th century. They have an exaggerated startle reflex and will jump to obey any abruptly uttered command however dangerous or foolish it may be.
- [Monoamine oxidase deficiency](#) - Monoamine oxidase is an enzyme which acts to destroy certain neurotransmitters such as dopamine. There is an association between deficiency for this enzyme (because of a gene mutation) and the committal of acts of criminal violence.

Ultimately, because almost all genes code for enzyme products, mutant genes give rise to their effects through the altered action of an enzyme. In some of the cases above we can guess what the responsible enzyme might be but in others we have no idea. In the cases below we have a clear understanding of the biochemical defect. In some cases this has led to either a cure or an alleviation.

examples of biochemical deficiencies:

- [Phenylketonuria](#) - is caused by the absence of an enzyme which converts phenylalanine to tyrosine. The result is a build up of phenylalanine in the blood to a level where it causes brain damage in infants. Excess phenylalanine is converted into several metabolites and which are excreted into urine where they cause a mousy smell. Understanding the condition has led to a treatment. Patients must be placed on a diet with very little phenylalanine until they are teenagers. The diet must be started within a few weeks of birth, the earlier the better. Everyone born in the UK since the 1950s has been screened within a day or two of birth to test for the presence of this disease. [See lecture 6](#)
- [favism](#) - an abnormal sensitivity to beans which cause a haemolytic anaemia in patients. Even to smell the flowers of the bean may be enough to trigger an attack. This disease is commoner in the Mediterranean region and was known to Pythagoras who exhorted his pupils to avoid beans. It is caused by deficiency for the enzyme glucose-6-phosphate dehydrogenase.
- [Adenosine deaminase deficiency](#) - causes a complete failure of the immune system. It is one of the few diseases which have been successfully treated by gene therapy. Bone marrow cells from patients have been removed, a working gene has been placed into them and the cells have been replaced restoring the immune system.

Although Online Mendelian Inheritance in Man, [OMIM](#), lists 8649 characteristics for which there is some evidence of straightforward, single gene inheritance, there are many more characteristics which do not conform to this simple pattern because they are the

result of the aggregate effects of several variable genes and also of interaction with the environment.

"The gene for has been cloned!"

We often seem to hear this. Much to my relief (after working on the project for ten years!) I was finally able to utter these words myself in 1997 about the gene *TSC1* which is responsible for a genetic disease, [Tuberous sclerosis](#). What is meant by this statement? And how will it help patients? We will go into fuller detail in [lecture 5](#). For the moment, suffice it to say that when we have identified a mutant gene by "cloning" it, we can often immediately deduce something about its protein product's structure and make a reasonable guess as to its function (see for example the [cystic fibrosis](#) story later on). Also, we are able to look for the mutation(s) which may be present in any family with the immediate benefit of being able to detect whether unaffected members of the family are "carriers" ([defined later](#)) for instance or to be able to carry out [antenatal](#) (or even preimplantation) testing of embryos.

The Human Genome Project (HGP) is a huge international effort to find out the complete DNA sequence of the human genome. The project does not stop there, the 3 billion nucleotides has to be annotated and the detailed structure of each gene has to be aligned to the genomic sequence. The entire DNA sequence of the human genome is scheduled for completion in 2004 and already useful information is starting to be generated. In the next few years more and more "disease" genes are going to be identified based on HGP data. In 1997, the identification both of genes responsible for many cases of breast cancer and the Tuberous sclerosis gene mentioned above were directly aided by the DNA sequence data of the HGP.

Genetics in the near future

One benefit of identifying a genetic disease gene is the potential to offer "gene therapy" - the replacement of the defective gene with a new, functional copy. This is by no means an easy procedure and as yet there have been few successes. However, in the next ten years we will see big advances in our abilities to treat some of the common genetic diseases such as Duchenne Muscular Dystrophy by gene therapy. Gene therapy is also being considered as an approach to fighting cancers and even HIV infection.

A newt can regrow an amputated limb. It would be convenient if humans could also regenerate damaged or missing tissues and organs. The "cloning" of 'Dolly' the sheep from one single cell of adult breast tissue has brought this exciting possibility one step nearer.

Cell Division

We begin with consideration of the mechanics of gene inheritance. You should already be aware that:

- genes are composed of DNA,
- DNA is present as very long molecules,
- the long DNA molecules are complexed with proteins and coiled and folded into a structure known as a chromosome
- the cell has a fairly simple method of ensuring the proper inheritance of its DNA (and hence genes) when it splits into two daughter cells.

An excellent (though slightly out of date) external web site to visit for a description of the basic molecular genetics is The US Department of Energy's [Primer of Molecular Genetics](#). (The Department of Energy is a big mover in the Genome World.)

It takes many cell divisions to make a person from a single celled egg. In each of those divisions all the genes must be replicated and passed to each daughter cell. For the remainder of this lecture we will concentrate on the process of cell division **mitosis** and on the structures known as **chromosomes**.

The Cell Cycle

Within a tissue which is growing individual cells each go through a regular pattern of growth and division known as the cell cycle. There are many interesting features of the cell cycle from, for instance, a cancer biologist's viewpoint, concerned with the regulation of cell division. Cancer is, after all, a disease of unregulated cell division. From our perspective as geneticists we can take a simplified view of the process and concentrate solely on that part of the cycle which actually takes the least time, the mitotic division.

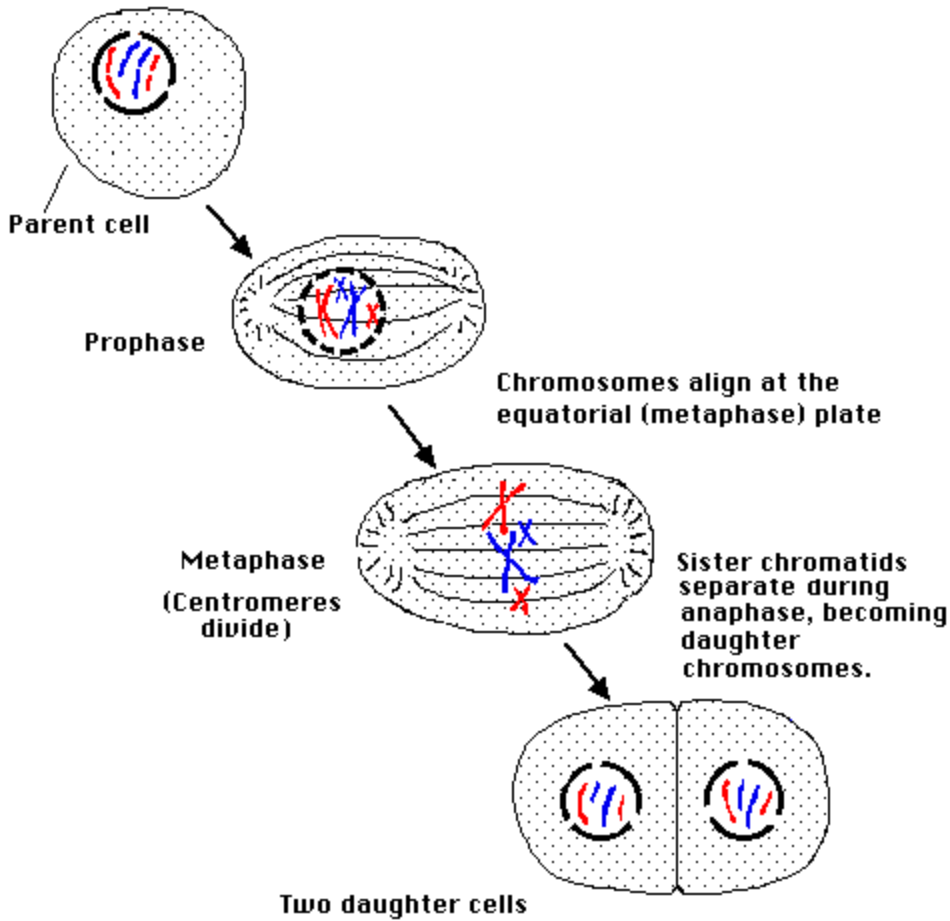
Interphase

Into this phase we can lump all the important events which we are not going to consider. During this phase, the cell nucleus is full of diffuse staining **chromatin**, DNA synthesis takes place (by semi conservative DNA replication), RNA and protein synthesis occurs and the cell doubles in size.

Mitosis

The act of mitosis can be conveniently divided into four phases.

Mitosis



prophase

- The replicated chromosomes become visible and can be seen to be comprised of two **sister chromatids** still joined together at the **centromere**,
- the nuclear membrane breaks down,
- **centrioles** move to opposite poles
- and the **spindle** forms between them.

metaphase

- The chromosomes attach to the spindle by their **kinetochores** (which are part of the centromere structure).
- The chromosomes are now at their most condensed.
- They line up at the equator of the spindle.

anaphase


- The centromeres divide
- the chromosomes are pulled apart to the two poles by contracting spindle fibres.

telophase

- Nuclear membranes reform
- the chromosomes decondense.
- The daughter cells return to interphase.

Chromosomes

Chromosomes serve to manoeuvre DNA through the difficulties of cell division where something like two metres of DNA has to be moved through a distance of only about 100 μm and be separated from a similar amount of DNA moving the other way.

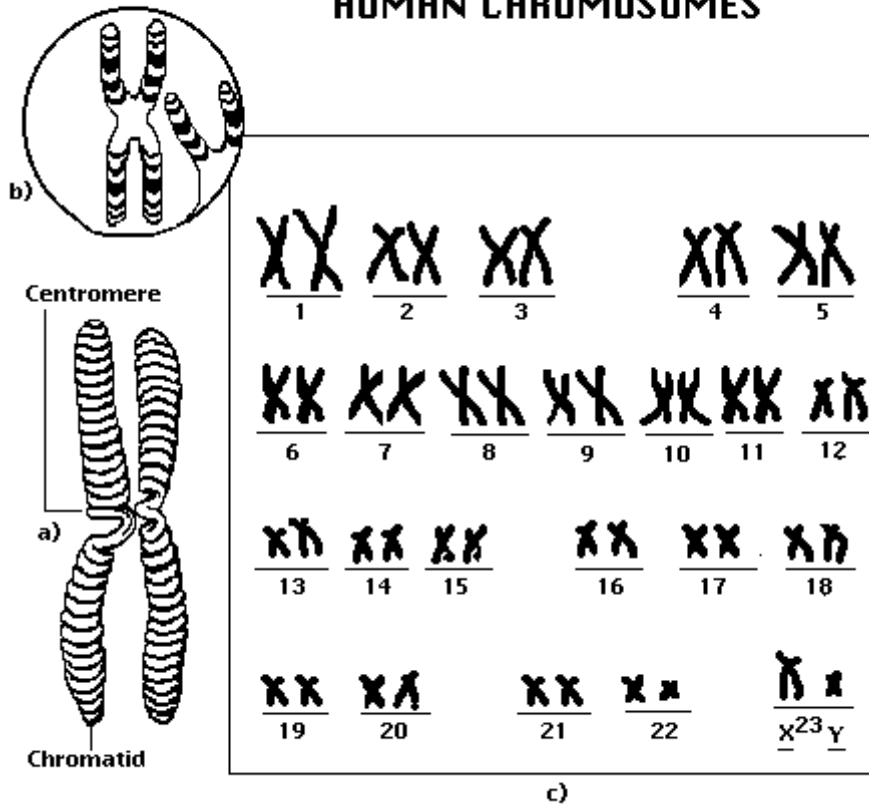


Humans have 23 pairs of chromosomes. Twenty two pairs, the **autosomes**, are the same in either sex and are numbered from 1 - 22 in order of diminishing size. One pair, the **sex chromosomes** are either a pair of X chromosomes (in females) or an X and the very much smaller Y chromosome (in males). One complete set of chromosomes i.e. autosomes 1-22 and a sex chromosome is known as a **haploid** set. Cells which contain two complete sets (i.e. most cells except for mature germ cells) are **diploid**. Each chromosome contains its own unique sequence of DNA. Consequently, when the chromosomal DNA and its associated histone and non-histone protein is at its most densely packed (i.e. at mitotic metaphase) it adopts a shape which is slightly different from any other chromosome.

Cytogeneticists study chromosomes microscopically. Cells are treated with a drug which prevents the spindle fibres forming. The chromosomes continue to be prepared for mitosis but because the spindle has not formed the division remains blocked and cells accumulate at metaphase. The cells can then be "fixed" i.e. treated with a chemical which denatures proteins causing the structures to be preserved, and then burst open on a microscope slide. After gentle treatment with a very small amount of proteinase, the chromosomes are stained. Each chromosome reveals a characteristic pattern of alternating dark and light bands which reflects in some way its underlying architecture. On the right for example is the **ideogram** which represents the characteristic banding pattern of chromosome 9.

The "**chromosome spread**" can be photographed, individual chromosomes cut out and paired and displayed as shown below.

HUMAN CHROMOSOMES



The sum of all the chromosome information is known as a **karyotype**.

All human chromosomes have two arms, the short arm is referred to as the **p** arm and the long arm as the **q**. The position of the **primary constriction** (another name for the centromere) defines whether the chromosome is **metacentric** (two substantial arms) or **acrocentric** (one very tiny arm and one which contains almost all the DNA). The acrocentric chromosomes are numbers 13, 14, 15, 21 and 22. At each end of the chromosome is a **telomere**, a structure designed to avoid problems with DNA replication right to the end of a linear molecule. Acrocentric chromosome short arms sometimes hardly seem to be attached, they can be linked via a short stalk known as a **secondary constriction**. The tiny short arm, bobbing about at a distance is known as a **satellite**.

Recommended reading

Topics

- Introduction to human genetics in medicine,
- chromosomes (the naming of parts) and
- mitosis.

Reading

The following are alternatives, read as many as you want - but at least one!

- Mange and Mange Chapters 1, 2 and 3 up to page 34
- Lewis Chapters 1, 2 (There's more in both these books than in the lecture but what the heck, it's all jolly interesting!)
- Mueller and Young Chapter 1 and Chapter 3 (pp 23 - 30 and p43)
- Jorde *et al.* Chapter 1 and Chapter 2 up to page 25.
- Thompson McInnes and Willard Chapters 1 and 2 (pp 13 - 23), chapter 3 (pp 31 - 40)

Chromosomes are well covered in

- Connor and Ferguson-Smith Chapter 4

The principals of Forensic medicine are well covered in:

- Molecular Medicine by R.J. Trent (2nd edition, Chapter 8 which would bear rereading after [lecture 5](#))

SAQs

1. What would happen to a chromosome with no telomeres?
2. If the DNA content of a haploid human germ cell is 3 pico grams, how much DNA is present in a normal diploid cell entering mitotic prophase?

Answers

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 2

Reproduction

Of the characteristics which distinguish the animate world from the inanimate, one in particular occupies the thoughts of medical students (particularly male medical students) more than any other, the need to reproduce. This is of course the core of the science of genetics.

The distinction between Somatic and Germ cells

Our bodies have evolved to carry out that one function as successfully as possible. However, most cells will do so only in a supporting role, they are not themselves destined to be transmitted. As geneticists, we draw a distinction between the **germ cells** which provide the continuity of life from one generation to the next, and the **somatic cells** which are all the rest. When a sperm fertilises an egg to create a **zygote**, the embryo begins to develop. Initially all the cells are capable of giving rise to any part of the embryo or its extraembryonic tissues. However, after about 6 or 7 divisions, some cells have become irreversibly programmed to give rise only to a subset of possible cell or tissue types and this process of irreversible differentiation continues until all the organs have been constructed. During this process, a small number of progenitor germ cells are sequestered until rudimentary gonads (testes or ovaries) have formed when the germ cells migrate into them. At this stage germ cells are neither sperm nor egg cells, they are precursors, spermatogonia or oogonia. In their own way they are every bit as differentiated as any other (somatic) cell of the body. There is no way that a spermatogonium will ever be able to differentiate into a liver cell for instance. However, the germ cells contain the potential to be transmitted to the next generation and contribute one half of the DNA of the next individual.

Testis and Ovary

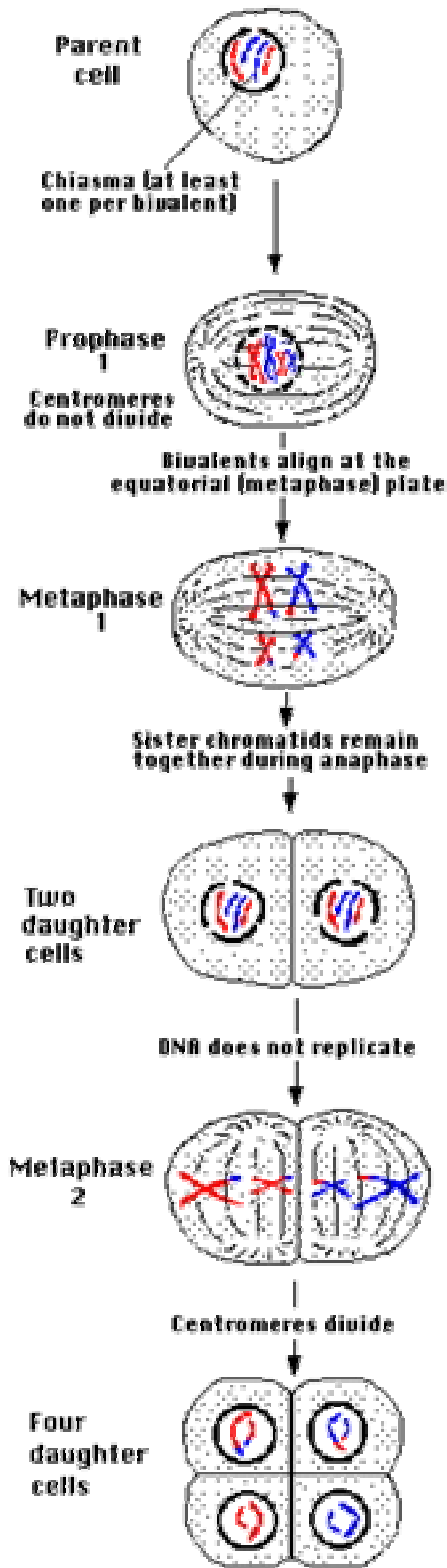
Under the influence of the surrounding somatic component of the testis or ovary the germ cells multiply by successive mitotic divisions. In the male this process continues throughout life but in females it stops prior to birth. In other courses you will learn about the hormonal influences acting on the cells, the female reproductive cycle etc., but this need not concern us here where we simply need to consider what happens to the chromosomes to ensure that only (and exactly) half of the genetic material is transmitted to the next generation via a sperm or an egg.

Meiosis

The reduction in chromosome number from diploid to haploid is accomplished by the specialised cell divisions of **meiosis**. Cells prepare for meiosis by replication of their genetic material as for mitosis. However, instead of a single division as in mitosis,

Meiosis

meiosis consists of two consecutive divisions as shown in the diagram:



- The first meiotic division

- **Prophase I**

- The chromosomes first become visible as thin threads within the nucleus at the stage called **leptotene**.
 - In the next phase, **zygotene**, the homologous pairs of chromosomes become closely associated along their lengths by a process called **synapsis** to form a structure comprised of two chromosomes, a **bivalent**.
 - The bivalents shorten and thicken throughout the next stage **pachytene**, synapsis is complete and the bivalents are held together throughout their length by a structure known as the [synaptonemal complex](#).
 - Crossing over - We cannot see it but at this point portions of one chromatid of one homologue may become exchanged with the corresponding region from a chromatid of the other homologue.
 - The chromosomes continue to condense and the synaptonemal complex breaks down. The homologous chromosomes appear to repel each other and remain held together only at **chiasmata**, the points where crossing over have occurred, and at the centromere. This phase is named **diplotene**. There is always at least one chiasma per chromosome arm. Some of the longer arms will have two or even three.
 - The final stage of prophase is **diakinesis** in which the nuclear membrane breaks down and the chiasmata slip to the ends of the chromosome arms.

- **Metaphase I** - the bivalents align on the spindle.

- **Anaphase I** - the first division begins, each bivalent divides so that one chromosome moves to one pole and the second chromosome moves to the other. (N.B. each chromosome is still comprised of two chromatids.)



- **Telophase I** - The nuclear membranes reform and the cells complete division (equally in the case of spermatogenesis but very unequally into egg and first polar body in the case of oogenesis).
- The second meiotic division
 - This follows immediately on the heels of the first division with no intervening round of DNA synthesis. **Prophase II, Metaphase II** and **Anaphase II** resemble mitosis but with only a haploid chromosome number. The second meiotic division in the egg is not completed until fertilisation and is again very unequal giving the mature egg and a small second polar body.

Mendelian Inheritance

Gregor Mendel is famous today but was relatively unknown outside Czechoslovakia in his lifetime. He was the first scientist to deduce clear and rational laws which could explain the process of inheritance. Unfortunately, few medical students are interested in the genetics of peas! However, it turns out that the rules which Mendel deduced from studies of peas are equally applicable to human inheritance and it is convenient to follow his train of logic beginning with characteristics determined by a single gene and moving on to the complications introduced by multiple genes. If you are interested to read a translation of his original paper then [click here](#).

Single gene

Mendel began by collecting varieties of pea which differed from each other in clearly defined ways. The pea flower has anthers and a stamen which are very close together. It will self fertilise in normal circumstances. It is possible to remove the anthers before they are ready to produce pollen and to cross fertilise the pea plant by bringing pollen from another plant on a paint-brush. Mendel allowed his plants to self fertilise for a number of generations until he was certain that they were **true breeding**, i.e. that the offspring always resembled the parent for the characteristics under consideration. Then he began his experiments.

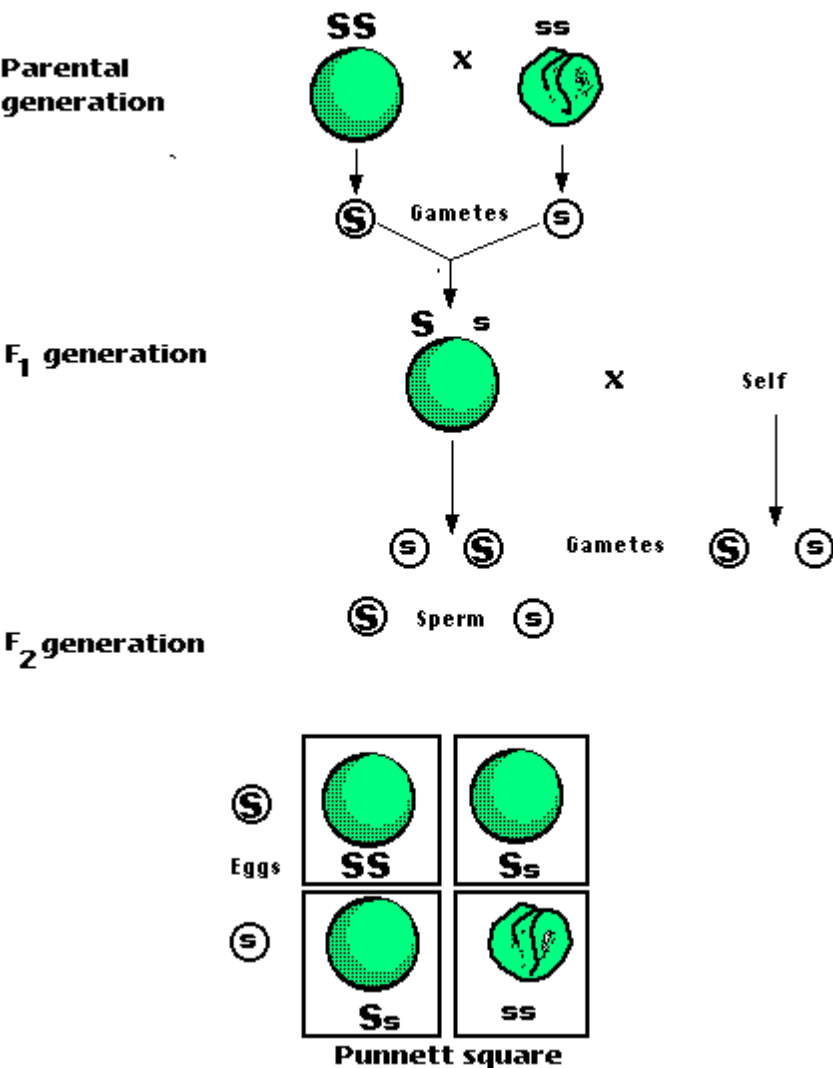
Characteristics studied by Mendel

Characteristic	Dominant allele
Axial or terminal flowers	axial flowers
round or wrinkled seeds	round seeds
yellow or green seed interiors	yellow interiors
violet or white petals	violet petals
tall or dwarf plants	tall plants

fat or shrunken ripe seed pods	fat pods
green or yellow unripe pods	green pods

wrinkled seed strain, and wisely he intercrossed them in each direction (i.e. "round" pollen onto "wrinkled" stigma and also "wrinkled" pollen onto "round" stigma).

Mendel: Monohybrid cross



First he took a pair of **parental** strains differing at a single characteristic, for instance, a round seed strain and a wrinkled seed strain, and wisely he intercrossed them in each direction (i.e. "round" pollen onto "wrinkled" stigma and also "wrinkled" pollen onto "round" stigma). In fact it made no difference that he tried each parent as the male or the female but it might have done. The plants resulting from this mating, the **first filial generation** or **F₁**, were all examined. All had the appearance of one of the parental strains, in this case, the round one. Mendel defined the visible characteristic as the **dominant** one. The F₁ plants were then allowed to self fertilise to produce a second filial generation or **F₂**.

A surprising result occurred in the F₂ generation, **wrinkled seeded plants reappeared!** Mendel counted more than enough F₂ plants to be able to say that the "wrinkled seed" plants were one quarter of the offspring. The characteristic which had disappeared and had now reappeared Mendel described as **recessive**. Mendel went on to allow each F₂ plant to self fertilise. He found that the wrinkled seed plants (and their offspring) were true breeding like the original wrinkled seed parental strain. One third of the round seeded plants were also true breeding (as were their offspring). The remaining two thirds of the round seeded plants behaved exactly like the F₁.

The Law of Segregation

The phenomenon could be explained if it were assumed that each plant had two copies of the factor influencing the trait. We now call the factor responsible a **gene**, we say that more than one form of the gene can exist and we call those alternative forms **alleles**. Mendel explained the results by suggesting that each plant contained two alleles which *did not blend together* but which remained unchanged. In the next generation the plants

passed one or other allele at random into a gamete to be combined with a gamete from the other parent. The non-blending followed by separation into the next generation is the *Rule of Segregation*.

We can distinguish in the above cross two sorts of individual, true breeding individuals with both alleles the same, and individuals in which the two alleles continue to segregate. We call the former **homozygotes** and the latter **heterozygotes** or **carriers**. The gene which we are considering is said to be **homozygous** or **heterozygous** respectively. We can also distinguish between an individual's outward appearance, its **phenotype**, and its inward genetic constitution, its **genotype**.

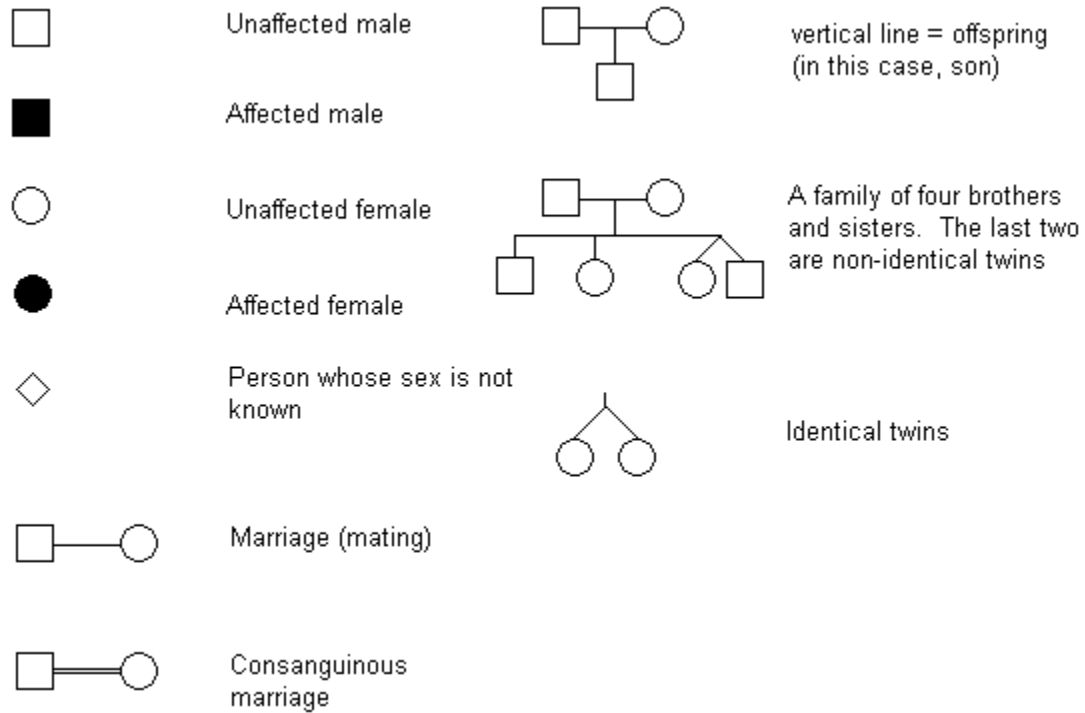
A convenient method of predicting the relative ratios of the progeny in any cross is by means of a **Punnett Square** an example of which is shown in the above diagram. The gametes from each parent are placed on the margins and at the intersections of the rows and columns are written the resulting offsprings' genotypes and, if we wish, their phenotypes.

The Relationship of Genes to Chromosomes

The alternation of two genes in individuals and one gene in gametes is of course reminiscent of the behaviour of chromosomes, diploid in most tissues but haploid in sperm and egg. We now know the reason why this is so, genes are carried on chromosomes, encoded in the DNA.

Human pedigrees

Before we consider human Mendelian inheritance it is convenient to consider the symbols used to draw pedigrees.



Generations are numbered from the top of the pedigree in uppercase Roman numerals, I, II, III etc. Individuals in each generation are numbered from the left in arab numerals as subscripts, III₁, III₂, III₃ etc.

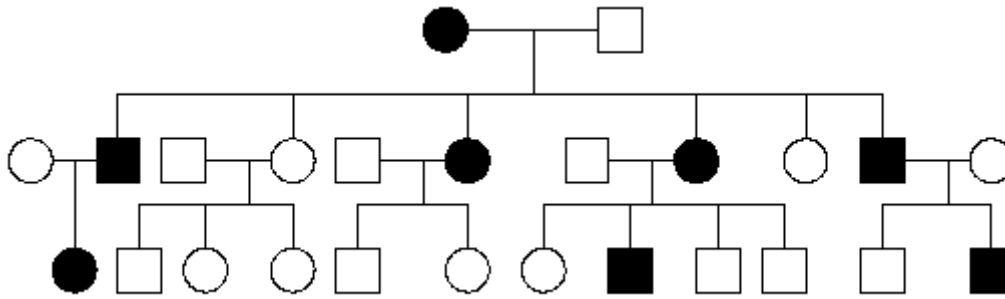
Modes of inheritance

Most human genes are inherited in a Mendelian manner. We are usually unaware of their existence unless a variant form is present in the population which causes an abnormal (or at least different) phenotype. We can follow the inheritance of the abnormal phenotype and deduce whether the variant allele is dominant or recessive.

autosomal dominant

A dominant condition is transmitted in unbroken descent from each generation to the next. Most matings will be of the form M/m x m/m, i.e. heterozygote to homozygous recessive. We would therefore expect every child of such a mating to have a 50% chance of receiving the mutant gene and thus of being affected. A typical pedigree might look

like this:

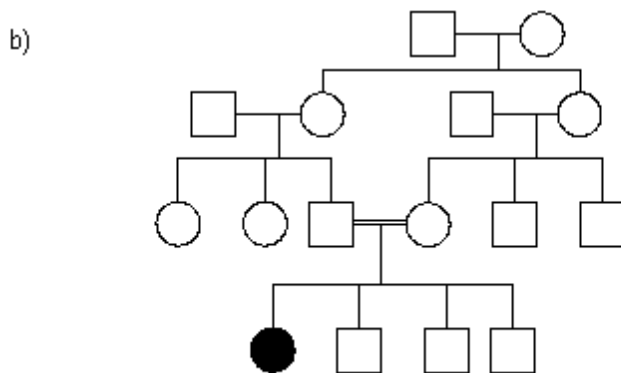
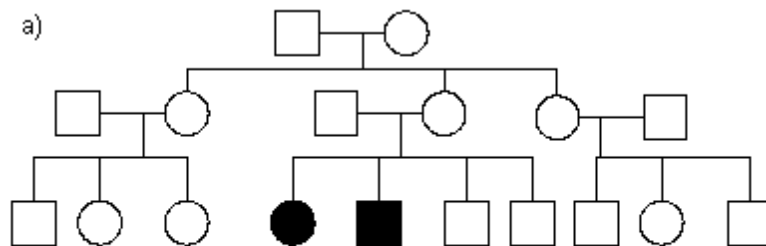


Examples of autosomal dominant conditions include [Tuberous sclerosis](#), [neurofibromatosis](#) and many other cancer causing mutations such as [retinoblastoma](#)

autosomal recessive

A recessive trait will only manifest itself when homozygous. If it is a severe condition it will be unlikely that homozygotes will live to reproduce and thus most occurrences of the condition will be in matings between two heterozygotes (or **carriers**). An autosomal recessive condition may be transmitted through a long line of carriers before, by ill chance two carriers mate. Then there will be a $\frac{1}{4}$ chance that any child will be affected. The pedigree will therefore often only have one 'sibship' with affected members.

a) A 'typical' autosomal recessive pedigree, and b) an autosomal pedigree with **inbreeding**:



If the parents are related to each other, perhaps by being cousins, there is an increased risk that any gene present in a child may have two alleles **identical by descent**. The degree of risk that both alleles of a pair in a person are descended from the same recent common ancestor is the degree of **inbreeding** of the person. Let us examine b) in the figure above. Considering any child of a first cousin mating, we can trace through the pedigree the chance that the other allele is the same by common descent. Let us consider any child of generation IV, any gene which came from the father, III₃ had a half chance of having come from grandmother II₂, a further half chance of being also present in her sister, grandmother II₄ a further half a chance of having been passed to mother III₄ and finally a half chance of being transmitted into the same child we started from. A total risk of

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = 1/16$$

This figure, which can be thought of as either

- the chance that both maternal and paternal alleles at one locus are identical by descent

or

- the proportion of all the individual's genes that are homozygous because of identity by common descent,

is known as **the coefficient of inbreeding** and is usually given the symbol **F**.

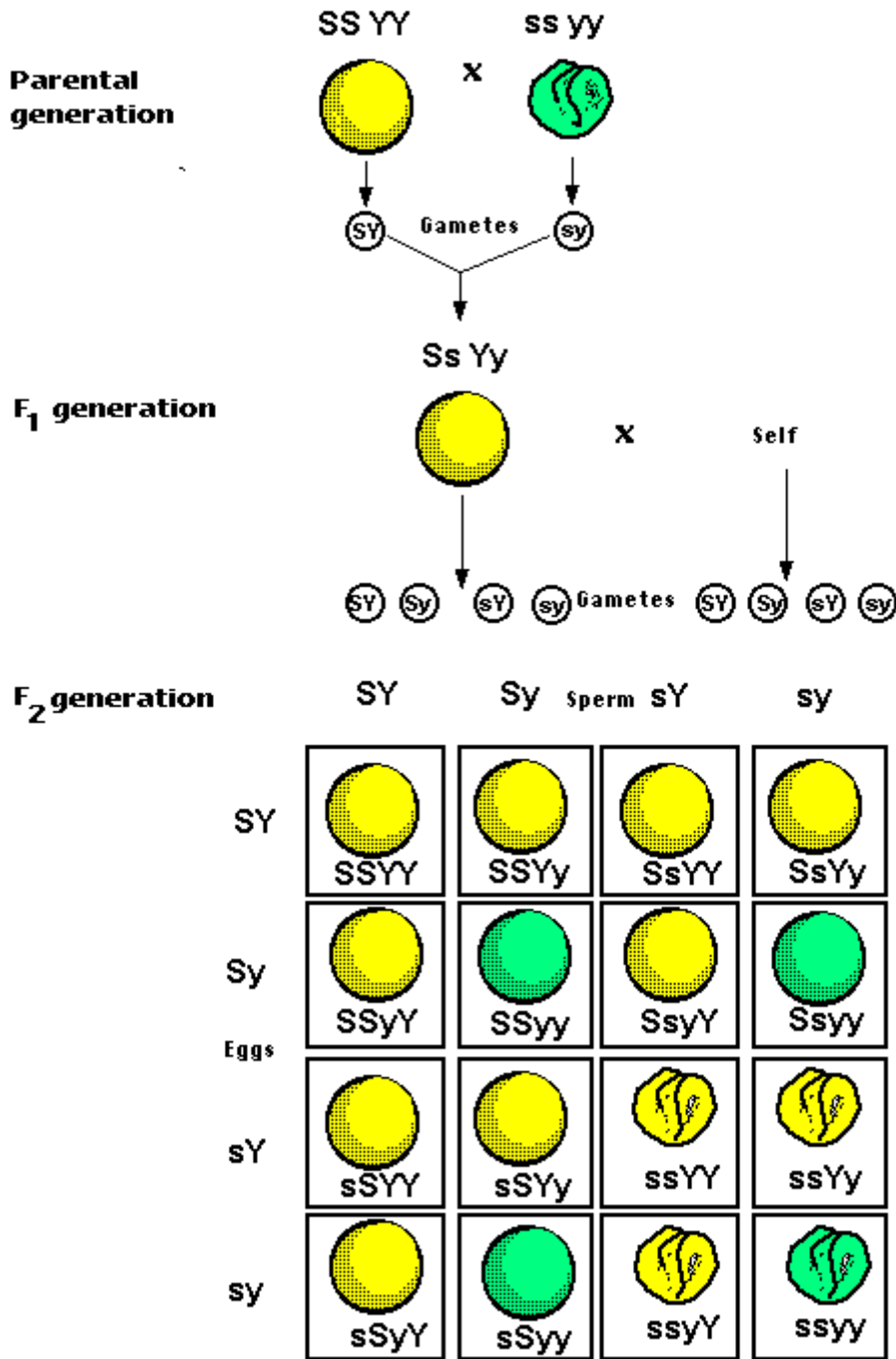
Two genes

Mendel went on to consider what happened if he crossed plants together which differed with respect to more than one character trait (a so-called dihybrid cross).

Independent assortment

What Mendel discovered can be put very simply, the two characteristics behaved completely independently of each other. He called this the *rule of independent assortment*. Here is an example of a cross between a strain which produced smooth yellow seeds and one with wrinkled green seeds.

Mendel: dihybrid cross



Punnett square

The classes of offspring in the F₂ occur in the well known 9:3:3:1 ratio, 9 Yellow smooth : 3 green smooth : 3 yellow wrinkled : 1 green wrinkled. This ratio is the result of the two genes behaving completely independently of each other in the cross.

Using probability

The Punnett square is a fine method of working out straightforward events. However, not all life is straightforward! Most of you have some background in mathematics and will have covered elementary probability. For those who have not, I strongly recommend reading Mange and Mange pp50-57.

The probability of an event is the chance that it will happen. The probability of tossing a coin to land heads up is just slightly less than $\frac{1}{2}$ (I did once have a coin stick on its edge in the mud and my unsporting opponent, instead of allowing the throw to be taken again like a gentleman, insisted that by calling "heads" I had lost the toss!).

- The probability of an impossible event is 0,
- the probability of a certain event is 1.
- If the probability of event x is p then the probability of 'not x' is 1-p.
- The probability of two independent events occurring is the product of their two individual probabilities. So, for example,
 - in the cross above, in the F2 the probability of a wrinkled seed is $\frac{1}{4}$, the probability of a green seed is also $\frac{1}{4}$ and the probability of being both green and wrinkled is therefore $\frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$.
 - The probability of being not wrinkled (i.e. smooth) is $1 - \frac{1}{4} = \frac{3}{4}$. The probability of being both smooth and green is therefore $\frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$ and so on.
 - In the [example above](#) about the coefficient of inbreeding of children from first cousin marriages, we considered a number of probabilities of $\frac{1}{2}$ which we multiplied together to reach a final probability of $\frac{1}{16}$ that any gene was homozygous by descent.

Recommended reading

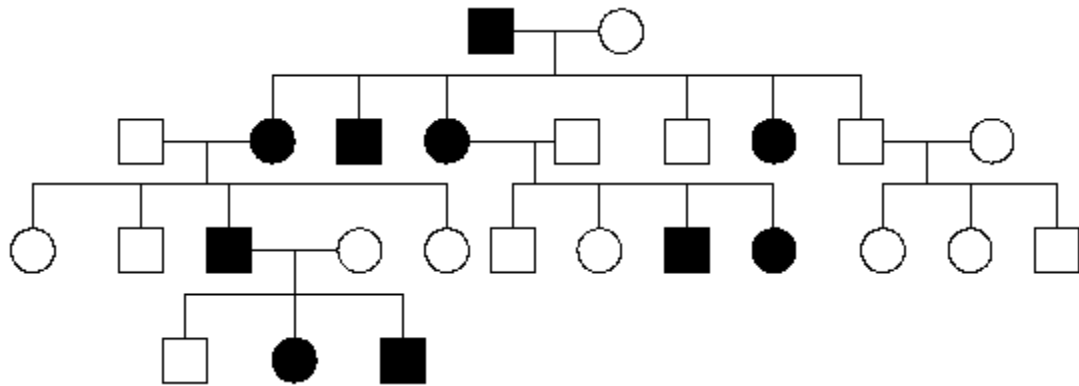
The topics include Meiosis, Mendel, pedigrees and autosomal dominant and recessive inheritance.

- Mange and Mange Chapters 3 (pp 34 - 43), 4 (pp 45 - 59), 5 (pp 67 - 81)
 - Lewis Chapters 3 (primarily pp42-53 but the rest of the chapter is also interesting) and 4
 - Mueller and Young Chapters 1 (pp 3 - 4), 3 (pp31 - 33), 6 (pp 77 - 82) (This book is rather too brief on this area)
 - Jorde *et al.* Chapter 2 pp 25 - 28, chapter 4 (all of it)
 - Thompson McInnes and Willard Chapters 2 (pp 23 - 30), 4 (pp 53 - 71) (Also rather brief)
 - Connor and Ferguson-Smith Chapters 5 (pp 46 - 50), 7
-

SAQs

The relevant questions at the end of chapters 3 and 4 in Mange and Mange and chapter 4 in Lewis are worth trying. In addition:

1. a) How many chromosomes are present in a normal human premeiotic germ cell?
b) How many chromosomes are present in a human gamete?
c) At what stage during gametogenesis did the number change?
2. The following pedigree could be the result either of the segregation of an autosomal dominant condition or of an autosomal recessive. In the former case what is the risk for individual III₆ of having a child affected with this condition. In the latter case, who in the pedigree is an obligate carrier? And which other members of the pedigree are at risk of being carriers. Write down their risks.



1. Coat colours in cats are caused by the interactions of several genes, some affecting the ability to produce the black pigment eumelanin, some the red pigment phaeomelanin and some the patterns of deposition of pigment. A brown tabby cat is crossed to a solid cinnamon cat. All the kittens resulting from this cross are brown tabbies. When adult these F1 were allowed to intercross freely with each other to produce an F2. In the resulting exploding cat population four classes of cat could be seen, brown tabbies, cinnamon tabbies, solid blacks and solid cinnamons in the relative proportions 9:3:3:1. Account for these observations.

[Answers](#)

SAQ answers

Lecture 1

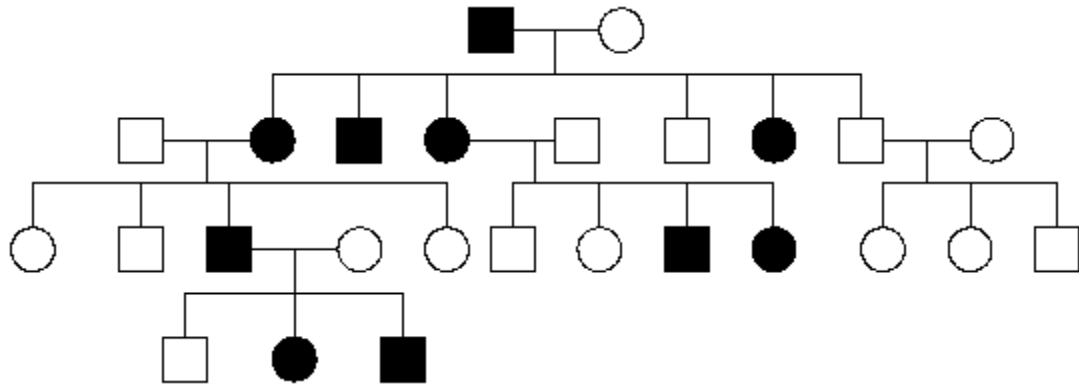
1. The trailing strand would be unable to be replicated all the way to the end (where would the RNA primer be placed to prime the Okazaki fragment synthesis?) and would shorten every cell generation. The leading strand would be replicated to the end of the trailing strand and so would also shorten (but one cell division cycle

behind the trailing strand. Eventually the chromosome would be lost. It has been suggested that the syndrome [progeria](#) which leads to premature aging, (the affected children die of diseases of old age in their teens) is caused by a mutation in the gene coding for the enzyme telomerase. The similar syndrome, [Werner's syndrome](#) has been proved to be caused by mutations in a DNA helicase gene (involved in DNA unwinding).

2. Four times that amount, or 12 picograms. (Diploid cells with replicated chromosomes)

Lecture 2

1. a) 46 chromosomes are present in a premeiotic germ cell. When meiosis begins each has replicated and when it condenses in division can be seen to be composed of two chromatids.
 b) Each human gamete contains 23 chromosomes, each at that point composed of a single chromatid.
 c) The number of chromosomes changed as a result of the first meiotic division in which homologous chromosomes go to opposite poles of the spindle.



2.
 - If dominant, then the chance that III₆ will have affected children is zero.
 - If recessive then obligate carriers are:
 - I₂,
 - II₁, II₅, II₆, II₈,
 - III₁, III₂, III₄, III₅, III₆, III₇,
 - IV₁

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 3

Some apparent exceptions to Mendelian rules

There are many reasons why the ratios of offspring phenotypic classes may depart (or seem to depart) from a normal Mendelian ratio. For instance:

- **Lethal alleles**

Many so called dominant mutations are in fact *semidominant*, the phenotype of the homozygote is more extreme than the phenotype of the heterozygote. For instance the gene T (Danforth's short tail) in mice. The normal allele of this gene is expressed in the embryo. T/+ mice develop a short tail but T/T homozygotes die as early embryos. Laboratory stocks are maintained by crossing heterozygotes,

$$\begin{array}{ccc} T/+ & \times & T/+ \\ & | & \\ & \downarrow & \\ T/T & T/+ & +/+ \\ 1 & : & 2 & : & 1 & \text{ratio at conception} \\ 0 & : & 2 & : & 1 & \text{ratio at birth} \end{array}$$

- **Incomplete or semi- dominance**

Incomplete dominance may lead to a distortion of the apparent ratios or to the creation of unexpected classes of offspring. A human example is [Familial Hypercholesterolemia](#) (FH). Here there are three phenotypes: +/+ = normal, +/- = death as young adult, -/- = death in childhood. The gene responsible codes for the liver receptor for cholesterol. The number of receptors is directly related to the number of active genes. If the number of receptors is lowered the level of cholesterol in the blood is elevated and the risk of coronary artery disease is raised.

- **Codominance**

If two or more alleles can each be distinguished in the phenotype in the presence of the other they are said to be codominant. An example is seen in the ABO blood group where the **A and B alleles are codominant**.

The *ABO* gene codes for a glycosyl-transferase which modifies the H antigen on the surface of red blood cells. The A form adds N-acetylgalactosamine, the B form adds D-galactose forming the A and B antigens respectively. The O allele

has a frameshift mutation in the gene and thus produces a truncated and inactive product which cannot modify H. A phenotype people have natural antibodies to B antigen in their serum and vice versa. O phenotype individuals have antibodies directed against both A and B. AB individuals have no antibodies against either A or B antigens.

ABO genotypes and phenotypes			
Genotype	Phenotype	red cell antigens	serum antibodies
AA	A	A	anti-B
AO	A	A	anti-B
BB	B	B	anti-A
BO	B	B	anti-A
AB	AB	A and B	neither
OO	O	neither	anti-A and anti-B

Codominance is the normal case for alleles which are revealed by direct study of an individual's DNA, see for example lecture 9, "[restriction fragment length polymorphism](#)" (RFLP).

- **Silent alleles**

In a multiple allele system, it is sometimes not obvious that a silent allele exists. This can give confusing results. Consider for example:

$$\begin{array}{r}
 \text{A/A} \quad \times \quad \text{A/B} \quad (\text{phenotype A crossed to phenotype AB}) \\
 | \\
 \text{V} \\
 \text{A/A} \quad : \quad \text{A/B} \\
 1 \quad : \quad 1
 \end{array}$$

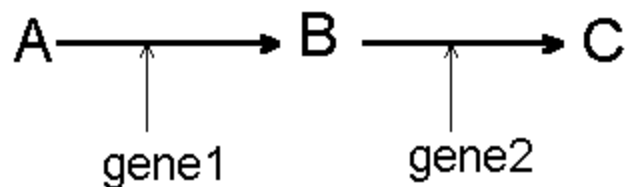
and compare with

$$\begin{array}{r}
 \text{A/O} \quad \times \quad \text{A/B} \quad (\text{phenotype A crossed to phenotype AB}) \\
 | \\
 \text{V} \\
 \text{A/A} \quad : \quad \text{A/O} \quad : \quad \text{A/B} \quad : \quad \text{B/O} \\
 1 \quad : \quad 1 \quad : \quad 1 \quad : \quad 1
 \end{array}$$

It would be important not to lump together these two different sorts of crosses but when there are only small numbers of offspring (which is the case in most human matings) some offspring classes may not be represented in a family and it may not be obvious which type of mating you are examining.

- **Epistasis**

This occurs where the action of one gene masks the effects of another making it impossible to tell the genotype at the second gene. The cause might be that both genes produce enzymes which act in the same biochemical pathway.



If the product of gene1 is not present because the individual is homozygous for a mutation, then it will not be possible to tell what the genotype is at gene2. The Bombay phenotype in humans is caused by an absence of the [H antigen](#) so that the ABO phenotype will be O no matter what the ABO genotype. A soap opera example of this is given on page 89 (figure 5.4) in Lewis.

- **Pleiotropy**

Mutations in one gene may have many possible effects. Problems in tracing the passage of a mutant allele through a pedigree can arise when different members of a family express a different subset of the symptoms. In the case of [Tuberous sclerosis](#), an autosomal dominant condition affecting about 1 in 6000 people in the UK, symptoms can include any subset of:

- small depigmented patches on the skin
- a disfiguring facial rash
- kidney cysts
- heart tumours
- mild to severe mental subnormality
- autism

among others.

Lewis gives the interesting example of [Tourette syndrome](#) which may cause strange behavioural problems.

Pleiotropy can occur whenever a gene product is required in more than one tissue or organ.

- **Genetic heterogeneity**

This is the term used to describe a condition which may be caused by mutations in more than one gene. Tuberous sclerosis again provides a good example of this, the identical disease is produced by mutations in either of two unrelated genes, *TSC1* on chromosome 9 or *TSC2* on chromosome 16. In such cases, presumably both genes act at different points in the same biochemical or regulatory pathway. Or perhaps one provides a ligand and one a receptor.

- **variable expressivity**

The degree to which a disease may manifest itself can be very variable and, once again, tuberous sclerosis provides a good example. Some individuals scarcely have any symptoms at all whereas others are severely affected. Sometimes very mild symptoms may be overlooked and then a person may be wrongly classified as non-affected. Clearly this could have profound implications for genetic counselling.

- **Incomplete Penetrance**

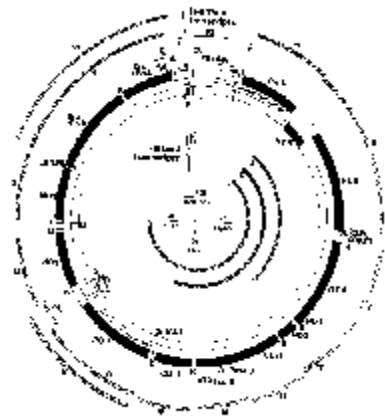
This is an extreme case of a low level of expressivity. Some individuals who logically ought to show symptoms because of their genotype do not. In such cases even the most careful clinical examination has revealed no symptoms and a person may be misclassified until suddenly he or she transmits the gene to a child who is then affected.

One benefit of gene cloning is that within any family in which a mutant gene is known to be present, when the gene is known, the mutation can be discovered and the genotype of individuals can be directly measured from their DNA. [See lecture 5](#). In this way diagnosis and counselling problems caused by non-penetrance can be avoided. The degree of penetrance can be estimated. If a mutation is 20% penetrant then 20% of persons who have the mutant genotype will display the mutant phenotype, etc.

- **Anticipation**

In some diseases it can appear that the symptoms get progressively worse every generation. One such disease is the autosomal dominant condition [myotonic dystrophy](#). This disease, which is characterized by a number of symptoms such as myotonia, muscular dystrophy, cataracts, hypogonadism, frontal balding and ECG changes, is usually caused by the expansion of a trinucleotide repeat in the 3'untranslated region of a gene on chromosome 19. The severity of the disease is roughly correlated with the number of copies of the trinucleotide repeat unit.

Number of CTG repeats	phenotype
5	normal



19 - 30	"pre-mutant"
50 - 100	mildly affected
2,000 or more	severely affected
myotonic dystrophy	

The "premutant" individuals have a small expansion of the number of trinucleotide repeats which is insufficient to cause any clinical effect in itself but it allows much greater expansions to occur during the mitotic divisions which precede gametogenesis. Mildly affected individuals can again have gametes in which a second round of expansion has occurred.

Germline Mosaicism

If a new mutation occurs in one germ cell precursor out of the many non-mutant precursors, its descendent germ cells, being diluted by the many non-mutant germ cells also present, will not produce mutant offspring in the expected Mendelian numbers.

- **Phenocopies**

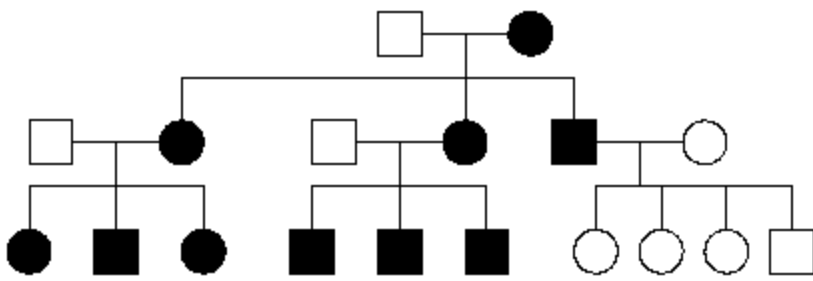
An environmentally caused trait may mimic a genetic trait, for instance a heat shock delivered to *Drosophila* pupae may cause a variety of defects which mimic those caused by mutations in genes affecting wing or leg development. In humans, the drug thalidomide taken during pregnancy caused phenocopies of the rare genetic disease [phocomelia](#), children were born with severe limb defects.

Real exceptions

- **mitochondrial inheritance**

The human mitochondrion has a small circular genome of 16,569 bp which is remarkably crowded. It is inherited only through the egg, sperm mitochondria never contribute to the zygote population of mitochondria. There are relatively few human genetic diseases caused by mitochondrial mutations but, because of their maternal transmission, they have a very distinctive pattern of inheritance.

A mitochondrial inheritance pedigree



All the children of an affected female but none of the children of an affected male will inherit the disease.

- **uniparental disomy**

Although it is not possible to make a viable human embryo with two complete haploid sets of chromosomes from the same sex parent it is sometimes possible that both copies of a single chromosome may be inherited from the same parent (along with no copies of the corresponding chromosome from the other parent.) Rare cases of [cystic fibrosis](#) (a common autosomal recessive disease) have occurred in which one parent was a heterozygous carrier of the disease but the second parent had two wild type alleles. The child had received two copies of the mutant chromosome 7 from the carrier parent and no chromosome 7 from the unaffected parent.

- **linkage**

When two genes are close together on the same chromosome they tend to be inherited together because of the mechanics of chromosome segregation at meiosis. This means that they do not obey the law of independent assortment. The further apart the genes are the more opportunity there will be for a chiasma to occur between them. When they get so far apart that there is always a chiasma between them then they are inherited independently. The frequency with which the genes are separated at meiosis can be measured and is the basis for the construction of genetic linkage maps (of which, more in lectures [8](#) & [9](#)).

Recommended reading

The topics include both apparent and real exceptions to Mendelian Inheritance.

- Mange and Mange Chapter 10 (pp 191 - 202)
- Lewis Chapter 5
- Mueller and Young Chapter 6
- Jorde *et al.* Chapter 4 (pp 69 - 84)
- Thompson McInnes and Willard Chapter 4 (especially pp 83 - 94)

SAQs

Questions 1, 2, 4, 5, 6, 7, 8, and 10 on page 216 of Mange and Mange and the questions at the end of chapter 5 in Lewis are worth trying. In addition:

1. Familial hypertrophic cardiomyopathy (FHC) frequently appears to be inherited as an autosomal dominant condition. The characteristic symptom is ventricular hypertrophy which can lead to left ventricular outflow obstruction and to cardiac failure. It is sometimes a cause of sudden death in apparently healthy young individuals and is a common postmortem finding in young sportsmen and women who die suddenly while competing. Mutations in a number of genes have been implicated in its etiology. These include beta myosin heavy chain gene on chromosome 14, the cardiac troponin gene on chromosome 1, alpha-tropomyosin on chromosome 15, myosin binding protein C and at least one more as yet undiscovered locus. The common factor is that all four discovered genes are components of the sarcomere. Many cases have no previous known family history. Most mutations are missense which means that they result from perhaps just a single nucleotide substitution. Some mutations are invariably of severe effect, others are much less likely to cause sudden death and they are sometimes passed on without having caused any symptom. What are the relevant terms from today's lecture to describe the problems of the genetics of FHC?
2. A male cinnamon tabby cat is crossed to a brown tabby cat. All the offspring are brown tabbies. The same male is crossed to another brown tabby. This time, although half the kittens are brown tabbies, half are chocolate tabbies. What problem above does this illustrate?

Genetics Lecture 4

Sex determination

Surprisingly, it is only in the last 50 years that we have begun to understand the nature of the biological events which determine our sex, (and for that matter, why we bother with sex at all and why two sexes are better than three or more). It is not so long ago that women were blamed if they failed to produce a son for their husband and clearly it was thought that the power of sex determination lay within the body of the woman. During this century the chromosomal basis of human sex determination has been demonstrated and in the last few years some of the genes responsible have been identified.

The sexual identity of an individual is determined at several levels, chromosomal sex, gonadal sex, somatic sex and sexual orientation.

sex chromosomes

The chromosomal basis of sex determination in humans was recognized when metaphase chromosomes from dividing male and female cells could be studied and counted. The [normal karyotype](#) contains 46 chromosomes including either two X chromosomes (46XX, females) or one X chromosome and one Y chromosome (46XY, males).

Individuals with [45X](#) or [47XXX](#) karyotypes are female, individuals with [47XXY](#) karyotype are male. Therefore it can be deduced that the Y chromosome is sex determining

sex determination

Experiments involving the removal of the embryonic gonad have revealed that in mammals, no matter what the chromosomal sex of the somatic cells, the body will develop as a female unless a male gonad is present to secrete mullerian inhibiting substance and testosterone. This can be partially mimicked in the genetic condition [testicular feminisation](#) in which the gene coding for the androgen receptor is not expressed so that, although the testis in an XY individual secretes testosterone, the somatic tissues are unable to respond to it. Consequently the individual's body develops as a woman but with internal testes instead of ovaries. In 1990, a Y encoded gene [SRY](#) was discovered which (at least in mice) is able to transform the sex of an XX embryo from female to male. Individuals with mutations in this gene develop as females despite having an XY chromosomal constitution. About one male in 10,000 does not appear to have an Y chromosome but instead has two X chromosomes. These XX males can frequently be shown to have inherited from their fathers an X chromosome onto which a little bit of the Y chromosome carrying *SRY* has been transferred by an "illegitimate" cross over. XX males are entirely normal except that they are infertile and their heights are in the normal female range rather than the male.

In other organisms things happen differently. Both *Drosophila* and the nematode *Caenorhabditis elegans* use a mechanism in which each cell measures the relative number of X chromosomes compared to the number of autosomes. However, the genes involved in the counting process and in its interpretation do not seem to be related in the two species.

Even within vertebrates there are a variety of sex determination mechanisms.

- Birds use a chromosomal system, however unlike the mammalian system, males are *homogametic ZZ* and females are *heterogametic ZW*.
- Alligators are chromosomally the same in both sexes, they determine sex by the temperature at which embryos are allowed to develop. If warm then males are formed and if cool then females are formed.

sexual identity

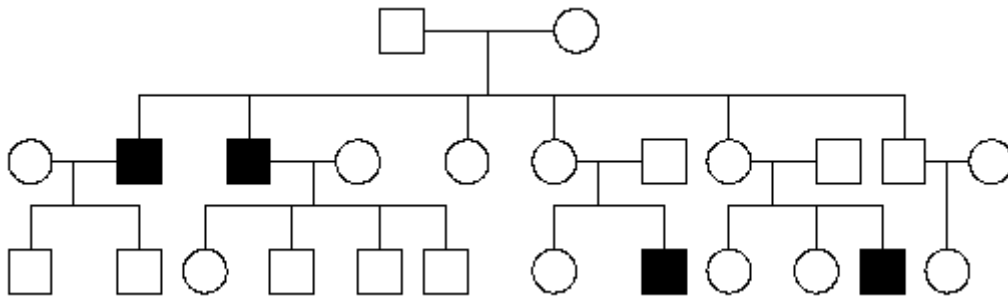
It is a much debated question as to whether our own sense of sexual identity is genetic or environmental in origin. Like most complex phenotypes it probably can be either but is usually both! Recent controversy centred on the [results obtained by Hamer](#) who showed evidence for one such gene, a recessive gene on the X chromosome, which when mutant may make its bearer, if male, more likely to be homosexual.

Sex linkage

sex linked recessive genes

Genes carried on the X chromosome have a distinctive pattern of inheritance. Because males are *hemizygous*, i.e. they have only one copy of the X chromosome, and because the Y chromosome carries very few genes (though those which it carries are often homologous to X linked genes) then recessive mutations manifest themselves in the phenotype of males. If the mutant gene is lethal (such as [Duchenne Muscular Dystrophy](#)) then it takes an unusual event to produce an affected female.

An X linked pedigree



A typical pedigree will show clusters of affected males (each brother will have a 50% chance of being affected) connected through unaffected carrier females. There will be *no cases of direct male to male transmission* because males transmit their X chromosomes to their daughters and not to their sons.

The following passage is quoted from [The history of haemophilia by Dr. P.L.F. Giangrande](#)

The story of Queen Victoria

Haemophilia is sometimes referred to as the Royal disease. Queen Victoria had no ancestors with the condition but soon after the birth of her eighth child, Leopold, in 1853 it became evident that he had haemophilia. Queen Victoria was thus an example of how the condition can arise as a spontaneous mutation. Leopold's medical condition was reported in the British Medical Journal in 1868, and it is clear that he was troubled by bleeds occurring at least once a month. He died at the age of 31 in 1884 from intracerebral haemorrhage after a fall. Leopold had married two years before his death. His daughter, Alice, was an obligate carrier and also went on to have a haemophilic son. Rupert, Viscount Trematon, was born in 1907 and died at the age of 21, also from an intracerebral haemorrhage. It also subsequently transpired that two of Queen Victoria's own daughters, Alice and Beatrice, were carriers of haemophilia. The condition was transmitted through them to several Royal families in Europe, including Spain and Russia. Perhaps the most famous affected individual was the son of Tsar Nicholas II of Russia. The story of the young Tsarevich Alexis, who was born in 1904, has been the subject of a Hollywood film as well as a novel by Dorothy Sayers ("Have his carcase": 1932). There has been speculation that the illness led to severe strain within the Royal family, and enabled Rasputin to gain influence on the family. Alexis and his family were murdered by the Bolsheviks in 1917. The haemophilic gene has now died out in these Royal families, emphasising the severity of the condition in the absence of effective medical treatment. Thus we do not know to this day if the condition was haemophilia A or B.

mutation rate of X linked genes

One third of all X chromosomes are present in males and hence one third of mutant X chromosomes are present in males. Consequently, if the condition is lethal, then one third of the mutant X chromosomes will be lost from the population each generation. If the frequency of the disease is not changing then the lost mutant chromosomes will have to be replaced by new mutation. Consequently, the mutation rate of a lethal X linked recessive disease is one third of the frequency of the disease.

sex linked dominant genes

Sex linked dominant conditions are extremely rare, examples include [incontinentia pigmenti](#) (which is lethal in males) and [congenital generalized hypertrichosis](#) (wolf man syndrome).

X inactivation

All cases of abnormal karyotypes in which a single autosome is missing (autosomal Genetics Lecture 4)

Sex determination

Surprisingly, it is only in the last 50 years that we have begun to understand the nature of the biological events which determine our sex, (and for that matter, why we bother with sex at all and why two sexes are better than three or more). It is not so long ago that women were blamed if they failed to produce a son for their husband and clearly it was thought that the power of sex determination lay within the body of the woman. During this century the chromosomal basis of human sex determination has been demonstrated and in the last few years some of the genes responsible have been identified.

The sexual identity of an individual is determined at several levels, chromosomal sex, gonadal sex, somatic sex and sexual orientation.

sex chromosomes

The chromosomal basis of sex determination in humans was recognized when metaphase chromosomes from dividing male and female cells could be studied and counted. The [normal karyotype](#) contains 46 chromosomes including either two X chromosomes (46XX, females) or one X chromosome and one Y chromosome (46XY, males). Individuals with [45X](#) or [47XXX](#) karyotypes are female, individuals with [47XXY](#)

karyotype are male. Therefore it can be deduced that the Y chromosome is sex determining

sex determination

Experiments involving the removal of the embryonic gonad have revealed that in mammals, no matter what the chromosomal sex of the somatic cells, the body will develop as a female unless a male gonad is present to secrete mullerian inhibiting substance and testosterone. This can be partially mimicked in the genetic condition [testicular feminisation](#) in which the gene coding for the androgen receptor is not expressed so that, although the testis in an XY individual secretes testosterone, the somatic tissues are unable to respond to it. Consequently the individual's body develops as a woman but with internal testes instead of ovaries. In 1990, a Y encoded gene [SRY](#) was discovered which (at least in mice) is able to transform the sex of an XX embryo from female to male. Individuals with mutations in this gene develop as females despite having an XY chromosomal constitution. About one male in 10,000 does not appear to have an Y chromosome but instead has two X chromosomes. These XX males can frequently be shown to have inherited from their fathers an X chromosome onto which a little bit of the Y chromosome carrying *SRY* has been transferred by an "illegitimate" cross over. XX males are entirely normal except that they are infertile and their heights are in the normal female range rather than the male.

In other organisms things happen differently. Both *Drosophila* and the nematode *Caenorhabditis elegans* use a mechanism in which each cell measures the relative number of X chromosomes compared to the number of autosomes. However, the genes involved in the counting process and in its interpretation do not seem to be related in the two species.

Even within vertebrates there are a variety of sex determination mechanisms.

- Birds use a chromosomal system, however unlike the mammalian system, males are *homogametic ZZ* and females are *heterogametic ZW*.
- Alligators are chromosomally the same in both sexes, they determine sex by the temperature at which embryos are allowed to develop. If warm then males are formed and if cool then females are formed.

sexual identity

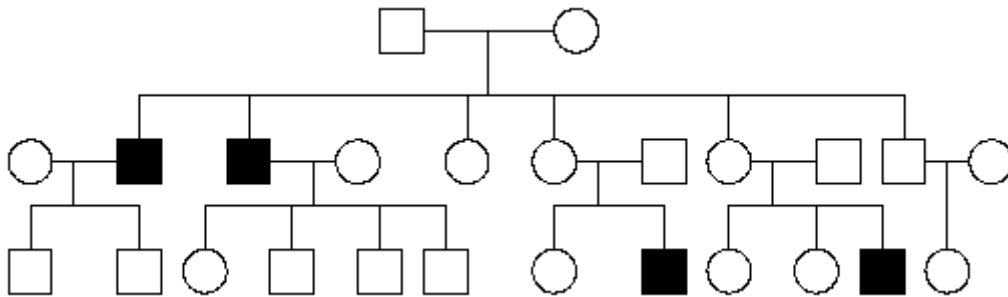
It is a much debated question as to whether our own sense of sexual identity is genetic or environmental in origin. Like most complex phenotypes it probably can be either but is usually both! Recent controversy centred on the [results obtained by Hamer](#) who showed evidence for one such gene, a recessive gene on the X chromosome, which when mutant may make its bearer, if male, more likely to be homosexual.

Sex linkage

sex linked recessive genes

Genes carried on the X chromosome have a distinctive pattern of inheritance. Because males are *hemizygous*, i.e. they have only one copy of the X chromosome, and because the Y chromosome carries very few genes (though those which it carries are often homologous to X linked genes) then recessive mutations manifest themselves in the phenotype of males. If the mutant gene is lethal (such as [Duchenne Muscular Dystrophy](#)) then it takes an unusual event to produce an affected female.

An X linked pedigree



A typical pedigree will show clusters of affected males (each brother will have a 50% chance of being affected) connected through unaffected carrier females. There will be *no cases of direct male to male transmission* because males transmit their X chromosomes to their daughters and not to their sons.

The following passage is quoted from [The history of haemophilia by Dr. P.L.F. Giangrande](#)

The story of Queen Victoria

Haemophilia is sometimes referred to as the Royal disease. Queen Victoria had no ancestors with the condition but soon after the birth of her eighth child, Leopold, in 1853 it became evident that he had haemophilia. Queen Victoria was thus an example of how the condition can arise as a spontaneous mutation. Leopold's medical condition was reported in the British Medical Journal in 1868, and it is clear that he was troubled by bleeds occurring at least once a month. He died at the age of 31 in 1884 from intracerebral haemorrhage after a fall. Leopold had married two years before his death. His daughter, Alice, was an obligate carrier and also went on to have a haemophilic son. Rupert, Viscount Trematon, was born in 1907 and died at the age of 21, also from an intracerebral haemorrhage. It also subsequently transpired that two of Queen Victoria's own daughters, Alice and Beatrice, were carriers of haemophilia. The condition was transmitted through them to several Royal families in Europe, including Spain and Russia. Perhaps the most famous affected individual was the son of Tsar Nicholas II of Russia. The story of the young Tsarevich Alexis, who was born in 1904, has been the subject of a Hollywood film as well as a novel by Dorothy Sayers ("Have his carcase": 1932). There has been speculation that the illness led to severe strain within the Royal family, and enabled Rasputin to gain influence on the family. Alexis and his family were murdered by the Bolsheviks in 1917. The haemophilic gene has now died out in these Royal families, emphasising the severity of the condition in the absence of effective medical treatment. Thus we do not know to this day if the condition was haemophilia A or B.

mutation rate of X linked genes

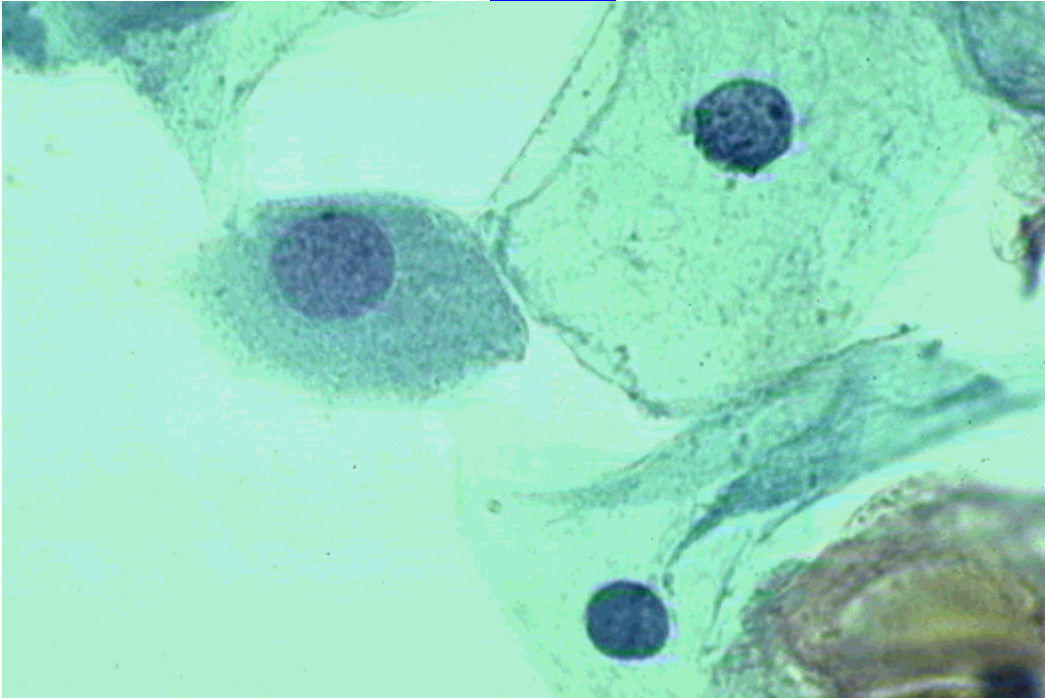
One third of all X chromosomes are present in males and hence one third of mutant X chromosomes are present in males. Consequently, if the condition is lethal, then one third of the mutant X chromosomes will be lost from the population each generation. If the frequency of the disease is not changing then the lost mutant chromosomes will have to be replaced by new mutation. Consequently, the mutation rate of a lethal X linked recessive disease is one third of the frequency of the disease.

sex linked dominant genes

Sex linked dominant conditions are extremely rare, examples include [incontinentia pigmenti](#) (which is lethal in males) and [congenital generalized hypertrichosis](#) (wolf man syndrome).

X inactivation

All cases of abnormal karyotypes in which a single autosome is missing (autosomal monosomy) are lethal during embryogenesis even for the smallest autosomes. Yet males with only one X chromosome (a medium sized chromosome) are (comparatively!) normal. How is this accomplished? The answer (as suggested by Mary Lyon in 1961) is by inactivation of one of the two X chromosomes in females so that the normal state for a cell is to have two active sets of autosomes and only one active X chromosome. The other X chromosome is condensed and inactive and is visible as a dark staining "Barr body" pushed against the nuclear membrane. The following image of cells from a female cheek swab was "borrowed" from the [US Army](#).



Three nuclei can be seen each with a characteristic dark blob (which is particularly clear in the centrally located nucleus).

Mary Lyon hypothesised that the X inactivation happened at random early in development so that each female is composed of two populations of cells. In one population one X chromosome is expressed and in the other, the second X chromosome is expressed. Females are thus *mosaics*, i.e. composed of two genetically distinct cell populations. For genes which are homozygous this will make no difference but for genes for which the female is heterozygous the two populations of cells will be of opposite phenotypes. In this way she explained the patterns of hair colouration in, for instance, the tortoiseshell cat which is always female except for the very rare occurrence of an XXY male, which exception proves the rule! (Those of you who are interested might like to try this well written page on [cat coat colour genetics](#).)

Sex limited traits

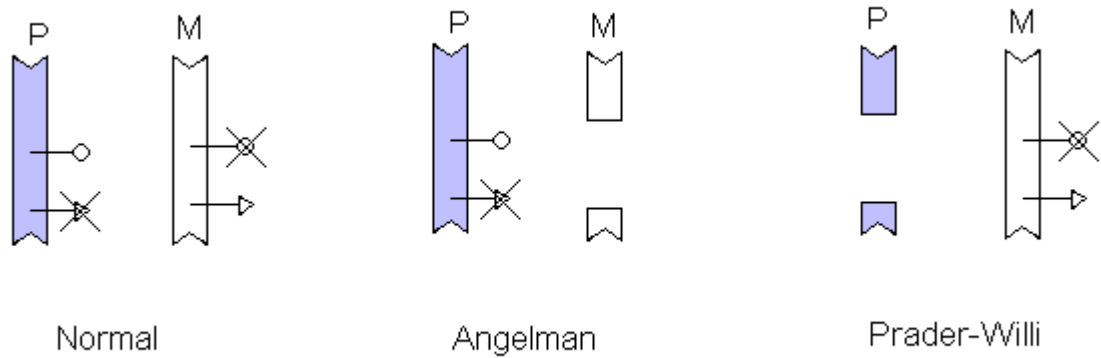
As we have seen, sex linked traits are generally expressed much more often in males than in females. You should be aware however that some traits which affect one sex more than another are not necessarily sex linked. Examples are cases of sex limited expression which might include genes affecting beard growth or breast size, and (in cattle), horn growth and milk yield. These genes have no visible effect in one sex because the necessary machinery to express them is not present.

Sex influenced traits

On the same general line as sex limited traits are sex influenced traits. Pattern baldness is a condition which is dominant in men but recessive in women.

Imprinting

Some genes are inactivated when transmitted through one sex. [Angelman syndrome](#) and [Prader-Willi](#) syndrome are two different conditions both of which seem to be caused by very similar deletions of a small part of chromosome 15.



In this diagram, two genes are shown in the critical region. Each is inactivated by *imprinting*, the Angelman syndrome gene is turned off on the chromosome inherited from the father while the Prader-Willi gene is turned off on the maternally transmitted chromosome. When a deletion covering the region is inherited on the other chromosome one syndrome or the other results.

Recommended reading

The topics include:

- sex determination
- sex chromosomes
- sexual identity
- sex linkage
- X inactivation
- sex limited and sex influenced traits
- imprinting

reading:

- Mange and Mange Chapters 4 (pp 59 - 63), 5 (pp 81 - 90), 10 (pp 202 - 214)
 - Lewis Chapter 6
 - Mueller and Young Chapters 5 (pp 91 - 94) and 6 (pp 102 - 106, 108 - 110) (very brief account)
 - Thompson McInnes and Willard Chapters 4 (pp 72 - 82, 92 - 93), and 10 (which we will cover again in Lecture 7)
 - Jorde *et al.* Chapter 5
 - Connor and Ferguson-Smith Chapters 5 (pp 50 - 54), and 8
-

SAQs

1. A woman with a father with deuteranopia marries a man with the same form of colourblindness. What is the likelihood that their first child will be
 1. a colourblind son
 2. a colourblind daughter
 3. a carrier daughter

How would these odds be changed if instead of having a deuteranopic father she had had a deuteranopic brother?

2. An infertile couple come to you for advice. The man is relatively short, only 1.60m tall. He has no other obvious symptoms. Preliminary chromosome tests reveal that he has a Barr body in his nuclei. What is the likely cause of the couple's infertility?

Answers

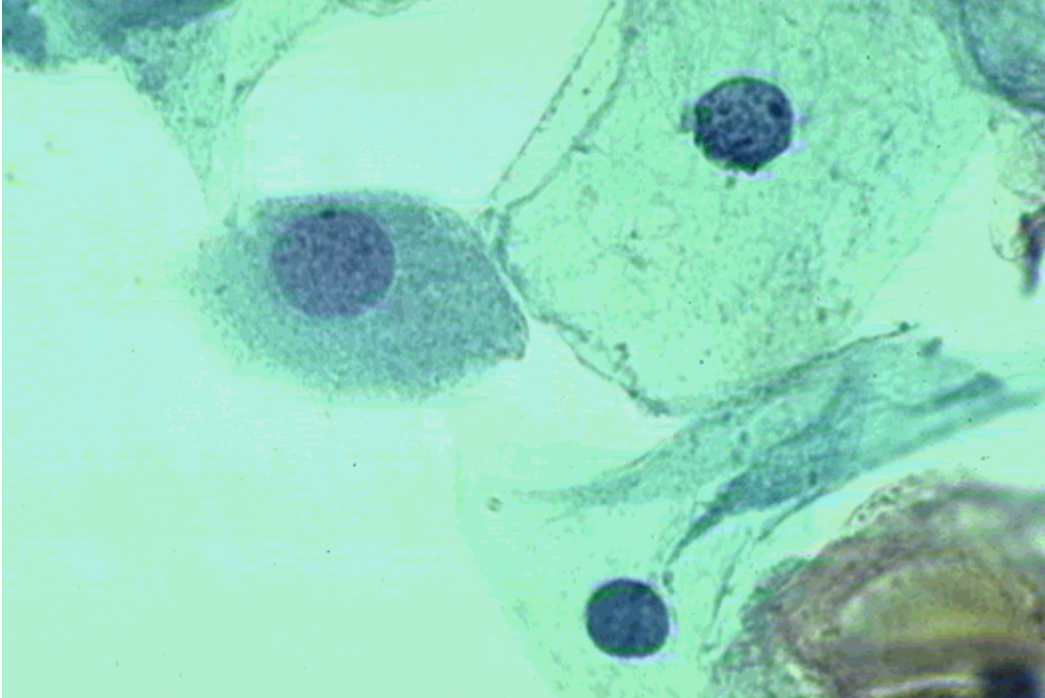
[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

monosomy) are lethal during embryogenesis even for the smallest autosomes. Yet males with only one X chromosome (a medium sized chromosome) are (comparatively!) normal. How is this accomplished? The answer (as suggested by Mary Lyon in 1961) is by inactivation of one of the two X chromosomes in females so that the normal state for a cell is to have two active sets of autosomes and only one active X chromosome. The other X chromosome is condensed and inactive and is visible as a dark staining "Barr body" pushed against the nuclear membrane. The following image of cells from a female

cheek swab was "borrowed" from the [US Army](#).



Three nuclei can be seen each with a characteristic dark blob (which is particularly clear in the centrally located nucleus).

Mary Lyon hypothesised that the X inactivation happened at random early in development so that each female is composed of two populations of cells. In one population one X chromosome is expressed and in the other, the second X chromosome is expressed. Females are thus *mosaics*, i.e. composed of two genetically distinct cell populations. For genes which are homozygous this will make no difference but for genes for which the female is heterozygous the two populations of cells will be of opposite phenotypes. In this way she explained the patterns of hair colouration in, for instance, the tortoiseshell cat which is always female except for the very rare occurrence of an XXY male, which exception proves the rule! (Those of you who are interested might like to try this well written page on [cat coat colour genetics](#).)

Sex limited traits

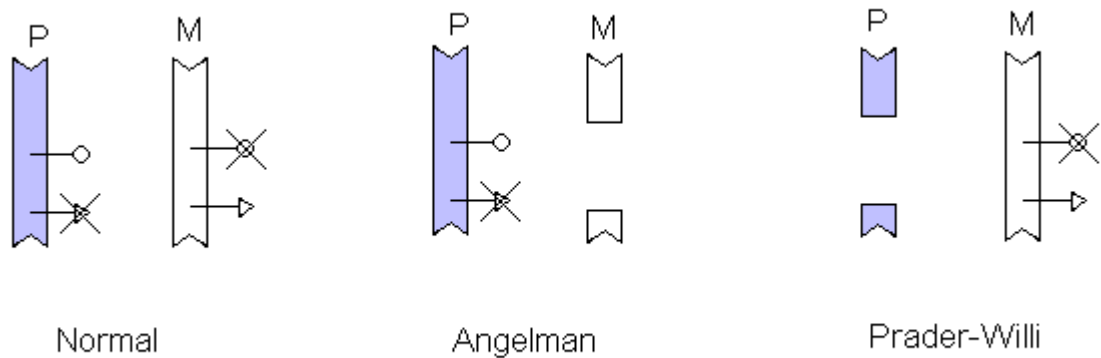
As we have seen, sex linked traits are generally expressed much more often in males than in females. You should be aware however that some traits which affect one sex more than another are not necessarily sex linked. Examples are cases of sex limited expression which might include genes affecting beard growth or breast size, and (in cattle), horn growth and milk yield. These genes have no visible affect in one sex because the necessary machinery to express them is not present.

Sex influenced traits

On the same general line as sex limited traits are sex influenced traits. Pattern baldness is a condition which is dominant in men but recessive in women.

Imprinting

Some genes are inactivated when transmitted through one sex. [Angelman syndrome](#) and [Prader-Willi](#) syndrome are two different conditions both of which seem to be caused by very similar deletions of a small part of chromosome 15.



In this diagram, two genes are shown in the critical region. Each is inactivated by *imprinting*, the Angelman syndrome gene is turned off on the chromosome inherited from the father while the Prader-Willi gene is turned off on the maternally transmitted chromosome. When a deletion covering the region is inherited on the other chromosome one syndrome or the other results.

Recommended reading

The topics include:

- sex determination
- sex chromosomes
- sexual identity
- sex linkage
- X inactivation
- sex limited and sex influenced traits
- imprinting

reading:

- Mange and Mange Chapters 4 (pp 59 - 63), 5 (pp 81 - 90), 10 (pp 202 - 214)
- Lewis Chapter 6

- Mueller and Young Chapters 5 (pp 91 - 94) and 6 (pp 102 - 106, 108 - 110) (very brief account)
 - Thompson McInnes and Willard Chapters 4 (pp72 - 82, 92 - 93), and 10 (which we will cover again in Lecture 7)
 - Jorde *et al.* Chapter 5
 - Connor and Ferguson-Smith Chapters 5 (pp50 - 54), and 8
-

SAQs

1. A woman with a father with deuteranopia marries a man with the same form of colourblindness. What is the likelihood that their first child will be
 1. a colourblind son
 2. a colourblind daughter
 3. a carrier daughter

How would these odds be changed if instead of having a deuteranopic father she had had a deuteranopic brother?

2. An infertile couple come to you for advice. The man is relatively short, only 1.60m tall. He has no other obvious symptoms. Preliminary chromosome tests reveal that he has a Barr body in his nuclei. What is the likely cause of the couple's infertility?

Answers

[Back to the top](#)

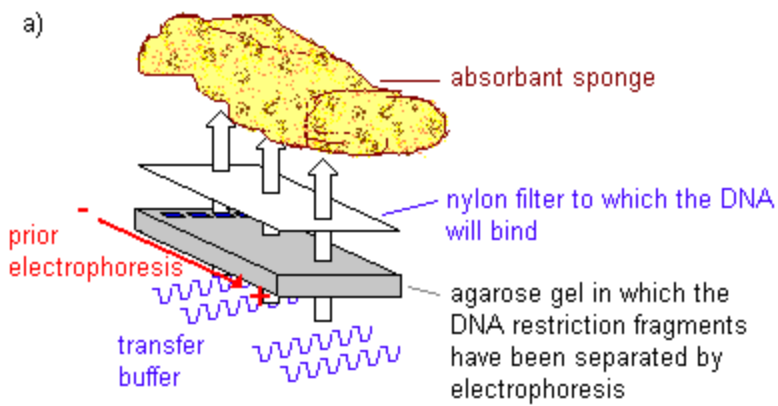
[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 5

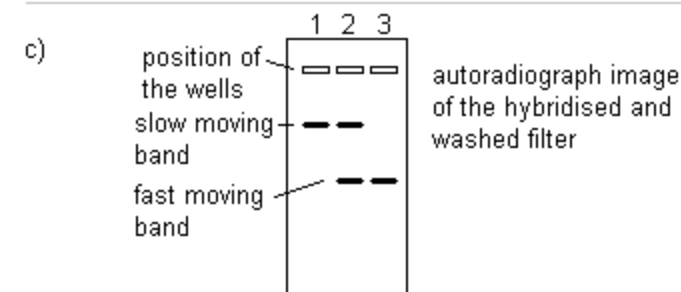
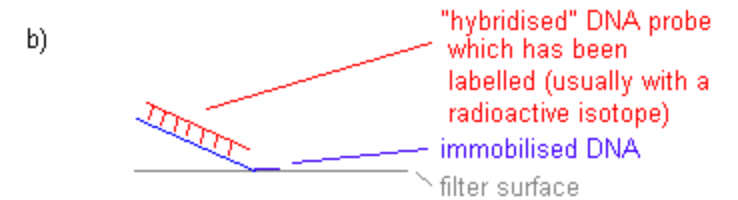
Tools of Molecular genetics

Reminder: an excellent (though slightly out of date) web site to visit for relevant information is The US Department of Energy's [Primer of Molecular Genetics](#). I have reproduced one or two figures from that primer here.



Gene Cloning

This is the general term given to the isolation of a specific fragment of target (e.g. human) DNA by propagating that fragment in a microorganism (usually a bacterium but sometimes a yeast). Using standard microbiological techniques, a single microorganism can be isolated and grown in culture carrying within its genome one small fragment of human DNA. The human DNA can readily be purified from the bacterium in gram quantities if necessary. For a more detailed explanation start [here](#)



All clones come from *libraries*. Essentially these may be only one or other of [cDNA](#) or [genomic](#), derived respectively from mRNA or genomic DNA, and it is vital to remember the distinction between them.

Southern and Northern Blotting

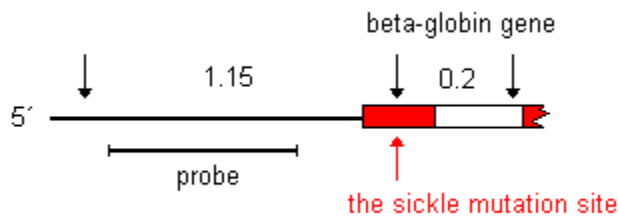
Before PCR and cheap fast sequencing changed our view of the universe that is genetics, the Southern Blot was a universal workhorse. There was not an experiment in molecular genetics which did not at some stage employ a Southern Blot. It is still a useful tool and you need to know about it so that you can interpret historical data.

Named after its inventor, Prof. Ed. Southern, the blot is a fast way of analysing a small number of DNA fragments which may be present in a complex mixture. For instance, suppose that we wish to ask whether the sickle cell mutation is present in an individual and the only material which we have available is a DNA sample. We could employ a Southern Blot:

- DNA is first digested with a **restriction endonuclease** and then separated according to the size of DNA fragments by electrophoresis through an agarose gel. Small fragments are able to migrate more quickly through the gel than are long fragments. Most restriction enzymes will digest human DNA into about a million fragments. We will want to compare only those which contain the human beta globin gene. To do this the DNA is first **denatured**, (made single stranded), by treatment with NaOH. It is then transferred to the surface of a nylon filter by blotting as in part a) of the figure.
- A probe is made. Usually this means incorporating a **label** into a fragment of DNA from the target sequence (beta globin in this case). The label may be either a radioactive isotope (often ^{32}P) or a chemical hapten such as biotin attached via a long side chain.

- The single stranded probe is **annealed** to its target sequence which is then **stringently washed** to remove any probe which has bound other than by perfect formation of a long run of complementary hydrogen bonds. See section b) of the figure.
- Finally the position of the annealed probe is determined, (by autoradiography in the case of a radioactive probe), see c).

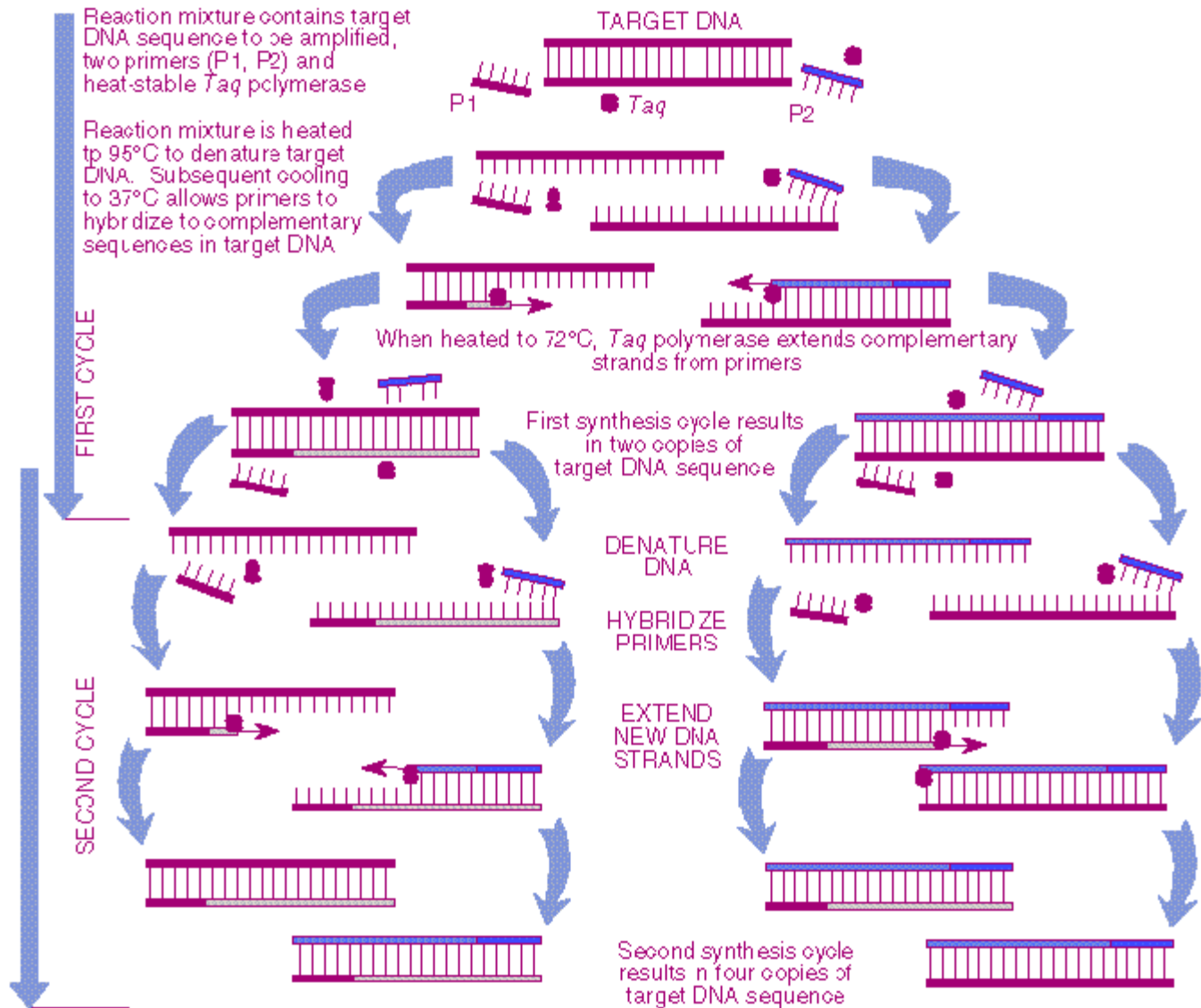
In the case of the sickle cell mutation, the single base change involved, as well as causing a missense mutation in the beta globin gene, also causes the disappearance of a restriction site for the enzyme *MstII*. As a consequence the size of the restriction fragment containing the 5' end of the gene is altered from 1.15kb to 1.35kb. See the figure below where *MstII* sites are shown as arrows.



The Polymerase Chain Reaction

This invention has revolutionised molecular genetics by doing away with the need to clone DNA in many circumstances where it used to be necessary. It is so powerful that it has made it possible to produce microgram amounts (that's a lot!) of DNA starting from just a single molecule. This has applications in forensic science, in archeology (Neandertal mitochondrial DNA amplified from ancient bones by PCR was recently sequenced), and in medicine where, for example, it can enable antenatal DNA tests to be performed in just a few hours work and large populations can be screened for particular mutations very quickly and cheaply.

DNA Amplification Using Polymerase Chain Reaction

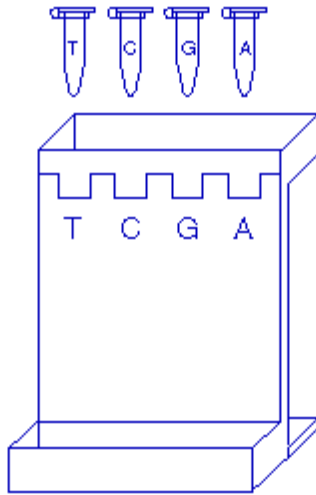


Source: *DNA Science*, see Fig. 13.

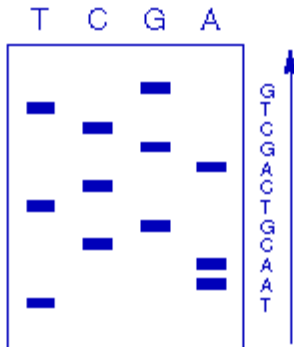
DNA sequencing

At the heart of the so called "new genetics" is our ability to sequence DNA rapidly and cheaply. With knowledge of sequence comes knowledge of gene structure and very often a beginning of understanding of gene function. The simplest method of sequencing DNA was invented by Dr. Fred Sanger for which he was awarded his second Nobel Prize. In

1. Sequencing reactions loaded onto polyacrylamide gel for fragment separation



2. Sequence read (bottom to top) from gel autoradiogram



purify DNA

separate strands

prime synthesis

elongate until

incorporate a chain terminating fluorescently tagged dideoxynTP

the UK our major national Human Genome Project DNA sequencing centre is named after him, the Sanger Centre in Cambridgehire. You can visit its home page [here](#) if you wish.

The Sanger method is also known as the "dideoxy chain termination

method".

- The DNA to be sequenced must be obtained in pure form, either by cloning or by PCR.
- The DNA strands must be separated (usually by cloning into a vector, M13 bacteriophage, which has a single stranded phase to its life cycle but sometimes by some other cunning technique or even just by heating).
- A primer is allowed to anneal to known sequence at one end of the target sequence. The known sequence may be part of the bacteriophage DNA flanking the cloned DNA.
- Deoxynucleotide triphosphates (dATP, dCTP, dGTP and dTTP) and DNA polymerase are added and the primer sequence is thus elongated along the target

"template" DNA.

- Also included in the mixture are small amounts of the four dideoxynucleotide triphosphates (ddNTPs) each tagged with a different coloured fluorescent dye. When a molecule of ddNTP is incorporated into an elongating DNA strand it prevents further chain elongation because it lacks the necessary terminal 3' OH group.
- The products of synthesis are then separated according to size by electrophoresis on a polyacrylamide gel. This figure shows an earlier method in which the four bases were read in separate reactions and a radioactive "label" was incorporated into the newly synthesised DNA. But the principle is the same.
- As the products pass a point on the gel one by one, the coloured dye tags can be read by a laser scanner and computer.

STSs and ESTs

PCR and sequencing together have made possible the creation of useful landmarks in the genome. These are several thousand short fragments of known DNA sequence whose presence in any DNA sample can be tested by PCR. They are known as STSs (Sequence Tagged Sites). If an STS is part of a transcribed sequence it is known as an EST (Expressed Sequence Tag). Hundreds of thousands of ESTs have been created and can be accessed by computer. One major attempt to classify them all is known as [Unigene](#)

Recommended reading

The topics include:

- gene cloning
- Southern and Northern blotting
- The Polymerase Chain Reaction
- DNA sequencing

Reading:

- Mange and Mange Chapters 6 and 8 (good chapters which cover the essentials and more)
- Lewis is weak on methodology but strong on applications. Partial information is in Chapters 8 and 17 (pp316 - 322 (the rest of the chapter is interesting but not directly relevant)), you could also read chapter 21
- Mueller and Young Chapter 4
- Jorde *et al.* Chapter 3 (pp 42 - 56)
- Thompson McInnes and Willard Chapter 5 (an excellent account)
- Connor and Ferguson-Smith Chapter 3 (too little (and too advanced?) - but have a look if its all that's available)

SAQs

1. If you wanted to study the structure and regulation of the chromosome 14 linked gene coding for alpha-1-antitrypsin which is expressed strongly in liver but weakly or not at all elsewhere in the body, which of the following resources might provide useful material?
 1. A human genomic library in a cosmid vector made from leukocyte DNA
 2. A cDNA library made from cardiac muscle
 3. A cDNA library made from foetal liver
 4. A cDNA library made from adult liver
 5. An arrayed human genomic library in a cosmid vector made from flow sorted chromosomes 9

6. A YAC clone containing STSs known to flank the alpha-1-antitrypsin gene on chromosome 14
2. How might you save yourself a great deal of unnecessary trouble by finding publically available clones to aid the study in the previous question?
3. The genes which when mutant cause the disease Tuberous sclerosis have finally been isolated after a long search. This could not have been accomplished without a very great deal of help from the families in which the disease is present. Apart from their natural curiosity into the nature of the disease which was afflicting them, what was in it for them? In other words, how does finding the genes help the families?

[Answers](#)

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 6

Mutation

Mutation is the alteration of DNA sequence, whether it be in a small way by the alteration of a single base pair, or whether it be a gross event such as the gain or loss of an entire chromosome. It may be caused through the action of damaging chemicals, or radiation, or through the errors inherent in the DNA replication and repair reactions. One consequence may be genetic disease. However, although in the short term mutation may seem to be a *BAD THING*, in the long term it is essential to our existence. Without mutation there could be no change and without change life cannot evolve. If it had not been for mutation the world would still be covered in primeval slime!

In this course we are not going to consider the molecular events involved in mutation but instead will concentrate on the genetic consequences of mutation.

Somatic or germinal?

The first point to consider is where is the mutation occurring? Most of our cells are somatic cells and consequently most mutations are happening in somatic cells. New mutation is only of genetic consequence to the next generation if it occurs in a germ line cell so that it stands a chance of being inherited. That is not to say that somatic mutation

is unimportant, [cancer](#) occurs as a direct consequence of somatic mutation and aging too may be caused at least in part by the accumulation of somatic mutations with time.

Different types of mutation occur at different frequencies:

type of mutation	mechanism	frequency per cell division
point mutation	1. mistakes in DNA replication 2. DNA damage by chemical mutagens (or by radiation) and misrepair	$\sim 10^{-10}$ /basepair $\sim 10^{-5}$ /gene ~ 0.5 /cell
submicroscopic deletion or insertion	1. unequal crossing over 2. misalignment during DNA replication 3. insertion of mobile element 4. DNA damage by chemical mutagens (or by radiation) and misrepair	included in the above
microscopically visible deletion, translocation or inversion	1. unequal crossing over 2. DNA damage by chemical mutagens (or by radiation) and misrepair	6×10^{-4}
loss of a whole chromosome	missegregation at mitosis	1 in 100

Most gross mutations will impose a survival or growth rate disadvantage on the cell in which they occur and on the descendants of that cell which consequently will not contribute significantly to the body. The obvious exception is again [cancer](#).

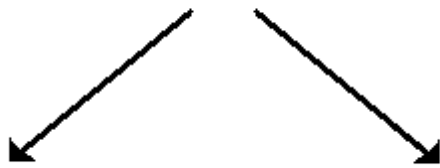
mosaicism

If a mutation such as a chromosome loss occurs early in development, the descendants of the cell may represent a significant fraction of the individual who, being composed of cells of more than one genotype is a genetic *mosaic*. (And see also [X inactivation](#) in lecture 4.) Mosaicism is not infrequent in for instance the cases of [Turner's syndrome](#) or [Down's syndrome](#) described in the next lecture.

Most mutations are recessive

Why is this? Because most genes code for enzymes. If one gene is inactivated the reduction in the level of activity of the enzyme may not be as much as 50% because the level of transcription of the remaining gene can possibly be up regulated in response to any rise in the concentration of the substrate. Also, the protein itself may be subject to regulation (by phosphorylation for instance) so that its activity can be increased to compensate for any lack of numbers of molecules. In any case, if the enzyme does not

Dietary Protein



Phenylalanine → Tyrosine

control the rate limiting step in the biochemical pathway a reduction in the amount of product may not matter.

phenylalanine hydroxylase

In the case of [phenylketonuria \(PKU\)](#) we can see that it is necessary to reduce the enzyme level to below 5% before any effect is apparent in the phenotype. This

genetic disease is caused by mutations in the gene coding for the enzyme phenylalanine hydroxylase which converts the amino acid phenylalanine to tyrosine. If an individual is homozygous for alleles which completely remove any enzyme activity, phenylalanine cannot be metabolised and it builds up in the circulation to a point where it begins to damage the developing brain. Newborn infants are routinely screened for this condition by the analysis of a tiny drop of blood from a heel prick (Guthrie test). This has revealed that there exist a few people with a condition known as benign hyperphenylalaninemia. These individuals have moderately elevated levels of phenylalanine in their serum. Their phenylalanine hydroxylase enzyme levels are about 5% of normal. Despite this they are apparently perfectly healthy and do not suffer from the brain abnormalities caused by the full blown disease.

What makes a mutation dominant?

- **Haploinsufficiency.**

In this case, the amount of product from one gene is not enough to do a complete job. Perhaps the enzyme produced is responsible for a rate limiting step in a reaction pathway. Hereditary hemorrhagic telangiectasia (HHT) is an autosomal dominant vascular dysplasia leading to telangiectases and arteriovenous malformations of skin, mucosa, and viscera. Death by uncontrollable bleeding occasionally occurs. It is caused by mutation in the gene *ENG*, which codes for the protein endoglin, a transforming growth factor-beta (TGF-beta) binding protein. Perhaps the TGF-beta is unable to exert sufficient effect on cells when only half the normal amount of receptor is present.

- **Dominant negative effect.**

The product of the defective gene interferes with the action of the normal allele. This is usually because the protein forms a multimer to be active. One defective component inserted into the multimer can destroy the activity of the whole complex. An example might be Osteogenesis imperfecta, [see below](#)

- **Gain of function**

It is possible to imagine that by mutation a gene might gain a new activity, perhaps an enzyme active site might be altered so that it develops a specificity for a new substrate. That this must be so is self evident, how else could evolution occur? Examples in human genetics of genes with two such different alleles are

rare but one example comes from the ABO locus. The difference between the A and B loci is determined by 7 nucleotide changes which lead to 4 amino acid changes. Probably only one of these changes is responsible for the change in specificity of the enzyme from alpha-3-N-acetyl-D-galactosaminyltransferase (A) to alpha-3-D-galactosyltransferase. There are also many examples from human evolution where many genes have duplicated and subsequently the two duplicates have diverged in their substrate specificities. On chromosome 14 is a little cluster of three related genes, alpha-1-antitrypsin, (*AAT*), alpha-1-antichymotrypsin, (*ACT*), and a related gene which has diverged to such an extent that it is probably no longer functional. The structural relationship between *AAT* and *ACT* is very obvious and both are protease inhibitors but they now clearly serve slightly different roles because they have different activities against a range of proteases and they are under different regulation.

- **Dominance at an organismal level but recessivity at a cellular level.**

Some of the best examples of this are to be found in the area of cancer genetics which we will consider in [lecture 10](#). A typical example of such mutant gene would be a tumour suppressor gene such as [retinoblastoma](#).

Point Mutation

'Point mutation' covers a variety of sins from genuine single base pair changes to small deletions and insertions (which I have however placed in the next section). A single base pair change may have no genetic consequences whatsoever but, on the other hand, it may cause a dominant lethal effect. The first alternative is far more likely. Why? Because 95% of DNA is non coding and a single base change occurring within it is unlikely to have any effect. In addition, because of the degeneracy of the genetic code, many mutations occurring within the third base position in a codon will have no consequence to the amino acid encoded.

The figure shows a short region of coding sequence with a variety of possible mutations.

Wild type allele:

M D D Q S R M L Q T L A G V N L
atggacgatcaatccaggatgctgcagactctggccgggggtgaacctg

silent (third base pair) mutation:

M D D Q S R M L Q T L A G V N L
atggacgatcaatccaggatgctgcaactctggccgggggtgaacctg

point mutation (missense):

M D D Q S R M L K T L A G V N L
atggacgatcaatccaggatgctgaagactctggccgggggtgaacctg

point mutation (nonsense):

M D D Q S R M L stop
atggacgatcaatccaggatgctgtagactctggccgggggtgaacctg

frameshift leading to premature termination:

M D D Q S R M L R L W P G stop
atggacgatcaatccaggatgctgagactctggccgggggtgaacctg

amino acid substitution

A single amino acid change may be unimportant if it is conservative and occurs outside the active site of the protein. On the other hand it can have severe effects.

- The single substitution of valine for glutamic acid at position six of the beta-globin polypeptide chain gives rise to sickle cell disease in homozygotes because the modified chain has a tendency to crystallise at low O₂ concentrations.
- Collagen is a family of related structural proteins which are vital to the integrity of many tissues including skin and bones. The mature collagen molecule is comprised of three polypeptide chains wound in a triple helix. The chains first associate at their C termini and then twist together in a C to N direction. To be able to accomplish this collagen polypeptide chains have a special repeating structure of three amino acids, glycine - X - Y (X is usually proline and Y can be any of a wide range of aminoacids). A point mutation which by changing a single amino acid disrupts either the association of chains at their C termini or which prevents the triple helix formation may have very severe consequences. One mutant chain can disrupt a triple helix with two wild type chains effectively 'mopping up' functional monomers. The small amount of working collagen produced, not being an enzyme, cannot be upregulated. The consequence can be the dominant lethal condition [osteogenesis imperfecta](#).

frame shift and nonsense

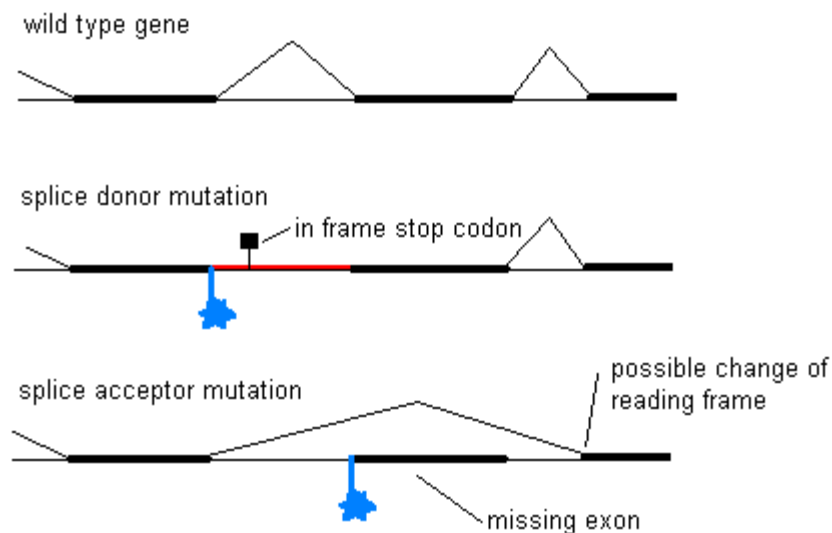
By nonsense is meant the premature insertion of a stop codon into the gene sequence. This might be by a single nucleotide change as in the figure above where the codon CAG which encodes glutamine (Q) has mutated to the stop codon TAG (UAG in the mRNA of course). Alternatively it can be as a consequence of a deletion or insertion of a number of nucleotides not divisible by three which shifts the reading frame and by chance will usually quickly lead to a stop codon. In the figure above, the same C nucleotide has been deleted in the last line leading to frameshift and a quick termination. The tuberous sclerosis gene TSC1 contains a direct repeat of four nucleotides, AAAGAAAG. Four independent mutations have been identified in which one repeat has been lost by deletion of AAAG. This leads to frameshift and premature chain termination.

triplet deletion

There is nothing particularly special about a triplet deletion which removes exactly one amino acid from the polypeptide (and which may change one amino acid at the mutation site). However, I include it because the most common mutation in [cystic fibrosis](#) is deltaF508 (i.e. deletion of amino acid number 508 (a phenylalanine, F)).

splice site

Frameshifts can also come about by mutations which interfere with mRNA splicing. The beginning and end of each intron in a gene are defined by conserved DNA sequences. If a nucleotide in one of the highly conserved positions is mutated then the site will no longer function with predictable consequences for the mature mRNA and the coded protein product.



There are many examples of such mutations, for instance, some beta thalassemia mutations in the beta globin gene are caused by splice junction mutations.

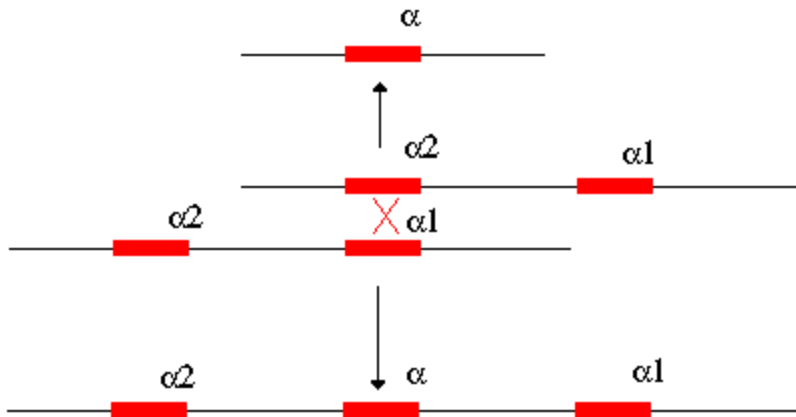
deletion and insertion

small i.e. less than 10kb

Deletions or expansions of a small number of nucleotides happen from time to time. Some deletions are entirely random but many are caused by misalignment of short repeats during DNA synthesis, so called **replication slippage**.

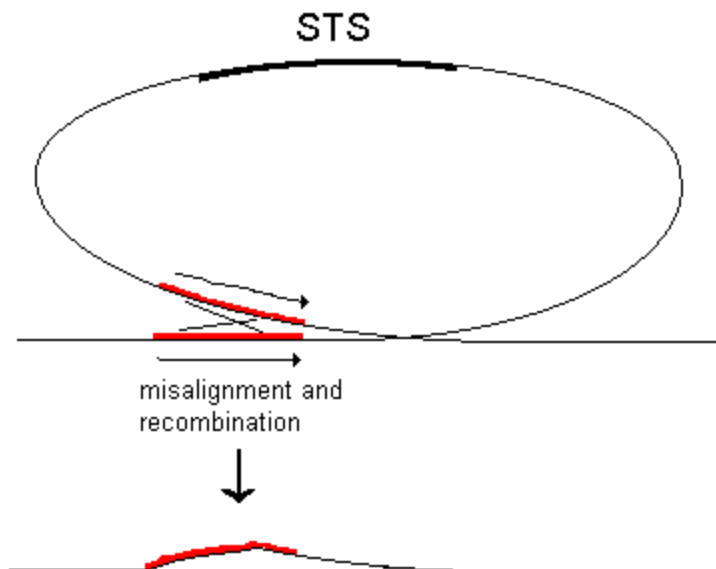
As much as 25% of the human genome is composed of repetitive DNA. Much of this is concentrated at chromosome centromeres and in heterochromatic regions but a great deal of it is in the form of **interspersed repetitive DNA**. There are several different types of this but the two principal types are **alu** elements and **L1** repeat elements. These interspersed repeats are very important from an evolutionary point of view. Working elements can transpose, via an RNA intermediate. New copies of the element can be inserted at random into the host genome. Sometimes, if the insertion takes place within an exon or within a control region, this can cause a mutation. Such fresh insertions are fairly rare events but one example has been described in the neurofibromatosis type 1 gene (*NFI*) where an alu has inserted within intron 5. This prevents proper splicing so that exon 6 is left out of the mature mRNA leading to a frameshift mutation and a premature stop codon. The patient was the result of a new mutation, neither parent carried the mutant allele. The presence of numerous repeats can however have other effects. If two alu elements are close together in the genome they may misalign at meiosis and if a recombination event takes place the consequence will be a gamete with either one fewer or one more copy of the element and the intervening DNA sequence.

Some genes are duplicated. We have already mentioned the alpha-1-antitrypsin gene cluster. A much more widely quoted example is that of the alpha globin genes which are present in a cluster on chromosome 16. There are two very closely related copies of the adult alpha globin gene, alpha1 and alpha2, which are separated by a gap of only 3.7kb. If insufficient alpha globin is produced there will be an excess of beta globin in the erythrocytes and haemoglobin (beta)₄ tetramers (known as Hb H) or, in the embryo, haemoglobin (gamma)₄ (Haemoglobin Barts) will be present. Neither of these will release oxygen to tissues causing the devastating disease alpha thalassaemia. This is one of the World's most common genetic diseases and the most common cause is homozygosity for a deletion of one alpha globin gene. This recurrent mutation is brought about by misalignment of the two alpha globin genes in meiotic prophase followed by unequal crossing over.



Large but still submicroscopic i.e. less than 1Mb

Almost all cases of X linked ichthyosis are brought about by total deletion of the gene *STS*, coding for steroid sulphatase. This is caused by a mutation which seems to be recurrent. The *STS* gene is flanked by two copies of a repetitive element. It is thought that sometimes in meiosis these two copies, on the same chromatid, align and recombine leading to excision of the DNA between them as in the figure below.



Sometimes deletions can cover more than one gene. When this occurs we have an example of a **contiguous gene syndrome**. For instance, some patients with Tuberous sclerosis also suffer from polycystic kidney disease. The two loci *TSC2* and *PKD* are adjacent on chromosome 16 and can be simultaneously deleted.

Cytogenetic

Some deletions become so large that they are visible on well stretched metaphase chromosomes. When enough of the genome is deleted to be visible in this way, the symptoms are usually very severe because a large number (probably more than a hundred) genes will have been lost. Contiguous gene syndromes will become likely. More on this in the [next lecture](#).

Trinucleotide expansion

The commonest inherited cause of mental retardation is a syndrome originally known as Martin-Bell syndrome. Patients are most usually male, have a characteristic elongated face and numerous other abnormalities including greatly enlarged testes. The pattern of inheritance of this disease was, at first, puzzling. It usually behaved as an X linked recessive condition but sometimes manifested itself in females and occasionally nonaffected transmitting males were found. In 1969 it was discovered that if cells from patients were cultured in medium deficient in folic acid their X chromosomes often displayed a [secondary constriction](#) near the end of the long arm. The name of the syndrome was changed to the [fragile X syndrome](#). The puzzling genetics remained unclear. Eventually the mutation was tracked down to a [trinucleotide expansion](#) in the gene now named *FMRI* (Fragile site with Mental Retardation) at the site of the secondary constriction. As in the case of [myotonic dystrophy](#) symptomless premutations could occur (and were the cause of the transmitting males). Only when the premutation chromosomes were transmitted through females did expansion to the full mutant allele and phenotype occur. A number of diseases have now been ascribed to trinucleotide expansions. These include [Huntington's disease](#).

Recommended reading

The topics include all aspects of mutation

Reading:

- Mange and Mange Chapter 7
- Lewis Chapter 10
- Mueller and Young Chapter 2 (pp 20 - 24), and, as further reading, chapter 9
- Jode *et al.* Chapter 3 (Some of this you've read before. **Read it again!!** Seriously, you'll be reading from a different viewpoint and I think that you will find this is worthwhile.)
- Thompson McInnes and Willard Chapter 6 (Chapters 11 and 12 make interesting further reading)
- Connor and Ferguson-Smith Chapter 2 (pp 17 - 23)

SAQs

The questions at the ends of chapter 10 in Lewis and chapter 12 of Mange and Mange are worth trying. In addition:

1. The following short sentence represents a gene:

WET WET WET ARE NOT ALL BAD

What types of mutations are the following?

- a. WEE WET WET ARE NOT ALL BAD
 - b. WET WET WET ARE COT ALL BAD
 - c. WEW ETW ETA REN OTA LLB AD
 - d. WET WET WET ART TEN OTA LLB AD
 - e. WET WET ARE NOT ALL BAD
 - f. WET WET WEG OOD MOR NIN GTA REN OTA LLB AD
 - g. WET WET WET WET WET WET ARE NOT ALL BAD
2. Full colour vision requires that the retinal cone cells be of three types differing in which of three opsin molecules they produce. One of the three opsin genes, that coding for a blue sensitive protein, is autosomal but the other two genes (red sensitive and green sensitive respectively) are adjacent to each other on the X chromosome. What are the consequences of this arrangement of genes for the mutation frequency of the X linked genes?

Answers

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

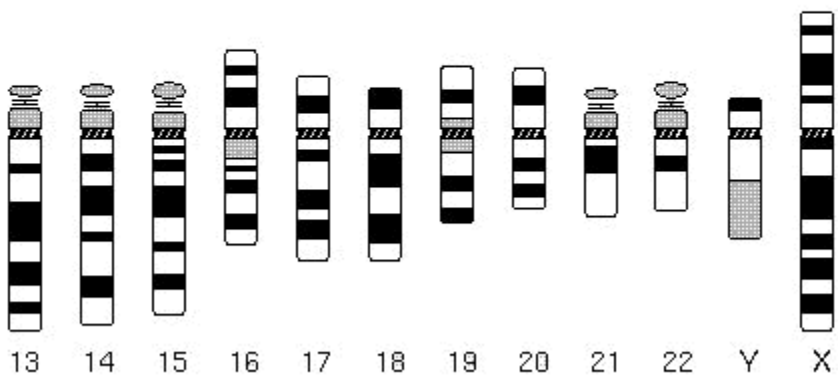
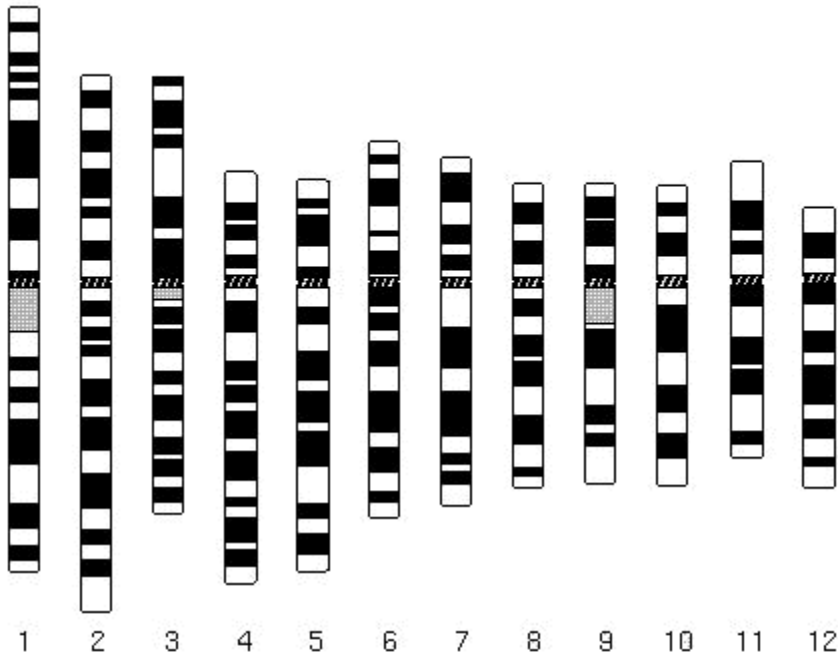
Genetics Lecture 7

Cytogenetics

The normal human chromosome complement

We already had a look at chromosomes in [lecture 1](#) and the following terms should be familiar:

46 XX and 46 XY



Human cells are diploid, that is they contain two of (almost) every gene. They do so by having two copies of each autosome, (chromosomes 1-22) and two sex chromosomes (either XX or XY). The normal human karyotype when viewed down the microscope at mitotic metaphase is thus either 46 XX or 46 XY. (Meaning 46 coloured blobs, two of which are XX or XY). This picture shows a normal male mitotic metaphase spread next to an interphase nucleus.

The **primary constriction** is the centromere, visible in the above picture as the point where the two chromatids remain attached, but also containing the **kinetochore**, the point of spindle attachment. Secondary constrictions are usually only found as the stalks connecting

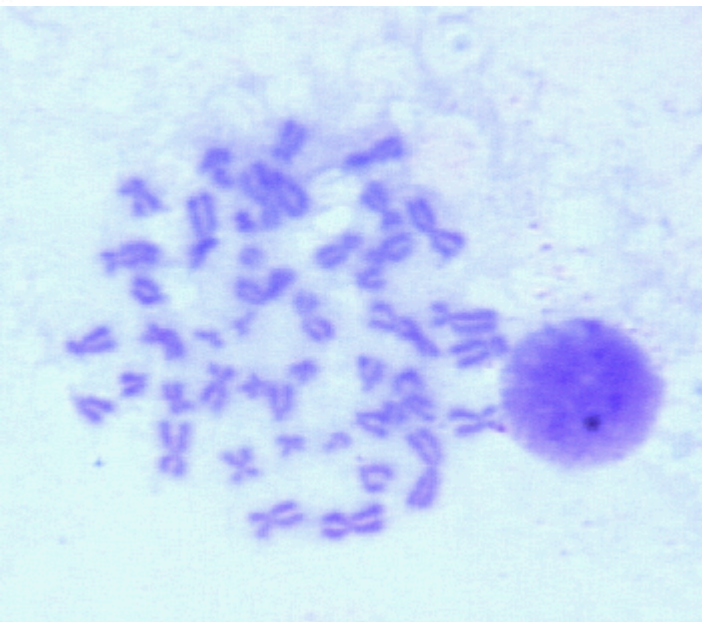
the short arms of the two groups of acrocentric chromosomes.

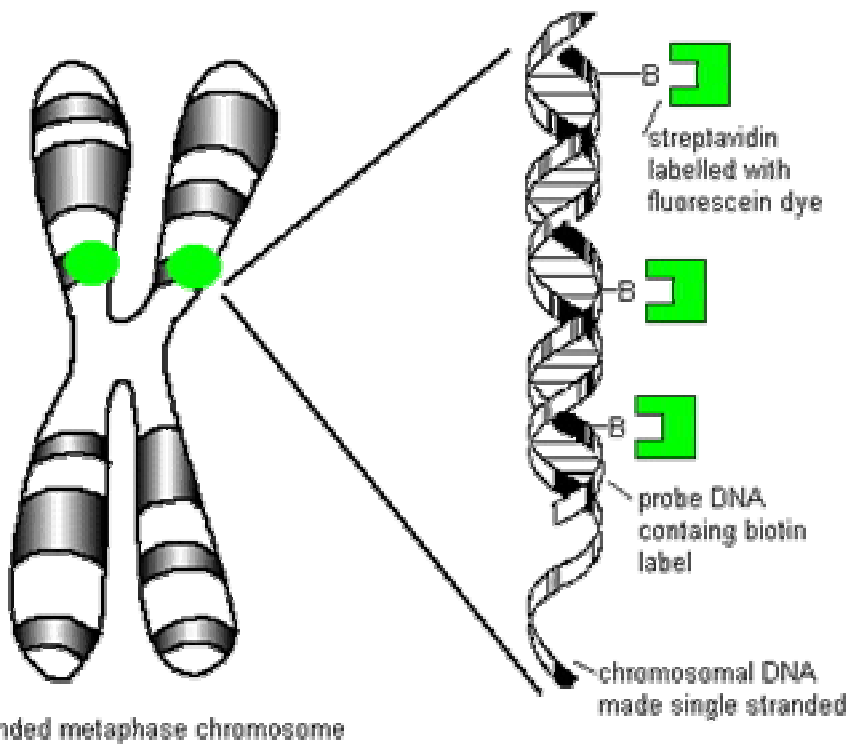
banding

When microscopes were improved to the point that the human karyotype could be reliably discerned (in the 1950s) the chromosomes could be grouped on the basis of their relative sizes and the relative lengths of their two arms, i.e. the positions of their centromeres. Now, banding techniques make it possible to identify each chromosome.

If chromosomes are treated briefly with proteinase before staining then each chromosome has a characteristic banding pattern.

The two chromosome arms are referred to as p and q (short and long respectively). Bands are numbered from the centromere. As microscopes improved the coarse banding patterns





were refined by the addition of further levels of numbering so that band 9q34.1 means the 1st subband of the 4th subband of the 3rd band of the long arm of chromosome 9! Some regions of the chromosomes are uniformly staining and late replicating. They contain few (if any) genes and are composed primarily of tandemly repeated DNA sequences (satellite DNA). These are called **constitutive heterochromatin**. (The inactive X chromosome in female cells is also late replicating and is called facultative heterochromatin.)

Centromeres, the points of attachment of the replicated chromatids to each other and to the mitotic spindle fibres are all within constitutive heterochromatin.

FISH



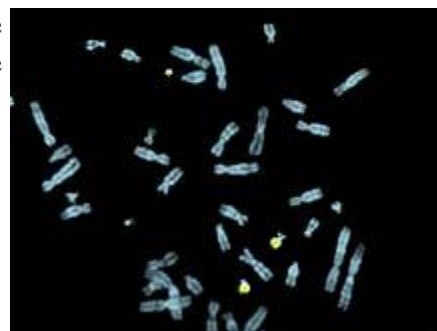
No, not this sort, but "**Fluorescent In Situ Hybridisation**".

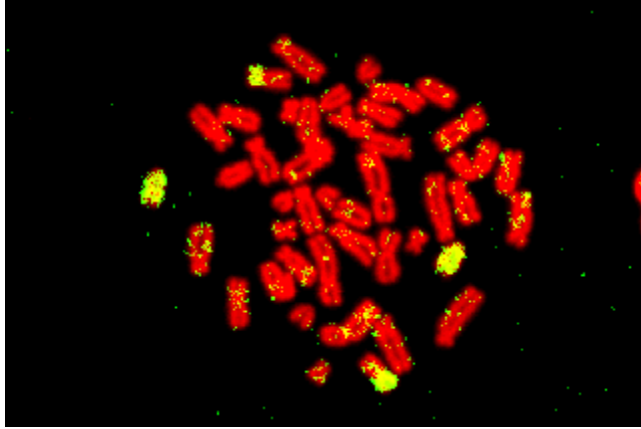
The technique of DNA-DNA hybridisation has been discussed in lecture 5 under the heading of [Southern blotting](#). The ability of a single stranded DNA molecule to find and bind to its complementary strand is pretty amazing. In FISH it is exploited to the utmost. It is possible to visualise by hybridisation, the site of a fragment of DNA of as little as 1 - 2 kb (but more usually 40 - 50 kb) at an efficiency approaching 100%. The probe molecule is labelled with a hapten such as biotin. The biotin is located with streptavidin. The streptavidin is located with antibodies. A fluorescent dye may be conjugated to the streptavidin and to the antibodies. When the spread chromosomes are illuminated by a UV lamp, a point of fluorescence can be seen where the probe / streptavidin / antibody / fluorescent dye multilayer sandwich has built up. From three to five layers of fluorescent antibodies are built up to amplify the signal.

For examples [click here](#)

chromosome painting

Complex probes made from entire chromosomes can be made. In this way the presence that chromosome can be ascertained and also whether it has been subject to any rearrangements. This picture shows a probe made from chromosome 22 which has been used to "paint" a cell





line which has a small mysterious extra chromosome. The probe hybridises to this as well as to the two normal chromosome 22s showing that the small "marker" chromosome is derived from chromosome 22.

prenatal diagnosis

If there is some reason to suspect that an embryo may have abnormal chromosomes, for instance maternal age or past history of early spontaneous abortions, it is usual to check.

- At about 12 - 14 weeks of gestation it is possible to obtain foetal cells by **amniocentesis**. In this technique a needle is inserted into the amniotic sac and fluid is drawn off. Foetal cells, amniocytes, are suspended in the fluid and these can be cultured and their metaphase chromosomes examined. One of the commonest chromosome abnormalities is Down syndrome, the figure links to the relevant passage later in this lecture.
- At 9-10 weeks of gestation, it is possible to aspirate via the cervix, a small amount of the placental tissue, the technique of **chorionic villus sampling**. This tissue can be cultured and chromosomes prepared.
- In fertility clinics embryos are often conceived by artificial insemination. A single, totipotent, cell from an 8 cell stage embryo can be examined directly and the information can be used to decide on which embryo to reintroduce into the mother.

relationship to other organisms

If a paint is made from a human chromosome it can be applied to the chromosomes of other species. This image is of marmoset chromosomes to which a human chromosome 8 paint has been applied. This shows homology with the short arm of marmoset chromosome 13, and the whole of marmoset chromosome 16. *Sherlock et al., (1996) Homologies between human and marmoset (callithrix jacchus) chromosomes revealed by comparative chromosome painting. Genomics 33: 214-9*

numerical abnormalities

One of the commonest mutations is a change in the chromosome number but it is also one of the most damaging occurrences. Very few mutations which cause visible changes in the autosomes are compatible with life. Connor and Ferguson-Smith make the analogy that if the length of the human haploid genome was drawn stretching from London to New York, the smallest visible deletion (about 4Mb) would represent about an 8 km gap and that on this scale, the average gene would be about 30 m long. So even the smallest gap will usually contain many genes. About 20% of conceptions have some sort of chromosomal disorder but because of the lethal effects of such disorders, the number actually born is only about 0.6%.

A wonderful resource for the next section is the [University of Tokyo Medical School](#) where someone has laboured long and hard to make all sorts of moving images to explain chromosome rearrangements (and lots of other things too, [BROWSE....](#))

chromosome abnormalities in early spontaneous abortions	
Defect	frequency
Triploidy	10%
tetraploidy	5%
Trisomy	30%
Turner syndrome (45 X)	10%
Other	5%
Total	60%

euploidy

Euploidy is the category of chromosome changes which involve the addition or loss of complete sets of chromosomes.

- triploidy

The possession of one complete extra set of chromosomes is usually caused by **polyspermy**, the fertilisation of an egg by more than one sperm. Such embryos will usually spontaneously abort.

- tetraploidy

This is usually the result of a failure of the first zygotic division. It is also lethal to the embryo. Any other cell division may also fail to complete properly and in consequence a very small proportion of tetraploid cells can sometimes be found in normal individuals.

aneuploidy

Aneuploidy is the category of chromosome changes which do not involve whole sets. It is usually the consequence of a failure of a single chromosome (or bivalent) to complete division.

- monosomies

All autosomal monosomies are lethal in very early embryogenesis. They do not even feature in the table above because they abort too early even to be recognised as a conception.

- Down syndrome, trisomy 21

The incidence of trisomy 21 rises sharply with increasing maternal age.

Incidence of Down's syndrome with increasing maternal age		
	Incidence	
Maternal Age	At chorionic villus sampling	At Birth
20	~1 in 750	1 in 1500
30	~1 in 450	1 in 900
35	1 in 240	1 in 400
40	1 in 110	1 in 100
45	1 in 13	1 in 30



Most cases arise from non disjunction in the first meiotic division, the father contributing the extra chromosome in 15% of cases. A small proportion of cases are mosaic and these probably arise from a non disjunction event in an early zygotic division. About 4% of cases arise by inheritance of a translocation chromosome from a parent who is a balanced carrier. The symptoms include characteristic facial dysmorphologies, and an IQ of less than 50. Down syndrome is responsible for about 1/3 of all cases of moderate to severe mental handicap.

- Trisomy 13

The incidence is about 1 in 5000 live births. 50% of these babies die within the first month and very few survive beyond the first year. There are multiple dysmorphic features. Most cases, as in Down's syndrome, involve maternal non-disjunction. Again, a significant fraction have a parent who is a translocation carrier.

- Trisomy 18 (Edward's syndrome)

Incidence ~1 in 3000. Again most babies die in the first year and many within the first month.

- sex chromosome aneuploidies



Because of X inactivation and because of the paucity of genes on the Y chromosome, aneuploidies involving the sex chromosomes are far more common than those involving autosomes.

- Turner syndrome, 45, X

The incidence is about 1 in 500 female births but this is only the tip of the iceberg, 99% of Turner syndrome embryos are spontaneously aborted. Individuals are very short, they are usually infertile. Characteristic body shape changes include a broad chest with widely spaced nipples and may include a webbed neck. IQ and lifespan are unaffected.

- Klinefelter's syndrome, 47, XXY

The incidence at birth is about 1 in 1000 males. Testes are small and fail to produce normal levels of testosterone which leads to breast growth (**gynaecomastia**) in about 40% of cases and to poorly developed secondary sexual characteristics. There is no spermatogenesis. These males are taller and thinner than average and may have a slight reduction in IQ. Very rarely more extreme forms of Klinefelter's syndrome occur where the patient has 48, XXXY or even 49, XXXXY karyotype. These individuals are generally severely retarded.

- 47, XYY

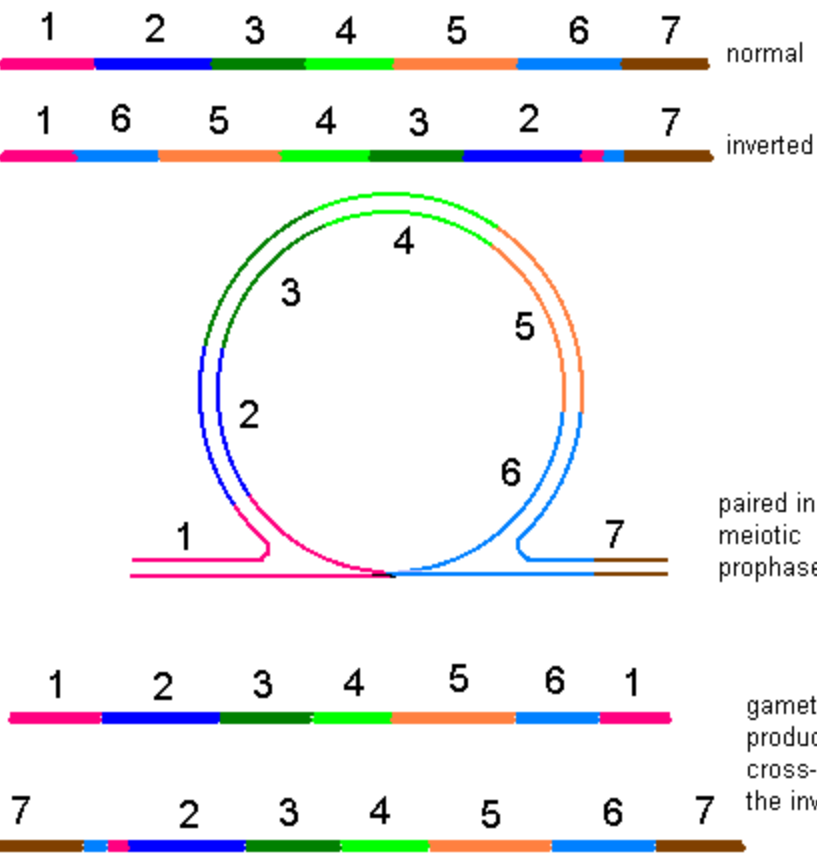
Incidence 1 in 1000 male births. May be without any symptoms. Males are tall but normally proportioned. 10 - 15 points reduction in IQ compared to sibs? More common in high security institutions than chance would suggest? Strangely, although they are fertile they do not seem to transmit the either this condition or Klinefelter's syndrome.

- XXX females

About one woman in 1000 has an extra X chromosome. It seems to do little harm, individuals are fertile and do not transmit the extra chromosome. They do have a reduction in IQ comparable to that of Klinefelter's males.

structural aberrations

From time to time a cell sustains damage from for instance an energetic cosmic ray particle passing through and leaving behind a trail of ionisation. This may lead to chromosome breakage. The repair systems in the nucleus will do their best to make good the damage and if this involves only one break they may be able to do so with no errors.



However, if more than one break has occurred they may become rejoined in the wrong combinations. This can lead to one of a spectrum of possibilities.

balanced rearrangements

Rearrangements where there is no visible loss or gain of genetic material are **balanced**. They include:

- inversions, peri- and para-centric

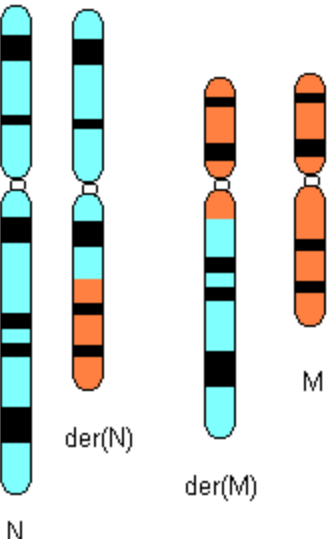
In an inversion, a piece of chromosome is lifted out, turned around and reinserted. If this includes the centromere

then the inversion is termed **pericentric**. If it excludes the centromere then it is a **paracentric** inversion. The two have slightly different genetic consequences. 1% of the UK population are heterozygous for a pericentric inversion of chromosome 9. This is absolutely without genetic consequences.

At meiosis heterozygous inverted chromosomes have difficulty pairing and can only do so by the formation of a loop. If a recombination event occurs within the inverted loop the consequence will be a duplication and a deletion, see the gametes drawn at the bottom of the figure. If the inversion is paracentric then the centromere itself may be duplicated (which gives rise to a dicentric fragment which will try to go to both poles at anaphase I with dire consequences) or deleted (giving rise to an acentric fragment which gets left behind on the metaphase plate).

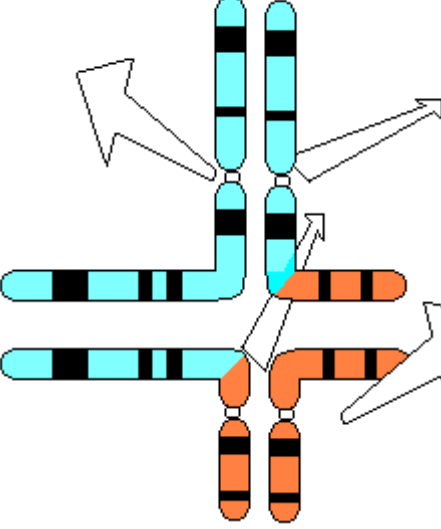
- translocations

reciprocal translocation between chromosomes N and M



In a balanced translocation there is no net gain or loss of chromosomal material, two chromosomes have been broken and rejoined in the wrong combination. The figure shows a translocation between the imaginary chromosomes "M" and "N". Balanced reciprocal translocation is unlikely to have any severe consequence for the cell because, even if one of the breakpoints lies within a gene, most mutations are recessive ([lecture 6](#)).

It is possible that an oncogene may be activated by the translocation and this can lead to cancer. The classic example is the chromosome 9 / chromosome 22



reciprocal translocation in chronic myeloid leukaemia. (Discussed in the [final lecture](#)).

Translocations can however give rise to difficulties with reproduction. During the first meiotic prophase, the chromosomes align in pairs to form bivalents. However, a heterozygote for a reciprocal translocation forms instead a 'tetravalent'. The chromosomes will segregate in the first meiotic division. Many possibilities are open, only one, the one shown in the figure, leads to balanced gametes. This is known as "alternate segregation" where the two intact chromosomes must both move to one pole and the two translocation chromosomes must both move to the other.

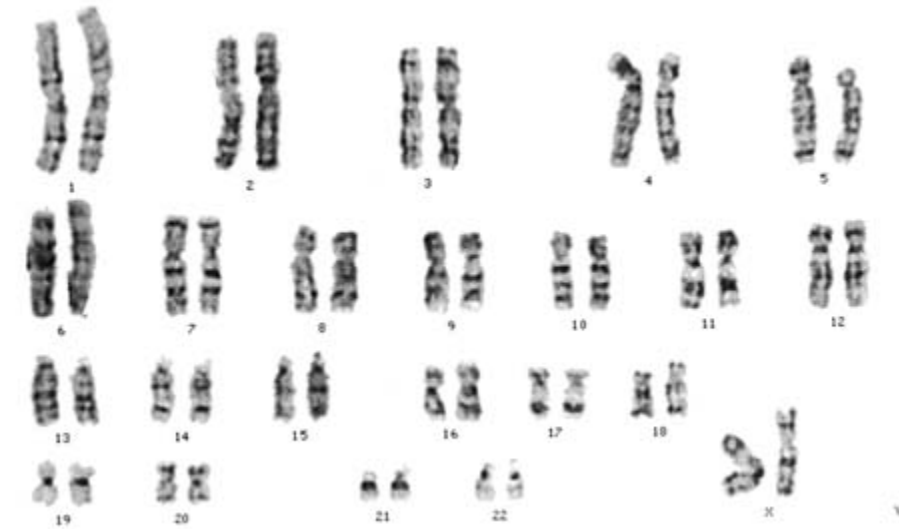
unbalanced rearrangements

- deletions

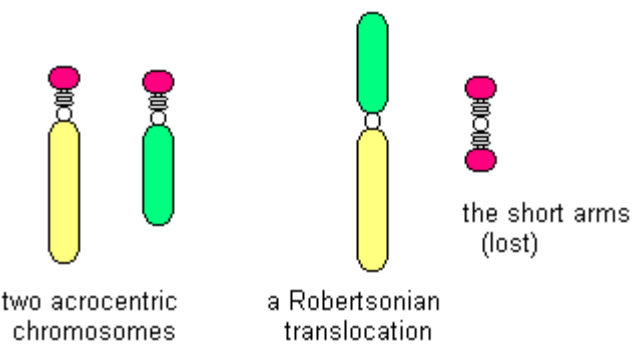
Deletions may either be either interstitial or terminal. If big enough to be visible a deletion must be removing many genes and will probably give rise to a severe phenotype. [See above](#). An example of an interstitial deletion is the 15q⁻ deletion which causes either [Angelman or Prader-Wili](#) syndromes. Many independent deletions in this region do indeed remove many genes including the two responsible for the two syndromes. You will find many more examples, try for instance [searching OMIM](#) with the acronym WAGR (which stands for **W**ilms tumour, **A**niridia, ambiguous **G**enitalia and mental **R**etardation). See also the table

on page 129 of Connor and Fergus on-Smith

Terminal deletions have only one breakpoint



oint, they extend to the telomere. For example the karyotype shown on the right is of a baby with Cri du chat syndrome in which a small part of the distal region of the short arm of chromosome 5 is deleted. Babies with this syndrome have a combination of symptoms which include pinched facial features, mental



retardation and developmental delay. The characteristic feature, for which the syndrome is named is a "mewing" cry. The characteristic cry may be separated genetically from the facial dysmorphism and developmental delay, since a small terminal deletion may have the cry only whereas a larger deletion extending further towards the centromere will include the genes whose hemizyosity is responsible for the other symptoms.

- translocations

An unbalanced translocation may arise spontaneously and is also likely to arise as an offspring of a balanced carrier. There are likely to be symptoms which may be severe. Their exact nature will be unpredictable. Such translocation chromosomes are extremely useful to science in helping to pinpoint genes responsible for the conditions expressed by their bearers.

- Robertsonian

Robertsonian translocations are a special case of 'almost balanced' translocations. Robertsonian translocations involve any two out of chromosomes 13, 14, 15, 21 and 22. These chromosomes are all **acrocentric**, that is, the centromere is very close to one end. The short arms contain few, if any, genes except for many tandemly repeated copies of the ribosomal RNA genes. Every diploid cell thus contains 10 copies of the block of repeated genes. A Robertsonian translocation is a fusion between the centromeres of two of these chromosomes with loss of the short arms forming a chromosome with two long arms, one derived from each chromosome. The loss of the short arms does not matter, each cell still has eight copies of the rRNA gene block and that, apparently, is enough. In the same way as balanced reciprocal translocation carriers have difficulties at meiosis in ensuring the correct segregation of the chromosomes to make a balanced set, so too do Robertsonian translocation carriers. In their case the chromosomes pair in meiotic prophase to form a trivalent and balanced gametes will only be formed when the translocation chromosome goes to the opposite pole to both of the normal chromosomes. Robertsonian carriers therefore suffer a similar reduction to their fertility as do carriers of reciprocal translocations and couples in which one of the partners has a translocation may have a number of early spontaneous abortions. Robertsonian translocations involving chromosome 21 possess a special problem of their own, one of the possible unbalanced gametes will contain effectively two copies of chromosome 21 (when the translocation chromosome and chromosome 21 segregate to the same pole). They are thus at risk of producing a baby with [Down syndrome](#).

- isochromosomes

A chromosome can split "the wrong way" in mitosis (or meiosis II) so that both long arms remain attached and move to one pole, and both short arms do likewise moving to the other pole. The consequence is the formation of an *isochromosome*. These are simultaneously duplicated for the genes in the retained arm and deleted for the genes in the other. The prognosis is poor except for iXq (isochromosome of the long arm of the X).

- ring chromosomes

A mutation event which removes both telomeres can be repaired by sealing the ends together forming a ring chromosome. This will be deleted for genes at both ends of the chromosome. the symptoms will depend on the extent of the deletion. Surprisingly, ring chromosomes are mitotically stable. One might expect them to get hopelessly entangled during DNA replication and at the very least to be concatenated when the time comes to separate to the two poles. While this undoubtedly happens, it is not a frequent occurrence.

Abbreviations used in describing chromosomes and karyotypes



Abbreviation	meaning	example	description
cen	centromere		
del	deletion	46, XX, del(5)(p15.3)	Cri du chat syndrome deletion of band 5p15.3
der	derivative chromosome	der(22)	a chromosome derived from chromosome 22 and containing the chromosome 22 centromere
dup	duplication	46, XY, dup(21)(q21)	A male with a duplication of band q21 on chromosome 21
inv	inversion	46, XX, inv(9)(p12q13)	A female with a pericentric inversion on chromosome 9
mar	marker chromosome	47, XX, +mar	A female with an extra (unidentified) chromosome
mat	of maternal origin	46, XY, del(15)(q12)(mat)	A boy with Angelman syndrome
pat	of paternal origin	46, XX, del(15)(q12)(pat)	A girl with Prader-Willi syndrome
t	translocation	46, XY, t(1;2)(p21;p24)	A boy with a reciprocal translocation between band p21 on chromosome 1 and band p24 on chromosome 2
ter	the end of the chromosome arm	pter or qter	
+	gain of material	47, XY, +21	A boy with trisomy 21
-	loss of material	46, XX, 5p-	A less precise definition of the girl with Cri du chat above
/	mosaic	46,XX / 45, X	A mosaic case of Turner's syndrome

Recommended reading

The topics include:

- The normal human chromosome complement
- FISH and chromosome painting
- Autosomal and sex chromosome aneuploidies
- Balanced and unbalanced structural rearrangements

Reading:

- Lewis Chapter 11 Another MUST READ chapter! 
- Mange and Mange Chapter 2 (pp24 - 27) but especially Chapters 13 and 14 - two very interesting and well written chapters. Well worth reading. 
- Mueller and Young Chapter 17 Drier account - but packed with facts.
- Jorde *et al.* Chapter 6
- Thompson McInnes and Willard Chapters 9 and 10 Well written and containing all the information and more.
- Connor and Ferguson-Smith Chapters 6 and 13 Chapter 6 is particularly clear, 13 is similar in content to Mueller and Young.

SAQs

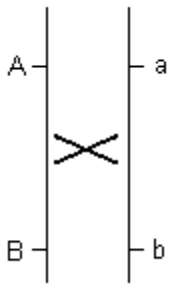
1. The Xg blood group gene, XG is located near the tip of the short arm of the X chromosome, just distal to the genes for [Kallmann syndrome](#) and [steroid sulphatase](#) (STS) gene. It has two alleles, one of which, XG^+ , codes for the production of an antigen on the surface of red blood cells, the other, XG^- , does not. A two year old boy, whose mother is Xg^+ , is unable to smell anything, has hypogonadism, a small penis and ichthyosis. On testing he turns out to be Xg^- . What possible cause can you suggest? What genetic tests might you carry out?
2. If heterozygosity for an inversion chromosome leads to problems with reproduction, would homozygosity be likely to give rise to worse problems?
3. Patients with the karyotype 46, X iXq have Turner's syndrome. Why?

[Answers](#)

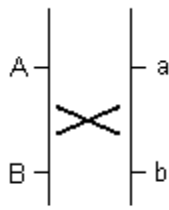
[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)



genes far apart,
crossovers are
relatively frequent



genes close together,
crossovers are less
likely to occur between
them

Genetics Lecture 8

Linkage

What is it?

Mendel's second law (of independent assortment) breaks down in one important way - when two genes lie close together on the same chromosome. In this case, alleles which were inherited from one parent tend to stick together when transmitted to the next generation because they are part of the same DNA molecule. However, the further apart two genes are along the molecule, the more likely it is that a meiotic crossover can occur between them and thus create a new combination of alleles. By measuring the frequency of recombinant chromosomes in the progeny, we can estimate the distance that separates the two genes and can make a **genetic map**.

Such maps are not made just to satisfy scientific curiosity, they have direct application both in science and in medicine.

- First, by mapping the gene responsible for a genetic disease, we can identify it even when we have little or no idea what it does. For examples, see the [cystic fibrosis cloning story](#) or the cloning of *TSC1* (*M. van Slegtenhorst et al. Identification of the tuberous sclerosis gene TSC1 on chromosome 9q34. Science 277:805-808, 1997.*).
- Also, even before a gene is identified, if it can be mapped with respect to other genes or to anonymous segments of DNA (known as **markers** or **marker loci**), then information gained from studying the linked loci can be used to deduce, for instance, whether or not a person is a carrier of a mutant gene or to give antenatal diagnosis as to whether an embryo is affected with the disease.

The mathematics of linkage analysis can be frightening. However, the fundamentals of the subject are easy to grasp particularly if we start by studying experimental organisms in which we can set up those crosses which we need. Later on we must consider how to make the same type of measurements in humans whose "crosses" cannot be experimentally manipulated but where we must make deductions from families in which genetic diseases (or interesting variants) are present.

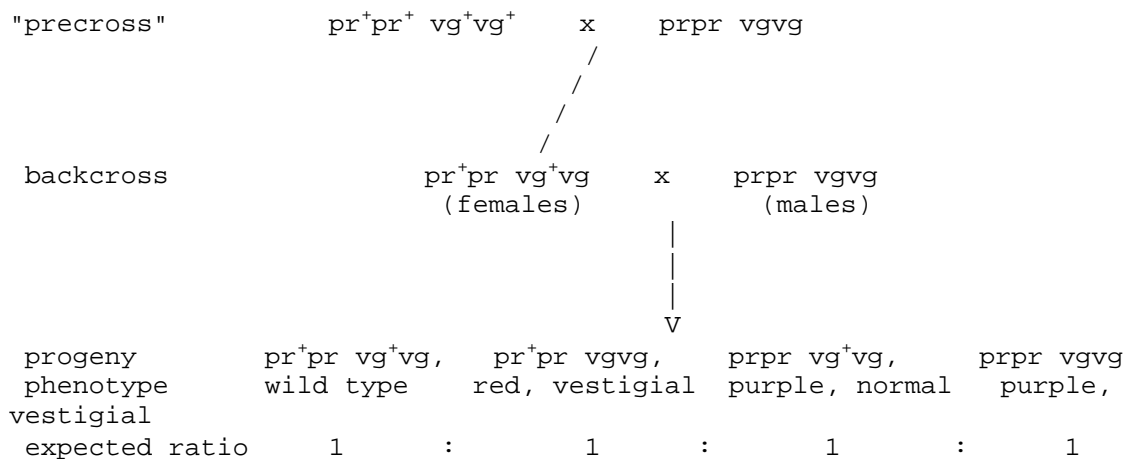
Measurement in experimental organisms such as *Drosophila*

The first observation of data from a dihybrid cross which did not fit the predicted Mendelian 9:3:3:1 ratio were made by Bateson and Punnett investigating genes affecting flower pigmentation and seed shape in sweet peas. However, the first thorough investigation was conducted by **Thomas Hunt Morgan** using the fruitfly *Drosophila melanogaster*. To minimize the variables he considered a cross between fruitflies in which the females were heterozygous at two loci, the genes *purple* and *vestigial* (one affecting eye colour and the other wing shape) while the males were homozygous for each mutant gene. The mutant alleles are written *pr* and *vg* respectively and the "wild type" alleles are written *pr*⁺ and *vg*⁺.



A female *Drosophila melanogaster* laying an egg. Note her (wild type) red eye colour. From *Microsoft Encarta Encyclopedia*

Morgan's First Cross:

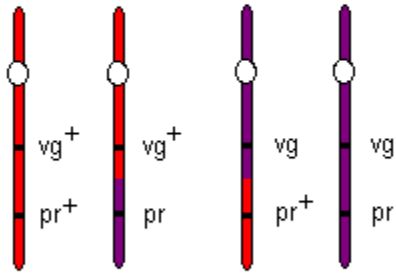
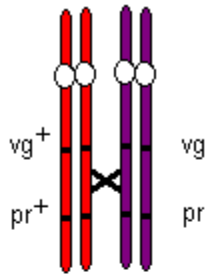


The table shows the actual progeny observed:

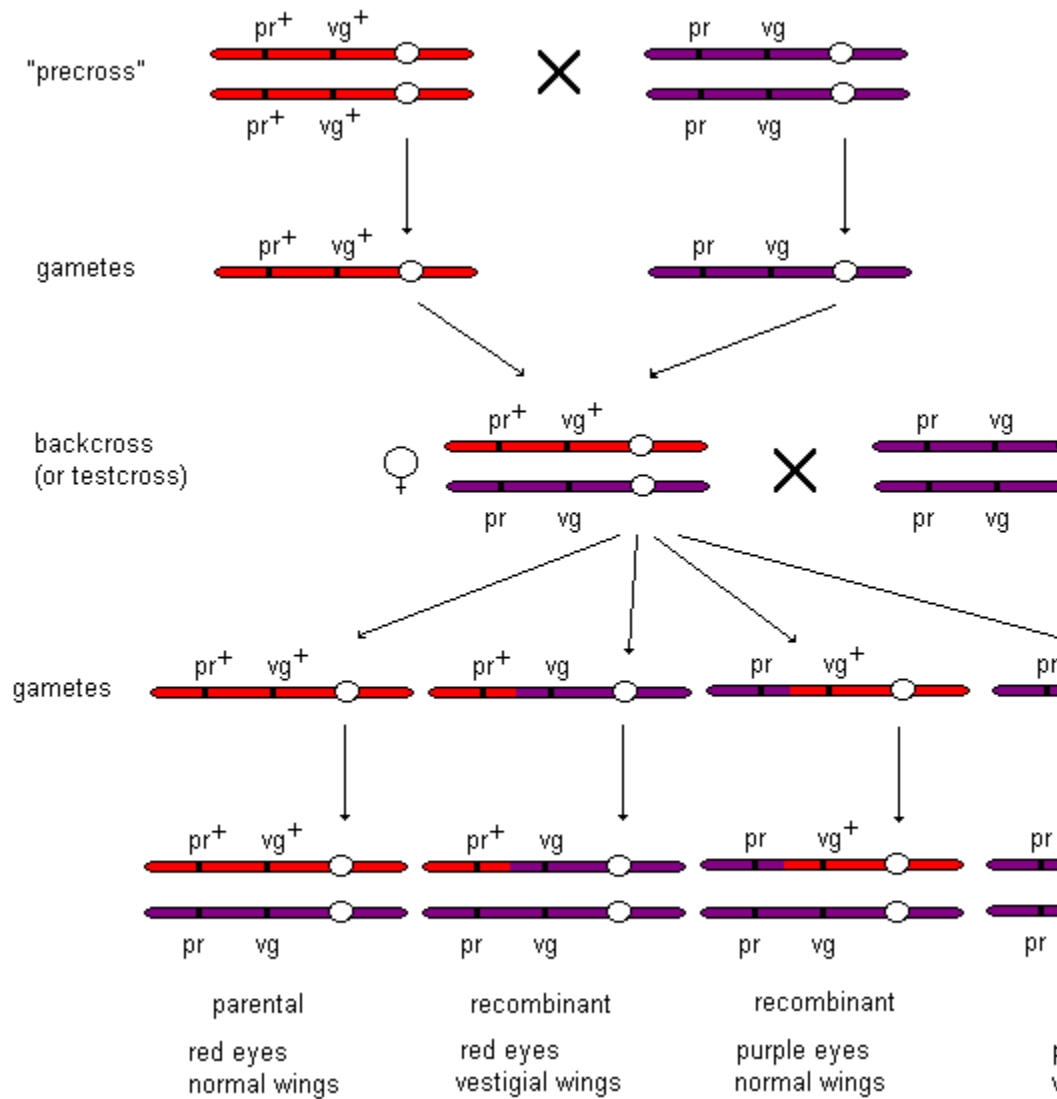
Backcross to measure linkage between <i>pr</i> and <i>vg</i>			
Phenotype	female gamete	Observed	Expected
wild type	pr ⁺ , vg ⁺	1339	709.75
red, vestigial	pr ⁺ , vg	151	709.75
purple, normal	pr, vg ⁺	154	709.75
purple, vestigial	pr, vg	1195	709.75

Even without doing any statistics on this result it is obvious that it is very far from the Mendelian expectation.

The reason is that the two genes under consideration are on the same chromosome as shown below in the diagram and the parental combinations of alleles remain together unless separated by a recombination event.



When there is a crossover, half the resulting gametes are still non-recombinant. So, even if a crossover occurred every meiosis, half of the gametes would still be "parental". Hence θ can at most be 0.5.



The closer together two genes are, the more likely they are to be inherited in the **parental** combination.

We can quantify this statement. We define the **Recombination Fraction**, theta.

$$\theta = \frac{\text{number of recombinant progeny}}{\text{total number of progeny}}$$

Theta must lie in the range from 0 to 0.5. A value of 0 means that the two genes are so close to each other that they never recombine, a value of 0.5 implies that the genes are unlinked. (The maximum value of 0.5 and not 1 comes from the fact that crossing over takes place after DNA replication and involves only one chromatid per chromosome of the pair, see diagram.) In this case, the recombination fraction between the *purple* and *vestigial* genes is $(151+154) / (1339+151+154+1195) = 0.107$

The linkage map

Because genes are organized along a linear structure, a DNA molecule, we could expect to create a linear map if we measure the frequencies of recombination between them. We make the recombination frequency the measure of the genetic distance between loci. The map unit corresponds to 1% recombination and is named a **centiMorgan** in honour of Thomas Hunt Morgan. So the *purple* and *vestigial* genes above are 10.7cM apart.

SAQ

- Here is a table of recombination frequencies for pairs of X linked genes in *Drosophila*. Use the information to construct a map of the chromosome.

gene 1	gene 2	recombination frequency
yellow	white	0.010
yellow	vermilion	0.322
white	vermilion	0.297
vermilion	miniature	0.030
white	miniature	0.337
white	rudimentary	0.452
vermilion	rudimentary	0.269

Answer

Three point crosses

The data in the question above come from seven separate crosses each involving only two genes. If a cross is set up involving more than one pair of genes it can confirm the approximately linear nature of the map and, by consideration of the relative frequencies of different classes of offspring it will also show directly the order of the genes.

Here is an example of a cross set up by Sturtevant between *Drosophila* females heterozygous for *yellow*, *white* and *miniature* and males hemizygous for the same three mutations. Notice that in this case the genetic notation has changed to an alternative form. When it is quite clear which gene we are referring to, it is acceptable simply to refer to the wild type allele as + instead of for instance, y^+ or w^+ etc.

female gamete	recombination?	number of offspring	frequency
+++	Parental	3501	0.664
ywm		3471	

++m	between w and m	1754	0.329
yw+		1700	
y++	between y and w	28	0.0057
+wm		32	
y+m	double recombination	6	0.00086
+w+		3	

The male gamete received by any progeny in the above cross is either an X chromosome bearing $y w m$ or a Y chromosome, in either case the genotype of the female gamete is obvious. The order of the genes is as written but it could have been deduced by considering which is the rarest class of female gametes. That must have been due to double recombination and its frequency will be, at most, the product of the frequencies of the two single crossovers. In fact, because the presence of one crossover reduces the likelihood of any other in its neighbourhood, the observed frequency of double crossovers is less than this.

The map distances deduced from the above data are:

$$y \text{ to } w, 100 * (28 + 32 + 6 + 3) / 10,495 = 0.66 \text{ cM}$$

$$w \text{ to } m, 100 * (1,754 + 1,700 + 6 + 3) / 10,495 = 33\text{cM}$$

$$y \text{ to } m, 100 * (1,754 + 1,700 + 28 + 32) / 10,495 = 33.48 \text{ cM}$$

Recommended reading

The topics include

- estimation of recombination frequency
- genetic maps
- the centiMorgan

Read the relevant chapter in any basic (not specifically aimed at human) genetics textbook. One good possibility is *Weaver and Hedrick, Genetics, 3rd edition, pub. W.C. Brown, ch. 5 pp104 - 113*. See the end of the next lecture for the relevant material from the recommended course text books.

More SAQs

2. Try setting up a cross using the "[Virtual Fly Lab](#)" to measure the genetic distance between *white* (eyes) and *miniature* (wings). (Or try [here](#) if the other URL is slow)

to reply.) This activity is best performed in the morning before America wakes up.

[Answers](#)

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 9

More about linkage

Human linkage maps

The size of the human genome is big enough that it took a long while to discover any evidence of genetic linkage. It was not until the 1950s that the first autosomal linkage groups were discovered. These all involved the polymorphic blood groups. One of the first was between one form of [hereditary elliptocytosis](#) (an anaemia caused by malformed erythrocytes) and the [rhesus blood group](#). This study was important in showing that linkage information could prove the existence of more than one form of the same disease, it showed that there were some families in which the disease elliptocytosis was clearly linked to rhesus but that there were others in which it was not. This implies the existence of at least two genes which when mutant could cause the same disease ([genetic heterogeneity](#)).

relationship to the human genome project

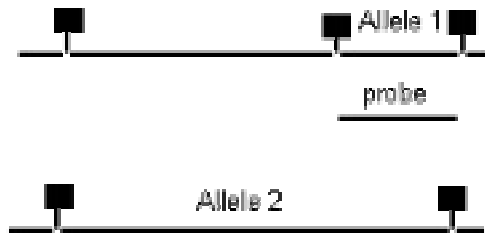
It was not until the discovery of extensive inherited variation in DNA sequence which could be traced in families using [Southern blotting](#) or by [PCR](#) that human linkage studies really took off.

- It became possible to construct detailed genetic maps of the human genome which have provided a crucial skeleton of genetic landmarks onto which the human genome project is fitting DNA clones which will be sequenced.
- The detailed genetic map has made it now a matter almost of routine to position a genetic disease gene onto the genome map. This is the first stage in *positional cloning*, the normal route these days to identifying a disease gene.

types of genetic 'marker'

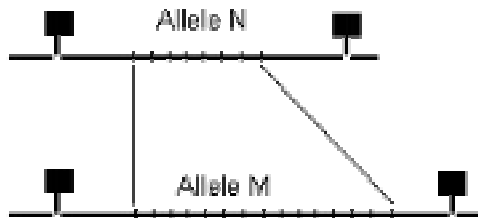
Any genetic variation can be studied and it is not necessary that the objects being measured in a linkage study are genes. Any *polymorphic* piece of DNA can be studied. Examples include

- Restriction Fragment Length Polymorphisms (RFLPs) caused by the presence or absence of a restriction site. The two alleles can be distinguished from each other by [Southern blotting](#).



- Variable Numbers of Tandem Repeats (VNTRs) which can include
 - mini satellite repeats (as discovered by Sir Alec Jeffreys) in which the alleles may be identified by Southern blotting or by [PCR](#).
 - microsatellite repeats, (di, tri and tetra nucleotide repeats) in which the alleles are almost always identified by [PCR](#).

All VNTRs have multiple alleles which is a good thing for genetic mapping, since it makes it more likely that you will be able to distinguish between both alleles in any person whom you wish to study.

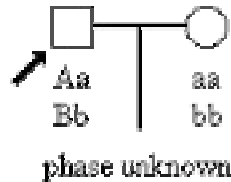
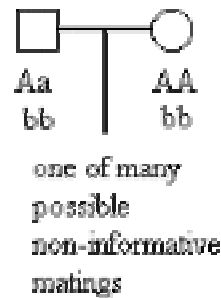
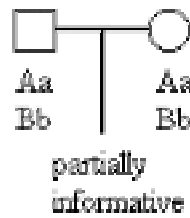
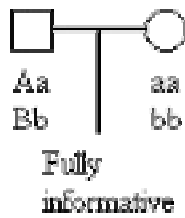


The use of Lod scores

What makes a cross informative?

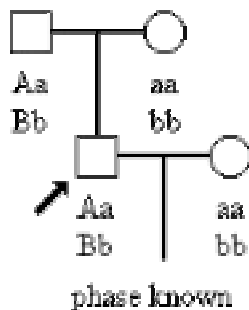
- There must be one (best) or two (not quite so good) parents who is (are) doubly heterozygous for the two loci.

The first thing to say here is that I do not expect you to be able to calculate a **Lod score**. However, you must have a general understanding of what the lod score is because it is an important and often quoted statistic.



In this mating, if genes A and B are on the same chromosome we do not know whether allele A is on the same chromosome as allele B or as allele b

Suppose that we wish to decide whether two genes / markers are linked.

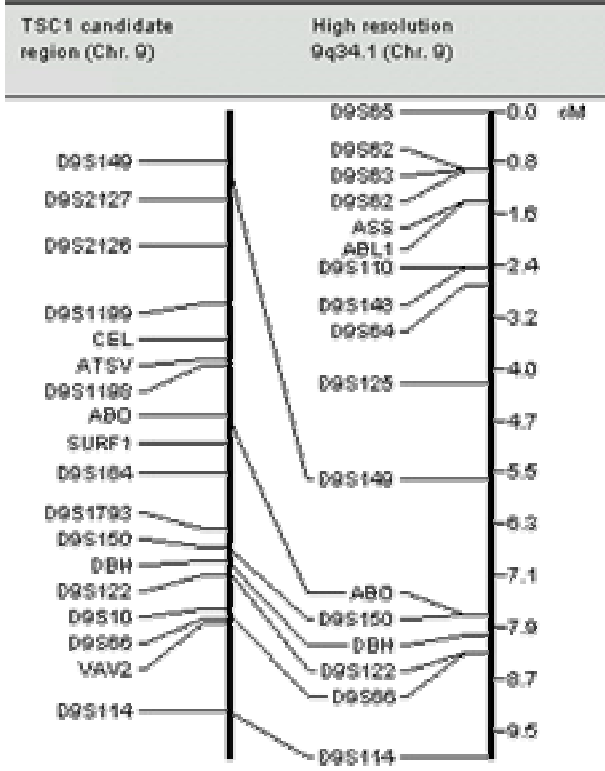
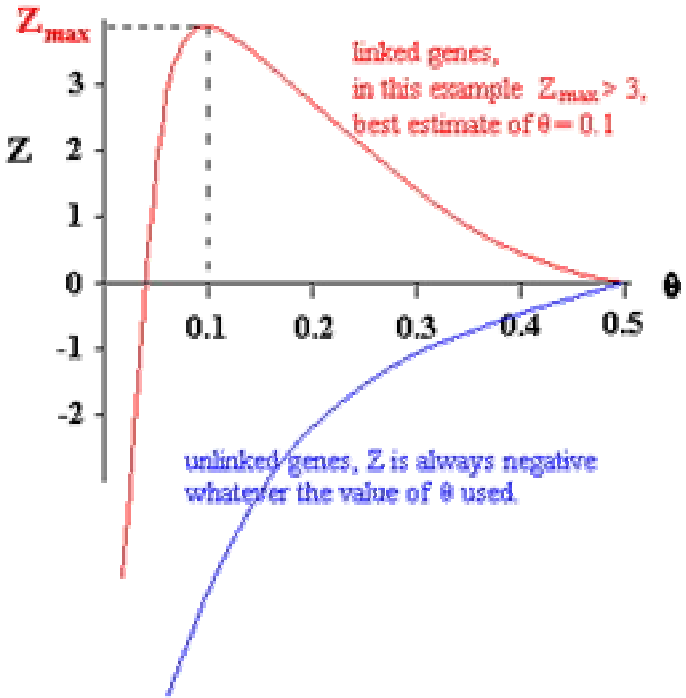


This time, if genes A and B are on the same chromosome we know that alleles A and B are on one chromosome of the pair and that alleles a and b are on the other.

- We examine a number of families which are **informative** for the loci in question.
- For each family we must find the genotype of each parent and of each of the offspring.
- We must try to deduce whether each child has the parental arrangement of alleles or is a recombinant. This is greatly helped if we know how the alleles at the loci which we are considering entered the parents i.e. do we know the **phase**.
- Now for the tricky part, we must decide whether it is more likely that we have observed the particular progeny in that family because the genes are linked and are tending to be inherited together than that we have observed those progeny through the chance (Mendelian) independent assortment of alleles at the two genes. For this we use a statistic known as the **logarithm of the odds** or **Lod score**.
- The Lod score,

$$Z = \log_{10} \left[\frac{\text{The probability of observing this family if the genes are linked}}{\text{The probability of observing this family if the genes are not linked}} \right]$$

- We can calculate this relative likelihood under the assumption that the two genes are tightly linked with no recombination occurring between them at all, i.e. with the recombination fraction $\theta = 0$. Or we can calculate the relative likelihood for any other value of θ that we care to choose. The value of θ for which Z reaches a maximum positive value is then our best estimate of the distance



b
e
t
w
e
e
n

t
h
e

t
w
o

loci. The value of Z_{max} is a measure of our confidence in the result. If $Z_{max} > 3$ then we conventionally take the linkage to have been proved. Conversely, if $Z < -2$ we take the linkage to have been disproved for that value of theta.

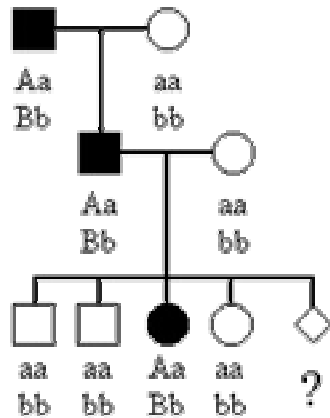
- The data are often displayed in the form of a table of Z calculated for a range of theta from 0 to 0.5. e.g.

θ	0.0	0.1	0.2	0.3	0.4	0.5
Z	$-\infty$	3.9	2.7	1.5	0.5	0.0

Notice in this example that Z has sunk to minus infinity at theta = 0. This will be the case if there has been at least one recombination event between the genes, they cannot then be zero distance apart. It is often helpful to display the data graphically as shown in the red curve on the right. The maximum value of Z occurs at a recombination fraction of 0.1 and it is greater than 3. Therefore we can feel confident that the two genes are indeed linked and our best estimate for the genetic distance between them is theta = 0.1 (equivalent to 10cM). Compare this with the blue curve plotted for two unlinked genes. In this case Z is always negative and the best estimate of theta is 0.5 i.e. unlinked.

The map of the human genome

Many thousands of genetic markers have now been mapped so that there is a minimum of about one marker per centiMorgan and in some regions of the genome there are as many as ten markers per centiMorgan. On the left are two maps covering the same small region of chromosome 9 (a region in which I have a particular interest). The map on the right is a genetic map made by considering the inheritance of many markers (with names such as D9S66) in about 60 large families. The map on the left is not a genetic map, it is a map based on a large number of overlapping pieces of cloned DNA (currently being sequenced) and so it is an accurate reconstruction of the actual DNA sequence of that region of chromosome 9. As is to be hoped, the genetic map gives the same order of



markers and approximately the same relative distances between them. There are some differences, The gene *ABO* is not resolved from marker locus *D9S150* on the genetic map whereas, in physical reality there is a good sized gap. Differences like this occur because the rate of genetic recombination is not absolutely even throughout the genome, some areas are hot spots and others are areas of reduced recombination. This is reflected in the genetic distances between pairs of markers. The total genetic map length of the human genome is about 3,000 cM and by a lucky coincidence, the total genome length is about 3,000 million basepairs. So on average, 1 cM is equivalent to 1 Mb (Mb = million basepairs). In this region of chromosome 9 there is an elevated recombination rate compared to the genome average and so here 1 cM = ~ 300 kb.

Diagnosis of genetic disease through linkage analysis

The identification of linkage to a marker locus is very often the first step on the way to cloning a disease gene (see the next section). However, it also immediately provides diagnostic opportunities even before the disease gene itself has been identified and often when absolutely nothing is known about the nature of the underlying genetic defect. In this pedigree (which you will recognise as being the phase known pedigree above with added children) an autosomal dominant mutation is present. The disease gene has not been identified but it is known to be closely linked to DNA marker polymorphisms Aa and Bb and to map between them. Clearly, in this family the mutant disease gene is



present on the chromosome which happens to have alleles A and B on it. [Remember that Aa and Bb have nothing whatsoever to do with the disease, they are simply two bits of the genome which are polymorphic and which are within about a million basepairs of the disease gene.] The unborn baby can be tested to see which alleles it has inherited from its father at the two marker loci. If it has inherited A and B then it almost certainly has also inherited the disease mutation.

The identification of 'positional candidate' genes

The identification of a disease gene proceeds by about 5 steps.

- First families are "collected" and the clinical investigation is repeated to ensure that there are no misdiagnoses of unaffected individuals, DNA samples are prepared (usually from a 10ml blood specimen).
- When there are enough families to give a chance of a significant positive Lod score, the DNA is amplified by PCR from about 200 polymorphic marker loci spread throughout the genome. That should ensure that one locus is within about 7.5 cM of the disease gene.
- When that initial linkage is found, many more loci from the same region are also tested in the families and recombinants are used to define the genetic interval within which the disease gene must lie. In the *TSC1* example above, the gene was initially defined as being between *D9S149* and *D9S114*. Later this interval was

- reduced to between *D9S2127* and *D9S150* because individuals were found who had inherited only part of the chromosome due to genetic recombination in the formation of the gamete from their affected parent.
- Most areas of the human genome have now been cloned. Clones covering the minimum area are identified and used to find all the genes in the area.
 - Mutations in these genes are sought in the affected patients. Hopefully this should lead to that "Eureka" moment.
 - As more and more of the genome is sequenced already, several of these steps may be bypassed. It may be possible to jump directly from the definition of the genetic interval within which the gene must lie to the mutation screen.


As well as the *TSC1* success story above there have been many other examples of successful identification of disease genes based solely, or almost solely, on positional information. The classic example is the [cystic fibrosis](#) gene *CFTR*.

Recommended reading

The topics include:

- Genetic Linkage
- Genetic maps
- Measurement in human studies
- lod scores
- 'positional candidate' genes
- relationship to the human genome project

Reading:

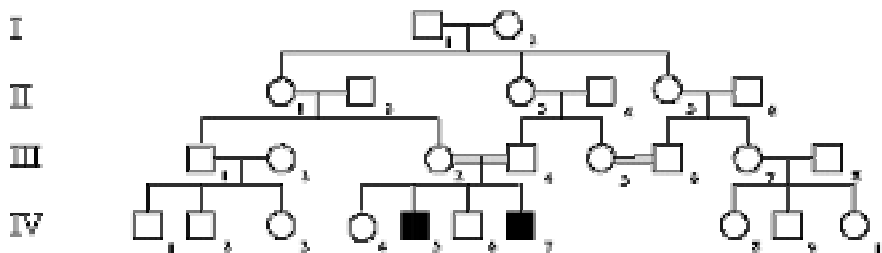
- Lewis Chapter 5 (pp94 - 98) gives information about linkage. It's one area that this otherwise excellent book is rather weak on. Chapter 21 for the Human genome project.
- Mange and Mange Chapter 9
- Mueller and Young Chapter 7 (pp 121 - 125) (It's here but rather condensed)
- Jorde *et al.* Chapter 8. An excellent chapter. 
- Thompson McInnes and Willard Chapter 8 The core material is on pp178 - 186 but the whole chapter is valuable background reading.
- Connor and Ferguson-Smith Chapter 9 The core material is on pp82 - 86 but the whole chapter is valuable background reading. Also read chapter 19, p200, figures 19.3, 19.4 and 19.6 but these might be difficult to understand unless you were keeping well up with the lecture.

SAQs

- 1.
- a) Define a centimorgan.
 - b) What do you understand by a lod score of 3?
 - c) Below is a table of lod scores obtained from a series of family studies. Does it provide any evidence that the genetic markers under consideration are linked in these families and if so, at what distance?

theta	0.0	0.1	0.2	0.3	0.4	0.5
Z	-infinity	-1.3	1.5	3.4	2.6	0.0

- a) What is an RFLP? What makes an RFLP useful for family studies?
 b) A DNA polymorphism with 6 alleles is linked to Vitamin D independent rickets (an autosomal recessive disease) at a distance of 1cM. What information does it provide when counselling the consanguineous couple III₅ and III₆ in the pedigree below who want to know whether their children are likely to inherit the disease?



Genotypes at the marker locus are given below where known:

I ₁	1,2	I ₂	1,3	II ₅	1,3	II ₆	2,4
II ₃	1,2	II ₄	1,3	III ₅	2,3	III ₆	2,3
III ₃	2,3	III ₄	1,2	IV ₆	2,3	IV ₇	2,2
IV ₄	1,3	IV ₅	2,2				

2.

Answers

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

Genetics Lecture 10

Genes in populations

One aspect of clinical genetics which makes it somewhat different to other branches of medicine is that results affecting one individual have ramifications extending to both the close family and to more distant relatives. But this is seen also in reverse; our ability to make predictive diagnoses of risk of affected offspring in a marriage will depend on our knowledge not just of the immediate families from which prospective parents are drawn but also of the populations.

why do genetic diseases vary in frequency between different populations?

founder effect

Some genetic diseases are present at a much higher frequency in one population rather than another. For instance in the Boer population of South Africa the disease [porphyria variegata](#) is much more common than in either the Dutch population from whom the Boers are descended or the surrounding African populations. This interesting disease (which is an example of [haploinsufficiency](#)) is familiar to those of us who have seen the film *'The Madness of King George'* (whose retrospective diagnosis was made in 1966). In South Africa it occurs at a frequency of 1 in 375 in the white population. The mutation responsible has been discovered and has been shown to be the same in almost all cases. All of these can trace their descent to one couple Gerrit Jansz and his wife, Ariaantje Jacobs who were some of the first settlers in the 17th Century. This is a classic example of the **founder effect**. Approximately forty founding families have now expanded until their one million descendants represent one third of the modern white population in the country. The defective gene originally present in either Gerrit or Ariaantje has similarly increased in frequency, as it has not seriously interfered with the ability to leave descendants. There has been some selection against the defective gene and during this time it has decreased slightly in relative frequency (from about 1 gene in 160 to about 1 gene in 250 in the one million descendants who trace their lineage back to the founders.)

A similar situation can be seen in Finland which was mostly settled only from the 17th Century. Today a number of distinct mutations which are rare elsewhere in Europe are present at a high frequency in the Finnish population. An example is one type of [familial colon cancer](#), another example is in one form of [Batten disease](#). The high incidence of [Huntington's Disease](#) in the Venezuelan population living around Lake Maracaibo is another example. All affected people are descendants of a Portuguese sailor who married a local woman in the 19th Century.

Natural selection

Many conditions which we recognise as genetic diseases are actually present in certain populations because, either presently or in the not too far distant past, they actually conferred a survival advantage on their carriers. In most cases the disease is manifested in the homozygote whereas the heterozygote is at an advantage. Malaria has been a powerful selective force. In countries where malaria is endemic are found high

frequencies of many such mutations, for example the sickle cell mutation in the beta globin gene, the thalassemia mutations affecting either alpha or beta globin genes and [favism](#) (glucose-6-phosphate dehydrogenase deficiency).

How are genetic diseases maintained in populations?

In the case of a dominant and a recessive allele at a locus, it is not immediately obvious what will happen to their frequencies in the population as generations go by. Do we expect the recessive allele to diminish in frequency? Also, there will be three genotypes, homozygotes for each of the alleles and the heterozygotes. What will be their relative frequencies?

The answer was provided by a physician, Weinberg and a mathematician, Hardy in 1908. They recognised that, provided certain conditions were met, the allele frequencies (sometimes rather misleadingly called the gene frequencies) would remain constant from one generation to the next and that no matter what the relative frequencies of the genotypes in the starting population, in all subsequent generations these too would be fixed at certain values determined by the allele frequencies. The conditions are listed in the box.

Conditions under which the Hardy Weinberg equilibrium will be met

- The population size must be infinite
- There must be random mating
- There must be no selection either in favour of or against particular genotypes
- There must be no introduction of new alleles by mutation
- There must be no net migration into or out of the population

In practice these conditions are impossible to fulfil exactly because:

- Population sizes are never infinite, however, for practical purposes, once the population size is greater than a few thousand individuals this will not lead to serious departures from the equilibrium conditions. Sometimes the population may be stratified (often invisibly to a casual outside observer) into sub-populations between which individuals rarely interbreed. This can have the effect of greatly reducing the true population size. In a small population, because of chance sampling of the gametes which make the next generation, there is a likelihood that allele frequencies will change every generation. This is known as **genetic drift**. There is also a finite chance every generation that one allele may not be transmitted to the next generation. When this happens it is lost from the population for ever unless reintroduced by mutation or by migration.
- Mating is not always random. For some traits such as blood groups it would be unlikely to find assortative (or disassortative) mating. However, some genetically determined (or partially genetically determined) traits show highly assortative mating. Height and IQ spring to mind. A gene closely linked to a gene

influencing such a trait might find itself subject to non random mating until sufficient time has passed to allow recombination to randomize its alleles with respect to those of the assortatively mating trait.

- As mentioned above, natural selection can act both to remove certain genotypes from the breeding population and also to improve the chances of others to leave offspring.
- Mutation is always occurring and the mutation rate from a functional allele to a non-functional allele will be greater than the reverse. A gain, in practice, because the mutation rate is relatively low this will not upset the equilibrium too much.
- Although it is possible to envisage a scenario in which it might occur, in practice, for most human genetic conditions, it is unlikely that one genotype will tend to migrate to or from a population more than any other.

The Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium can be stated thus:

For a gene with just two alleles **A** and **a**, if the frequency of allele **A** in the population is **p** and the frequency of allele **a** is **q** then

Hardy Weinberg frequencies	
Genotype	frequency
AA	p^2
Aa	$2pq$
aa	q^2

- first, the sum of the frequencies of the alleles must equal 1 ($p + q = 1$) because that's all the alleles added up.
- Secondly, (and this is the important bit) the frequencies of the genotypes **AA**, **Aa** and **aa** will be p^2 , $2pq$ and q^2 respectively. [And incidentally $p^2 + 2pq + q^2$ will equal 1 because that's all the possible genotypes added together and because it's an inevitable mathematical consequence of the fact that $p + q = 1$ (try squaring both sides!).]

Some applications

The Hardy Weinberg equilibrium can be used to estimate genotype and gene frequencies from limited data.

- For example, for a recessive condition: if the frequency of affected people is **f** then the gene frequency $q = f^{1/2}$ and the frequency of carriers will be $2pq$ (and because p is close to 1 this will be more or less $2q$). For cystic fibrosis, the frequency of affected children is about 1 in 2000, therefore $q = [1/2000]^{1/2} = 0.0223$, therefore the frequency of carriers is approximately $2q = 0.0446$ or one in 22.

SAQs

1. In two populations, the gene frequencies of the MN blood group were found to be different as shown in the table:

Population	Frequency of M allele	Frequency of N allele
UK (Somerset)	0.49	0.51
Japan (Kyoto)	0.35	0.65

2. a) What frequency of MN heterozygotes would you expect to find in the Somerset population?
3. b) What frequency of NN homozygotes would you predict in Kyoto?
4. Show your reasoning.
5. The allele frequency of an autosomal DNA polymorphism with two alleles (1 and 2) was investigated in a sample of 100 elderly individuals and 100 infants of an African tribe. The following table of results was obtained:

Genotype	number of adults	number of infants
1,1	33	44
1,2	67	45
2,2	0	11

6. Which of these results fits those predicted by the Hardy Weinberg equilibrium equations? (Show your working.) Suggest a possible explanation why the other might not.

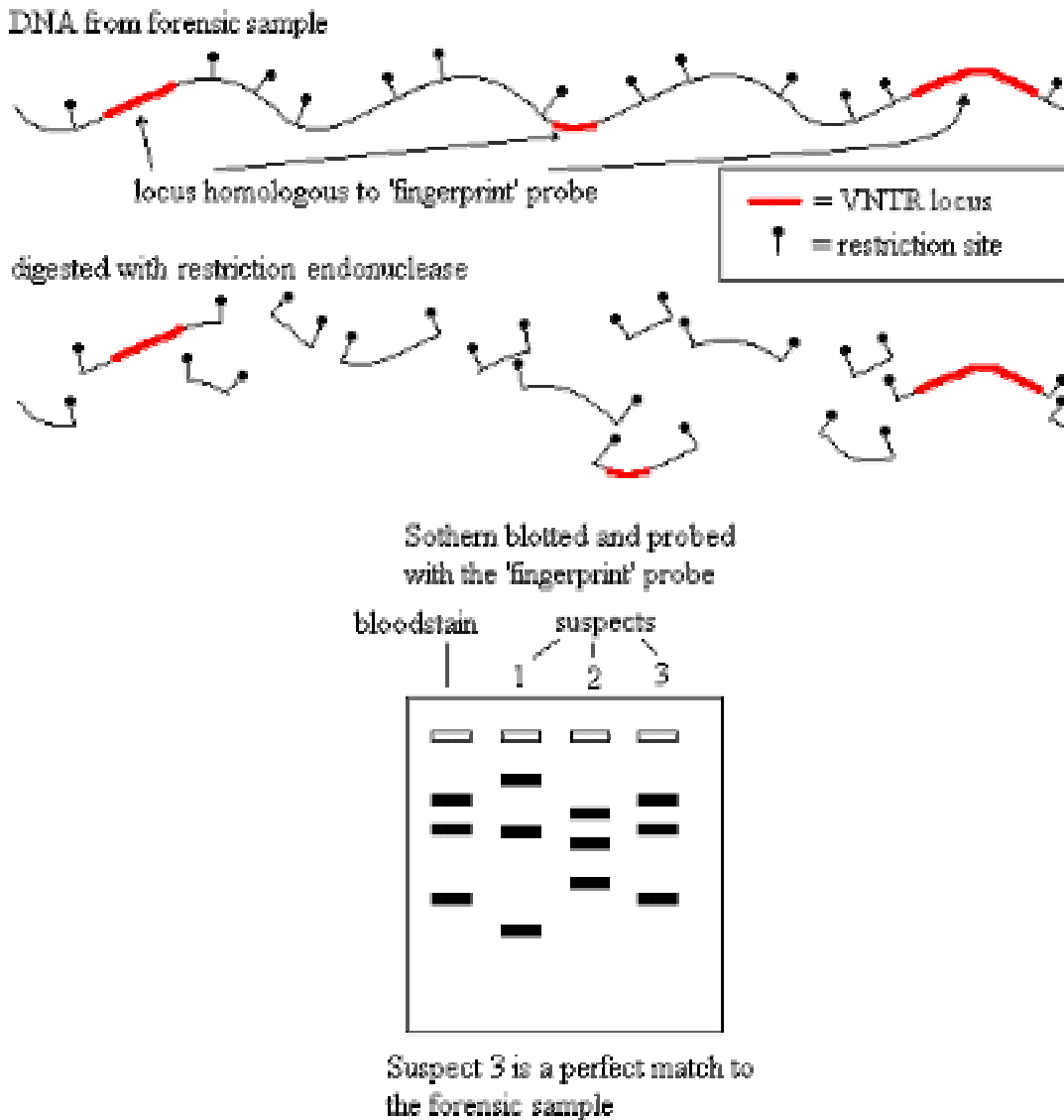
Answers

Forensic Population Genetics

The use of DNA evidence in both criminal and civil trials has been called into question for one of two reasons, either the poor quality control and procedures of the forensic laboratory concerned or the lack of proper control data from the population from which the DNA sample has been drawn. The former is none of our concern here but the latter is highly relevant.

What is a DNA fingerprint? Essentially, it is an attempt to examine the alleles present at sufficient polymorphic loci in a DNA sample to be able to give a unique identity to that sample. Sometimes this can be done by examining a single Southern blot if the probe

recognises many loci simultaneously (which is where the term fingerprint came into the jargon).



If the DNA probe recognises one locus and if that locus has more than one allele then the chance of a suspect having the same alleles as a forensic sample will depend on the relative frequency of the genotype of the sample. If the sample and suspects can be typed for many loci simultaneously then the probability of a perfect match diminishes to the product of the frequencies of the genotypes at all the loci. However, between different populations the frequencies of different genotypes are likely to be different. It may be, for instance, that all the members of a small tribe in New Guinea might be likely to have the same combination of alleles because genetic drift has greatly reduced the variety of alleles present at each locus. Forensic laboratories are thus creating databases of genetic

variation within many different populations so that the right control data will always be available for any case.

Cancer

The aim of this small section is to introduce the idea of cancer as a genetic disease. Although probably a relatively small proportion of all cases of cancer are caused by the inheritance of faulty genes, all cases of cancer are caused by mutation of one (more often several) genes. In addition, for all types of cancer there exist a small and instructive proportion of cases in which the mutation was inherited.

Loss of cell cycle control

Cancer is caused by the loss of control of cell division. Many cells need to divide for tissue growth or maintenance but this ability must be strictly regulated. When the regulation breaks down and a cell is able to divide unchecked then cancer is the result. The cause of the breakdown is mutation

- either to cause a gene whose normal function is to stimulate a cell to divide to be over expressed or inappropriately expressed
- or to inactivate both alleles of a gene which functions to prevent cell division.

The former class of genes are known as oncogenes and the latter as tumour suppressor genes.

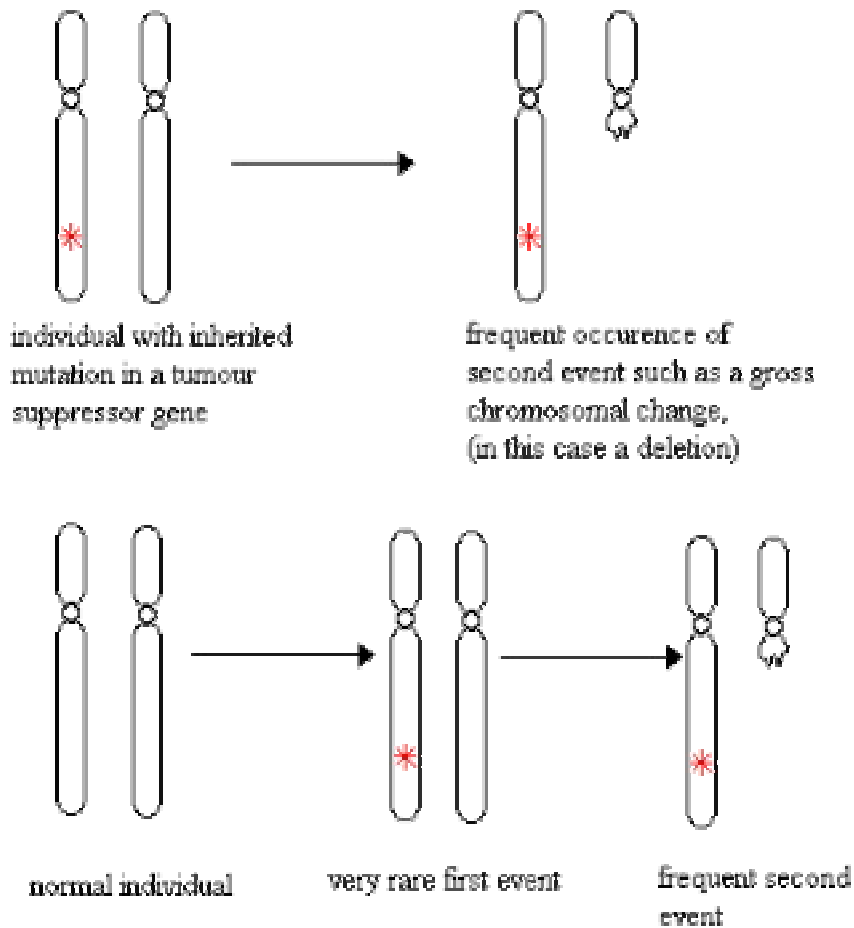
Oncogenes

Most examples of oncogenes were discovered by molecular virologists working in the 1970s and 80s. They were studying RNA viruses which had the property of transforming cells which they infected from a normal to a cancerous state. The viruses had acquired copies of host genes (protooncogenes) which had mutated (into oncogenes). The normal roles of the genes were varied. Some were growth factors involved in the normal stimulation of cells to divide. Some were growth factor receptor proteins which would now stimulate their targets even without binding external growth factors. Some were themselves transcription factors, directly stimulating other genes.

tumour suppressor genes

Tumour suppressor genes act normally to prevent cells dividing. If they are inactivated by mutation then cell division may be triggered. It is always necessary to inactivate both copies of a tumour suppressor gene. Sometimes one copy is lost through an inherited mutation. In this case every cell in the body will have only one active copy of the gene. It will not be an infrequent event that the second copy be removed because of a second somatic mutation event (remember for instance that loss of a chromosome occurs at a

frequency of 1 in 100 cell divisions). Patients will show multiple tumours caused by many such second events. There are many examples of inherited genes of this type such as [retinoblastoma](#) and the breast cancer genes [BRCA1](#) and [BRCA2](#). The concept of the first (rare) event which must be followed by a second (much more frequent) event to initiate a cancer was first put forward by Knudsen in 1972.



Recommended reading

The topics include:

- Hardy -Weinberg
- Natural Selection
- Founder Effects
- Cancer

Reading:

- Mange and Mange Chapters 17 (Cancer) and 12 (population and evolution)
- Lewis Chapters 12, 13 (populations) and 16 (cancer). Chapter 12 makes interesting reading because it is very oriented towards forensic DNA fingerprint testing. Chapter 13 covers similar ground to Mange and Mange or Thompson McInnes and Willard
- Jorde *et al.* Chapter 3 (again), Chapter 4 (pp 60 - 63), Chapter 13 (genetic screening), Chapter 11 (Cancer)
- Mueller and Young Chapters 7 (pp 113 - 126) (populations, medically oriented) and 13 (cancer).
- Thompson McInnes and Willard Chapters 7 (populations) and 16 (cancer).
- Connor and Ferguson-Smith Chapters 11 (populations, very little here) and 17 (cancer, some basic information but mostly specific information about different types of cancer).

[Back to the top](#)

[Back to the introduction](#)

[Answers](#)

SAQ answers

Lecture 1

1. The trailing strand would be unable to be replicated all the way to the end (where would the RNA primer be placed to prime the Okazaki fragment synthesis?) and would shorten every cell generation. The leading strand would be replicated to the end of the trailing strand and so would also shorten (but one cell division cycle behind the trailing strand. Eventually the chromosome would be lost. It has been suggested that the syndrome [progeria](#) which leads to premature aging, (the affected children die of diseases of old age in their teens) is caused by a mutation in the gene coding for the enzyme telomerase. The similar syndrome, [Werner's syndrome](#) has been proved to be caused by mutations in a DNA helicase gene (involved in DNA unwinding).
2. Four times that amount, or 12 picograms. (Diploid cells with replicated chromosomes)

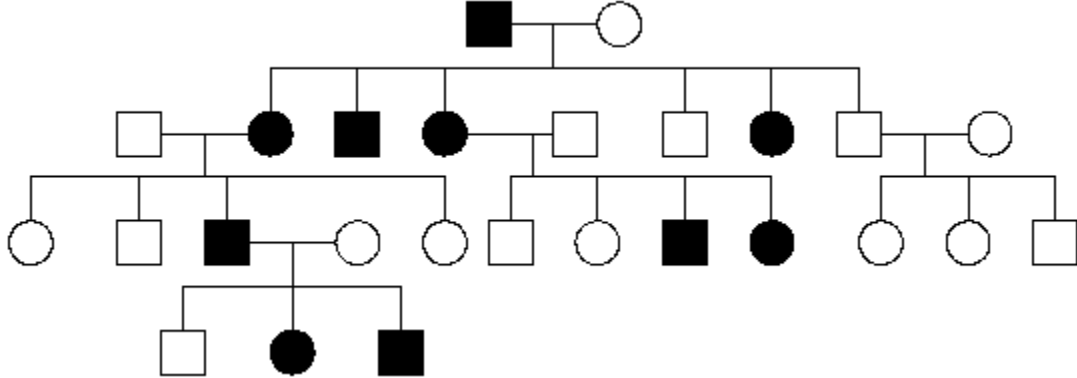
Lecture 2

1. a) 46 chromosomes are present in a premeiotic germ cell. When meiosis begins each has replicated and when it condenses in division can be seen to be composed

of two chromatids.

b) Each human gamete contains 23 chromosomes, each at that point composed of a single chromatid.

c) The number of chromosomes changed as a result of the first meiotic division in which homologous chromosomes go to opposite poles of the spindle.



2.

- If dominant, then the chance that III₆ will have affected children is zero.
- If recessive then obligate carriers are:
 - I₂,
 - II₁, II₅, II₆, II₈,
 - III₁, III₂, III₄, III₅, III₆, III₇,
 - IV₁

The following are at risk of being carriers:

- III₁₀, III₁₁, III₁₂ all with chance 50%

1. Several genes are involved in determining cat coat colour but only two are segregating in this cross, the genes Black (alleles B and b¹) and A gouti (alleles A and a).

The initial cross is:

BB AA x b¹b¹ aa
brown tabby cinnamon solid

↓
V

F1 Bb¹ Aa
brown tabby

These cats are intercrossed to give:

F2

BB AA		
Bb ¹ AA (x2)		
BB Aa (x2)		9 brown tabby
Bb ¹ Aa (x4)		
b ¹ b ¹ AA		
b ¹ b ¹ Aa (x2)		3 cinnamon tabby
BB aa		
Bb ¹ aa (x2)		3 black

Lecture 3

1. [FHC](#) is **genetically heterogeneous**. It is also a disease with **reduced penetrance**. The high frequency of sporadic cases might be caused by a high mutation frequency and, given the number of possible gene targets this could be so. However, as the causative mutations are being found, most can in fact be traced back to subclinically affected parents so this is a case of either **low expressivity** (minor symptoms) or reduced penetrance. It is not known what is the phenotype of a homozygote, might it be lethal? Nor is it known what is the result of simultaneous heterozygosity for more than one of the relevant genes.
2. The tabby gene is irrelevant here. The first cross was of the form:

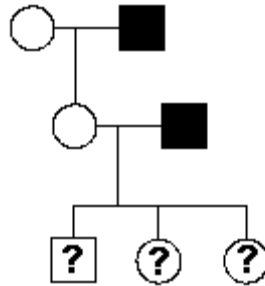
$$\begin{array}{rcc}
 b^1b^1 & \times & BB \\
 \text{cinnamon} & & \text{black} \\
 & | & \\
 & \vee & \\
 & b^1B & \\
 & \text{black} &
 \end{array}$$

The second cross was:

$$\begin{array}{rcc}
 b^1b^1 & \times & Bb \\
 \text{cinnamon} & & \text{black} \\
 & | & \\
 & \vee & \\
 & b^1B & b^1b \\
 & \text{black} & \text{chocolate} \\
 & 1 & : & 1
 \end{array}$$

The allele b is recessive to B but dominant to b¹. This illustrates the point that in a multiple allele system it may not be obvious which alleles are present.

Lecture 4



1. The pedigree looks like this and the woman must be a carrier.
 1. The chance that any son will be affected is $\frac{1}{2}$. The chance that the first child will be an affected son is $\frac{1}{4}$, (half a chance of being a son, half a chance of being affected).
 2. Similarly the chance of any daughter being affected is also $\frac{1}{2}$ (the daughter must receive one affected X chromosome from her father), the chance of the first child being an affected daughter is thus $\frac{1}{4}$.
 3. Again, the chance that any daughter will be a carrier is $\frac{1}{2}$ and so the chance that the first child is a carrier daughter is $\frac{1}{4}$

If the woman had had an affected brother rather than an affected father, then her chance of being a carrier goes from 1 to $\frac{1}{2}$. All the odds in the question thus reduce by a factor of 2 to $\frac{1}{8}$

2. Perhaps the husband is an XX male. His karyotype should be examined more closely. DNA tests for the presence of Y chromosome DNA and for the gene *SRY* could be carried out.

Lecture 5

1. Answer:
 1. A whole genome cosmid library should contain clones covering the desired region. How will you screen it to find the correct few clones out of the several hundred thousand clones in the library?
 2. This is unlikely to be useful.
 3. This may be useful. Is alpha-1-antitrypsin expressed in the embryo? If not the cDNA library will not contain any relevant clones.
 4. This will certainly contain alpha-1-antitrypsin cDNA clones at a good frequency. Again, how will you screen it? Is it in an expression vector to enable you to screen with antibodies?
 5. Wrong chromosome - so not useful. It would be nice to have access to a chromosome 14 library though.
 6. This might be useful, but I would be reluctant to place too much reliance on any DNA which has at some stage in its evolution passed through a

YAC. These clones are very prone to internal deletions and rearrangements. It might be a good source of DNA fragments to probe the first cosmid library to obtain cosmids from the area.

2. Begin with:
 1. A Search of the [Genome Database](#) (GDB)
 2. A search of EST databases such as [UniGene](#)

Follow up any clones listed. Sometimes it will be necessary to contact authors but, since the advent of gridded libraries it is often possible to obtain named clones from public depositories (for instance in the UK the [Human Genome Resource Centre](#) (on the same research campus as the Sanger Centre).

3. When the gene (in this case - genes) has been identified, the DNA from members of a family can be studied in order to find the mutation(s) which is (are) causing a problem in that family. Once this is done all family members can, if they wish, be screened to confirm whether or not they are carrying a mutation. Advice can then be given on the likelihood of transmitting the disease to their offspring. It will also be possible to offer antenatal diagnosis of genetic disease by chorionic villus sampling or by amniocentesis, followed by PCR amplification of the mutation site and then by a mutation detection technique (possibly DNA sequencing).

Lecture 6

1. Stealing ourselves to ignore the questionable message expressed by this gene:

a. WEE WET WET ARE NOT ALL BAD

point mutation

b. WET WET WET ARE COT ALL BAD

point mutation

c. WEW ETW ETA REN OTA LLB AD

deletion of one nucleotide leading to frameshift

d. WET WET WET ART TEN OTA LLB AD

insertion of two nucleotides leading to frameshift

e. WET _WET ARE NOT ALL BAD

deletion of a trinucleotide

f. WET WET WEG OOD MOR NIN GTA REN OTA LLB AD

insertion (of a repetitive element)

g. WET WET WET WET WET WET

[Back to the top](#)

[Back to the introduction](#)

[Next lecture](#)

[Back to Genetics](#)

[Back to Evolution](#)

[Back to Biostatistics](#)

[Back to HLA](#)

[Back to MHC](#)

[Glossary](#)

[Homepage](#)

BASIC POPULATION GENETICS

M.Tevfik Dorak, M.D., Ph.D.

G.H. Hardy (the English mathematician) and W. Weinberg (the German physician) independently worked out the mathematical basis of population genetics in 1908. Their formula predicts the expected genotype frequencies using the allele frequencies in a diploid Mendelian population. They were concerned with questions like "what happens to the frequencies of alleles in a population over time?" and "would you expect to see alleles disappear or become more frequent over time?"

Hardy and Weinberg showed in the following manner that if the population is very large and random mating is taking place, allele frequencies remain unchanged (or in equilibrium) over time unless some other factors intervene. If the frequencies of allele A and a (of a biallelic locus) are p and q, then $(p + q) = 1$. This means $(p + q)^2 = 1$ too. It is also correct that $(p + q)^2 = p^2 + 2pq + q^2 = 1$. In this formula, p^2 corresponds to the frequency of homozygous genotype AA, q^2 to aa, and $2pq$ to Aa. Since 'AA, Aa, aa' are the three possible genotypes for a biallelic locus, the sum of their frequencies should be 1. In summary, Hardy-Weinberg formula shows that:

$$p^2 + 2pq + q^2 = 1$$

AA Aa aa

If the observed frequencies do not show a significant difference from these expected frequencies, the population is said to be in Hardy-Weinberg equilibrium (HWE). If not, there is a violation of the following assumptions of the formula, and the population is not in HWE.

The assumptions of HWE

1. Population size is effectively infinite,
2. Mating is random in the population (the most common deviation results from inbreeding),
3. Males and females have similar allele frequencies,
4. There are no mutations and migrations affecting the allele frequencies in the population,
5. The genotypes have equal fitness, i.e., there is no selection.

The Hardy-Weinberg law suggests that as long as the assumptions are valid, allele and genotype frequencies will not change in a population in successive generations. Thus, any **deviation from HWE** may indicate:

1. Small population size results in random sampling errors and unpredictable genotype frequencies (a real population's size is always finite and the frequency of an allele may fluctuate from generation to generation due to chance events),
2. Assortative mating which may be positive (increases homozygosity; self-fertilization is an extreme example) or negative (increases heterozygosity), or inbreeding which increases homozygosity in the whole genome without changing the allele frequencies. Rare-male mating advantage also tends to increase the frequency of the rare allele and heterozygosity for it (in reality, random mating does not occur all the time),
3. A very high mutation rate in the population (typical mutation rates are $< 10^{-5}$ per generation) or massive migration from a genotypically different population interfering with the allele frequencies,
4. Selection of one or a combination of genotypes (selection may be negative or positive). Another reason would be unequal transmission ratio of alternative alleles from parents to offspring (as in mouse *t*-haplotypes).

The implications of the HWE

1. The allele frequencies remain constant from generation to generation. This means that hereditary mechanism itself does not change allele frequencies. It is possible for one or more assumptions of the equilibrium to be violated and still not produce deviations from the expected frequencies that are large enough to be detected by the goodness of fit test,
2. When an allele is rare, there are many more heterozygotes than homozygotes for it. Thus, rare alleles will be impossible to eliminate even if there is selection against homozygosity for them,
3. For populations in HWE, the proportion of heterozygotes is maximal when allele frequencies are equal ($p = q = 0.50$),

4. An application of HWE is that when the frequency of an autosomal recessive disease (e.g., sickle cell disease, hereditary hemochromatosis, congenital adrenal hyperplasia) is known in a population and unless there is reason to believe HWE does not hold in that population, the gene frequency of the disease gene can be calculated (for an example visit the [Cancer Genetics](#) website and choose Topics).

It has to be remembered that when HWE is tested, mathematical thinking is necessary. When the population is found in equilibrium, it does not necessarily mean that all assumptions are valid since there may be counterbalancing forces. Similarly, a significant deviance may be due to sampling errors (including **Wahlund effect**, see below and [Glossary](#)), misclassification of genotypes, measuring two or more systems as a single system, failure to detect rare alleles and the inclusion of non-existent alleles. The Hardy-Weinberg laws rarely holds true in nature (otherwise evolution would not occur). Organisms are subject to mutations, selective forces and they move about, or the allele frequencies may be different in males and females. The gene frequencies are constantly changing in a population, but the effects of these processes can be assessed by using the Hardy-Weinberg law as the starting point.

The direction of departure of observed from expected frequency cannot be used to infer the type of selection acting on the locus even if it is known that selection is acting. If selection is operating, the frequency of each genotype in the next generation will be determined by its relative fitness (W). Relative fitness is a measure of the relative contribution that a genotype makes to the next generation. It can be measured in terms of the intensity of selection (s), where $W = 1 - s$ [$0 \leq s \leq 1$]. The frequencies of each genotype after selection will be $p^2 W_{AA}$, $2pq W_{Aa}$, and $q^2 W_{aa}$. The highest fitness is always 1 and the others are estimated proportional to this. For example, in the case of heterozygote advantage (or overdominance), the fitness of the heterozygous genotype (Aa) is 1, and the fitnesses of the homozygous genotypes negatively selected are $W_{AA} = 1 - s_{AA}$ and $W_{aa} = 1 - s_{aa}$. It can be shown mathematically that only in this case a stable polymorphism is possible. Other selection forms, underdominance and directional selection, result in unstable polymorphisms. The weighted average of the fitnesses of all genotypes is the mean fitness. It is important that genetic fitness is determined by both fertility and viability. This means that diseases that are fatal to the bearer but do not reduce the number of progeny are not genetic lethals and do not have reduced fitness (like the adult onset genetic diseases: Huntington's chorea, hereditary hemochromatosis). The detection of selection is not easy because the impact on changes in allele frequency occur very slowly and selective forces are not static (may even vary in one generation as in antagonistic pleiotropy).

All discussions presented so far concerns a simple biallelic locus. In real life, however, there are many loci which are multiallelic, and interacting with each other as well as with the environmental factors. The Hardy-Weinberg principle is equally applicable to multiallelic loci but the mathematics is slightly more complicated. For multigenic and multifactorial traits, which are mathematically continuous as opposed to discrete, more complex techniques of quantitative genetics are required.

In a final note on the practical use of HWE, it has to be emphasised that **its violation** in daily life is most frequently due to genotyping errors. Allelic misassignments, as frequently happens when PCR-SSP method is used, sometimes due to allelic dropout are the most frequent causes of the Hardy-Weinberg disequilibrium. When this is observed, the genotyping protocol should be reviewed. In a case-control association study, it is of paramount importance that the control group is in HWE to rule out any technical errors. The violation of HWE in the case group, however, may be due to a real association.

Some concepts relevant to HWE

Wahlund effect: Reduction in observed heterozygosity (increased homozygosity) because of considering pooled discrete subpopulations that do not interbreed as a single randomly mating unit. When all subpopulations have the same gene frequencies, no variance among subpopulations exists, and no Wahlund effect occurs ($F_{ST}=0$). **Isolate breaking** is the phenomenon that the average homozygosity temporarily increases when subpopulations make contact and interbreed (this is due to decrease in homozygotes). It is the opposite of Wahlund effect.

F statistics: The F statistics in population genetics has nothing to do the F statistics evaluating differences in variances. Here **F** stands for fixation index, fixation being increased homozygosity resulting from inbreeding. Population subdivision results in the loss of genetic variation (measured by heterozygosity) within subpopulations due to their being small populations and genetic drift acting within each one of them. This means that population subdivision would result in decreased heterozygosity relative to that expected heterozygosity under random mating as if the whole population was a single breeding unit. Wright developed three fixation indices to evaluate population subdivision: F_{IS} (interindividual), F_{ST} (subpopulations), F_{IT} (total population).

F_{IS} is a measure of the deviation of genotypic frequencies from panmictic frequencies in terms of heterozygous deficiency or excess. It is what is known as the **inbreeding coefficient (f)**, which is conventionally defined as the probability that two alleles in an individual are identical by descent (autozygous). The technical description is the correlation of uniting gametes relative to gametes drawn at random from within a subpopulation (**Individual within the Subpopulation**) averaged over subpopulations. It is calculated in a single population as $F_{IS} = 1 - (H_{OBS} / H_{EXP})$ [equal to $(H_{EXP} - H_{OBS}) / H_{EXP}$] where H_{OBS} is the observed heterozygosity and H_{EXP} is the expected heterozygosity calculated on the assumption of random mating. It shows the degree to which heterozygosity is reduced below the expectation. The value of F_{IS} ranges between -1 and +1. Negative F_{IS} values indicate heterozygote excess (outbreeding) and positive values indicate heterozygote deficiency (inbreeding) compared with HWE expectations.

F_{ST} measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. F_{ST} is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called **coancestry coefficient (θ)** [Weir & Cockerham, 1984] or '**Fixation index**' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (**Subpopulation within the Total population**). It is calculated as using the subpopulation (average) heterozygosity and total population expected heterozygosity. F_{ST} is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the population) and 1 = complete isolation (extreme subdivision). F_{ST} values up to 0.05 indicate negligible genetic differentiation whereas >0.25 means very great genetic differentiation within the population analyzed. F_{ST} is usually calculated for different genes, then averaged across all loci, and all populations. F_{ST} can also be used to estimate gene flow: $0.25 (1 - F_{ST}) / F_{ST}$. This highly versatile parameter is even used as a genetic distance measure between two populations instead of a fixation index among many populations (see [Weir BS, Genetic Data Analysis II, 1996](#); and [Kalinowski ST, 2002](#)).

F_{IT} is rarely used. It is the overall inbreeding coefficient (F) of an individual relative to the total population (**Individual within the Total population**).

Detecting Selection Using DNA Polymorphism Data

Several methods have been designed to use DNA polymorphism data (sequences and allele frequencies) to obtain information on past selection events. Most commonly, the ratio of non-synonymous (replacement) to synonymous (silent) substitutions (d_N/d_S ratio; see below) is used as evidence for overdominant selection (balancing selection) of which one form is heterozygote advantage. Classic example of this is the mammalian **MHC** system genes and other **compatibility systems** in other organisms: the self-incompatibility system of the plants, fungal mating types and invertebrate allorecognition systems. In all these genes, a very high number of alleles is also noted. This can be interpreted as an indicator of some form of balancing (diversifying) selection. In the case of neutral polymorphism, one common allele and a few rare alleles are expected. The frequency distribution of alleles is also informative. Large number of alleles showing a relatively even distribution is against neutrality expectations and suggestive of diversifying selection.

Most tests detect selection by rejecting neutrality assumption (observed data is deviate significantly from what is expected under neutrality). This deviation, however, may also be due to other factors such as changes in population size or genetic drift. The original neutrality test was **Ewens-Watterson homozygosity test** of neutrality (see [Glossary](#)) based on the comparison of observed homozygosity and

predicted value calculated by [Ewens's sampling formula](#) which uses the number of alleles and sample size. This test is not very powerful.

Other commonly used statistical tests of neutrality are Tajima's D (theta), Fu & Li's D, D* and F. Tajima's test ([Tajima F, Genetics 1989](#)) is based on the fact that under the neutral model estimates of the number of segregating/polymorphic sites and of the average number of nucleotide differences are correlated. If the value of D is too large or too small, the neutral 'null' hypothesis is rejected. [DnaSP](#) calculates the D and its confidence limits (two-tailed test). Tajima did not base this test on coalescent but Fu and Li's tests ([Fu & Li, Genetics 1993](#)) are directly based on coalescent. The tests statistics D and F require data from intraspecific polymorphism and from an outgroup (a sequence from a related species), and D* and F* only require intraspecific data. [DnaSP](#) uses the critical values obtained by [Fu & Li, Genetics 1993](#) to determine the statistical significance of D, F, D* and F* test statistics. [DnaSP](#) can also conduct the Fs test statistic ([Fu YX, Genetics 1997](#)). The results of this group of tests (Tajima's D and Fu & Li's tests) based on allelic variation and/or level of variability may not clearly distinguish between selection and demographic alternatives (bottleneck, population subdivision) but this problem only applies to the analysis of a single locus (demographic changes affect all loci whereas selection is expected to be locus-specific which are distinguishable if multiple loci are analysed). Tests for multiple loci include the **HKA test** described by Hudson, Kreitman and Aguade ([Genetics, 1987](#)). This test is based on the idea that in the absence of selection, the expected number of polymorphic (segregating) sites within species and the expected number of 'fixed' differences between species (divergence) are both proportional to the mutation rate, and the ratio of them should be the same for all loci. Variation in the ratio of divergence to polymorphism among loci suggests selection.

A different group of neutrality tests that are not sensitive to demographic changes include McDonald-Kreitman test ([McDonald JH & Kreitman M, Nature 1991](#)) and d_N/d_S ratio test. McDonald-Kreitman test compares the ratio of the number of nonsynonymous to synonymous 'polymorphisms' within species to that ratio of the number of nonsynonymous to synonymous 'fixed' differences between species in a 2x2 table (see a worked example [here](#)). The most direct method of showing the presence of positive selection is to compare the number of nonsynonymous (d_N) to the number of synonymous (d_S) substitutions in a locus. A high (>1) value of (d_N/d_S) substitutions suggest fixation of nonsynonymous mutations with a higher probability than neutral (synonymous) ones. Statistical properties of this test are given by [Goldman N & Yang Z, Mol Biol Evol, 1994](#) and by [Muse SV & Gaut BS, Mol Biol Evol, 1994](#). The d_N/d_S ratio tests take into account of transition/transversion rate bias and codon usage bias.

For other tests and software to perform these statistics, see [DNA Sequence Polymorphism, DnaSP](#). See also: [Statistical Tests of Neutrality \(Lecture Note by P Beerli\)](#); [Statistical Tests of Neutrality of Mutations against Excess of Recent Mutations \(Rare Allels\)](#); [Statistical Tests of Neutrality of Mutations against an Excess of Old Mutations or a Reduction of Young Mutations](#); [Estimation of theta](#); Properties of statistical tests of neutrality for DNA polymorphism data, [Simonsen et al Genetics 1995;141:413-29](#). Review of statistical tests of selective neutrality on genomic data, [Nielsen R, Heredity 2001](#); and a [Lecture Note by Gil McVean](#).

Linkage disequilibrium (LD)

The tendency for two 'alleles' to be present on the same chromosome (positive LD), or not to segregate together (negative LD). As a result, specific alleles at two different loci are found together more or less than expected by chance. The same situation may exist for more than two alleles. Its magnitude is expressed as the delta (Δ) value and corresponds to the difference between the expected and the observed haplotype frequency (see [Measures of LD by Devlin & Risch, 1995](#) for further details). It can have positive or negative values. LD is decreased by recombination. Thus, it decreases every generation of random mating unless some process opposing the approach to linkage 'equilibrium'. Permanent LD may result from natural selection if some gametic combinations confer higher fitness than other combinations. For more on LD, see [Statistical Analysis in HLA and Disease Association Studies](#). Link to a lecture on [linkage disequilibrium](#); [online LD analysis](#). Software to perform LD analysis: [Genetic Data Analysis, EH, 2LD, MLD, PopGene, Arlequin 2000](#), and [Online Easy LD](#). Please note that LD has nothing to do HWE and should not be confused with it (see [Possible Misunderstandings in Genetics](#)).

Genetic distance (GD)

Genetic distance is a measurement of genetic relatedness of samples of populations (whereas genetic diversity represents diversity within a population). The estimate is based on the number of allelic substitutions per locus that have occurred during the separate evolution of two populations. (See lecture notes on [Genetic Distances](#), [Estimating Genetic Distance](#); and [GeneDist: Online Calculator of Genetic Distance](#). The software [Arlequin](#), [PHYLIP](#), [GDA](#), [PopGene](#), [Populations](#) and [SGS](#) are suitable to calculate population-to-population genetic distance from allele frequencies.

Genetic Distance can be computed on freeware PHYLIP. Most components of PHYLIP are available [on the web](#). One component of the package [GENDIST](#) estimates genetic distance from allele frequencies using one of the three methods: Nei's, Cavalli-Sforza's or Reynold's (see papers by [Cavalli-Sforza & Edwards, 1967](#), [Nei et al, 1983](#), [Nei M, 1996](#) and [lecture note \(1\)](#) and [\(2\)](#) for more information on these methods). GENDIST can be run [online](#) using default options ([Nei's genetic distance](#)) to obtain genetic distance matrix data. The PHYLIP program [CONTML](#) estimates phylogenies from gene frequency data by maximum likelihood under a model in which all divergence is due to genetic drift in the absence of new mutations (Cavalli-Sforza's method) and draws a tree. The program is also available [on the web](#) and runs with default options. If new mutations are contributing to allele frequency changes, Nei's method should be selected on GENDIST to estimate genetic distances first. Then a tree can be obtained using one of the following components of PHYLIP: [NEIGHBOR](#) also draws a phylogenetic tree using the genetic distance matrix data (from GENDIST). It uses either [Nei and Saitou's](#) (1987) "[Neighbor Joining \(NJ\) Method](#)," or the [UPGMA](#) (unweighted pair group method with arithmetic mean; average linkage clustering) method ([Sneath & Sokal, 1973](#)). [Neighbor Joining](#) is a distance matrix method producing an unrooted tree without the assumption of a clock (the evolutionary rate does not have to be the same in all lineages). Major assumption of UPGMA is equal rate of evolution along all branches (which is frequently unrealistic). NEIGHBOR can be run [online](#). Other components of PHYLIP that draw phylogenetic trees from genetic distance matrix data are [FITCH](#) / [online](#) (Fitch-Margoliash method with no assumption of equal evolutionary rate) and [KITSCH](#) / [online](#) (employs Fitch-Margoliash and Least Squares methods with the assumption that all tip species are contemporaneous, and that there is an evolutionary clock -in effect, a molecular clock). ([Mathematical formulae of various genetic distance measures.](#)) Another freeware [PopGene](#) calculates Nei's genetic distance and creates a tree using UPGMA method from genotypes. For genetic distance calculation on Excel, try freeware [GenAIEx](#) by Peakall & Smouse.

Because of different assumptions they are based on the NJ and UPGMA methods may construct dendrograms with totally different topologies. For an example of this and a review of main differences between the two methods, see [Nei & Roychoudhury, 1993 \(free full-text\)](#). Both methods use distance matrices (also Fitch-Margoliash and Minimal Evolution methods are distance methods). The principle difference between NJ and UPGMA is that NJ does not assume an equal evolutionary rate for each lineage. Since the constant rate of evolution does not hold for human populations, NJ seems to be the better method. For the genetic loci subject to natural selection, the evolutionary rate is not the same for each population and therefore UPGMA should be avoided for the analysis of such loci (including the HLA genes). The leading group in HLA-based genetic distance analysis led by [Arnaiz-Villena](#) proposes that the most appropriate genetic distance measure for the HLA system is the DA value first described by [Nei et al, 1983](#). Unlike UPGMA, NJ produces an unrooted tree. To find the root of the tree, one can add an outgroup. The point in the tree where the edge to the outgroup joins is the best possible estimate for the root position. One persistent problem with tree construction is the lack of statistical assessment of the phylogenetic tree presented. This is best done with widely available bootstrap analysis originally described in Felsenstein J: Evolution 1985;39:783-791 (available through [JSTOR](#) if you have access) and [Efron et al. 1996](#); and reviewed in [Nei M, 1996](#)). For a discussion of statistical tests of molecular phylogenies, see [Li & Gouy, 1990](#) and [Nei M, 1996](#). For the topology to be statistically significant the bootstrap value for each cluster should reach at least 70% whereas 50% overestimates accuracy of the tree. Bootstrap tests should be done with at least 1000 (preferably more) replications.

Nei noted that some genes are more suitable than others in phylogenetic inference and that most tree-building methods tend to produce the same topology whether the topology is correct or not [Nei M, 1996](#). He also added that sometimes adding one more species/population would change the whole tree for unknown reasons. An example of this has been provided in a study of human populations with genetic distances [Nei & Roychoudhury, 1993](#). The properties of most popular genetic distance measures have been reviewed ([Kalinowski, 2002](#)). Whichever is used, large sample sizes are required when populations are relatively genetically similar, and loci with more alleles produce better estimates of genetic distance. However, in a simulation study, Nei et al concluded that more than 30 loci should be used for making phylogenetic trees ([Nei et al, 1983](#)). There seems to be a consensus that estimated trees are nearly always erroneous (i.e., the topological arrangement will be wrong) if the number of loci is less than 30 ([Nei M, 1996](#); Jorde LB. Human genetic distance studies. *Ann Rev Anthropol* 1985;14:343-73; available through [JSTOR](#) if you have access). If populations are closely related even 100 loci may be necessary for an accurate estimation of the relationships by genetic distance methods. Cavalli-Sforza et al have noted important correlations between the genetic trees and linguistics evolutionary trees with the exceptions for New Guinea, Australia and South America ([Cavalli-Sforza et al. 1994](#)).

Especially for the HLA genes, phylogenetic trees can be constructed by using the Nei's DA genetic distance values and NJ method with bootstrap tests on [DISPAN. Correspondence analysis](#), a supplementary analysis to genetic distances and dendrograms, displays a global view of the relationships among populations ([Greenacre MJ, 1984](#); [Greenacre & Blasius, 1994](#); [Blasius & Greenacre, 1998](#)). This type of analysis tends to give results similar to those of dendrograms as expected from theory ([Cavalli-Sforza & Piazza, 1975](#)), and is more informative and accurate than dendrograms especially when there is considerable genetic exchange between close geographic neighbors ([Cavalli-Sforza et al. 1994](#)). In their enormous effort to work out the genetic relationships among human populations, Cavalli-Sforza et al concluded that two-dimensional scatter plots obtained by correspondence analysis frequently resemble geographic maps of the populations with some distortions ([Cavalli-Sforza et al. 1994](#)). Using the same allele frequencies that are used in phylogenetic tree construction, [correspondence analysis](#) using allele frequencies can be performed on the [ViSta \(v7.0\)](#), [VST](#), [SAS](#) but most conveniently on Multi Variate Statistical Package [MVSP](#). Link to [a tutorial on correspondence analysis](#).

Internet Links

[History of Population Genetics and Evolution in A History of Genetics](#) by AH Sturtevant

[ASHI 2001 Biostatistics and Population Genetics Workshop Notes](#)

[Microsatellites and Genetic Distance \(Primer on Genetic Distance\)](#)

[HWE in Kimball's Biology Pages](#) [Online HWE Test](#) [Online GD Calculation](#) [Online Easy LD](#)

[Population Genetics Simulations](#) [Molecular Evolution / Computational Pop Genet Course](#)

[Lectures on Population Genetics \(1\) & \(2\) & \(3\) & \(4\) & \(5\)](#)

[Statistical Genetics Websites](#)

[Freeware Population Genetic Data Analysis Software \(List of Features\):](#)

[Arlequin 2000](#) [PopGene](#) [GDA](#) [Genetix](#) [GenePop](#) [GeneStrut](#) [SGS](#) [TFPGA](#) [MVSP](#)
[PHYLIP\(Online\)](#) [DISPAN](#) [ViSta](#) [GenAlEx](#) [CLUMP](#) [TDT](#) [HAPLOTYPER](#) [PHASE v2.0](#)
[EasyLD](#)

[MSA](#) (for microsatellite data) [POPULATIONS](#) [WINPOP v2.0](#) [QUANTO](#)

[GSF: Genetic Software Forum](#) [Partition for Online Bayesian Analysis](#)

[Comprehensive List of Genetic Analysis Software \(1\)](#) [\(2\)](#) [\(3\)](#)

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated March 19, 2004

[Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#)
[Homepage](#)

[Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#) [Homepage](#)

ESSENTIALS OF GENETICS

Notes for an evening class originally organised by [Cardiff University Centre for Lifelong Learning \(1999/2000\)](#)

The following notes are being updated regularly:

[Landmarks in the History of Genetics](#)

[Basic Genetic Terms and Rules](#) [Glossary](#)

[Chromosomes and Genes](#) [Gene Expression](#)

[Clinical Genetics](#) [Cancer Genetics](#) [Transplantation Genetics](#)

[Viral and Bacterial Genetics](#) [Plant Genetics](#)

[Biotechnology](#) [Genetic Engineering & Cloning](#) [Bioethics](#)

[Basic Population Genetics](#) [Evolution](#)

[Possible Misunderstandings in Genetics](#)

[Resources](#)

Relevant Internet Links

[Genetics Virtual Library](#) [Biomedical Life Long Learning \(Online Genetics Courses\)](#) [BBC Adult Learning](#)

[Dictionary of Genetic Terms](#) [NHGRI Talking Glossary of Genetic Terms](#) [Genetics Image Library](#)
[Genetic Animations](#)

Online Books: [Introduction to Genetic Analysis](#) [Modern Genetic Analysis](#) [Mouse Genetics](#)

[Companion to Genetics by Klug & Cummings](#) [Companion to Genetics: Principles and Analysis](#)

[Classic Papers in Genetics](#) [Electronic Manuscripts in Genetics](#) [Medline](#)

[Notes on Evolutionary Biology](#) [Ecology & Evolution Course](#)

[Intermediate Genetics](#)

[Primer on Molecular Genetics](#) [Biology Hypertextbook](#)

[Genetic Science Learning Center](#) [Human Genome Project](#) [Cancer Genetics](#)

[The Biology Project: Mendelian Genetics](#) [Genetics and Ethics](#)

[New Scientist: Genetics Links](#) [BBC Education: Gene Genomics Links](#) [Science: Genetics Collection](#) [Nature:](#)

[Biotech Graphics Gallery](#) [DNA Animation Laboratory](#) [DNA from the Beginning](#) [Genetics Education Center](#)

[Genetics Databases](#)

M. Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Jan 27, 2004

[Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#) [Homepage](#)

LANDMARKS IN THE HISTORY OF GENETICS

M. Tefvik Dorak, MD, PhD

Robert Hooke (1635-1703), a mechanic, is believed to give '**cells**' their name when he examined a thin slice of cork under microscope, he thought cells looked like the small, rectangular rooms monks lived.

1651 William Harvey suggests that all living things originate from eggs

1694 JR Camerarius does pollination experiments and discovers sex in flowering plants

1735 CV Linnaeus (originally Linne) proposes the taxonomic system including the naming of Homo sapiens.

1761-7 JG Kolreuter finds in experiments on Nicotiana that each parent contributes equally to the characteristics of the offspring.

1798 TR Malthus publishes *An Essay on the Principle of Population* (foundations of the struggle for existence and the survival of the fittest).

1800 Karl Friedrich Burdach coins the term biology to denote the study of human morphology, physiology and psychology.

1809 JB de Monet Lamarck puts forward his ideas on evolution

1818 WC Wells suggests natural selection in African populations (for their relative resistance to local diseases)

1820 CF Nasse describes the sex-linked transmission of haemophilia

1822-1824 TA Knight, J Goss, and A Seton independently do studies in peas and observe the dominance, recessiveness and segregation in the first filial generation, but did not detect regularities

1828 Karl Ernst von Baer publishes *The Embryology of Animals*

1830 GB Amici shows that the pollen tube grows down the style and into the ovule of the flower; Charles Lyell publishes his multi-volume *Principles of Geology*

1831 Robert Brown notes nuclei within cells; Charles Darwin starts his voyage on HMS Beagle (returns in 1836)

1839 MJ Schleiden & T Schwann develop the cell theory [all animals and plants are made up of cells. Growth and reproduction are due to division of cells]

1840 Martin Barry expresses the belief that the spermatozoon enters the egg

1855 Alfred Russell Wallace publishes *On the Law Which Has Regulated the Introduction of New Species*

1858 Alfred Russell Wallace sends to Darwin a manuscript "*On the Tendency of Varieties to Depart Indefinitely from the Original Type*"

1859 Darwin publishes *The Origin of Species*

Darwin's five theories:

1. *The organisms steadily evolve over time (evolution theory)*
2. *Different kinds of organisms descended from a common ancestor (common descent theory)*
3. *Species multiply over time (speciation theory)*
4. *Evolution takes place through the gradual change of populations (gradualism theory)*
5. *The mechanism of evolution is the competition among vast numbers of unique individuals for limited resources under selective pressures, which leads to differences in survival and reproduction (natural selection theory).*

1864 Ernst Haeckel (Häckel) outlines the essential elements of modern zoological classification

1865 Gregor Johann Mendel presents his principles of heredity [particulate inheritance] to the Brunn Society for Natural History and publishes in the Proceedings of the Brunn Society for Natural History in the following year [CPG p.1] (Brunn is now Brno in Czech Republic)

Mendel's work showed that:

1. *Each parent contributes one factor of each trait shown in offspring*
2. *The two members of each pair of factors segregate from each other during gamete formation*
3. *The blending theory of inheritance was not correct*
4. *Males and females contribute equally to the traits in their offspring*
5. *Acquired traits are not inherited.*

Mendel had referred to the genes as 'particles of inheritance'

1866 EH Haeckel (Häckel) hypothesizes that the nucleus of a cell transmits its hereditary information

1869 Francis Galton publishes *Hereditary Genius* (study of human pedigrees)

1871 Darwin publishes *Descent of Man* (principles of sexual selection)

1875 F Galton demonstrates the usefulness of twin studies for elucidating the relative influence of nature (heredity) and nurture (environment) upon behavioural traits; Oscar Hertwig concludes from a study of the reproduction of the sea urchin that fertilisation consists of the physical union of the two nuclei contributed by the male and female parents

1876 J Horner shows that colour-blindness is an inherited disease

1877 Fleming visualized chromosomes

1882 August Weismann notes the distinction between somatic and germ cells; chromosomes observed by Walther Flemming in the nuclei of dividing salamander cells. He uses the word mitosis

1887 A Weismann postulates the reduction of chromosome number in germ cells

- 1888** W Waldeyer coins the word chromosome
- 1889** Johann Miescher isolates DNA from salmon sperm; F Galton publishes *Natural Inheritance* (biometry)
- 1892** A Weismann's book *Das Keimplasma (The Germ Plasm)* emphasizes meiosis as an exact mechanism of chromosome distribution
- 1894** William Bateson's *Materials for the Study of Variation* emphasizes the importance of discontinuous variations; Karl Pearson publishes his first contribution to the mathematical theory of evolution (he develops the Chi-squared test in 1900)
- 1896** EB Wilson publishes *The Cell in Development and Heredity*
- 1899** The First International Congress of Genetics held in London
- 1900** The Dutch botanist Hugo de Vries and two others discover Mendel's principles; W Bateson publishes its translation to English in the following year
- 1901** Hugo de Vries adopts the term mutation
- 1902** WS Sutton and T Boveri (studying sea urchins) independently propose the chromosome theory of heredity [full set of chromosomes are needed for normal development; individual chromosomes carry different hereditary determinants; independent assortment of gene pairs occurs during meiosis] [CPG p.27]
- 1905** W Bateson gives the name genetics (means 'to generate' in Greek) to this branch of science, and introduces the words allele (allelomorph), heterozygous (impure line) and homozygous (pure line); W Bateson & RC Punnett work out the principles of multigenic interaction (linkage) and heredity [CPG p.42]
- 1908** GH Hardy and W Weinberg independently formulate the Hardy-Weinberg principle of population genetics [CPG p.60]
- 1909** AE Garrod publishes *Inborn Errors of Metabolism* [biochemical genetics of albinism, cystinuria, pentosuria and alkaptonuria]; W Johannsen uses the words phenotype, genotype and gene for the first time in his studies with beans CPG p.20]; CC Little produces the first inbred strain of mice (DBA)
- 1910** Thomas Hunt Morgan discovers the white-eye and its sex-linkage in *Drosophila* (the beginning of *Drosophila* genetics) [CPG p.63] [receives the Nobel prize in 1933]; J Herrick describes sickle cell anaemia
- 1911** TH Morgan shows the first example of chromosomal linkage in the X chromosome of *Drosophila* [Nobel prize 1933]; EB Wilson shows that the gene for colour-blindness is on the X chromosome (first gene identified on a chromosome); Davenport founds the first US genetic clinic
- 1912** TH Morgan shows that genetic recombination does not take place in males in *Drosophila* and also discovers the first sex-linked lethal gene [Nobel prize 1933]
- 1917** S Wright works out the biochemical basis of coat colour inheritance in animals [CPG p.78]
- 1919** A Hungarian engineer, Karl Ereky, coins the term biotechnology (to mean production of beer, cheese, bread etc with the help of living organisms)
- 1925** CB Bridges proposes the balanced chromosome determination of sex theory [relationship between the autosomes and sex chromosomes] [CPG p.117]

1927 HJ Muller demonstrates that X-rays are mutagenic in *Drosophila* [CPG p.149] [receives the Nobel prize in 1946]

1928 F Griffith discovers type-transformation in pneumococci

1941 George Wells Beadle & Edward Lawrie Tatum proposes the one gene - one enzyme (polypeptide) concept [CPG p.166] [Tatum receives the Nobel prize in 1958]

1944 Oswald Theodore Avery *et al* describe the DNA as the hereditary material [Pneumococcus transformation experiments] [CPG p.173]

1946 J Lederberg & EL Tatum demonstrate genetic recombination (conjugation) in bacteria [CPG p.192] [they receive the Nobel prize in 1958]

1949 L Pauling shows that a defect in the structure of hemoglobin causes sickle cell anemia

1950 E Chargaff *et al* demonstrate for DNA that the numbers of adenine and thymine groups are always equal, so are the numbers of guanine and cytosine groups; B McClintock discovers the transposable elements in maize [CPG p.199] [she receives the Nobel prize in 1983]

Early 1950s Rosalind Franklin and Maurice HF Wilkins at King's College, London show by X-ray crystallography that DNA exists as two strands wound together in a spiral or helical shape

1952 Frederick Sanger *et al* work out the amino acid sequence of insulin [Sanger receives his first Nobel prize in 1958]; AD Hershey & M Chase demonstrate that the genetic material of bacteriophage T2 is DNA and the DNA enters the host but not the protein [AD Hershey receives the Nobel prize in 1969]; ND Zinder & J Lederberg discover phage-mediated transduction in *Salmonella* [CPG p.221] [Lederberg receives the Nobel prize in 1958]

1953 On the basis of Chargaff's chemical data (1950; numbers of A and T, and C and G are the same in DNA), and Wilkins and Franklin's already available X-ray diffraction data, James D Watson & Francis HC Crick describe the DNA's double helix structure by inference [CPG p.241] [they share the Nobel prize in 1962]

1956 JH Tijo & A Levan show that the diploid chromosome number for humans is 46 ([Hereditas 1956;42:1-6](#)); S. Ochoa's laboratory discovers RNA polymerase and A Kornberg's group DNA polymerase and synthesize nucleic acids *in vitro* [they receive the Nobel prize in 1959]

1957 VM Ingram reports the amino acid sequence of HbS; H Frankel-Conrat, A Gierer and G Schramm independently demonstrate that the genetic information of tobacco mosaic virus is stored in RNA [CPG p.264]

1958 MS Meselson & FW Stahl demonstrate that DNA replication is semiconservative (in *E.coli*)

1959 J Lejeune *et al* show that Down's syndrome is a chromosomal abnormality [trisomy of a small telocentric chromosome] as the first identification of the genetic basis of a disease; PA Jacobs & JA Strong identify the chromosomal basis of Klinefelter's syndrome as XXY

1960 Riis & Fuchs performs the first prenatal sex determination; Moorhead performs the first chromosome analysis

1961 MF Lyon and LB Russell independently show that one of the X chromosomes is inactivated in females; SB Weiss & T Nakamoto isolate RNA polymerase; MW Nirenberg starts experiments to unveil the genetic code [gets the Nobel prize in 1968 together with Khorana]; F Jacob & J Monod publish *Genetic Regulatory Mechanisms in the Synthesis of Proteins* in which they propose the operon model for regulating gene expression in bacteria [they receive the Nobel prize in 1965]. Robert Guthrie in New York performs first genetic screening of newborns (for phenylketonuria).

1962 Werner Arber notices that E.coli extracts restrict viral replication with some enzymatic activity, hence the name restriction endonucleases. He later shared the 1978 Nobel prize with Smith and Nathans.

1964 DJL Luck & E Reich isolate mitochondrial DNA from *Neurospora*; WD Hamilton proposes the genetical theory of social behaviour; F Lilly *et al* shows the genetic basis of susceptibility to leukemia in mice (first documented MHC - disease association)

1965 S Brenner *et al* discovers the stop codons; RW Holley works out the first complete nucleotide sequence of a tRNA (yeast alanine tRNA) [receives the Nobel prize in 1968 together with Nirenberg and Khorana]

1966 B Weiss & CC Richardson discover DNA ligase; VA McKusick publishes *Mendelian Inheritance in Man* which is available [online](#).

1967 CB Jacobson & RH Barter use amniocentesis for prenatal diagnosis of a genetic disorder; MC Weiss & H Green works out the autosomal chromosomal assignment of a human gene for the first time [thymidine kinase gene]; HG Khorana *et al* establishes the genetic code [receives the Nobel prize in 1968 together with MW Nirenberg]; amniocentesis and chromosome analysis are developed

1968 RT Okazaki *et al* report the discontinuous synthesis of the lagging DNA strand; M Kimura proposes the Neutral Gene Theory of Molecular Evolution; RP Donahue *et al* assigns the Duffy blood group locus to chromosome 1; S Wright publishes the first volume of *Evolution and the Genetics of Populations*

1969 HA Lubs finds a fragile site on the X chromosome and its clinical correlation (mental retardation in males); M Delbruck, SE Luria and AD Hershey receive the Nobel Prize for their contributions to viral genetics

1970 M Mandel & A Higa develop a method for transformation of bacteria [CaCl₂ method]; D Baltimore & HM Temin isolate reverse transcriptase from two oncogenic RNA viruses; R Sager & Z Ramanis publish the first genetic map of non-mendelian genes (chloroplast genes of *Chlamydomonas*); T Caspersen *et al* do the first chromosome banding. [Hamilton Smith and Daniel Nathans](#) successfully used *HindIII* to manipulate DNA sequences reproducibly (and received the 1978 Nobel Prize along with Arber (see 1962)).

1971 ML O'Riordan *et al* shows all 22 pairs of human autosomal chromosomes by quinacrine dyeing and identifies the Philadelphia chromosome as an aberrant chromosome 22; AG Knudson suggests that the retinoblastoma locus acts as a dominant anti-oncogene.

1972 DE Kohne *et al* studies the evolution of primates by DNA:DNA hybridisation; In P Berg's laboratory, the first recombinant DNA is produced in vitro [P Berg receives the Nobel prize in 1980]

1974 RD Kornberg describes the chromatin structure (nucleosomes); RW Hedges & AE Jacob discover the bacterial plasmid as an ampicillin-resistant gene (transposon)

1975 EM Southern describes the Southern transfer method; F Sanger & AR Coulson develop the DNA sequencing method; R Dulbecco, H Temin, and D Baltimore receive Nobel prizes for their studies on oncogenic viruses

1976 [WY Kan *et al*](#) perform the first DNA diagnosis (prenatally) in alpha-thalassaemia by RFLP analysis; JM Bishop & HE Varmus demonstrate the protooncogene to oncogene relationship [they receive the Nobel prize in 1989]

1977 JC Alwine *et al* describe the Northern blotting method; RJ Roberts and PA Sharp separately describe split genes in adenovirus; J Collins & B Holm develop cosmid cloning technique; K Itakura *et al* chemically synthesize a gene for human somatostatin and express it in E Coli, thus produce the first human protein *in vitro*; W Gilbert induces bacteria to synthesize insulin and interferon; Sanger *et al* publish the complete sequence of phage $\Phi X174$ (5387 nucleotides) [Sanger & Gilbert receive the Nobel prize in 1980, second for Sanger]

1978 W Gilbert coins the terms intron and exons; T Maniatis *et al* develop the genomic library screening technique. Boyer & Swanson cofound the first biotechnology company, Genentech. The first test-tube baby is born in the UK

1979 Edwards & Steptoe achieve in vitro fertilisation; DV Goeddel *et al* produce human growth hormone using recombinant DNA technology

1980 JW Gordon *et al* produce the first transgenic mouse; Dr Chakrabarty is awarded the first patent for a genetically engineered (unicellular) organism; GD Snell, J Dausset, and B Benacerraf receive the Nobel prize for their studies on the MHC

1981 Identification of the first cancer causing gene; DNA analysis is developed for diagnosis of sickle cell trait

1982 Sanger *et al* publish the complete sequence of phage lambda (48,502 nucleotides). The first transgenic mouse is created (carrying a rat growth hormone gene)

1983 Gene for Huntington's disease is located to chromosome 4

1984 Alec Jeffreys develops genetic fingerprinting. The first American test-tube baby is born

1985 Cystic fibrosis gene is located to chromosome 7

1986 RK Saiki, KB Mullis and five colleagues describe the polymerase chain reaction [Mullis receives the Nobel prize in 1993]; muscular dystrophy gene is identified

1987 R Sinsheimer proposes the human genome project; the US Patent Office rules that multicellular organisms produced by genetic engineering may be patented

1989 LC Tsui, F Collins and co-workers clone the cystic fibrosis gene; TR Cech & S Altman receive the Nobel Prize for establishing the existence of catalytic RNAs

1990 WF Anderson in the USA reports the first gene successful therapy in humans (in ADA deficiency causing SCID); the California Hereditary Disorders Act comes into force; human genome project begins

1993 Huntington's disease gene is identified; gene therapy for SCID and cystic fibrosis begins in the UK

1994 The FlavrSavr tomato is approved by the FDA as the first GM food to go on the market (now discontinued)

1997 Complete *Saccharomyces cerevisiae* genome is sequenced; complete *E.coli* genome is sequenced [*Science* 277;1453-74]

1998 *Caenorhabditis elegans* becomes the first animal whose genome is totally sequenced [*Science* 282:2012-8]

1999 A human MHC (HLA-DR52) haplotype is totally sequenced (October). Human chromosome 22 becomes the first one to be sequenced completely (November)

2003 Complete sequence of human Y-chromosome is published [*Nature* 423:825-38]

Compiled from:

Classic Papers in Genetics. JA Peters (Ed). Englewood Cliffs: Prentice-Hall, Inc. 1959 [CPG]

Genetics in Context [timeline]: <http://www.esp.org> (scroll down and choose Chronology)

AH Sturtevant's [A History of Genetics: Chronology](#) Appendix

King RC & Stansfield WD. [A Dictionary of Genetics](#) (5th Edition, 1997). New York: Oxford University Press

[A Student's Guide to Biotechnology](#): Words and Terms. Greenwood Press, 2002

[Dynamic Timeline](#) from Human Genome Project Website.

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on March 9, 2004

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

BASIC GENETIC CONCEPTS

M.Tevfik Dorak, MD, PhD

Dominant vs Recessive

Mendelian inheritance patterns usually apply to traits governed by a single gene. Most characters, however, are determined by multiple genes or by an interaction of genes and the environment. The inheritance of quantitative characteristics that depend on several genes is called **polygenic** inheritance. A combination of genetic and environmental influences is known as **multifactorial** inheritance. Diseases with multifactorial inheritance are called **complex genetic diseases**.

A gene may act as **dominant** or **recessive**, although dominance or recessiveness concerns the phenotype. But these are not the only modes of action (incomplete

dominance and co-dominance are possible). Examples of dominant characters include dark hair (to blonde hair); brown eyes to blue eyes (as originally though by Davenport & Davenport in their 1907 Science paper 'available through [JSTOR](#), but also see more recent views [here](#)); and lobed ears to unlobed ears. Possible examples of recessive characters are, therefore, blonde hair and blue eyes. Ability to taste a chemical (PTC) is dominant to inability to taste it. The grey body colour of *Drosophila* is dominant to black. Yellow is dominant to green in seed cotyledons of *Pisum sativum* used in Mendel's experiments. A dominant allele usually codes for a functional product and continues to do so in heterozygous form (however, beware of dominant negative and haploinsufficiency; see [glossary](#) and [Clinical Genetics](#)). A recessive allele usually directs the synthesis of a non-functional product causing the lack of the product in homozygous form. For example, a green seed is always homozygous but a yellow one may be heterozygous (yellow is dominant to green).

Recessive sex-linked genes do not require homozygosity for expression (coat color in cats, various sex-linked diseases such as hemophilia and red-green color-blindness in humans). A rare X-linked dominant trait is the blood type X_g^a (males expressing this trait would transmit it to all daughters but not to a son).

Incomplete dominance (also called blending, semi-dominance): The four-o'clock plant (the snapdragon, *Antirrhinum*), for example, may have flowers that are red, white, or pink. Plants with red flowers have two copies of the allele *R* for red flower color (pure line for red) and hence are homozygous *RR*. Plants with white flowers have two copies of the allele *r* for white flower color (pure line for white) and are homozygous *rr*. Plants with one copy of each allele, heterozygous *Rr*, are pink—a blend of the colors produced by the two alleles. Fur color of Andalusian Fowl is another example: black and white are partial-dominant, therefore heterozygotes are grey.

Co-dominance: ABO blood groups (A and B are co-dominant; O is recessive) and HLA transplantation antigens (HLA-A,B,C,DR,DQ,DP) are co-dominantly expressed. The gene responsible for sickle cell anemia has two co-dominant alleles: HbA (normal gene) and HbS (sickle cell gene). HbAA is normal, HbSS is diseased and HbAS is mildly anemic and protective for malaria (hence its selection). White or Dutch clover (*Trifolium repens*) leaf patterns are co-dominantly expressed. A heterozygous plant shows both patterns superimposed, while homozygotes for the other two patterns show a single pattern.

Sex-influenced dominance: A dominant expression that depends on the sex of the individual. An example is the horns in sheep (dominant in males, recessive in females). Another one is the plumage in domestic fowl: autosomal alleles whose expression is modified by sex hormones; the same genotype (*hh*) results in long, more curved and pointed plumage in cocks; shorter and more rounded in hens. Certain coat patterns in cattle, baldness patterns, breast development and facial hair (as well as all secondary sex characters) in humans are also sex-influenced traits. An autosomal dominant trait that controls precocious puberty is expressed in heterozygous males but not in heterozygous females. Affected males undergo puberty at 4 years of age or earlier. Heterozygous females are unaffected but pass this trait on to half of their sons (confusing the pattern with a sex-linked trait).

A different phenomenon is **temperature-influenced expression** (in Himalayan rabbit and Siamese cat fur color). The temperature-sensitive allele causes say, darker patches in extremities, ears and nose.

Further points: As in rabbit fur and mouse coat color determination, dominance relationships between alleles may be more complicated. The reasons for dominance may be related to the activity of an enzyme coded by the relevant gene. Mendel's yellow pea allele is dominant because the gene involved codes for the breakdown of chlorophyll (which is green). In the homozygous recessive case, no functional enzyme will be present, chlorophyll breakdown cannot occur and the seeds remain green. In a situation called **epistasis**, one gene affects the expression of another gene that is not linked (multigenic determination of a phenotype). The masked gene is said to be **hypostatic** to the **epistatic** gene. An epistatic-hypostatic relationship between two loci is similar to a dominant-recessive relationship between alleles at a particular locus.

Genomic imprinting, methylation and penetrance are other factors that may influence the expression of characters. If one of the possible genotypes is an **embryonic lethal**, the observed proportions would be different from expected ones (also remember the **cytoplasmic male sterility** in plants).

Having a disease-causing gene does not always mean having the disease. In [phenylketonuria](#) (PKU), avoiding phenylalanine-containing food prevents ill effects of the mutant gene; and in [xeroderma pigmentosum](#), avoiding ultraviolet radiation prevents the development of melanoma. On the other hand, having a protective gene may prevent the ill effects of an environmental agent. The gene for the xenobiotic enzyme CYP1A1 on chromosome 15 may activate the polycyclic aromatic hydrocarbons (PAH) depending on the allele. PAHs are present in cigarette smoke and those who have the non-activating allele may not be affected from the carcinogenic effects of PAH.

SuperLectures on Genetics by RM Fineman: [Part I](#) and [Part II](#)

[Genetics: A Beginner's Guide](#)

[M.Tevfik Dorak, B.A., M.D., Ph.D.](#)

Last updated on March 8, 2004

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Evolution](#) [Back to Genetics](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Homepage](#)

Common Terms in Evolutionary Biology and Genetics

M.Tevfik DORAK

[PBS Evolution](#) [On Line Biology Book - Glossary](#) [Glossary of Genetic Terms](#) [Talking Glossary \(Genetics\)](#) [Life Science Dictionary](#) [Cytogenetics Glossary](#)
[UCMP Glossary \(Evolution\)](#) [BTO \(Genetics\)](#) [Population Genetics Glossary](#) [Molecular Biology Glossary \(ASH\)](#) [Molecular Biology Glossary \(UM\)](#)
[More Human Genetics Glossaries](#)

[For best results, please use the FND option by pressing "CTRL + F" to locate the word you are looking for]

Acrocentric chromosome: A chromosome with its **centromere** towards one end. Human chromosomes 13,14,15,21,22 are acrocentric.

Adaptation: Adjustment to environmental demands through the long-term process of natural selection acting on genotypes.

Adaptive immunity: A collective term for the long-lasting and specific response of lymphocytes to antigens. Requires the **MHC**, T-cell receptors (**TCR**) and immunoglobulins (Ig) as well as enzymes with a **recombinase** activity (for the rearrangements at **TCR** and Ig gene loci). Present in all vertebrates except **jawless fish** (see **innate immunity**).

Additive and non-additive components: In studies of heredity, the portions of the genetic component that are passed and not passed to offspring, respectively.

Advanced (synonym: derived; opposite: primitive): In phylogenetic studies, an organism or character further removed from an evolutionary divergence than a more primitive one.

Agnatha (means jawless): The Class Agnatha represents the most primitive 'jawless' vertebrates. MHC genes have been cloned from all vertebrate classes except Agnatha.

Agrobacterium tumefaciens: A soil bacterium that causes a cancer-like plant disease (crown gall) in dicotyledonous plants (all agricultural crops except cereals). It contains the **Ti plasmid**. The tumor induction ability of the bacterium spreads to neighboring cells via the plasmid.

α -helix: Common secondary structure of proteins in which the linear sequence of amino acids is folded into a spiral that is stabilized by hydrogen bonds between the carboxyl oxygen of each peptide bond.

Algae: A heterogeneous group of aquatic, unicellular, colonial or multicellular, eukaryotic and photosynthetic organisms. They belong to the Kingdom Protista and include the multicellular red (rhodophyte), green and brown (kelp) algae. They are not plants but all land plants evolved from the green algae (see also **Chromista**).

Allee effect: The benefit individuals gain from the presence of conspecifics. Link to a brief explanation of [Allee effect](#).

Allele: A known variation (version) of a particular gene. Formerly called allelomorph.

Allelic association: see **linkage disequilibrium**.

Allelic exclusion: Expression of only one of the two homologous alleles at a locus in the case of heterozygosity. This usually occurs at loci such as immunoglobulin or T cell receptor (TCR) genes where a functional rearrangement among genes takes place. One of the alleles is either non-functionally or incompletely rearranged and not expressed. This way, each T-cell expresses only one set of TCR genes.

Allelopathy: The influence exerted by a living plant on other plants nearby or microorganisms through production of a chemical.

Allogeneic: Two genetically dissimilar individuals of the same species like any two human beings except monozygotic twins.

Allophenic: Chimeric, i.e., composed of cells of two different genotypes (also called hybrid).

Allorecognition: Recognition by T cells of the **MHC** molecules on an allogeneic individual's antigen-presenting cells which results in allograft rejection *in vivo* and **mixed lymphocyte reaction** (MLR) *in vitro*.

Altered self: A term used to describe the MHC molecule associated with a peptide rather than in its native form. Thus, a native MHC molecule does not induce an immune reaction except when it is presenting a peptide.

Alternation of generations: An alternation of sexual (haploid) and asexual (diploid) form of generations in a life cycle (example: aphids). The relative dominance of each phase is variable in each organism (mosses have a dominant haploid phase whereas angiosperms have a dominant diploid phase). Besides aphids, Daphnia (water flea), rotifers, Hydra have alternation of generations in response to environmental conditions.

Alternative splicing: Formation of diverse mRNAs through differential splicing of the same RNA precursor.

Altruism: Helping others without direct benefit, and sometimes harm, to oneself.

Amphipathic: A molecule that has both a hydrophobic and a hydrophilic part.

Analogy: A similarity due to convergent evolution (common function) but not inheritance from a common ancestor (bat's wings and bird's wings). See also **homology**.

Angiosperm: The most recently evolved and the largest group of plants whose reproductive organs are in their flowers (flowering plants). A superclass in the sperm plants (Spermatophyta) division belonging to the vascular plants (Tracheophyta) phylum of the plant kingdom. They are divided into two subclasses: Dicots (Magnoliopsida such as magnolia, dandelion, roses, violet) and Monocots (Liliopsida such as lily, iris, orchid, grasses). Their ovules are enclosed in the carpel and pollen travels through the pollen tube to reach it. Angiosperms evolved in the Cretaceous era together with the Mammals. **Angiosperms** in Tree of Life.

Anisogamy: Sexual reproduction in which one sex produces sex cells much larger (egg) than those of the other (sperm).

Antagonistic pleiotropy: The effects of a gene which are beneficial early in life (i.e., increasing fitness) but deleterious later in life (no change in fitness after the reproductive age). Such genes will be maintained by selection, because by the time the gene exerts its damage, its bearers will already have had more offspring than other individuals.

Anthropology: The study of human kind.

Antigen: Any macromolecule that triggers an immune response. Antigenicity depends on the ability of the peptide fragments to be presented by the **MHC** molecules.

Antisense DNA/RNA: Single stranded nucleic acid that is complementary to the coding/sense strand of a gene. It is then also complementary to the mRNA produced from the same gene.

Apes: Species belonging to the Family Pongidae of the Order **Primates**: Gibbon (genus Hylobates), Orangutan (genus Pongo), Chimpanzee (genus Pan), and Gorilla (genus Gorilla). They have no tails.

Apoptosis: The genetically programmed death of cells at specific times during embryonic morphogenesis and development, metamorphosis, and during cell turnover in adults including the maturation of T and B cells of the immune system. Defects in apoptosis are associated with maintenance of the transformed state and cancer. Anti-apoptotic proteins include Bcl-2 and **HSP** families (see also **caspase**). Apoptosis is often induced by activation of death receptors (DR) belonging to the tumor necrosis factor receptor (TNFR) family. Examples are Fas (CD95), TNFR-1 and TNFR-related apoptosis-mediated protein (TRAMP). Death signals are conducted through

a cytoplasmic motif (death domain - DD) - death-inducing signaling complex (DISC) and caspase-8 that leads to the activation of caspase cascade and eventual death of the cell.

Aposematic colouring: Colouration that warns the predators about poisonous or distasteful nature of the organism. Usually black and yellow stripes in animal. Aposematic colouring of poisonous organisms is sometimes mimicked by others to gain advantage against predators (**mimicry**). (Link to an interesting article by Lev-Yadun on **aposematic colouring in plants**.)

Arabidopsis thaliana: A small member of the mustard family (kitchen cress). It has a very small genome (130-140 Mbp), five chromosomes and contains almost no repetitive DNA. Its genome will be completely sequenced by the end of 2000. It is a plant model system of choice because of the additional advantages of short generation time (about five weeks), high seed production (up to 40,000 seeds per plant) and natural self-pollination (as opposed to natural cross-pollination in maize). It has five small chromosomes. Link to **Arabidopsis website**.

Arboreal: Tree-living (like monkeys).

Archaea: A prokaryote kingdom that has not diverged much from the ancestral prokaryote stock. Contemporary species of Archeobacteria live in extreme conditions. The three major groups are halobacteria, sulphobacteria and methanogens. All other prokaryotes are grouped in Eubacteria.

Archezoa: One of the kingdom level taxa proposed by Cavalier-Smith which consists of the most ancient unicellular eukaryotes with a nucleoid and rod shaped chromosome but no mitochondria or plastid, thus believed to be the intermediate stage between prokaryotes and eukaryotes. They are also used as evidence for the evolution of nucleoid before the organelles. The intestinal parasite Giardia lamblia (a protist) is an example.

ARS: Autonomously replicating sequence. ARS is the origin of replication in yeast.

Artificial selection: Selective evolutionary pressure imposed by humans to obtain breeds with certain features (such as breeding cows, dogs, chicken).

Asexual reproduction: Any form of reproduction not depending on a sexual process. It involves a single individual. Reproduction by cell division, fragmentation or budding.

Association (genetic): Association refers to a concurrence greater than predicted by chance between a specific allele and another trait (for example, a disease) that may or may not have a genetic basis. Evaluation of association requires the study of unrelated individuals. Association studies may prove useful in identifying a genetic factor in a disease. Except when linkage disequilibrium exists, association is not due to genetic **linkage** and should not be confused with it.

Assortative matings: Reproduction in which mate selection is not random but is based on physical, cultural, or religious grounds (see **negative and positive assortative mating**).

Atomic mass unit (amu or dalton): The basic unit of mass on an atomic scale. One amu or dalton is one-twelfth the mass of a carbon 12 atom (in other words, the mass of a hydrogen atom, 1.66×10^{-24} g). Therefore, there are 6.023×10^{23} amu in one gram (Avogadro number).

Australopithecus: The extinct genus of Plio-Pleistocene hominids found in South and East Africa. The evolutionary link between **apes** and humans.

Automorphy: Unique derived characteristic; a trait present in only one member of a lineage or in only one lineage among many.

Autosome: Any chromosome except a sex chromosome.

Axolotl (Ambystoma mexicanum): Literally meaning water monster (in Aztec), axolotl is a salamander (amphibian) extensively used in the evolutionary immunogenetic studies.

Bacillus thuringiensis: This bacteria is pathogenic to insects and the gene for its toxin is used to create transgenic plants with their own insecticide.

Background extinction: The normal constant occurrence of extinctions. See **extinction files** in BBC Education.

Bacteriophage: A virus that infects a bacterium.

Balanced lethal: Lethal mutations in different genes on the same pair of chromosomes that remain in repulsion because of close linkage or crossover suppression. In a closed population, only the trans-heterozygotes ($I_1 + / + I_2$) for the lethal mutations survive.

Balanced polymorphism: The maintenance of two or more alleles in a population due to a selective advantage of the heterozygote.

Balancing selection: Selection involving opposing forces in which selective advantages and disadvantages cancel each other out. Heterozygote advantage (or overdominant selection) is an example in which an allele selected against in the homozygous state is retained because of the superiority of heterozygotes. Other balanced states may occur including when: an allele is favoured at one developmental stage and is selected against at another (antagonistic pleiotropy); an allele is favoured in one sex and selected against in another (sexual antagonism); an allele is favoured when it is rare and selected against when it is common (negative frequency dependent selection).

Barr body: Also called sex-chromatin body, which represents the inactivated X chromosome in the nucleus of somatic mammalian cells. Normally only seen in female cells and not in male cells. It is the result of the process called **dosage compensation**.

Base: A compound, usually containing nitrogen, that can accept a H^+ . It is used to describe the non-sugar components of nucleotides (despite the basic nature of nucleotides, nucleic acids are acidic due to the phosphate atoms they contain). The five bases that form the nucleic acids are adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U).

B cells: A major family of small lymphocytes that are responsible for antigen-specific humoral immunity as part of the adaptive immunity. Their antigen receptors are surface immunoglobulins (antibodies). They recognize peptides directly and secrete antibodies by differentiating into plasma cells. They also exist as long-lived memory cells.

B Factor: A fungal incompatibility factor. It operates in the Basidiomycetes species **Schizophyllum commune** (not to be confused with Factor B of the immune system).

Binary fission: Mode of reproduction not involving any sex but division of a parent cell into two equally sized offspring.

Biome: A grouping of plant ecosystems into a large distinct group occupying a major terrestrial region. They are created and maintained by climate. See examples of **biomes**.

Biosphere: The geographical region of the Earth where life is found.

Bipedal: Two-footed posture and locomotion like a human standing upright on the hindlimbs.

Bony fish (class Osteichthyes): The vertebrate Class evolved after **jawless** and **cartilaginous fishes**. They have jaws, their skeleton is made up of bone and their body is covered with overlapping scales. Most familiar freshwater and sea water fishes belong to this group. The living fossil Coelacanth is a bony fish whose relatives (lobe-finned fishes) can be traced back to the Devonian geological period (363-409 Mya).

Botryllus schlosseri: A colonial hemaphrodite (Phylum Chordata, Subphylum Tunicata, Class Ascidian=sea squirts). It grows fast, reproduces weekly, and thus, a good model for genetic studies of **protochordates**. It also has a well-characterized allorecognition system called **Fu/HC** whose functions are prevention of fusion with non-kin and selective fertilization by sperm bearing **Fu/HC** alleles different from that of the egg. See a webpage about **Botryllus**.

Bottleneck: A drastic reduction in the population size followed by an expansion. This often results in altered gene pool as a result of genetic drift.

β -pleated sheet: A planar secondary structure element of proteins. It is created by hydrogen bonding between the backbone atoms in two different polypeptide chains or segments of a single folded chain.

CAAT box: A highly conserved DNA sequence found about 75 bp 5' to the site of transcription in eukaryotic genes. Its specific (trans-acting) transcription factor is CTF-1 (NF-1) (see also **TATA / Goldberg-Hogness box**).

Caenorhabditis elegans: A normally self-fertilizing hermaphrodite soil nematode whose developmental genetics has been extensively studied. It is no more than 1 mm long. Loss of an X chromosome by meiotic disjunction leads to the production of males. The genetic basis of apoptosis was first shown in *C.elegans* in 1986. It has five equally sized chromosomes and it is the first animal whose whole genome has been sequenced (in 1998). The 97 Mbp genome contains 19,000 genes on 6 chromosomes. About 74% of human genes have their homologues in the *C.elegans* genome. Links to the [C.elegans](#) website and an introduction to [C.elegans](#).

Cambrian: Cambria is the old name for Wales (UK) where the first skeletalized animal fossils were found. The first period in the Palaeozoic era marked by the occurrence of many forms of invertebrate life (540-500 Mya). The sudden appearance of the major animal phyla in the fossil record during the Cambrian period is called Cambrian explosion.

Cap: A methylated guanine residue (**GTP**) which is added to the 5' end of eukaryotic mRNAs in a post-transcriptional reaction. It protects the mRNA against 5'-exonuclease.

Cap site: The initiation site of transcription in a eukaryotic gene. The initiation of translation of most eukaryotic mRNAs involves recognition of the cap followed by either the first downstream AUG or by a 5' proximal AUG with a consensus sequence surrounding it (like the bacterial **Shine-Dalgarno** or the viral **Kozak** sequence). Such a consensus sequence has not been recognized in eukaryotes yet.

Carrier: A healthy person who is a heterozygote for a recessive trait. Also includes persons with balanced chromosomal translocations. The unfortunate use of 'carrier' to describe individuals positive for a genetic marker is wrong, and the use of 'carrier frequency' in that context should be replaced by 'marker frequency'.

Carter Effect: Higher incidence of a genetically determined condition in relatives when the index case is the less commonly affected sex. This phenomenon was first demonstrated in Dr **Cedric Carter's** study of pyloric stenosis, where the incidence is highest in the sons of affected women and lowest in daughters of affected men.

Cartilaginous fish (Class Chondrichthyes): The most primitive 'jawed' vertebrates evolved about 400 Mya. Their skeleton is composed of entirely cartilage. The Class includes the sharks, rays and skates (subclass Elasmobranchii) and the ratfish (subclass Holocephali). The earliest taxon which has both **MHC** class I and class II genes. The next step in the evolutionary ladder is the **bony fish**.

Caspase: Cysteine-containing aspartic acid-specific proteases involved in the execution phase of apoptosis. Fas/Fas ligand system is one activator of caspase-dependent apoptotic cell death.

Catarrhini: One of the two divisions (suborder) of Primates containing the old world monkeys and apes (extinct and extant). The other division is **Platyrrhini** (new world monkeys).

Centromere: Constricted region where sister chromatids are attached in mitotic chromosomes. The centromere is generally flanked by repetitive DNA sequences and it is late to replicate. The centromere is an A-T region of about 130 bp. It binds several proteins with high affinity to form the kinetochore which is the anchor for the mitotic spindle.

Centric fusion: Fusion of the long arms of two acrocentric chromosomes [13,14,15,21,22] into a single chromosome having lost the short arms at the same time. Most often occurs as 21/21, **13/14**, and 14/21 translocations. Apart from being an important cause of **uniparental disomy**, it may cause **trisomy 21** (Down's syndrome) in the offspring. Human chromosome 2 is a result of a centric fusion between two ancestral ape chromosomes (gorillas have 24 pairs of chromosomes).

Cetacea: An Order of marine mammals. The suborder Odontoceti includes dolphins, killer whales and toothed whales. Many of the great whales (such as the blue whale) belong to a different suborder (baleen whales or Mysticeti).

Chaperone: Any cellular protein that binds to an unfolded or partially folded target protein to prevent misfolding, aggregation, and/or degradation of it. Chaperones also facilitate the target protein's proper folding, translocation and assembly within cells, preventing inappropriate interactions with other proteins.

Character displacement: Forced evolution of dissimilar characters in related species where their ranges overlap. Elsewhere, where they exist on their own, their similarities are maintained.

Character release: This is the opposite of character displacement. Two closely related species become more alike in regions where their ranges do not overlap than in regions where they do.

Chiasma (plural **chiasmata**): The points of physical overlap of nonsister chromatids crossing-over in meiosis.

Chi-like Sequence: An octamer nucleotide sequence (A/G - C/T - A/T - A/G - G - A/T - G - G) that creates a recombinational hotspot in the genome (originally discovered in coliphage lambda). MHC class I transmembrane domain length variation, frequent gene conversions and deletions in the MHC-linked 21-hydroxylase gene (CYP21), gene conversions within the MHC class II genes in mice and humans, many oncogene translocations (BCL2 for example) are attributed to chi-like sequences at the breakpoint region. It acts like a restriction site for recombinase.

Chlamydomonas: The unicellular green alga that is probably the closest living organism to the ancestor of green plants. It reproduces both asexually and sexually (two mating types). When reproduces sexually, the mitochondria are inherited from the (-) mating type and chloroplasts from the (+) mating type.

Chloramphenicol acetyl transferase (CAT): The bacterial gene for chloramphenicol, CAT, is commonly used as a reporter gene for investigating physiological gene regulation. Beta-galactosidase and luciferase genes can also be used for the same purpose.

Chloroplast: A major component of a **plastid** in green plants and eukaryotic algae of any colour. It is involved in photosynthesis. Prokaryotic photosynthetic organisms do not have chloroplasts.

Chordata: A major phylum in the Kingdom Animalia. A chordate is characterized by the presence of a dorsal notochord at some stage of development and a dorsal hollow nerve chord.

Chromatid: One of two copies of a replicated chromosome during mitosis. Together they are called sister chromatids. Each one becomes a daughter chromosome at anaphase of mitosis and at the second meiotic division.

Chromatin: The complex of DNA and associated histone and non-histone proteins that represents the normal state of genes in the nucleus. It exists in two forms: euchromatin can be transcribed, and heterochromatin is highly condensed and cannot be transcribed.

Chromista: A major taxon in the Kingdom Protocista (also called Heterokonta). They are mostly photosynthetic algae but distinct from the rest of the algae. They have genetic material derived from four different ancestors (purple bacteria, cyanobacteria, red alga and green alga). Link to [Chromista](#).

Chromosome: Structure in a cell nucleus that carries the genes. Each chromosome consists of one very long strand of DNA, coiled and folded to produce a compact body. They become more compact and visible during metaphase of cell division. In interphase chromosomes, chromatin fibers are organized into 30 to 100 kb loops anchored in a supporting matrix within the nucleus. The length of each DNA molecule must be compressed about 8000-fold to generate the structure of a condensed metaphase chromosome.

Ciliates: This group of protists is most like animals in their behaviour and complexity. The **Paramecium** is the representative of the Ciliates.

Cis-acting gene: A gene acting on or co-operating with another gene on the same chromosome (see **trans-acting gene**).

Cistron: A DNA segment coding for a specific polypeptide, and includes its own start and stop codons. When an mRNA encodes two or more proteins, it is called polycistronic.

Clade: All descendants of any given species. A single whole branch of a phylogeny.

Class: A category of classification (taxon); a subdivision of subphylum. The classes in the Subphylum Vertebrata are: Pisces (Fishes), Amphibia, Reptilia, Avis (Birds) and Mammalia.

Class switching: The process by which a IgM or IgD producing B lymphocyte switches to produce one of the secondary immunoglobulins (IgG, IgA or IgE) with the same antigen binding specificity.

Clone - All the cells derived from a single cell by repeated cell division and having the same genetic constitution.

Cnidaria: See Coelamata.

Coadaptation: Interaction of genes at the genotypic level. Natural selection acts on the complex product of such interactions rather than on individual locus. The correlated variation and adaptation present in two mutually dependent organisms.

Coalescence: Growing into each other, uniting into one whole. Link to a lecture on [coalescence](#).

Coalescence theory: The evolutionary theory that estimates the time for divergence from the last common ancestor. A lecture on [Coalescent Theory](#).

Codominance: Equal effect on the phenotype of two alleles of the same locus (as opposed to **recessive** and **dominant**).

Codon bias: Although several codons code for a single amino acid, an organism may have a preferred codon for each amino acid. This is called codon bias.

Coelacanth: Any member of a zoological family of largely extinct lobe-finned fish (like the living genus Latimeria).

Coelomata (Cnidaria): A phylum in the Animal Kingdom. They are invertebrate aquatic animals showing radial or biradial symmetry. Examples are corals, sea anemones, jelly fish. See also [Cnidaria](#) in the Shape of Life (PBS).

Co-evolution: Joint evolution of two unrelated species that have a close ecological relationship resulting in reciprocal adaptations as happens between host and parasite, and plant and insect.

Cognate molecule: A relative descended from a common ancestor. Usually used to describe the corresponding partner in a receptor-ligand complex.

Cohesive end: Also known as sticky end. Overhanging ends of a double-stranded DNA molecule that are capable of hybridizing with complementary ends.

Complementary (copy) DNA (cDNA): Single-stranded DNA produced from an RNA template (usually mRNA) by **reverse transcriptase** in vitro. It lacks the introns present in corresponding genomic DNA. It is most commonly made to use in PCR to amplify RNA (RT-PCR).

Compound heterozygote: An individual who is affected with an autosomal recessive disorder having two different mutations in homologous alleles.

Concerted (coincidental) evolution: The preservation of sequence homology among members of a multigene family within the same species.

Congenic: Animals which have been bred to be genetically identical except for a single gene locus. This is achieved by superimposing the locus of interest on the genetic background of another by first crossing two inbred lines followed by extensive (about 20 generations) backcrossing hybrids to one parental line (the background strain) while selecting for the alleles of the locus of interest of the other. The result is an inbred strain uniquely identified by a difference at a single locus.

Conifers: Cone bearing trees. A Class of the Gymnospermae which includes needle-leaved trees such as pines and cypresses. Their flowers are in cones, and male and flower cones are separate. The oldest (bristlecone pine) and the largest (sequoia) extant organisms belong to this Class. Their unique feature is the inheritance of cytoplasmic DNA (chloroplasts) via pollens.

Conjugation: In unicellular organisms, temporary cell contact between complementary genders and exchange of genetic material, as in the ciliate **Paramecium aurelia**, or one-way transfer of genes as in bacteria (see **F Factor**) and green alga Spirogyra.

Consanguineous matings: Matings between two individuals who share a common ancestor in the preceding two or three generations.

Consensus sequence: The nucleotides or amino acids most commonly found at each positions of the sequences of related molecules.

Convergent evolution: Evolution of two or more different lineages towards similar morphology due to similar adaptive pressures. Examples of convergence are: fins or fin-like structures in fish, cuttlefish and whales; extreme similarity in alarm calls by five small birds; endothermy in dogs and ducks, wings of butterflies and birds.

Coefficient of relatedness: $r = n(0.5)^L$ where n is the alternative routes between the related individuals along which a particular allele can be inherited; L is the number of meiosis or generation links.

Correspondence analysis: A complementary analysis to genetic distances and dendrograms. It displays a global view of the relationships among populations (**Greenacre MJ, 1984**). This type of analysis is more informative and accurate than dendrograms especially when the number of loci is fewer than 30 (when phylogenetic analysis tends to be inaccurate (**Nei M, 1996**)). Using the same allele frequencies that are used in phylogenetic tree construction, **correspondence analysis** can be performed on the **ViSta** v6.4 or **MVSP** computer programs.

Coupling (cis-arrangement): The condition in which a double heterozygote has received two linked mutations from one parent and their wild-type alleles from the other parent, e.g., $a b / ++$ (as opposed to $a + / + b$; see also **repulsion**).

CpG island: Repetitive CpG doublets creating a region of DNA greater than 200 bp in length with a G+C content of more than 0.5 and an observed/expected presence of CpG more than 0.6. Usually associated with transcription-initiation regions of (housekeeping) genes transcribed at low rates that do not contain a TATA box. The CpG-rich stretch of 20-50 nucleotides occurs within the first 100-200 bases upstream of the start site region (where **promoter-proximal elements** reside). A trans-acting transcription factor called SP1 recognizes the CpG islands (see also **Htf islands**). In vertebrates, many of the nontranscribed genes (and the genes on the inactivated X chromosome) have a 5-methyl group on the C residue in CpG dinucleotides in transcription-control regions. On the other hand, many genes with restricted expression patterns have (methylated) CpG islands located downstream of transcription initiation does not block elongation of the transcript (see also **Methylation paradox**).

Cretaceous: The youngest of the three geologic periods of the Mesozoic era (145 to 65 Mya).

Crossing-over (recombination): The exchange of genetic material between non-sister chromatids of homologous chromosomes (i.e., between maternal and paternal chromosomes) during meiosis. This results in a new and unique combination of genes on the daughter chromosome which will be passed on to the offspring (if that particular gamete is involved in fertilization).

Crustacean: Any of a zoological class (Crustacea) that have a chitinous and/or calcareous exoskeleton (lobsters, shrimps, crabs).

Cryptic female choice: Besides precopulatory female sexual selection, there are also postcopulatory selection processes going on in the female reproductive system. This less appreciated mechanism is the basis for differential fertilization which includes sperm selection as

opposed to pollen selection in plants. This should not be confused with **sperm competition / pollen tube competition** (link to a **book** by Tim Birkhead on [*Sperm Selection*](#)).

C value: The amount of DNA comprising the haploid genome for a given species (picograms per cell; 2-3 pg in mammals). The C value paradox is the lack of correlation between the C values of species and their evolutionary complexity. For example, some amphibians have 30 times as much DNA as we have but not more complex than humans.

Cyanobacteria: Unicellular, photosynthetic (photo-autotroph) prokaryote (in the Kingdom Monera). Formerly known as blue-green algae. It contains chlorophyll *a* but not chloroplast. They reproduce by fission and never sexually.

Cytogenetics: The study of the structure, function, and abnormalities of human chromosomes ([basics](#), [cytogenetics information](#), [cytogenetics glossary](#), [cytogenetics gallery](#)).

De novo: Literally, 'from new' as opposed to inherited.

Degeneracy: A feature of the genetic code that more than one nucleotide triplet codes for the same amino acid. The same applies to the **termination signal**, which is encoded by three different **stop codons**.

Denaturation: Reversible disruption of hydrogen bonds between nucleotides converting a double-stranded DNA molecule to single-stranded molecules. Heating or strong alkali treatment result in denaturation of DNA.

Devonian Period: A geologic period in the Phanerozoic Eon (410 to 360 Mya).

Dicentric chromosome: One chromosome having two centromeres.

Dikaryotic: A cell that contains two separate haploid nuclei ($n+n$) which is different from being **haploid** (n) or **diploid** ($2n$). Naturally seen in fungal **heterokaryons**. Dikaryosis is a significant genetic peculiarity of the fungi.

Diploblast: A lower invertebrate such as jelly fish that are composed of two tissue layers (ectoderm and endoderm) and lacking the third layer (mesoderm) present in higher invertebrates and vertebrates.

Diploid number ($2n$): The full complement of chromosomes in a somatic cell (or a sex cell before meiosis). In humans, the diploid number is 46.

Diploten (diplonema): The stage of meiosis I in which recombination between homologous chromosomes occurs. In females, oocytes are frozen at this stage at birth. Only one proceeds to the completion of meiosis every month during reproductive years.

Disjunction: Separation of homologous chromosomes during anaphase of mitotic or meiotic divisions (see also **nondisjunction**).

Disposable soma theory: A theory on the evolution of ageing and death suggesting that organisms derive little benefit from investing resources in increasing their lifespan beyond a certain point. It originated from the economic phenomenon that manufacturers invest minimum in durability.

Disruptive selection: Selection against the middle range of variation causing an increase in the frequency of a trait showing the extreme ranges of its variation. Disruptive selection might cause one species to evolve into two.

Disulfide bond (-S-S-): A covalent linkage between two cysteine residues in different parts of a protein or between two different proteins. Insulin (a small protein having two polypeptide chains) and immunoglobulin molecules, for example, have interchain and intrachain disulfide bonds. Endothelin and HLA molecules also have disulfide bonds. The C282Y mutation removes one of the disulfide bonds in the HLA class I-like HFE protein and abolishes its surface expression.

Divergent evolution: A kind of evolutionary change that results in increasing morphological difference between initially more similar lineages.

DNA (deoxyribonucleic acid): The large double-stranded molecule carrying the genetic code. It consists of four bases (adenine, guanine, cytosine and thymine), phosphate and ribose.

DNA binding motif: Common sites on different proteins which facilitate their binding to DNA. Examples are leucine zipper and zinc finger proteins. Any such protein is called DNA-binding protein.

DNA Polymerase: A group of enzymes mainly involved in copying a single-stranded DNA molecule to make its complementary strand. Eukaryotic DNA polymerases participate in chromosomal replication, repair, crossing-over and mitochondrial replication. To initiate replication, DNA polymerases require a priming RNA molecule. They extend the DNA using deoxyribonucleotide triphosphates (dNTP) as substrates and releasing pyrophosphates. The dNTPs are added to the 3' OH end of the growing strand (thus, DNA replication proceeds from 5' to 3' end).

DNA repair: Restoration of the correct nucleotide sequence of a DNA molecule that has acquired a mutation or modification. It includes proofreading by **DNA polymerase** (see **helicase**).

dn/ds ratio: In molecular phylogenetic studies, the ratio of the number of non-synonymous nucleotide substitutions to the number of synonymous nucleotide substitutions. In the case of functionally important (or otherwise constrained) genes, ds is expected to exceed dn ($dn/ds < 1$). Because most amino acid changes will disrupt protein structure and those non-synonymous substitutions (dn) causing them will not be maintained. In a non-functional pseudogene, there will be no discrimination between them and equal numbers of dn and ds are expected ($dn/ds = 1$). When natural selection is acting to favour changes at the amino acid level, it is predicted that dn will exceed ds, hence a high dn/ds ratio. In classical MHC loci, in the peptide binding regions (allele-specific sequences) because of heterozygote advantage/frequency-dependent selection, there is always a high dn/ds ratio (> 1) whereas in the remainder of the gene $dn/ds < 1$ (due to functional constraints). This suggests balancing selection is acting on peptide binding regions.

Domain: Region of a protein with a distinct tertiary structure and characteristic activity (for example, the membrane distal and membrane proximal domains of an MHC molecule).

Dominance: The property possessed by some alleles of determining the phenotype for any particular gene by masking the effects of the other allele (when heterozygous). Thus, **homozygosity** or **heterozygosity** for the dominant allele result in the same genotype in complete dominance (if red is dominant over white, the petals of a flower heterozygous for red and white would be red). Incomplete dominance appears as a blend of the phenotypes corresponding to the two alleles (like pink petals as opposed to red or white). In co-dominance, both alleles equally contribute to the phenotype (red and white petals occur together).

Dominant allele: An allele that masks an alternative allele when both are present (in heterozygous form) in an organism (see **recessive**).

Dominant-negative mutation: A (heterozygous) dominant mutation on one allele blocking the activity of wild-type protein still encoded by the normal allele causing a loss-of-function phenotype. The phenotype is indistinguishable from that of homozygous dominant mutation. [p53](#) mutations may act as dominant-negative (see also **haploinsufficiency**).

Dosage compensation: The phenomenon in women, who have two copies of genes on the X chromosome, of having the same level of the products of those genes as males (who have a single X chromosome). This is due to the process of random inactivation of one of the X chromosomes in females (Lyonisation). See **Lyon's hypothesis**.

Double fertilization: The union of one male gametophyte nucleus with the ovum nucleus and another male gametophyte nucleus with the polar nuclei to form the endosperm in seed plants. See also **pollen grain**.

Double heterozygote: An individual who is heterozygous at two loci under investigation.

Double stranded RNA (dsRNA): In eukaryotes, it is an accidental byproduct of transcriptional process. It may occur as the genome of certain viruses (such as reovirus) or may be produced

during viral replication as a general marker for viral infection. It is believed that dsRNA is the toxic substance responsible for general symptoms of viral infection as it induces cytokine production. dsRNA is the major activator of the PKR enzyme which is the major agent of anti-viral innate immunity.

Downstream: The direction which RNA polymerase moves during transcription (5' to 3') and ribosomes moves during translation. By convention, the +1 position of a gene is the first transcribed nucleotide, nucleotides downstream from the +1 position are designated +2, +3, etc.

Drift evolution: A high rate of immunologically significant mutations in certain viruses. This results in drifting away from recognition by the immune system by antigenic change. Influenza virus, HIV and HCV constantly change their antigenic structure through drift evolution.

Drosophila melanogaster: Common fruit fly. It contributed heavily to the study of genetics because of its ease and speed to breed. It contains only four pairs of chromosomes. Link to [Drosophila](#) website.

Ecogenetics: The branch of genetics that studies how (inherited or acquired) genetic factors influence human susceptibility to environmental health risks. It studies the genetic basis of environmental toxicity to develop methods for the detection, prevention and control of environment-related disease. Ecogenetics interacts with ecology, molecular genetics, toxicology, public health medicine and environmental epidemiology.

Ecological genetics: The analysis of genetics of natural populations and of the adaptations of them to the environment.

Ecology: The study of the interrelationships among living organisms and their environment. Human ecology means the study of human groups as influenced by environmental factors, including social and behavioural ones.

Ediacaran fauna: E Vendian age assemblage of soft-bodied multicellular animals. The oldest fauna known.

Effective population size (N or N_e): number of individuals contributing alleles to the next generation (N_f = number of mothers in a population; relevant in the calculation of number of generations for the fixation of a mitochondrial allele).

Embryo: A developing offspring during the period when most of its internal organs are forming. It is called fetus in the next stage of development.

Endonuclease: A nuclease which cuts a nucleic acid molecule by cleaving the phosphodiester bonds between two internal residues. Best known examples are restriction endonucleases.

Enhancer: A cis-acting (on either side of a gene) enhancer of **promoter** function. They are located 10 to 50 kb downstream or upstream of a gene. They may be tissue-specific. The enhancer effect is mediated through sequence-specific DNA-binding proteins (see also **silencer**).

Eocene epoch: A temporal subdivision (epoch) of the Tertiary period (58 to 37 Mya).

Epidemiology: The study of the distribution and determinants of health-related events (including but not only disease epidemics) aiming to trace down their cause and subsequently to control health problems.

Epigenesis: The theory that the development of an embryo consists of the gradual production and organization of parts.

Epigenetics: The study of heritable changes in gene expression that occur without a change in DNA sequence. Epigenetic phenomena such as **imprinting** and **paramutation** violate Mendelian principles of heredity. Epigenetic studies link genotype to phenotype working out the chain of processes (mainly in [developmental biology](#)) (see [Epigenetics: Special Issue of Science, 2001](#)).

Epistasis: The nonreciprocal interaction between nonallelic genes. This may result in masking of one and in this case, the masked gene is said to be hypostatic. An epistatic-hypostatic

relationship between two loci is similar to a dominant-recessive relationship between alleles at a particular locus.

Escherichia coli: A gram-negative bacterium whose genome has been sequenced in its entirety. It is model organisms for the study of the prokaryotes. Link to the [E.coli](#) genome project website.

Ethology: Study of animal behaviour under normal conditions. Link to [animal behaviour resources](#).

Eugenics: The idea of improving the quality of human species by selective breeding. Encouraging breeding of those with supposedly good genes is positive eugenics, whereas discouraging those with genes for undesirable traits is negative eugenics.

Eukaryotic cell: The DNA lies within a true nucleus (eu-karyon). May be unicellular (protist, some fungi) or multicellular (most fungi, plants, animals). Among eukaryotes, most fungi are haploid.

Eutheria: Placental mammals. A subclass of the Class Mammalia (others are monotremes and marsupials). Embryo and fetus are nourished by a placenta. They are viviparous (producing young alive rather than laying eggs), have a long period of gestation and the young is born immature.

Excinuclease: The excision nuclease involved in nucleotide exchange repair of DNA.

Exon: The coding sequence of a eukaryotic gene (see also **open reading frame**).

Exon - intron boundary: Introns end with the dinucleotide ApG [3' splice site / acceptor] and start with the dinucleotide GpT [5' splice site / donor].

Exonuclease: A nuclease which degrades a double-stranded DNA molecule by removing nucleotides from its two ends.

Expressivity: The range of phenotypes resulting from a given genotype (cystic fibrosis, for example, may have a variable degree of severity). This is different from **pleiotropy** which refers to a variety of different phenotypes resulting from the same genotype.

Extended phenotype: All effects of a gene upon the world where the effects influence the survival chance of a gene [Richard Dawkins].

Extra-chromosomal inheritance: Non-Mendelian inheritance due to extra-nuclear DNA (mitochondrial DNA in animals). The transmission of the trait only occurs from mothers.

Evolution: The process that results in heritable changes in a population spread over many generations (change in allele frequencies over time). Biological evolution refers to populations and not to individuals and that the changes must be passed on to the next generations. Genes mutate, individuals are selected, and populations evolve. Link to [evolution-related links](#).

Evolutionarily stable strategy (ESS): A strategy such that, if most members of a population adopt it, would give a reproductive fitness higher than any mutant strategy.

Ewens-Watterson Neutrality Test: Also called E-W homozygosity statistics. Described by Ewens (1972) and Watterson (1978). A widely used test in population genetics to estimate the selection acting on a locus. It compares the sum of observed homozygosity for each allele of a given locus (F_o) with the expected F_e value based on the number of alleles in the locus of interest, neutrality expectations and random mating assumption. A test of comparison yields an F value. Values close to zero mean that the locus is evolving under neutrality (genetic drift only) and there is no selection. Values of F significantly different from zero suggest selection. When $F_o > F_e$, the locus is undergoing purifying selection, and when $F_e > F_o$, the locus is under balancing selection (very common for HLA loci). Alternative tests for neutrality are Tajima's D (Tajima F: Genetics 1989;123:585) and Slatkin's Exact Test for Neutrality (Slatkin M: Genet Res 1996;68:259). See also Slatkin & Muirhead: Genetics 2000;156:2119).

F₁: First filial (son or daughter) hybrids arising from a first cross. Subsequent generations are denoted by F₂, F₃ etc.

F Factor (Fertility Factor): Transmissible plasmid (episome) in bacteria (such as E. coli) that acts as a sex factor. It is a circular DNA about 94 kb long. Conjugation and chromosomal gene transfer occur from F⁺ (male) to F⁻ (female) bacterium.

F' (F-prime) factor: Normally, the **F factor** contains genes related to conjugation/mating. The F' factor contains an additional portion of the bacterial genome.

F+ strain: E.coli strain behaving as donors during conjugation (male). It has the **F factor**.

F- strain: E.coli strain behaving as recipients during conjugation (female). It lacks the **F factor**.

Fauna: A certain species of animals occurring in a particular region or period.

Fertilization: Fusion of female and male haploid gametes to form a diploid zygote from which a new individual develops.

Fingerprinting: The use of RFLPs or repeat sequence DNA to establish a unique individual-specific pattern of DNA fragments.

F.I.S.H. (fluorescence in situ hybridisation): One of the more modern methods in cytogenetics, which uses fluorescence-labelled chromosome-specific DNA, probes to detect translocations, inversions, deletions, amplifications and other structural or numerical chromosomal abnormalities. FISH permits analysis of proliferating (**metaphase cells**) and non-proliferating (**interphase nuclei**) cells and is useful in determining the percentage of neoplastic cells before and after therapy (minimal residual disease) (see examples at [Cytocell](#) website; a review by [Mathew & Raimondi, 2003](#); a review of the use of FISH in childhood leukemia, see [Harrison CJ, 2001](#)).

Fisher's Fundamental Theorem: The rate of increase in fitness is equal to the additive genetic variance in fitness. This means that if there is a lot of variation in the population the value of S will be large.

Fisher's Theorem of the Sex Ratio: In a population where individuals mate at random, the rarity of either sex will automatically set up selection pressure favouring production of the rarer sex. Once the rare sex is favoured, the sex ratio gradually moves back toward equality.

Fitness: Lifetime reproductive success of an individual (i.e., the total number of offspring who themselves survive to reproduce). It can be seen as the extent to which an individual successfully passes on its genes to the next generation. It has two components: survival (viability) and reproductive success (fecundity). Variation in fitness is the major driving force in biological evolution (see also **genetic fitness**).

Fixed: The establishment of a single allelic variant at a locus as a result random genetic drift.

Five-prime (5') end: The end of a DNA or RNA strand with a free 5' phosphate group corresponding to the transcription initiation (see also **three-prime end**).

FlavrSavr (a short for flavor saver): The first genetically modified food (GM food) to go on the market. This kind of tomato was modified not to ripen so fast.

Flower: The part of an angiosperm containing the organs of reproduction (male stamen and female stigma as well as the ovary).

Footprinting, DNAase: DNA with protein bound is resistant to digestion by DNAase. When a sequencing reaction is performed using such DNA, a protected area representing the footprint of the bound protein will be detected. This permits identification of the protein binding regions of the DNA.

Founder effect (Sewall Wright effect): A type of genetic drift in which allele frequencies are altered in a small population, which is a nonrandom sample of a larger (main) population.

Frameshift mutations: Mutations, usually deletions or insertions, that change the reading frame of the codon triplets.

Fruit: Mature ovary with seeds inside. Its function is seed protection and dispersal. Fruits are a development of the ovary wall and sometimes the other flower parts as well. Its formation is induced by the plant hormone auxin, which is released by the maturing seeds.

Fugu: The puffer fish, *Fugu rubripes*, has essentially the same number of genes as the human genome, but its genome is eight times more compact than human genome (about 400 Mb as opposed to 3 Gb). A project to sequence the whole *Fugu* genome is underway. Link to [Fugu Website](#) which has the draft sequence (Oct 25, 2001).

Fu/HC: The fusion/histocompatibility system of the [Ascidians](#). It is involved in self - nonself recognition regulating the fusion between compatible organisms and prevention of self-fertilization. Link to an [abstract](#) on Fu/HC research (see also [Botryllus](#) and [protochordates](#)).

Fungus: A Kingdom made up of a diverse group of unicellular or multicellular, eukaryotic organisms which are not plants or animals. Many are parasitic or saprophytic. Both asexual and sexual reproductions are possible. The Kingdom includes five phyla: Zygomycetes (conjugating fungi, black bread molds), Deuteromycetes (reproduce only asexually, *Aspergillus* 'brown mold' and *Penicillium*), Basidiomycetes (incl. mushrooms), Ascomycetes (incl. *Neurospora* 'bread mold' and *Saccharomyces* 'baker's yeast') and Mycophycophyta (incl. lichens). Some of them (Basidiomycetes) have one of the most ancient pheromone-based mating-type recognition systems. See [Fungi](#) in Kimball's Biology Pages and [Fungus](#) in Tree of Life. See also [dikaryosis](#), [heterokaryon](#) and [mating types](#).

Gaia hypothesis: A hypothesis developed by Lovelock, Watson, Margulis and others. In its form, -based on a mathematical model- this theory proposes that the physical and chemical condition of the surface of the Earth, of the atmosphere, and of the oceans has been and is actively made fit and comfortable by the presence of life itself. This is in contrast to the conventional wisdom which held that life adapted to the planetary conditions and they evolved their separate ways. Lovelock and Watson developed the [Daisyworld model](#) - an imaginary planet, which maintains conditions for its survival simply by following its own natural processes. This simple model has since become an integral part of the debate about the Gaia Hypothesis. The Daisyworld planet contains only two species of life: light daisies and dark daisies and an example of a self-regulating system. See also <http://www.magna.com.au/~prfbrown/gaia.html> and Lovelock's book: [Gaia: A New Look at Life on Earth](#). Oxford University Press.

Galton's Regression Law: Individuals differing from the average character of the population produce offspring, which, on the average, differ to a lesser degree but in the same direction from the average as their parents.

Gamete: A haploid reproductive cell such as sperm (or [pollen](#)) and egg (oocyte).

Gametophyte: The haploid, gamete-forming (sexual) generation in plants with [alternation of generations](#). Typically it is produced from a haploid spore. See also [sporophyte](#).

Gametic association: see [linkage disequilibrium](#).

Gender: Differences between any two complementary organisms of the same species that render them capable of mating (see [mating types](#)).

Gene: Physical and functional unit of heredity that carries information from one generation to the next. The entire DNA sequence necessary for the synthesis of a functional polypeptide or RNA molecule. In addition to the coding regions ([exons](#)), a gene may have non-coding intervening sequences ([introns](#)) and transcription-control regions.

Gene conversion: Partial sequence transfer from one allele to another (interallelic recombination) converting one gene or allele to another one. It is the most common mechanism, especially for the HLA-B locus, in the generation of new [MHC](#) alleles. Less common are conversions between alleles of different [MHC](#) loci (intergenic conversion).

Gene flow: The movement of genes within a population or between two populations following genetic admixture. Gene flow creates new combinations of genes or alleles in individuals that can be tested against the environment. This way it is one of the sources of variation in the process of natural selection.

Genetic anticipation: The progressive shift of the age of onset of a hereditary disease to earlier ages in successive generations. It may occur because a parent is a mosaic, and the child has the full mutation in all cells. Triplet repeat expansion may demonstrate anticipation when the number of repeats increases with each generation.

Genetic determinism: The (incorrect) belief that genes alone form all characteristics of an individual organism.

Genetic distance: A measurement of genetic relatedness of populations. The estimate is based on the number of allelic substitutions per locus that have occurred during the separate evolution of two populations. Link to a lecture on [Estimating Genetic Distance](#) and [GeneDist: Online Calculator of Genetic Distance](#). The software [Arlequin](#), [PHYLIP](#), [GDA](#), [PopGene](#), [Populations](#) and [SGS](#) are suitable to calculate population-to-population genetic distance from allele frequencies. [GenAIEx](#) can be used to calculate genetic distance on Excel.

Genetic Distance Estimation by PHYLIP: The most popular (and free) phylogenetics program [PHYLIP](#) can be used to estimate genetic distance between populations. Most components of PHYLIP can be run [online](#). One component of the package [GENDST](#) estimates genetic distance from allele frequencies using one of the three methods: Nei's, Cavalli-Sforza's or Reynold's (see papers by [Nei et al, 1983](#), [Nei M, 1996](#) and a [lecture note](#) for more information on these methods). [GENDST](#) can be run [online](#) using the default options ([Nei's genetic distance](#)) to obtain genetic distance matrix data. The PHYLIP program [CONTML](#) estimates phylogenies from gene frequency data by maximum likelihood under a model in which all divergence is due to genetic drift in the absence of new mutations (Cavalli-Sforza's method) and draws a tree. The program comes as a freeware as part of PHYLIP or this program can be run [online](#) with default options. If new mutations are contributing to allele frequency changes, Nei's method should be selected on [GENDST](#) to estimate genetic distances first. Then a tree can be obtained using one of the following components of PHYLIP: [NEIGHBOR](#) also draws a phylogenetic tree using the genetic distance matrix data (from [GENDST](#)). It uses either Nei's "[Neighbor Joining Method](#)," or the [UPGMA](#) (unweighted pair group method with arithmetic mean; average linkage clustering) method. Neighbor Joining is a distance matrix method producing an unrooted tree without the assumption of a clock (UPGMA does assume a clock). [NEIGHBOR](#) can be run [online](#). Other components of PHYLIP that draw phylogenetic trees from genetic distance matrix data are [FITCH](#) / [online](#) (does not assume evolutionary clock) and [KITSCH](#) / [online](#) (assumes evolutionary clock).

Genetic drift: Evolutionary change over generations due to random events in small populations (not to be mixed with sampling error due to a small sample size). It operates unless overcome by strong selective forces. Wildly different HLA allele frequencies among South AmerIndian tribes are believed to be result of probable genetic drift in each small tribe. Link to a lecture on [genetic drift](#) and a [simulation](#).

Genetic fitness: Classic genetic fitness is the average direct reproductive success of an individual possessing a specific genotype in comparison to others in the population. Inclusive fitness is described as the classic fitness plus the probability that an individual's genotype may be passed on through relatives.

Genetic heterogeneity: Presence of several different genotypes contributing to the genetic component of a disease on their own.

Genetic linkage: The situation referring to segregation of two or more genes together as a unit. Genetic linkage is thought to arise to accommodate genes that function best in each other's company, i.e., to provide a necessary cooperative effect that enhances survival. Genetic linkage reflects a lack of meiotic crossovers between two genes (see a [tutorial](#)).

Genetic load: The average number of lethal equivalents (or any recessive mutant lowering fitness) per individual in a population which are propagated by heterozygotes in a masked state.

Genetic relatedness (r): A quantitative measure of genetic relatedness between individuals. In diploid species, $r=1/2$ between full siblings, or parent and child; $r=1/4$ for half siblings or aunt/uncle versus niece/nephew, or for grandparents versus grandchildren; $r=1/8$ for first cousins; $r=0$ for non-relatives (see also **coefficient of relatedness**).

Genetic variance: The phenotypic variance in a population that is due to genetic heterogeneity.

Genetics: Study of variation and heredity and their physical basis in DNA.

Genocopy: A gene/genotype causing the same phenotype as another gene/genotype. Genocopies are the basis of genetic heterogeneity and important in genetic diagnosis and counseling.

Genome: Total genetic material in a set of haploid chromosomes as in a germ cell. The human genome contains 3,000 Mbp whereas the E.coli genome has 4.6 Mbp (see also **C value**). Link to the [Genome Catalogue](#).

Genomic imprinting: Differing expression of genetic material dependent on the parent-of-origin. This is due to **methylation** of one of the alleles depending of its origin. A very illustrative example is the inherited neck tumor paraganglioma for which the susceptibility gene is only active if inherited from the father. Genomic imprinting must be considered in disorders that appear to have skipped a generation. Link to a [genomic imprinting](#) website.

Genomic instability: One of the first phenomena in the formation of malignancies. It is due to defects in DNA repair and cell cycle controls. This can happen by gain-of-function mutations in proto-oncogenes or loss-of-function mutations in tumor suppressor genes.

Genotype: The diploid genetic formula at one or more loci.

Genotype-environment (GxE) interaction effect: This term refers both to the modification of genetic risk factors by environmental risk and protective factors and to the role of specific genetic risk factors in determining individual differences in vulnerability to environmental risk factors. For a review, see [Heath & Nelson, 2002](#)).

Geological timescale: The period between the origin of earth (4,500 Mya) and the beginning of the Cambrian period (540 Mya) is called the Precambrian Eon. The last 540 million years (Phanerozoic Eon) are divided into three eras: Palaeozoic (540-245 Mya); Mesozoic (245-65 Mya); Cainozoic. The geological periods (included in an era, longer than an epoch) are as follows: Vendian (immediately before the Cambrian; 610-540 Mya); Cambrian (540-510 Mya); Ordovician; Silurian; Devonian; Carboniferous; Permian; Triassic / Jurassic / Cretaceous (altogether the Mesozoic Era); Tertiary (65-1.64 Mya) and Quaternary. An epoch is a subdivision of a period. See the [geological table](#) in BBC Education.

Germ line: Genetic material transmitted from one generation to the next through the gametes. A germ line mutation exists in all cells of the offspring formed from that gamete.

Germinal mosaicism: A mixture of gonadal cells with different numbers of chromosome numbers or other chromosomal abnormalities. It can lead to aneuploid offspring from phenotypically normal parents with an unpredictable recurrence risk.

Gnathostomata: Jawed vertebrates, evolved following the jawless vertebrates (Class **Agnatha**). The oldest extant branch of jawed vertebrates is the **cartilaginous fish**.

Gonadal (germline) mosaicism: If a mutation selectively affects the cells destined to become gonadal cells during early embryogenesis, the affected individual will be phenotypically normal, the somatic cells will be free of the mutation but all or some gonadal/germ cells will have the mutation. The end result is transmission of a genetic disorder by a healthy person causing sporadic form a genetic disease in the offspring and a higher than the general population risk in the following siblings.

Great Ape: Chimpanzees (including bonobos), gorillas, and orangutans.

GTP (guanosine 5'-triphosphate): A nucleotide that is a precursor in RNA synthesis, which plays a role in protein synthesis (as well as in signal transduction and microtubule assembly). See also **cap**.

Gymnosperm (Gr. gymnos=naked; sperm=seed): Woody plants whose life histories include alternation of generations and ovules are not enclosed in a carpel. The pollen typically germinates on the surface of the ovule. A superclass in the sperm plants (Spermatophyta) division. Examples are cycad, conifer and ginkgo. Cycads are the most primitive ones evolved in the Devonian period about 400 Mya. There are about 700 extant species in Gymnosperms. Link to a lecture on **Gymnosperms**.

Gyrase: One of the bacterial DNA topoisomerases that functions during DNA replication to reduce molecular tension caused by supercoiling (super-twisting). DNA gyrase produces, then seals, double-stranded breaks.

H-2 complex: The major histocompatibility complex (**MHC**) of the mouse. It is the first MHC discovered in 1937 by Peter Gorer.

Hair pin loop: Binding of complementary sequences to each other to form a hair pin loop (also called stem loop). If happens in a PCR primer, it will not function.

Haploid number (n): The number of chromosomes in the gamete after meiosis. In humans, the haploid number is 23.

Hamilton's Altruism Theory: If selection favoured the evolution of altruistic acts between parents and offspring, then similar behaviour might occur between other close relatives possessing the same altruistic genes, which were identical by descent. In other words, individual may behave altruistically not only to their own immediate offspring but to others such as siblings, grandchildren and cousins (as happens in the bee society).

Hamilton's Rule (theory of kin selection): In an altruistic act, if the donor sustains cost C, and the receiver gains a benefit B as a result of the altruism, then an allele that promotes an altruistic act in the donor will spread in the population if $B/C > 1/r$ or $rB - C > 0$ (where r is the **coefficient of relatedness**).

Haploinsufficiency: Situation where one normal copy of a gene alone is not sufficient to maintain normal function. It is observed as a dominant mutation on one allele (or deletion of it) resulting in total loss-of-function in a diploid cell because of the insufficient amount of the wild-type protein encoded by the normal allele on the other haplotype (see also **dominant negative**). A recent example of haploinsufficiency by **Kurotaki et al, 2002** in **Sotos syndrome**.

Haplotype: The particular combination of alleles in a linked group encoded by genes in close vicinity on the same chromosome.

Hardy-Weinberg equilibrium (HWE): In an infinitely large population, gene and genotype frequencies remain stable as long as there is no selection, mutation, or migration. For a bi-allelic locus where the gene frequencies are p and q: $p^2 + 2pq + q^2 = 1$ (**Online HWE Analysis**; lectures on HWE: **1** & **2**).

Heat Shock Response: Heat shock response is ubiquitous and highly conserved defence mechanism for protection of cells from harmful conditions such as heat shock, UV irradiation, toxic chemicals, infection, transformation and appearance of mutant and misfolded proteins. Heat Shock Proteins (HSPs) also function as accessory molecules in antigen presentation. HSP70 genes are within the MHC in most vertebrates. High levels of HSP70 prevent stress-induced apoptosis, and may have a transforming potential.

Helicase: An enzyme that unwinds the double DNA helix near the replication fork before **DNA polymerase** acts on it. Replication fork moves from 3' to 5' of the leading strand. Unwinding is also necessary for **DNA repair**. Mutations in the helicase genes on chromosome 2q and 19q are

one group of causes of the DNA repair defect xeroderma pigmentosum (an autosomal recessive disease). See also **primosome**.

Hemizygous: As in any X-linked trait in males, absence of a homologous counterpart for an allele. It may also result from deletion.

Heritability: The proportion of the total phenotypic variance that is attributable to genetic causes ($h^2 = \text{genetic variance} / \text{total phenotypic variance}$).

Hermaphroditism: Having both male and female sexual organs in one individual. Most invertebrates and plants are hermaphrodites. Union of the gametes of the same individual (self-fertilization) is the most extreme example of inbreeding.

Heterogametic sex: The sex, which has the two different sex chromosomes (XY). Human and Drosophila males are the heterogametic sex, whereas, in birds, moths, some fish and amphibians, females are the heterogametic sex (ZW).

Heterokaryon: A cell containing more than one genetically different nucleus. Naturally occurs in fungi as long as their fungal (heterokaryon) incompatibility types are identical (see also **dikaryotic**).

Heterogeneous nuclear RNA (hnRNA): RNA products immediately synthesized from the DNA template in the nucleus (sometimes called DNA-like RNA or dRNA). This RNA species has a short half-life, is very heterogeneous and very large (molecular weight in excess of 10^7). hnRNA molecules are processed to generate the mRNA molecules (molecular weight generally less than 2×10^6) before leaving the nucleus.

Heterothallic: Organisms (fungi, algae, plants) that can only undergo sexual reproduction with another bearing a different mating/compatibility type (**self-incompatible**). See also **homothallic**.

Heterozygosity: Presence of two different alleles at a locus in a diploid organism (see **homozygosity**). It is the result of inheritance of different alleles from parents.

Hfr: A male bacterial cell that has the F factor integrated into its chromosome is an Hfr (high frequency of recombination) cell. Crosses between Hfr cells and F- females produce far more recombinant progeny than do crosses between F+ males and F- females.

Histones: Basic proteins that are involved in the packing of DNA. They bind to the phosphate groups of DNA. There are five major types of histone proteins.

HLA complex: The human major histocompatibility complex (**MHC**). An HLA haplotype has been totally sequenced in 1999.

Holandric gene: A gene carried on the Y chromosome and therefore transmitted from father to son.

Homeobox: Conserved protein sequence, which forms a DNA-binding domain in a class of **transcription factors**.

Hominid: A member of the Hominidae family.

Homologous chromosomes: Chromosomes that occur in pairs one having come from the male parent and the other from the female parent. They pair and participate in crossing-over during meiosis. Homologous chromosomes contain the same array of genes but may contain different alleles at those loci.

Homology: A similarity due to inheritance from a common ancestor (see also **analogy**). An example is mammals' back legs. Homology may be due to **orthology** (between species) or **paralogy** (within a species).

Homothallic: Organisms (fungi, algae, plants) that can undergo sexual reproduction with a similar strain including the self (self-compatible) (see also **heterothallic**).

Homozygosity: Presence of two identical alleles at a locus in a diploid organism (see **heterozygosity**). It is the result of inheritance of identical alleles from both parents.

House-keeping genes: Genes which are constitutively expressed in most cells because they provide basic functions.

Htf island: *Hpa* *Tiny Fragment* island which are unmethylated CpG-rich regions in the genome. Eighty percent of these occur at or near genes, particularly housekeeping genes. Many of the **MHC** genes discovered were not near Htf islands.

Hybrid: The offspring of two distinct species.

Hybridization: The specific reassociation of complementary strands of nucleic acids.

Hybrid vigor (heterosis): Unusual growth, strength, and health of heterozygous offspring from two less vigorous homozygous parents.

Hypothesis: An unproven but testable scientific proposition. A theory is a statement with some confirmation.

Hypotrich: A protozoan of the Ciliate order which reproduces sexually or by asexual binary division. Sexual reproduction can be via conjugation (nuclear exchange) or gamete fusion. In the former, multiple mating types are involved. In gametic fusion there are only two types one of which is the only source of intracellular organelles (see also **mating type**).

Ichthyosaur: An extinct group of marine fish-like or porpoise-like reptiles abundant in Mesozoic seas.

Idiomorph: This term is used to describe the fungal mating types which are extremely dissimilar from each other and do not show homology between strains of the opposite sex (as opposed to the allelic relationship in most polymorphic systems). Also used as ideomorph.

Imprinting: See genomic imprinting.

Inbreeding: Production of offspring by (blood) related parents. Its most extreme form is self-fertilization in hermaphrodites (most invertebrates and plants).

Inbreeding depression: Reduction in offspring fitness resulting from mating between blood relatives.

Incest: Sexual relationships between parents and children, or between brothers and sisters.

Incomplete dominance: One allele is not expressed, but the other allele expresses itself normally so that the phenotype gets half the dose of the effect.

Initiation complex: A multi-protein complex that forms at the site of transcription initiation and is composed of RNA polymerase II, ubiquitous or general **transcription or initiation factors** (TFII or IF/eIF) and gene-specific **enhancers/silencers**.

Innate immunity: Pre-existing and non-specific defence immunity with a very low memory component if any. As the primitive immune response against bacteria, it is present in invertebrates and vertebrates.

Integrase: An enzyme that catalyzes a site-specific recombination (integration or excision) involving a **prophage** and a bacterial chromosome.

Intron: A non-coding section of DNA within a gene that is not translated to a peptide. Intervening sequences between **exons**. Introns are featured in the primary transcript (pre-mRNA) but removed by splicing during nuclear RNA processing/editing.

Invertebrate: All animals other than those in the phylum **Chordata**; lower metazoans. They do not possess a notochord or vertebral column. Examples are worms, corals, sponges, etc. The protochordates are sometimes called higher invertebrates.

In vitro: Literally, 'in glass' meaning in the laboratory.

In vivo: Literally, 'In the living organism'.

Iterative evolution: Repeated origination of lineages with generally similar morphology at different times in the history of a clade.

Jacobson's organ: In some vertebrates, an accessory olfactory organ developed in connection with the roof of the mouth.

Jawless vertebrates (Class **Agnatha**): The most primitive Vertebrates evolved about 500 Mya. The extant species hagfish and lampreys in the order **Cyclostomata** belong to this group.

Karyotype: A photomicrograph of metaphase chromosomes arranged in standard order. Normal human karyotype consists of 46 chromosomes, of which 44 are somatic (autosomal) and 2 are sex chromosomes.

Kingdom: The major taxonomic group in the current classification of living organisms with the exception of informal division of prokaryotic and eukaryotic empires. The five Kingdoms are Monera, Protocista, Fungi, Plants and **Animals**. In the late 1980s Cavalier-Smith proposed that within the Eukaryota there are six kingdoms: Archezoa, Protozoa, Chromista, Plants, Fungi, and **Animals** (see also **taxonomy**).

Kin recognition: Ability to distinguish the incest from unrelated members of a species.

Kin selection: Natural selection for behaviours that lowers an individual's own chance of survival but raises that of a relative.

Kozak sequence: In some viral mRNAs, the consensus sequence surrounding the initiating AUG 5' ACCAUGG 3'. It facilitates ribosomal binding and therefore, protein synthesis. The most consistent position is located three nucleotides before the initiation codon (ATG) and is almost always an adenine nucleotide (see also **Shine-Dalgarno sequence**).

lac operon: A structural unit in the E.coli genome that consists of three structural genes (encoding different enzymes involved in sugar metabolism) transcribed together and their common promoter and operator genes. Provides a good model for studying the interactions between promoters and repressors.

Last male sperm precedence: A situation that results in fertilization of the ovum by the sperm of the last male in multiply inseminated female. This is due to sperm incapacitation by the semen and sperm displacement. This well-documented form of **sperm competition** is best known in *Drosophila*.

Latency (viral): The state of viral infection in which the virus exists in host cells without reproducing itself. This is slightly different from viral persistence when basal replication continues.

Leader sequence: A sequence at the 5' (N-terminal) of the DNA and mRNA that leads the newly synthesized mRNA to the ribosome (it is not translated). It is also used to mean the **signal sequence**, which is translated but is subject to post-translation cleavage when the final destination is reached or following secretion.

Lek: A special site where communal courtship display takes place by swarms of animals. Lekking is best known in male birds to attract female mates. Rare form of lekking by females is seen in the African butterfly (***Acraea encedon***), the Dance fly (*Empis borealis*) and the European dotterel bird (***Eudromias morinellus***).

Ligase: An enzyme which is of vital importance in recombinant DNA technology. It joins nucleotides together by a phosphodiester bond between the 5'-P end of a polynucleotide chain and the 3'-OH end of another one.

LINES (long interspersed elements): One of the abundant intermediate (6 to 7 kb) repeat DNA sequences in mammals (see also **SINES**).

Linkage: The tendency of 'genes' on the same chromosome to segregate together. This means that linked genes are transmitted to the same gamete more than 50% of the time. Genetic linkage reflects a lack of meiotic crossovers between two genes.

Linkage disequilibrium (LD): The tendency for two 'alleles' to be present on the same chromosome (positive LD), or not to segregate together (negative LD). As a result, specific alleles at two different loci are found together more or less than expected by chance. The same situation may exist for more than two alleles. Its magnitude is expressed as the delta (Δ) value and corresponds to the difference between the expected and the observed haplotype frequency. It can have positive or negative values. LD is decreased by recombination. Thus, it decreases every generation of random mating unless some process opposing the approach to linkage equilibrium. Permanent LD may result from natural selection if some gametic combinations result in higher fitness than other combinations. Link to a lecture on [linkage disequilibrium; online linkage disequilibrium analysis](#).

Liposome: Small spheres of phospholipid bilayers (just like cell membranes). They are used for gene transfer into cells. The DNA to be transferred is attached to them. The DNA-liposome complex fuses with the cell membrane to enter the cell and releases the DNA into the cell.

Living fossil: An extant species which is morphologically very similar to a species from the ancient past. Despite apparent lack of change, they seem to have escaped extinction. Coelacanth (a 350 million-years-old lobe-finned fish), Horseshoe Crab (a 510 million-years-old marine arthropod), Amazon River Dolphin, Gingko (maidenhair tree, a gymnosperm), and Metasequoia (Metasequoia glyptostrobodes, a conifer) are examples.

Lobe-finned fish: A group of fish that have bone and muscle in their limbs as opposed to simple fins as in most teleosts. The group includes coelacanth and lungfish.

Locus: The position on a chromosome occupied by a particular gene (plural: loci). For information and official nomenclature of human genetic loci, see [LocusLink](#). [UniGene](#), [GenAtlas](#).

Lod score: Logarithm of the odds favouring linkage obtained from the statistical analysis of linkage. The lod score (Z) of +3 is considered evidence for linkage.

Loss-of-heterozygosity (LOH): Refers to the disappearance of polymorphic marker alleles when constitutional DNA and tumor DNA from cancer patients are compared. The consequence is usually genomic deletion discarding the normal copies of tumor suppressor genes. Such deletion (or functional deletion through methylation) may uncover existing mutations in the homologue copy.

Lyon hypothesis: The proposition by Mary F Lyon that random inactivation of one X chromosome in the somatic cells of mammalian females is responsible for **dosage compensation** and mosaicism. When randomisation is skewed, an X-linked recessive disease may be seen in a female. Also, discordance in female twins for an X-linked trait may be due to atypical Lyonisation. When a part of X-chromosome is translocated to an autosomal chromosome, it escapes inactivation.

Maize (Zea mays): A crop plant used as animal and human food. It features frequently in scientific literature because it is the only diploid plant among economically important food crops such as oats and wheat (which are **polyploids**). Its disadvantages are that it has a relatively long generation time (a maximum of three generations per year), and a very large and complex genome with a lot of repetitive sequences. Natural cross-pollination ability of maize is another problem. Its meiotic chromosomes are, however, excellent subjects for cytogenetic studies. At present, the model system in plant genetics is **Arabidopsis thaliana**. Link to [Maize](#) website.

Major Histocompatibility Complex (MHC): A genetic complex of vertebrates consisting of around 100 genes including the extremely polymorphic cell surface molecules called **HLA** in humans and **H-2** in mice. These molecules provide an immunological marker for selfness and a genetic self-identity to the individual. This information is used in mate choice, union of gametes, maintenance of pregnancy, and immune response against nonself (including a transplanted graft). These molecules are the most polymorphic ones in vertebrates. The polymorphism arises from point mutations not at an unusually higher rate than other genes, and mainly from interallelic gene conversion events. The polymorphism is maintained through pathogen and non-pathogen

driven mechanisms via heterozygous advantage (overdominant selection) and negative frequency dependent selection. A 3.6 Mb long human MHC haplotype and the 92 kb chicken MHC have been totally sequenced (see Nature 1999(Oct 28);401:921-925) ([dbMHC](#), [MHC-X](#), [MHC Map](#)).

Mammals: One of the eight Classes in the Phylum **Chordata** which contains approximately 4500 species in 15 Orders. In mammals, the fertilization of the egg is internal, the young develops within the body of the mother, and is fed by milk produced by the mammary glands. The mammals are warm-blooded and the body is covered with hair. In Mammals, female is the heterogametic sex (XY) and thus male-to-male competition is the predominant form of sexual selection.

Mapping: Determining the physical location of a gene or a genetic marker on a chromosome. Used to be achieved by linkage and association studies besides other methods but genome projects have mapped more or less all genes in respective genomes.

Marine Species Database

Maternal effect lethals: One form of selfish/parasitic DNA that facilitates its own propagation. They are post-zygotic distorters that kill progeny lacking the factor. *Medea* in beetles and *Scatin* mice are the known examples. Progeny of the heterozygous mothers that are homozygous for wild-type are killed. Progeny carrying a copy of the lethal are actually protected.

Maternal inheritance: Diseases due to mutations in mtDNA are transmitted only by mothers because all mitochondria are inherited via the egg. Thus, all offspring of an affected female are at risk of inheriting the abnormality, whereas no offspring of an affected male are at risk. Clinical manifestations are variable and may be due to variable mixtures of mutant and normal mitochondrial genomes (heteroplasmy) within cells and tissues (see [Clinical Genetics](#)).

Mating type: Genetically determined characteristics of bacteria, ciliates, fungi and algae, determining their ability to conjugate and undergo sexual reproduction with other members of the species. In yeasts (*S.cerevisiae*) which have only two types, only cells of opposite types can conjugate. The common mushroom *Schizophyllum* has more than 50,000 mating types (genders) encoded in two separate loci. In species where the organelles are inherited uniparentally and reproduction is by the union of gametes, there are only two mating types. In species reproducing via sexual conjugation (nuclear exchange) so that each cell preserves its own organelles, there can be multiple types.

Meiosis: Cell division with two phases resulting four haploid cells (gametes) from a diploid cell. In meiosis I, the already doubled chromosome number reduces to half to create two diploid cells each containing one set of replicated chromosomes. Genetic recombination between homologous chromosome pairs occurs during meiosis I. In meiosis II, each diploid cell creates two haploid cells resulting in four gametes from one diploid cell.

Melting temperature (T_m): The temperature at which the two strands of a double-stranded DNA molecule come apart. A short (<18 nucleotides) oligonucleotide T_m value ($^{\circ}\text{C}$) is estimated by the formula: $T_m = (\text{number of A} + \text{T}) \times 2 + (\text{number of G} + \text{C}) \times 4$.

Mendelian inheritance: Inheritance of traits mediated by nuclear genes (as opposed to mitochondrial DNA) according to the laws defined by Gregor Mendel.

Mendel's first law (law of segregation): The two alleles received one from each parent segregate independently in gamete formation, so that each gamete receives one or the other with equal probability.

Mendel's second law (law of recombination): Two characters determined by two unlinked genes are recombined at random in gametic formation, so that they segregate independently of each other, each according to the first law (note that recombination here is not used to mean crossing-over in meiosis).

Metacentric chromosome: A chromosome with its centromere near the center. If the centromere is slightly off-center, the chromosome is said to be submetacentric (see also **acrocentric** and **telocentric**).

Metaphase: Mitotic phase at which replicated chromosomes are fully condensed and become visible under the light microscope.

Metazoa: A major division in the Animal Kingdom consisting of multicellular animals.

Methylation: The addition of a methyl group (-CH₃) to DNA. Methylated DNA is inactivated and not transcribed. Most frequently occurs at CpG doublets (see **genomic imprinting** and **CpG islands**).

Methylation paradox: Methylation of the CpG islands in the transcribed region is often correlated with transcription but an inverse correlation is seen at the CpG islands at the transcription initiation site (link to a review on **Methylation Paradox** by PA Jones).

MHC: **Major Histocompatibility Complex**. **H-2 complex** in mice, **HLA complex** in humans.

MHC restriction: The phenomenon that a T cell can only recognize a peptide when it is (processed and) presented by another cell sharing the same MHC type. The only exception is that a foreign MHC antigen itself does not have to be presented by a cell but is able to induce a reaction directly (as happens in MHC - mismatched transplantation).

Microlymphocytotoxicity (MLC) assay: An assay used in the typing of **HLA** molecules (serological typing).

Microsatellite repeat sequences: Sequences of 2 to 5 bp repeated up to 50 times such as a TA dinucleotide repeat polymorphism. The variable number of repeats creates the polymorphism. They may occur at 50 - 100 thousand locations in the human genome. Microsatellites mutate faster than nonrepeat polymorphisms and can be used to estimate evolutionary relationships over shorter time scales (see the **abstract** of a study by Goldstein et al).

Mimicry: the resemblance of one kind of organism to another to make the organism difficult to find, to discourage the potential predators, or to attract potential prey. The common kinds of mimicry are Batesian and Mullerian mimicry (see **Evolutionary Biology Notes**). See also **molecular mimicry**.

Missense mutation: A mutation that causes the substitution of one amino acid for another (non-synonymous change). An example is the major HFE mutation C282Y in which results in an amino acid change at position 282.

Missing link: An absent member needed to complete an evolutionary lineage.

Mitochondrial DNA: The maternally inherited nucleic acid found in cytoplasm whose homologue in plants is chloroplast DNA. This small circular DNA codes for tRNAs, rRNAs and some mitochondrial DNAs. It is more closely related to bacterial DNA than to eukaryotic nuclear DNA. Mitochondrial DNA mutates 10-20 times faster than nuclear DNA.

Mitosis: Cell division into two identical daughter cells with the same chromosome number as the mother cell (see also **meiosis**). Replicated chromosomes separate and each chromatid goes to a daughter cell.

Mitotic recombination: During mitosis sister chromatids freely exchange material without changing anything in genetic material because they are identical. Very rarely, and by chance, homologous chromosomes come very close to each other and exchange material as in **meiosis** which results in a recombinant chromosome.

Mixed lymphocyte reaction (MLR): The activation of T cells *in vitro* by other (allogeneic) lymphocytes due to differences in their MHC molecules.

Molecular mimicry: Resemblance of a DNA sequence or a polypeptide by an unrelated sequence at the nucleotide or amino acid level, respectively. Mimicry of MHC proteins is an immunoevasion mechanism used by pathogens.

Monera: One of the five Kingdoms that contains all prokaryotes. It contains archaeobacteria, eubacteria and cyanobacteria. The first life form emerged over 3,500 Mya were the members of this Kingdom from which eukaryotes evolved.

Monotreme: A subclass of the Class Mammalia consisting of animals that lay eggs.

Mosaicism: Mosaicism is the presence of more than one cell lines differing in genotype or karyotype but derived from one zygote. Post-zygotic new mutations result in mosaic individuals who may not be clinically affected themselves, but are at risk of bearing multiple affected offspring. Mosaicism is well recognized in Duchenne muscular dystrophy and in autosomal dominant disorders with high new mutation rates (see [Clinical Genetics](#)).

mRNA: Messenger RNA. It is the first product of the DNA transcription by RNA polymerase. mRNA forms 1-5% of the total cellular RNA. Its molecular weight is generally less than 2×10^6 . At any time, there are about 10^5 species of mRNA in a cell.

Muller's ratchet: The continual decrease in fitness due to accumulation of (usually deleterious) mutations without compensating mutations and recombination in an asexual lineage. Recombination (sexual reproduction) is much more common than mutation, so it can take care of mutations as they arise. This is one of the reasons why sex is believed to have evolved.

Multivariate analysis: A statistical analysis of several variables assessed simultaneously. Should be preferred over individual analyses of pairs of variables.

Mutation: Any heritable change (not only point mutation) brought about by an alteration in the genetic material. Includes gene conversion, deletion, duplication, insertion and so forth. Link to [Human Gene Mutation Database](#) (Cardiff, UK).

Mutation pressure: Evolution by different mutation rates alone.

Mutation rate: The number of mutations at a particular locus, which occur per gene per cell generation. This is the only source of variation in asexual organisms. The mutation rate is the likelihood of parentage when findings suggest otherwise. Beware of the different units in different mutation rates. In humans, the mutation rate is 1 bp per 10^9 bp per cell division. This corresponds to 10^{-6} mutations per gene per cell division and because there are 10^{16} divisions in a lifetime, 10^{10} mutations per gene per lifetime.

Mya: Million years ago.

Natural killer (NK) cell: Bone marrow-derived, mononuclear white blood cells (large granular lymphocytes) that are able to kill invading microorganisms without activation by cells of the immune system. They are, therefore, part of the innate immune system. They are specialized in killing virus-infected cells and cells transformed to develop cancer.

Natural selection (Darwin's definition, 1859): "As many more individuals of each species are born than can possibly survive; and as, consequently, there is a frequent recurrent struggle for existence, it follows that any being, if it varies however slightly in any manner profitable to itself, under the complex and sometimes varying conditions of life, will have a better chance of surviving, and thus be naturally selected." Link to a simulation on [natural selection](#).

Nature-nurture debate: The debate on the relative contributions of genetics (nature) and environment (nurture) to the characteristics of an organism. An example is the debate on whether gene(s) and/or environmental factors determine the sexual orientation of an individual. Finding a gene playing a role in the development of a condition does not necessarily mean it is a purely genetic trait.

Negative assortative mating: A type of nonrandom mating in which individuals of unlike phenotype mate more often than predicted under random mating conditions.

Neoteny: Retention of juvenile features in sexually mature adult animals. Neoteny frequently correlates with recent evolution of the species (like Homo sapiens).

Neurospora crassa: Haploid, heterothallic, filamentous Ascomycete fungus (bread mold). It has two mating types (A and a) operating as sexual compatibility system, and 11 het loci operating as heterokaryon compatibility system in vegetative phase. Link to [Neurospora](#) website.

Neutral theory: Link to a lecture on [neutral theory](#) of evolutionary change.

Non-coding region: Parts of a gene that include sequences, which are not translated. Both 5' and 3' untranslated regions (UTRs), upstream promoter region and introns are classified as non-coding regions.

Non-disjunction: Due to failure in pairing of homologous chromosomes in meiosis, the two members of one pair migrate to the same pole, giving rise to unbalanced gametes, one of which contain both homologous chromosomes, and the other none (most frequent in sex chromosomes). The non-disjunction event is much more frequent in maternal meiosis I. This may be due to the fact that in a mature woman, oocytes have been held in the ovary for a very long time at prophase I of meiosis from before her birth to shortly before ovulation of the oocyte in question.

Non-overlapping generation model: This population biology model assumes death of all members of a generation (in the cycle of birth, maturation and death) to die before the next generation reaches maturity. This assumption is necessary to (mathematically) simplify the models.

Nonsense mutation: A mutation that changes an amino acid specifying codon to one of the three termination (**stop**) codons.

Notochord: A rod that forms in the embryonic mesoderm and which establishes the front-to-back orientation of vertebrate embryos. It also initiates the formation of the nervous system, the skeleton and most muscles.

Nuclease: An enzyme that breaks bonds in nucleic acids. Deoxyribonuclease (DNAase) and ribonuclease (RNAase).

Nucleoid: The loosely tangled clump of DNA within the cytoplasm of a prokaryotic cell.

Nucleolar organizer: A region on a chromosome that is associated with formation of a new nucleolus following cell division. It contains the genes for several species of **ribosomal RNA** (rRNA), i.e., 18S, 5.8S, 5S and 28S in eukaryotes.

Nucleolus: The site of synthesis of rRNA within the nucleus of a eukaryotic cell.

Nucleoside: A small molecule composed of a purine or pyrimidine base linked to a five-carbon sugar (pentose: ribose or deoxyribose). With the addition of a phosphate group, it becomes a **nucleotide**. Nucleosides in RNA are adenosine, guanosine, cytidine and uridine; in DNA, they are (d)adenosine, (d)guanosine, (d)cytidine and (d)thymidine.

Nucleosome: A beadlike structure of eukaryotic chromosomes. It consists of a core of eight histone molecules and a DNA segment of about 150 base pairs. Each nucleosome is separated from another by a linker DNA sequence of about 50 base pairs. Nucleosome structure helps to fold DNA into a compact form in the interphase nucleus. Otherwise the length of a chromosome, when linear, is many orders of magnitude greater than the diameter of the nucleus.

Nucleotide: The monomeric unit that makes up the DNA or RNA, formed by a phosphate group, a pentose and one of the nitrogenous bases (A, T/U, C, G). Nucleotides in RNA are adenylate, guanylate, cytidylate and uridylylate; in DNA, they are (d)adenylate, (d)guanylate, (d)cytidylate and thymidylate.

Obligate carrier: An individual who must carry a recessive mutant gene based on analysis of the family history. This definition usually applies to disorders inherited in an autosomal recessive and X-linked recessive manner such as parents of a child with an autosomal disease or mothers of boys with an X-linked disease.

Okazaki fragment: The small (<1 kb), discontinuous strands of DNA produced using the lagging strand as template during DNA synthesis. DNA ligase links the Okazaki fragments to give rise to a continuous strand.

Olfactory: Related to smell.

Oligonucleotide: A short, synthetic DNA string used as a probe (as in SSOP) or primer (as in SSP) in molecular genetic studies.

Oligonucleotide ligation assay (OLA): A PCR-based method for SNP typing. It is a ligase mediated gene detection system which uses exact 3' matching of a primer to one of the SNP allele. If this happens the other labelled oligonucleotide which binds to the nucleotide immediately next to the SNP on the other side would be joined to the primer by ligase. The resulting sample can then be tested for the presence of the label (for example biotin). Unless controls are included, false positives are possible (link to a book chapter explaining [OLA](#)).

Oncogene: A gene capable of causing malignant transformation. Link to an overview of [oncogenes](#).

One gene - one (enzyme) polypeptide hypothesis: The hypothesis that each gene controls the synthesis of a single polypeptide which may be a subunit of a complex protein.

Oocyte: Female sex cell which undergoes meiosis and produces an egg (ovum).

Open reading frame (ORF): A nucleotide sequence encoding a polypeptide starting with a start and ending with a stop codon.

Operon: A type of genetic unit which consists of one or more transcription units that are transcribed together into a polycistronic mRNA. The transcription of each operon is initiated at a promoter region and controlled by a neighboring regulatory gene (an operator which binds to a repressor or an apoinducer, to repress or induce the transcription, respectively). An example is the *lac operon* of *E. coli*.

Ori: The origin of replication in prokaryotes.

Orthology: Being homologous by descent between species. In other words, descendants from a common ancestor. An example is the MHC class II genes in different species that all descended from a common ancestral class I gene.

Overdominance: See **balancing selection**. See also **underdominance**.

Oviparity: Egg birth as opposed to live birth (viviparity).

Paleontology: The study of fossils.

Parallel evolution: Evolution of roughly similar changes in two or more closely related lineages.

Paralogy: Being homologous due to a recent or past duplication within the same species. An example is the chromosomal regions 1q22-q23;6p21.31; and 9q33-q34 in humans. These regions contain very similar genes including some of the **MHC** class III genes.

Paramecium: A unicellular Protocist belonging to the group Ciliates. Although normally reproduces asexually, they also undergo sexual **conjugation** in which **mating types** play a role. *Paramecium aurelia* has 34 hereditary mating types that form 16 distinct mating groups (link to [Paramecium](#) in Encyclopaedia Britannica).

Paramutation: In paramutation, two alleles of a gene interact so that one of the alleles is epigenetically silenced. The silenced state is then genetically transmissible for many generations.

Parapatric speciation: Speciation that occurs as a result of two populations diverging in adjacent geographical areas.

Parental imprinting: see **genomic imprinting**.

Parsimony: The scientific convention whereby the simplest explanation is preferred over the others. This is usually a phylogenetic tree requiring the fewest evolutionary steps.

Parthenogenesis (virgin birth): Reproduction involving unfertilized eggs. The offspring of parthenogenetic parents are less diverse than those of sexual parents. Parthenogenesis often alternates with sexual reproduction. Two examples are the desert grassland whiptail lizard (*Cnemidophorus uniparens*) and Caucasian rock lizard. Never seen in birds and mammals.

PCR: Polymerase chain reaction. A technique that allows amplification of specific DNA segments in a very short time. Link to an article on [PCR](#).

Pedomorphosis: The retention of infantile, fetal or embryonic characteristics into the adult.

Penetrance: The proportion of individuals with a given genotype (heterozygotes for a dominant gene) who express an expected trait, even if mildly. If a disease gene is not causing the disease in all its carriers, its penetrance is low [not to be mixed with **variable expression**]. BRCA1 mutations show both age-dependent penetrance and overall reduced penetrance, the lifetime risk for a female mutation carrier being estimated at around 70%. Breast cancer is also an example of an autosomal condition where penetrance is sex-dependent. While male mutation carriers can develop breast cancer (particularly with BRCA2 mutations), females are at much greater risk.

Phenocopy: A non-genetic condition resembling a genetically determined one. Such conditions confound the interpretation of pedigrees and therefore genetic counseling. Some teratogens may cause congenital anomalies mimicking genetically caused anomalies.

Phenotype: The visible or measurable (i.e., expressed) characteristics of an organism (see genotype).

Pheromone (formerly called mating-type factors, sex factors or gamones): Species- or mating-type-specific chemical produced by an animal to communicate with an effect on their behaviour without being consciously perceived as smell. Probably the most ancient communication system in living organisms. It has been noted in bacteria (*Streptomyces faecalis*); protists such as ciliates (*Euplotes raikovi*); amoeba (*Dictyostelium*); algae (*Fucus vesiculosus*); and in fungi (for example Basidiomycetous). It is mainly used to attract opposite sex in insects (in cockroaches, moths, beetles and bees) but social insects use them for communication other than reproduction, and mammals use them for territory marking. In vertebrates, they are also used for kin recognition and mate selection. Specific examples of pheromones include a substance produced by male cockroaches that orients females in the correct mating positions and a desert locust pheromone that accelerates sexual maturation in adults of both sexes. Male-attracting pheromones are produced by the females of many species of beetles, bees, and moths. The polyphemus moth needs red oak leaves for mating as the leaves release a volatile aldehyde that stimulates the female to produce her male-attracting pheromone. Links to an essay on [Pheromones](#), [Human Pheromone Research](#) and [pheromone-related links](#).

Phyletic gradualism: A model of evolutionary mode characterized by slow and gradual modifications of biological structures leading to speciation. This is the opposite of **punctuated equilibrium**.

Phylogenetic footprinting: The use of phylogenetic comparisons to reveal conserved functional elements.

Phylogenetics: Study of reconstructing evolutionary genealogical ties between taxa and line of descent of species or higher taxon.

Phylogeny: An evolutionary tree showing the inferred relationships of descent and common ancestry of any given taxa. Link to the [Tree of Life](#), [Spectrum of Life](#), [Lecture on Tree Construction](#), Freeware Phylogenetic Data Analysis Download Pages: [Phylogenetics Software](#), [Phylogeny Programs](#)

Plasmid: A transferable extrachromosomal genetic element found in some bacteria. They are 1 to 200 kb long, double-stranded, circular DNA molecules. They replicate independently of the bacterial chromosomes and usually confer an advantage to the bacteria (such as antibiotic or

heavy metal resistance). Plasmids are popular vectors in recombinant DNA technology. They can carry up to 10 kb foreign DNA.

Plastid: Chloroplast and related organelles in the cells of eukaryotic algae and green plants. Together with mitochondria they are responsible for energy production.

Platyrrhini: New world monkeys found in South America with widely spaced nostrils. They last had a common ancestor with the old world monkeys (**Catarrhini**) about 55 million years ago.

Pleiotropy: More than one effect of a gene on the phenotype. The effects may occur simultaneously or sequentially. An example is the determination of the colour pattern and the shape of the eyes by a single allele in Siamese cats. Another example is the **DNA repair** enzymes which have several other functions (transcription, cell cycle regulation, regulation of gene rearrangements).

Point mutation: A single nucleotide change in the DNA sequence. Even if it is in the coding region of a gene, it may or may not change the amino acid sequence. The rates of point mutation for **MHC** genes are not unusually high. The extensive polymorphism results from their accumulation over many millions of years of transspecies evolution.

Pollen grain: The microspores of seed plants. It germinates to form the male gametophyte. The male gametophyte contains three haploid nuclei. One of these fertilizes the ovum, a second fuses with the two polar nuclei to form the triploid endosperm, and the third degenerates once **double fertilization** occurs.

Pollen (tube) competition: The plant equivalent of **sperm competition** (see also **cryptic female selection**). A type of **sexual selection**.

Poly (A) tail: A sequence of 20 - 200 adenylic acid residues which is added to the 3' end of most eukaryotic mRNAs. It increases their stability by making them resistant to nuclease digestion.

Polygenic: Traits controlled by two or more genetic loci. They are usually influenced by environment as well (multifactorial).

Polymorphism: Presence of discreetly different forms of a gene or a character. It is defined as a Mendelian trait that exists in the population in at least two phenotypes, neither of which occurs at a frequency of less than 1%. Polymorphism at a genetic locus is due to either balanced polymorphism (heterozygous advantage, frequency-dependent selection) or unequilibrium states (temporary polymorphism) as occurs during frequency-dependent selection and genetic drift (alleles becoming fixed or extinct).

Polyploidy: The situation in which the organism has more than two ($2n$) sets of chromosomes. It could be $3n$, $4n$ or more. A common situation in earthworms and plants. About half of **angiosperms** are polyploid. It arises as a result of meiotic irregularities and gives rise to sterile progeny, which can still reproduce asexually. The original South American potato is a tetraploid ($4n$). Many of the common food plants are polyploid as this results in larger flowers and fruits (as well as larger cells, thicker and fleshier leaves). The wheat now grown for bread *T. aestivum* is hexaploid ($6n = 42$ chromosomes). Polyploidy is a common mechanism for sympatric speciation which played a role in angiosperm evolution. Link to a mini-essay on **polyploidy**.

Population biology: The study of the patterns in which organisms are related in time and space. It is a combination of disciplines such as population genetics, ecology, taxonomy, ethology and others. Link to a **population biology** website.

Population genetics: The branch of genetics that deals with frequencies of alleles and genotypes in breeding populations. It also deals with selective influences on the genetic composition of the population (links to **freeware population genetic data analysis software: Arlequin 2000, PopGene, GDA, Genetix, Tools for Population Genetic Analysis, GenePop, GeneStrut, SGS, GenAlEx, WinPop, Quanto, features of data analysis software**; lectures on **population genetics** and **population biology**). See also **Basic Population Genetics**.

Position effect: A difference in phenotype that is dependent on the position of a gene or a group of genes, often caused by heterochromatin nearby. Thus, the change in a gene's location may cause a change in its expression (a problem that has to be overcome in gene therapy).

Positive assortative mating: A type of nonrandom mating in which individuals of similar phenotype mate more often than predicted under random mating conditions.

Post-translational modifications: Cleavage of amino terminal peptide, hydroxylation and oxidation of amino acids in the polypeptide chain for cross-linking, covalent modifications by acetylation, phosphorylation and glycosylation.

Pre-Cambrian Eon: The whole of geological time before the **Cambrian** period (< 540 Mya).

Pre-mRNA (precursor mRNA): The primary transcript and intermediates in RNA processing that yield functional (mature) mRNA.

Primates: One of the Mammalian Orders which includes Lemurs (suborder Prosimii), old world and new world Monkeys, Apes and Humans. The suborder Anthroidea covers all Primates except Prosimii. Prosimians and Anthropoids diverged from each other 65 Mya; Apes and Old World Monkeys 35 Mya. The separation of the two groups of Anthropoids (**Platyrrhini** and **Catarrhini**) occurred about 55 Mya.

Primase: An enzyme that makes the RNA primer required by DNA polymerase in DNA replication. See also **primosome**.

Primer: A short nucleic acid molecule which, when annealed to a complementary template strand, provides a 3' terminus suitable for copying by a DNA polymerase.

Primosome: The mobile complex of **helicase** and **primase** that is involved in DNA replication.

Prion: An infectious agent, which does not have any nucleic acid (but just protein). Responsible for scrapie in sheep, kuru and Creutzfeldt-Jakob disease in humans.

Processed pseudogene: A pseudogene, derived from a retrotranscript of mRNA of any expressed gene and inserted back into the genome. A processed pseudogene is intronless, usually flanked by the repeat sequence (GC/AGCTCTCC), and rich in multiple genetic lesions including substitution, deletion and/or insertion events that modify the reading frames. Because of the lack of its original promoter and the genetic lesions it has accumulated, a processed pseudogene is not normally expressed.

Prokaryotic cell: The cell type in which the DNA is not enclosed in a nucleus. Consists of Eubacteria and Archaeobacteria. Always unicellular. When the cell has a proper nucleus (eukaryon), it is **eukaryotic**. Link to **Prokaryotes** website.

Promoter: Initial binding site for RNA polymerase in the process of gene expression. First **transcription factors** bind to the promoter which is located 5' to the **transcription** initiation site in a gene. General and tissue/cell-specific promoters stimulate the expression of a gene under the control of **enhancers**.

Promoter-proximal element: Any regulatory sequence in eukaryotic DNA that is located within 200 bp of a **promoter** and binds a specific protein to modulate transcription of the associated gene.

Proofreading: In DNA synthesis, the ability of DNA polymerase to recognize mismatched bases. DNA polymerase corrects mistakes with its exonuclease activity. RNA editing is also possible at the mRNA level in some simple organisms.

Prophage: The phage DNA inserted into a bacterial chromosome.

Prosimians (Prosimii): The suborder of **Primates** including Lemurs, Lorises and Tarsiers.

Protein clock hypothesis: The idea that amino acid replacements occur at a constant rate in a given protein family (ribosomal proteins, cytochromes, etc) and the degree of divergence between two species can be used to estimate the time elapsed since their divergence.

Proteomics: Similar to the genomic microarray technique, separation, identification and characterization of the complete set of proteins in the cells in order to see how they affect cell function.

Protochordata: A division of chordata phylum including subphylum hemichordata, urochordata (tunicates) and cephalochordata (ancelets). Immediate ancestors of **vertebrates**.

Protoctista: The modern name for the Kingdom Protista. Probably the first double-chromosomed beings. Well-known members are: seaweeds (algae), amoebas, ciliates, mastigotes, water molds, slime molds and slime nets.

Protozoa (meaning "first animals"): A collective name for several phyla of unicellular and eukaryotic, the most animal-like, mainly parasitic organisms belonging to the Kingdom Protista (modern name: Protoctista meaning "first beings"). This Kingdom includes Protozoa and other not necessarily single-celled but eukaryotic organisms (algae and others formerly known as fungi) except fungi. They reproduce by fission or conjugation, and move by cilia, flagella or pseudopodia. The taxonomic Kingdom Protista is thought to be the ancestor of the other eukaryotic kingdoms Fungi, Plants and Animals. Familiar examples of protozoa are flagellates (incl. **Chlamydomonas**), Amoeba, Plasmodium, Ciliates (incl. **Paramecium**). Link to a discussion of the **Protists**.

Pseudoalleles: Genes that behave like alleles but can be separated by crossing over. The eye colour genes on the X chromosome of Drosophila are for example closely adjacent but separable loci and not alleles of a single gene.

Pseudoautosomal inheritance: The X and Y chromosomes share a common ancestor. There is a part of X chromosome, which has its homologous counterpart on the Y chromosome. The pattern of inheritance for a gene located on both the X and Y chromosomes may appear to be autosomal.

Pseudogenes: A gene which has acquired a nonsense mutation and lost its transcription ability.

Pseudomonas syringae: The genetically engineered strain of this bacteria lack a cell-surface protein that helps ice crystals to form. Spraying these bacteria on crops may prevent freezing-related damages.

Pulse field gel electrophoresis (PFGE): A form of gel electrophoresis that allows extremely long DNA molecules to be separated from one another.

Punctuated equilibrium: Put forward by Niles Eldredge and Stephen Jay Gould in 1972 as a counter theory to Darwin's gradualism in speciation (see **phyletic gradualism**). It suggests that new species may have arisen rapidly over a few thousand years and then remained unchanged (stasis) for many millions of years. Punctuated equilibrium postulates that change occurred in only a small part of the population (rather than the whole population is evolving gradually). The most plausible explanation for a sudden and drastic change would be mutations in regulatory sequences that affect a whole operon (see Gould SJ & N Eldredge. Punctuated equilibria: the tempo and mode of evolution reconsidered. Paleobiology 1977;3:115-51). Link to an essay on **punctuated equilibrium**.

Quantitative character: A character displaying a continuous phenotypic range rather than discrete classes (characters taking any value within a limit; characters measured rather than counted such as metabolic activity, height, length, width, body fat content, growth rate, milk production). The genetic variation underlying a continuous character distribution may be the result of segregation at a single genetic locus or more frequently, at numerous interacting loci which produce cumulative effect on the phenotype. A gene affecting a quantitative character is a quantitative trait locus, or QTL.

Quantitative genetics: The statistical study of the genetics of quantitative characters (biometrical genetics) as opposed to Mendelian (discrete) characters. Quantitative genetic characters are those that do not assort in a simple way in crosses. Examples are physiological activity, reproductive rate, behaviour, size, and height. A major task of quantitative genetics is to

determine the ways in which genes interact with the environment to contribute to the formation of a given quantitative trait distribution (and the estimation of genetic and environmental variance). Link to a lecture on [quantitative genetics](#).

Quasidominance: Direct transmission, generation to generation, of a recessive trait giving the impression of dominance. It happens if the recessive gene is frequent or inbreeding is intense.

Quasispecies: The whole population of phylogenetically related (virus) variants observed within a single (infected) individual. Viruses with high mutation rates such as human immunodeficiency virus (HIV) and hepatitis C virus (HCV) occur like this. As a comparison, HIV variation within a single infected individual can be as great as the variation of influenza throughout the worldwide-infected population in a flu season. A review on quasispecies by [DB Smith](#) (1997).

Race: Described in population genetics as a geographic subdivision of a species distinguished from others by the allele frequencies of a number of genes. A beautiful discussion that there are no genetically defined races within *Homo sapiens* can be found in Cavalli-Sforza's book [Genes, Peoples, and Languages](#) (2000).

Random mating: Mating without any preference for mates. One of the assumptions of Hardy-Weinberg equilibrium (nonrandom mating may be due to **assortative or disassortative mating**).

Recessive: A trait that is not expressed in heterozygotes (i.e., that can only be expressed in the homozygous state).

Recombinase: A group of enzymes that catalyze the joining of two DNA molecules after recognizing the recombination sites. See also **integrase** and **transposase**.

Recombination (crossing-over): The exchange (reshuffling) of genetic material between a homologous pair of chromosomes during meiosis (see also **somatic recombination**; **sister chromatid exchange**).

Regulators of Complement Activation (RCA): Membrane proteins that inhibit formation of and promote decay of C3-activating convertases, and prevent formation of membrane attack complex. Their main role is to prevent host cells from complement attack.

Repetitive DNA: Non-coding DNA, which consists of nucleotide sequences repeatedly occurring in chromosomal DNA. They do not normally have any function but those capping the chromosomes prevent the loss of genetic information after each replication (as this would cause a 3' overhang). In human genome, at least 20% of the DNA consists of repetitive sequences.

Replicon: A unit of genetic material, which behaves autonomously during replication of DNA. In bacteria, a whole chromosome is a replicon. In eukaryotes, chromosomes are divided into hundreds of replicons. Each replicon contains a segment beginning with a binding site for RNA polymerase.

Repulsion (trans-arrangement): The condition in which a double heterozygote has received a mutant and a wild-type allele from each parent, e.g., $a + / + b$ (see also **coupling**).

Residue: A compound such as an amino acid or a nucleotide when it is part of a larger molecule.

Reverse transcriptase: RNA-dependent DNA polymerase.

RFLP (restriction fragment length polymorphism): Genetic polymorphism as revealed by the sizes of fragments generated with a particular restriction endonuclease enzyme (such as *EcoRI*, *PstI*, *BglII*).

Ribosomal RNA (rRNA): One of the four types of RNA that exist in a eukaryotic cell. It has a very slow mutation rate which is useful in phylogenetic analysis of kingdoms and phyla.

Ribosome: A small cytoplasmic organelle that is the site of mRNA translation, thus protein synthesis.

Ribozyme: RNA molecules with enzymatic activity. Its presence in organelles from plants, yeast, viruses and eukaryotic cells revolutionized the ideas about the origin of life.

Robertsonian translocation: see **centric fusion**.

RNA (ribonucleic acid): A single-stranded nucleic acid that is found both in nucleus and cytoplasm. Other differences from DNA are: it contains uracil instead of thymine, it is single-stranded, and its sugar molecule is ribose. Total cellular RNA is made up of ribosomal RNA (rRNA, 80-85%), transfer RNA (tRNA, 15-20%) and messenger RNA (mRNA, 1-5%). See also **small nuclear RNA** and **heterogeneous nuclear RNA**.

RNA interference (RNAi): The use of double stranded RNA to interfere with gene expression. RNAi is usually mediated by approximately 21-nt small interfering RNAs. See a review by [Zhang & Hua, 2004](#).

RNA polymerase: An enzyme that transcribes an RNA molecule from the template strand of a DNA molecule. It adds to the 3' end of the growing RNA molecule one nucleotide at a time using ribonucleotide triphosphates (rNTPs) as substrates (this reaction releases pyrophosphates). RNA polymerase I is dedicated to the synthesis of only one type of RNA molecule (pre-rRNA). RNA polymerase II is required for general transcription reactions. RNA polymerase III produces small RNAs such as tRNAs and 5S rRNA.

Rotifer: Microscopic aquatic organisms (pseudocoelomates). Their peculiar features are ancient asexuality, parthenogenetic reproduction and dormancy. Despite being diploid, they don't have homologous chromosomes (see Aydin Orstan's [Bdelloid Rotifers](#) website, and [Rotifera](#) by UCMP).

Saccharomyces cerevisiae: Unicellular Ascomycete yeast known as the baker's or brewer's yeast. Widely used as a simple eukaryotic model, particularly in recombinant DNA and cell cycle studies as well as in **mating type** and **heterokaryon** compatibility studies. It has most advantages of a prokaryotic system but is a true **eukaryote**. It is considered as the E.coli of the eukaryotes. S.cerevisiae can reproduce both asexually and sexually, and can be cultured in either the haploid or the diploid state. One major advantage of yeast is the ease with which specific gene disruptions, gene replacements, and gene retrievals can be accomplished. Its complete genome was sequenced in 1997 and contains 12,057,500 bp, 6,000 genes in 16 chromosomes. It is used in the creation of YACs. Link to [S.cerevisiae](#) website.

SAGE (serial analysis of gene expression): A high-throughput method that uses 10-14 bp-long tags from each cDNA expressed in a cell. The concentration of each tag sequence is proportional to the level of its mRNA in the original sample. This method is used to explore gene regulation in cell populations. Link to the [NCBI SAGE](#) website.

Schizophyllum commune: A fungus species of Basidiomycetes group. It has thousands of mating types in a multiallelic locus, a pheromone receptor system and a pheromone system. Because of this, it has the maximal outbreeding rates (98.8%) in nature.

Sea Urchin: A small spiny marine invertebrate belonging to the phylum [Echinodermata](#). It is a model animal for the study of fertilization and development. Because it is a spawner, its gametes can be obtained in large quantities. Simple mixing of sperms and eggs causes synchronous mitosis and cytokinesis. Furthermore, its eggs are large and clear. Links to [Sea Urchin Embryology](#) (Stanford) and [Sea Urchin Genome Project](#) (CalTech) websites.

Seed: A fertilized and mature ovule which will develop into a new plant if sown. It contains usually one fertilized ovum (embryo) and endosperm (or perisperm) which is the nutritive cell surrounding the embryo.

Segregation: The separation of members of a gene pair from each other during gamete formation.

Segregation distortion: Violation of Mendel's first law which results in unequal segregation of a pair of alleles.

Selection Differential (S) and Response to Selection (R): Following a change in the environment, in the parental (first) generation, the mean value for the character among those individuals that survive to reproduce differs from the mean value for the whole population by a

value of (S). In the second, offspring generation, the mean value for the character differs from that in the parental population by a value of R which is smaller than S. Thus, strong selection of this kind (directional) leads to reduced variability in the population. Link to a lecture on [selection](#).

Self-fertilization (also selfing, self-pollination): The fusion of male and female gametes produced by the same (hemaphrodite or bisexual) individual. Self-fertilization allows an individual to create a local population, but it fails to provide variability within a population and limits the possibilities for adaptation to environmental change. Some plants reproduce by self-fertilization but most hermaphroditic animals rarely use self-fertilization, since many of them have adaptations encouraging cross-fertilization.

Self-incompatibility system (SI system): The genetic complex of plants that prevents self-fertilization. See also **MHC** and **mating types**.

Sense mutation: A mutation that changes a termination (stop) codon into one that codes for an amino acid. Such a mutation results in an elongated protein.

Sequence tagged site (STS): Short (200-500bp) DNA sequence with known location and sequence and only occurs once in the human genome. For a catalogue of human STSs, see [UniSTS](#).

Serological typing: Identification of **MHC** molecules expressed on cells using either naturally occurring antibodies in multiparous women or by alloantiserum raised in animals.

Sex: Formation of new organism containing genetic material from more than a single parent.

Sex-influenced dominance: A dominant expression that depends on the sex of the individual. For example horns in sheep are dominant in males and recessive in females.

Sexual dimorphism: The existence, within a species, of differences in morphology between the sexes. Examples are greater size in males of gorilla, baboon and elephant seals.

Sexual reproduction: Reproduction requiring the union of sex cells (gametes) which are themselves products of meiotic division. Each offspring has a unique genetic composition due to independent assortment of chromosomes during meiosis, recombination and union of gametes.

Sexual selection: Natural selection operating on factors that contribute to an organism's mating success. Described by Darwin as natural selection in relation to sex. Links to an [essay](#) and a [lecture note](#) on sexual selection.

Sexually antagonistic: Having opposite effects in the two sexes.

Shine-Dalgarno (S-D) sequence: An eight nucleotide consensus sequence 5' UAAGGAGG 3' found in bacterial mRNAs five to ten bases before the translation initiation codon (AUG). It is thought to be involved in initiation of translation by helping the mRNA bind to the ribosome (16S rRNA), thus it can be called the ribosomal binding site (see also **Kozak sequence**). In eukaryotic DNA, there is no such sequence. The 5' cap present on all eukaryotic mRNAs seems to be the first signal to start protein synthesis.

Sibling species: Two species evolved from a common ancestor and are genetically distinct but morphologically similar.

Signal sequence: A stretch of hydrophobic amino acids at the amino-terminal of a protein that guides protein translocation through cellular membranes such as lysosomal membrane. It helps the protein to pass through the membrane via interaction with its receptor on the membrane and is usually cleaved off at the final destination by an endopeptidase. Sometimes used interchangeably with **leader sequence**.

Signal transduction: A complex multistep pathway by which extracellular signals are transduced from plasma membrane receptors to the transcription machinery in the nucleus and the translation machinery in the cytoplasm, subsequently to regulate cell proliferation and differentiation. The components are growth factors, growth factor receptors, membrane and

cytoplasmic tyrosine kinases, GTP-binding (G) proteins, nuclear binding proteins and **transcription factors**.

Silencer: A DNA sequence which acts in the opposite direction of an **enhancer** to inhibit the transcription of a gene.

Silent mutation: Base-pair substitution, which alters a codon but does not result in altered phenotype due to the degeneracy of the genetic code (synonymous mutation).

SINES (short interspersed elements): An abundant intermediate DNA sequence in mammals about 300 bp long (see also **LINES**).

Single Nucleotide Polymorphism (SNP): A single nucleotide change in the DNA code. It is the most common type of stable genetic variation and usually bi-allelic. SNPs may be silent -no change in phenotype- (sSNP), may cause a change in phenotype (cSNP) or may be in a regulatory region (rSNP) with potential to change phenotype. Anonymous SNPs are the most common ones. These are in non-coding regions and used as genetic markers. On average, each 1 kb of human genome contains 2-10 SNPs, i.e., one in every 100-500 nucleotides is polymorphic most frequently a C to T substitution (links to a [lecture](#), [SNP Consortium Website](#), [dbSNP](#), [GeneSNPs](#), [WIAF SNP](#)).

Sister chromatid exchange: An exchange (crossing-over) of genetic material between the two (identical) chromatids of a chromosome in mitosis (mitotic recombination). Normally, genetic **recombination** takes place between homologous chromosomes in meiosis I. Sister chromatid exchange may be a sign of chromosomal instability but has no genetic consequences as long as the exchange is the result of an equal crossover.

Small nuclear RNA (snRNA): Small (90 to 300 nucleotides) RNA molecules that are not directly involved in protein synthesis but may have roles in RNA processing (splicing) and the cellular architecture. There are six types of snRNA: U1 to U6.

Somatic recombination: Rearrangement of genes in cells other than germ cells which happens to generate the extreme diversity of T-cell receptors and immunoglobulins (see **adaptive immunity**).

Species: A group of individuals, which can successfully breed with each other to produce offspring which can still breed with each other. There are evolutionary, biological, and recognition species concepts.

Speciation: It is now almost universally agreed that the prevailing process of speciation is geographical (allopatric) speciation. (There are also **parapatric** and **sympatric speciation** concepts.) According to biological species concept, however, species are defined as aggregation of populations that are reproductively isolated from one another. Link to a lecture on [speciation](#).

Sperm competition: Competition not for access to females but for fertilization of egg. Equivalent to **pollen tube competition** in plants and a type of **sexual selection**. See Birkhead's book [Promiscuity](#) (2000) for a detailed study of sperm competition in nature.

Spermatocyte: A male germ cell that undergoes meiosis and produces a haploid spermatid and subsequently a sperm.

Spermatophyta (sperm plants; Gr. sperm=seed; phyton=plant): A major subdivision within the vascular plants division of the plant Kingdom characterized by reproduction by seeds. Sperm plants include angiosperms and gymnosperms.

Splicing: An event which takes place within the nucleus whereby introns are removed from the precursor mRNA and the exons are joined together as a post-transcriptional modification.

Spore: In plants with alternation of generations, small reproductive bodies capable of giving rise to a new offspring either immediately or after a period of dormancy. It can be produced asexually or sexually. It usually germinates without fusing with another cell. Sexual spores of plants are haploid cells produced by meiosis.

Sporophyte: The diploid (asexual) spore-producing generation in plants with **alternation of generations**. A sporophyte is typically formed by the union of sexual cells produced by the gametophyte. In higher plants, the sporophyte is the conspicuous plant. In lower plants (such as mosses), the **gametophyte** is the dominant generation.

SSOP: Sequence-specific oligonucleotide probe. Together with PCR-SSP, commonly used to type classical MHC genes following amplification by PCR reaction (see also serological typing).

SSP: Sequence-specific primer. PCR-SSP is a common method to type classical MHC genes.

Stabilizing selection: Natural selection against extreme deviations from the average (like low and high birth weight).

Stop (termination) codon: Codons that signal the end of a growing polypeptide chain. These are UAA, UGA and UAG.

Superantigen: An antigen of a virus or bacteria, which bind to non-polymorphic parts (outside of the antigen-binding cleft) of MHC class II molecules and interact with the V β domain of the T-cell receptor. This way, they activate an entire subgroup of T cells (rather than a specific one) expressing the appropriate V β and this is followed by deletion of the activated T cells upon exposure to the superantigen. One superantigen can activate up to 20% of the helper T cell repertoire. The prototype bacterial superantigen is the staphylococcal enterotoxins.

Supercoiled/supertwisted DNA: A closed circular DNA molecule in which the DNA molecule is further twisted on itself to form a more compact molecule. Left-handed (negative) supercoiling leads to a loosening of the strands of the double helix (underwinding). Positive supercoiling is not seen in vivo.

Symbiosis: Two organisms living together and both benefit from this. An example is the coral and algae (zooxanthellae).

Sympatric speciation: Speciation that occurs as a result of divergence of two populations occupying the same geographical area. It is due to reproductive isolation of subpopulations of a species without physical isolation. One well-established mechanism is **polyploidy**.

Syngamy: The union of the (haploid) nuclei of two gametes following fertilization to form a single (diploid) nucleus for the zygote.

Syngenic: Genetically identical (isogenic) members of the same species like monozygotic twins.

Synonymous (silent) base change: A change in the nucleotide sequence that does not cause an amino acid change. **Non-synonymous** changes replace the amino acid and are called replacement change.

Synteny: Refers to two genomes in which certain groups of genes are conserved in similar regional maps. Parts of mouse chromosome 17 and human chromosome 6 are syntenic.

Synthetic theory of evolution: Proposed to explain the transformation of a species by natural selection and for the splitting of a species into reproductively isolated subgroups.

Systematics: Classification of living things with regard to their evolutionary relationship. Link to a lecture on **systematics**.

TATA box (Goldberg-Hogness box): A short nucleotide sequence in the promoter 25 to 35 bp upstream to the transcription initiation (cap) site of eukaryotic genes to which RNA polymerase II binds. The consensus sequence is TATAA/TA. The TATA box binds the general transcription factor TFIID (see also **CAAT box**).

Taxon: Any group of organisms to which any rank of taxonomic name (classification) is applied. Plural: taxa.

Taxonomic hierarchy: All taxa are classified within the following groups (starting from the most inclusive): kingdom, ('division' in plants), phylum, class, order, family, genus, species, subspecies (race). See **taxonomy** in BBC Education.

T cells: A subgroup of T lymphocytes characterized by having T-cell receptor (**TCR**) complex and CD3 surface marker. T cells are roughly subdivided into CD4+ helper T cells and CD8+ cytotoxic and suppressor T cells.

TCA cycle: The tricarboxylic acid cycle (also known as the Krebs cycle). A major metabolic pathway involved in aerobic cellular respiration (energy production) which takes place in the mitochondria of animal and plant cells. Pyruvic acid produced during glycolysis is converted to acetyl CoA which is then oxidized to CO₂. The reducing power of the end products NADH and FADH₂ are used in the synthesis of ATP by oxidative phosphorylation.

Teleology: Teleology, from the Greek word telos (purpose), asserts that there is an element of purpose or design behind the workings of nature. Attributing any purposeful direction to evolutionary change would be called teleological.

Teleost: Bony fishes with well-developed bone structure. **Teleostei in Tree of Life.**

Telocentric: A chromosome with a terminal **centromere** (like chromosome 21 in humans).

Telomerase: A **reverse transcriptase** (h *TERT*) containing an RNA molecule (h *TR*) that functions as the template for the tandem repeat at **telomere**. It synthesizes telomere to maintain its length after each cell division. It is active in embryonic cells and gametes, inactive in differentiated somatic cells, and reactivated in malignant cells. Telomerase can add one base at a time to the telomeric end of a chromosome. This maintenance work is required for cells to escape from replicative senescence. Telomerase activity is the most general molecular marker for identification of human cancer.

Telomere: The end of eukaryotic chromosome consisting of tandemly repeated sequences. Chromosomes lose about 100 bp from telomere every time the cell divides. The enzyme **telomerase** can add the lost bases.

Testicular feminization (Androgen Insensitivity Syndrome): An X-linked trait that causes XY individuals to develop into phenotypic females. A mutation causes loss of sensitivity to testosterone (see **Clinical Genetics**).

Tetrapods: Vertebrate animals other than fishes (amphibians, reptiles, birds, and mammals).

Tetrapolar: Used for the mating types of Basidiomycete to describe four distinct ways of interactions between haploid mycelia. These fungi have two mating type loci and there are four degrees of matching: fully compatible at both loci, fully incompatible at both loci, semicompatible (compatible only at locus 1), and semicompatible (compatible only at locus 2). In Ascomycete, the mating type locus is biallelic and mating types are bipolar.

Three-prime (3') end: The end of a DNA or RNA strand with a free 3' hydroxyl group corresponding to the end of transcription (see also **five-prime end**).

Ti (tumor-inducing) plasmid: A plasmid of **Agrobacterium tumefaciens** that is often used as a vector in genetic engineering of plants. During tumor induction, a transposon of the plasmid (T-DNA) integrates into the host chromosome.

Topoisomerase: A class of enzymes that convert DNA from one topological form to another. During DNA replication, they facilitate the untwisting of supercoiled DNA (see also **gyrase** and **helicase**).

Trans-acting gene: A gene acting on or co-operating with another gene on a different chromosome (see also **cis-acting gene**).

Transcription: As the first step in protein synthesis, transfer of genetic information from the DNA template to RNA molecule mediated by RNA polymerase (followed by **translation**). A

transcription unit is a segment of DNA between the sites of initiation and termination of transcription. It may contain more than one gene.

Transcription factors: Proteins that are directly involved in regulation of **transcription** initiation by binding to the control elements and allowing RNA polymerase to act. There are ubiquitous transcription factors as well as cell and tissue-specific ones. Several families have been identified including helix-loop-helix proteins, helix-turn-helix proteins, leucine zipper proteins and zinc finger proteins.

Transcription start site: The position in a gene where the mRNA synthesis starts. The first nucleotide transcribed is denoted +1.

Transcription unit: The region of DNA that extends between the promoter and the termination codon.

Transduction: Transfer of genes from one bacterium to another by means of a bacteriophage.

Transfection: Addition of foreign DNA into a eukaryotic cell by exposing them to naked DNA (i.e., not in a bacteriophage as in **transduction**). In bacterial genetics, it is also called transformation.

Transformation: In bacterial transformation, it means the transfer of genes from one bacterium to another in the forms of soluble fragments of DNA; in malignant transformation, it means conversion of normal animal cell state to unregulated growth.

Transgenic: An organism (animal or plant) that contains genes from another species. This is achieved by introducing the foreign gene into the germline.

Transition: A nucleotide substitution between two purine nucleotides (A and G) or between pyrimidines (C and T/U) in DNA or RNA. Transition-type substitution is more common than **transversion**.

Translation: The process of converting the RNA sequence into the linear sequence of amino acids in a protein product. Translation start site contains a codon (AUG) for methionine but not all proteins start with a methionine as most of the time it is cleaved off post-translationally.

Translocation: Transfer of chromosomal material between chromosomes (usually reciprocal).

Transposase: An enzyme that catalyzes the insertion of a **transposon**.

Transposon: A long mobile DNA element that moves in the genome by a mechanism involving DNA synthesis and transposition.

Transspecies evolution: The favoured type of evolution of the **MHC** allelic diversity. The age of an allele or an allelic lineage is greater than the species. Therefore, common allelic lineages have been inherited from a common ancestor and species-specific mutational diversification occurred within these lineages. For example, no single **MHC** class I allele is shared between humans and chimpanzees, but numerous similarities in lineage, polymorphic motifs and individual substitutions can be observed. One consequence is that certain alleles will be more similar to their correspondent alleles in another species than to the other alleles of the same locus in the same species. The long-term persistence of families of MHC alleles whose origins predate speciation events is called 'trans-species evolution'.

Transversion: A mutation caused by the substitution of a purine (A and G) for a pyrimidine (C and T/U) or vice versa in DNA or RNA (see also **transition**).

Triplet repeat: In this situation, a triplet of nucleotides increases in number within a gene. A mutation especially occurring in central nervous system disorders is the increased number of triplets repeats. Examples include myotonic dystrophy, Huntington's disease, **Friedreich's ataxia** and **fragile X syndrome**. Also in polycystic ovary syndrome, androgen receptor gene has increased number of CAG repeats (**Hickey et al, 2002**). Expansion may be greater depending on the transmitting parent (eg, the mother in myotonic dystrophy, the father in Huntington's disease); thus, a parent-of-origin effect and genetic anticipation can be observed. Increased number of

repeats of a triplet may trigger methylation of the gene that causes the disease (see [Mitas M, 1997](#) for a review). (See also [Clinical Genetics](#).)

Ubiquitin: A small protein that becomes covalently linked to a protein targeted for degradation.

Underdominance: Also called heterozygous disadvantage. This unusual selection process occurs when heterozygotes are less fit than either homozygote. This situation is likely to arise when two adjacent populations are isolated and become homozygous for different alleles, and then come into secondary contact at the borders of their ranges. This is the opposite of **overdominance**.

Uniparental disomy: Inheritance of both homologues of a chromosome from one parent, with loss of the corresponding homologue from the other parent. Hydatidiform mole is a parental disomy disorder.

5' untranslated region: The short sequence between the transcription initiation site and the start of translation that is retained in mRNA but not translated. It contains the ribosomal binding site (leader sequence) and signal sequence. It is the beginning of exon 1.

Upstream: Sequences located to the opposite direction to transcription (which runs from 5' to 3' on the sense strand of the DNA). A nucleotide 25 bp upstream to the first transcribed nucleotide is at position -25.

Ustilago: A genus in Basidiomycetes Phylum of the Fungal Kingdom. It represents the smut fungi that are well known for their highly polymorphic mating types.

Variable expression: A variation in phenotype between affected members of the same family (i.e. individuals carrying identical mutations). It occurs in many dominant conditions and may be associated with reduced penetrance [see also **penetrance**].

Variant: Although commonly used to mean allele, mutant or to denote any genetic polymorphism, its precise use should be reserved for genetic markers that occur in less than 1% frequency in the population (hence cannot be called polymorphism).

Vector: A plasmid, phage, or cosmid into which foreign DNA may be inserted for cloning.

Vertebrates: A subphylum in the Phylum **Chordata** of the Kingdom Animalia. All members have a **notochord** and a cranium (skull). Includes the Classes: Fishes, Amphibians, Reptiles, Birds, and Mammals (monotremes, marsupials, placentals). Link to the [vertebrates](#) page in the tree of life.

Viroid: A disease-causing agent consisting of only a single-stranded, short (270 to 380 nucleotides) RNA molecule.

Virus: An entity that is capable of reproducing only by infecting a bacterial or eukaryotic cell. Viruses are incapable of autonomous replication and have to use a host cell's translational system. They consist of a nucleic acid molecule and protein coat. The genetic material of a virus may be DNA or RNA. If it is RNA, it will have to be converted to DNA first by the **reverse transcriptase** enzyme encoded by the viral nucleic acid. These viruses are called retrovirus.

Wahlund effect: The finding of excess homozygosity in a large sample of population consisted of several subpopulations. It is due to differences in gene frequencies in the subpopulations and purely a mathematical complication. Link to lectures on **Wahlund effect (1)**, **(2)** and a [simulation](#).

Wild type: The customary phenotype or standard for comparison. Deviants from this type are said to be mutant.

Wobble hypothesis: Hypothesis to explain how one tRNA can recognize two codons. The third base in the anticodon can pair with more than one bases. This is due to the degeneracy of the genetic code, which results in more than one triplet codes for some amino acids.

W, Z chromosomes: Sex chromosomes in species (like snakes, birds, moths) where the female is the **heterogametic** sex (WZ).

Xenopus: An amphibian (frog) who shared a common ancestor with mammals about 350 million years ago. The oldest species in which all three regions of the **MHC** are linked. Its eggs are very large and have front-to-back orientation even before they are fertilized. See [Xenopus](#) website.

Yeast: The genus *Saccharomyces* of the unicellular fungi. See [Yeast](#) website.

Yeast artificial chromosomes (YAC): An artificial chromosome created from DNA, centromere and telomere of yeast chromosomes. Heavily used in cloning of very large genomic fragments.

Zebrafish: A model organism (*Danio rerio*) to study vertebrate biology, physiology and human disease. Its high fecundity and short generation time make it useful for genetic studies as well. Another useful feature is that their fry are transparent. Hundreds of mutants resembling human diseases have been identified. Link to [Zebrafish](#) website. See also **Fugu**.

Zinc finger protein: A DNA-binding domain of a protein that has a characteristic pattern of cysteine and histidine residues that complex with zinc ions. This motif occurs in several types of eukaryotic **transcription factors**.

Zygote: A cell formed by the fusion of sperm and egg. It develops to become an **embryo**.

Some laws and principles in evolutionary biology

Allen's Rule: Within species of warm-blooded animals (birds + mammals) those populations living in colder environments will tend to have shorter appendages than populations in warmer areas.

Allometry Equation: Most lines of relative growth conform to $y = bx^a$ where y and x are the two variates being compared, b and a are constants. The value of a , the allometric exponent, is 1 one the growth is isometric; allometry is said to be positive when $a > 1$ and negative when $a < 1$.

Biejernik's Principle (of microbial ecology): Everything is everywhere; the environment selects.

Bergmann's Rule: Northern races of mammals and birds tend to be larger than Southern races of the same species.

Coefficient of Relatedness: $r = n(0.5)^L$ where n is the alternative routes between the related individuals along which a particular allele can be inherited; L is the number of meiosis or generation links.

Cope's 'law of the unspecialized': The evolutionary novelties associated with new major taxa are more likely to originate from a generalized member of an ancestral taxon rather than a specialized member.

Cope's Rule: Animals tend to get larger during the course of their phyletic evolution.

There is a gradient of increasing species diversity from high latitudes to the tropics (see [New Scientist](#), 4 April 1998, p.32).

Fisher's Fundamental Theorem: The rate of increase in fitness is equal to the additive genetic variance in fitness. This means that if there is a lot of variation in the population the value of S will be large.

Fisher's Theorem of the Sex Ratio: In a population where individuals mate at random, the rarity of either sex will automatically set up selection pressure favouring production of the rarer sex. Once the rare sex is favoured, the sex ratio gradually moves back toward equality.

Galton's Regression Law: Individuals differing from the average character of the population produce offspring, which, on the average, differ to a lesser degree but in the same direction from the average as their parents.

Gause's Rule (competitive exclusion principle): Two species cannot live the same way in the same place at the same time (ecologically identical species cannot coexist in the same habitat).

This is only possible through evolution of niche differentiation (difference in beak size, root depths, etc.).

Haeckel's 'Biogenetic Law': Proposed by Ernst Haeckel in 1874 as an attempt to explain the relationship between ontogeny and phylogeny. It claimed that ontogeny recapitulates phylogeny, i.e., an embryo repeats in its development the evolutionary history of its species as it passes through stages in which it resembles its remote ancestors (embryos, however, do not pass through the adult stages of their ancestors; ontogeny does not recapitulate phylogeny. Rather, ontogeny repeats some ontogeny - some embryonic features of ancestors are present in embryonic development (L. Wolpert: *The Triumph of Embryo*. Oxford University Press, 1991). Also discussed in detail with original pictures by Haeckel in D Bainbridge: *Making Babies*. Harvard University Press, 2001).

Haldane's Hypothesis (on recombination and sex): Selection to lower recombination on the Y-chromosome causes a pleiotropic reduction in recombination rates on other chromosomes [hence, the recombination rate is lower in heterogametic sex such as males in humans, females in butterflies].

Hamilton's Altruism Theory: If selection favoured the evolution of altruistic acts between parents and offspring, then similar behaviour might occur between other close relatives possessing the same altruistic genes, which were identical by descent. In other words, individual may behave altruistically not only to their own immediate offspring but to others such as siblings, grandchildren and cousins (as happens in the bee society).

Hamilton's Rule (theory of kin selection): In an altruistic act, if the donor sustains cost C, and the receiver gains a benefit B as a result of the altruism, then an allele that promotes an altruistic act in the donor will spread in the population if $B/C > 1/r$ or $rB - C > 0$ (where r is the relatedness coefficient).

Hardy-Weinberg Law: In an infinitely large population, gene and genotype frequencies remain stable as long as there is no selection, mutation, or migration. In a panmictic population in infinite size, the genotype frequencies will remain constant in this population. For a biallelic locus where the allele frequencies are p and q:

$p^2 + 2pq + q^2 = 1$ (see [Notes on Population Genetics](#) for more).

Selection Coefficient (s): $s = 1 - W$ where W is relative fitness. This coefficient represents the relative penalty incurred by selection to those genotypes that are less fit than others. When the genotype is the one most strongly favoured by selection its s value is 0.

Selection Differential (S) and Response to Selection (R): Following a change in the environment, in the parental (first) generation, the mean value for the character among those individuals that survive to reproduce differs from the mean value for the whole population by a value of (S). In the second, offspring generation, the mean value for the character differs from that in the parental population by a value of R which is smaller than S. Thus, strong selection of this kind (directional) leads to reduced variability in the population.

Heritability: the proportion of the total phenotypic variance that is attributable to genetic causes:

$h^2 = \text{genetic variance} / \text{total phenotypic variance}$

Natural selection tends to reduce heritability because strong (directional or stabilizing) selection leads to reduced variation.

Lyon hypothesis: The proposition by Mary F Lyon that random inactivation of one X chromosome in the somatic cells of mammalian females is responsible for dosage compensation and **mosaicism**.

Muller's Ratchet: The continual decrease in fitness due to accumulation of (usually deleterious) mutations without compensating mutations and recombination in an asexual lineage (HJ Muller, 1964). Recombination (sexual reproduction) is much more common than mutation, so it can take care of mutations as they arise. This is one of the reasons why sex is believed to have evolved.

Protein clock hypothesis: The idea that amino acid replacements occur at a constant rate in a given protein family (ribosomal proteins, cytochromes, etc) and the degree of divergence between two species can be used to estimate the time elapsed since their divergence.

Red Queen theory: An organism's biotic environment consistently evolves to the detriment of the organism. Sex and recombination result in progeny genetically different from the previous generations and thus less susceptible to the antagonistic advances made during the previous generations, particularly by their parasites.

Tangled Bank Theory. An alternative theory to the **Red Queen theory** of sex and reproduction. This one states that 'sex and recombination' function to diversify the progeny from each other to reduce competition among them (link to an [abstract](#) by Burt & Bell on the Tangled Bank Theory).

van Baer's Rule: The general features of a large group of animals appear earlier in the embryo than the special features.

Weismann's hypothesis: Evolutionary function of sex is to provide variation for natural selection to act on (link to an [abstract](#) of a review by Burt discussing Weismann's hypothesis).

[PBS Evolution](#) [On Line Biology Book - Glossary](#) [Glossary of Genetic Terms](#) [Talking Glossary \(Genetics\)](#) [Life Science Dictionary](#) [Cytogenetics Glossary](#)
[UCMP Glossary \(Evolution\)](#) [BTO \(Genetics\)](#) [Population Genetics Glossary](#) [Molecular Biology Glossary \(ASH\)](#) [Molecular Biology Glossary \(UM\)](#)
[More Human Genetics Glossaries](#)

Compiled by [M.Tevfik Dorak, B.A. \(Hons\), M.D., Ph.D.](#)

Last updated on March 12, 2004

[Back to Evolution](#) [Back to Genetics](#) [Back to HLA Homepage](#) [Back to MHC](#) [Back to Biostatistics](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA Glossary](#) [Back to MHC Homepage](#) [Back to Biostatistics](#)

CHROMOSOMES and GENES

M.Tevfik Dorak, MD, PhD

CHROMOSOMES

Descriptive and comparative statistics

* Number of chromosome pairs: humans 23; gorilla 24; cattle 30; dog 39; mouse 20; goldfish 47; tobacco plants 24; peas 7; Drosophila 4; Parascaris (a nematode roundworm) 1; S.cerevisiae 16; Arabidopsis thaliana: 5; hermit crab (Eupagurus) 127; some types of fern >250. In Muntiacus muntjac (a small SE Asian deer), the number of chromosomes differs between species: the Chinese subspecies has a haploid number of 23 (like humans) but the Assam subspecies has only 3 pairs of chromosomes. In C.elegans (a nematode), the sexes differ in their chromosome

numbers: the male is haploid for the sex chromosome (X,O) and the female is diploid (X,X) resulting in a total of 11 diploid chromosomes in males and 12 in females. Note that plants do not have sex chromosomes.

* Chromosomes differ in their sizes. The smallest human chromosome is chromosome 21 (50 Mb) and the largest one is chromosome 1 (263 Mb). This is one reason why Down's syndrome (trisomy 21) is the most common trisomy, which is not a tolerable condition (See [Human Chromosome Maps](#)).

Definitions (see [Glossary](#))

The word 'chromosome' means colored body. This naming is due to the capacity of chromosomes to take up histological stains more effectively than other cell structures. Chromosomes are usually (in the interphase) dispersed throughout the nucleus but become compacted during metaphase of cell division. This is the state the chromosomes are depicted. At this stage, they are also replicated as sister chromatids (the arms of the X shape). This is different from the pair of homologous chromosomes, which represents the chromosomes inherited from the father and the mother. The point the two sister chromatids join together is called centromere, and the ends of chromosomes are called telomere. Telomeres have important functions such as preventing end-to-end fusion of chromosomes, assisting with chromosome pairing in meiosis, and ensuring complete replication of chromosome extremities. The staining pattern of each chromosome is unique and helps to identify individual chromosomes (along with the size). The densely stained bands (with Giemsa) are called G-bands, which correspond to AT-rich segments of DNA. Lightly stained bands are R-bands that are GC-rich and transcriptionally more active.

Haploid (n) number is the number of chromosomes in germ cells (23 in humans), diploid ($2n$) number is the number of chromosomes in somatic cells (46 in humans).

Chromosomes vary in shape; they may be metacentric (the arms are equal in size), submetacentric when centromere is off center, and acrocentric if centromere is close to the end.

Extra-chromosomal (cytoplasmic) DNA is still called "nucleic acid". Mitochondrial DNA (in mammals) is inherited only through the maternal lineage (see [mtDNA](#)).

Physical (kbp, Mbp) distance is the number of base pairs between two loci but genomic distance (cM) is the recombination fraction between two loci. Generally 1 Mbp corresponds to 1 cM but this varies hugely depending on the part of the genome. The human genomic average is 0.89 cM per 1 Mbp.

Chromosomal aberrations may be structural and numerical (discussed in [Clinical Genetics](#)).

Cell division: mitosis (in somatic cells) and meiosis (in germ cells)

** Key points about meiosis: it halves the number of chromosomes per cell and it gives rise to new gene combinations (via crossing-over within the chromosomes and chromosomal re-assortment). In mitosis, totally identical two daughter cells are formed (as in asexual reproduction).*

Mendel's first principle, *segregation*, is the direct result of the separation of homologous chromosomes during anaphase I of meiosis. Mendel's second principle, *independent assortment*, occurs because each pair of homologous chromosomes line up at the metaphase plate in meiosis I independently of all other pairs of homologous chromosomes. This results in a brand new set of mixture of paternal and maternal origin chromosomes each one of which may have undergone rearrangement.

Sex chromosomes X and Y are the 23rd pair in humans. There are two Xs in females but only a single X in males, whereas the autosomal chromosomes are present in duplicate in both sexes. The presence of a single autosome (a monosomy) is invariably an embryonic lethal event but monosomy for the X chromosome is viable because of

dosage compensation, which assures equality of expression of most X-linked genes in females and males. In mammals, the dosage compensation system involves silencing of most of the genes on one X chromosome; it is called X chromosome inactivation (Lyonisation). Divergent sex chromosome pairs are thought to have evolved from homologous autosomes. During evolution, the Y chromosome has retained little coding capacity, leaving the male with reduced gene dosage for many functions encoded by the X chromosome.

Human [sex chromosomes](#) have homologous region at the tips of their short and long arms. These are called [pseudoautosomal regions](#) (PAR). PAR-1 is at the tip of the short arms, and PAR-2 is at the tip of the long arms. PAR-1 consists of about a quarter of Xp and almost all of Yp (2.6Mbp). The smaller Xq/Yq pseudoautosomal region (PAR-2) is 320kb. It is believed that this region is duplicated onto the Y chromosome (from X) during primate evolution as a terminal interchromosomal rearrangement. [X-linked pseudoautosomal Hodgkin's disease](#) has a susceptibility locus within PAR-1, probably [MIC2](#) encoding [CD99](#). The blood group [Xg\(a\)](#), which behaves like an X-linked dominant trait, is also encoded within PAR-1. Polymorphism at the Xg locus and the Yg locus shows similar allele frequencies. This could be due to chance, to selection, or to recombination between the X and Y chromosomes ([Burgoyne PS, 1982](#)). The genes within PAR on X chromosome are not subject to inactivation by Lyonisation. This escape from inactivation results in an equal dosage of expressed sequences between the X and Y chromosomes. Despite morphological dissimilarity, human sex chromosomes pair also in *male* meiosis and a single obligatory recombination event takes place in the short arm pseudoautosomal region (PAR-1). The crossover point is at variable locations but mostly in the terminal third of the Xp/Yp pairing segment. Recombination at male meiosis in the terminal regions of Xp and Yp is up to 20-fold higher than between the same regions of the X chromosomes during female meiosis. The overall recombination fraction per unit of physical distance within PAR is 3- to 70-fold greater than the genome-average rate ([Lien S et al, 2000](#)). Thus, in this region LD exists only in short (~3kb) fragments ([May CA et al, 2002](#)). The consequences of the obligatory recombination within PAR-1 are that genes show only partial sex linkage and are passed equally to XX and XY offspring by male carriers. Another consequence is that a mutation favourable in males but disadvantageous in females will increase in frequency on the Y chromosomes, while remaining rare on the X chromosomes, only if the recombination rate is smaller than the fitness advantage of the mutation. The high recombination activity of the pseudoautosomal region at male meiosis sometimes results in unequal crossover, which can generate various sex-reversal syndromes (such as [XX male syndrome](#) and maybe [XY female type gonadal dysgenesis](#)). Interleukin-9 receptor ([IL9R](#)) gene is located at Xq28 and Yq12 and was the first gene to be mapped to the PAR-2. For reviews on PAR, see [Rappold GA, 1993](#) and [Meller & Kuroda, 2002](#).

Chromosome abnormalities

Chromosome abnormalities may be numerical (aneuploidy: monosomy or trisomy) or structural: deletion, inversion (pericentric or paracentric), translocation, duplication, isochromosome, ring chromosome etc. In general, detection of a structural anomaly in a child should trigger chromosome analysis of parents to rule out a carrier state but numerical anomalies are presumed to be due to sporadic cell division errors. Maternal age effect is seen in trisomies due to nondisjunction (whereas paternal age effect is more relevant in conditions due to de novo mutations). Risk of having an offspring with a chromosomal anomaly for a parent with balanced (pericentric fusion type) translocation (Robertsonian) depends on which parent is a carrier: if the mother has it, the risk is 8%;

it is 4% when the father has it. Disorders caused by chromosomal deletions are clinically more severe than those caused by duplications.

GENES

Descriptive and comparative statistics

* All of the DNA in one cell measure about 1.7m

* Estimated number of structural genes: humans 30,000; mouse 30,000 (all will be sequenced by 2005); Drosophila 13,600 (complete sequencing has been finished, see Science March 24, 2000). The yeast *S.cerevisiae* has 6,000, the bacteria *E. coli* has 4,377, and the nematode (roundworm) *C.elegans* has 19,000 genes.

* Total genome size: 3,000 Mbp in humans; 100 Mbp in *C.elegans*; 12.05 Mbp in *S.cerevisiae*; 4.64 Mbp in *E.coli*, 1.83 Mbp in another bacteria *Haemophilus influenzae* (the first fully genome sequenced free-living organism); 1.045 Mbp in the parasitic bacterium *Chlamydia trachomatis*; 130-140 Mbp in *A. thaliana* (see [Genome Sizes](#)).

* *Mycoplasma genitalium* has the smallest known genome capable of independent replication. It has 517 genes.

* The largest gene identified so far is the dystrophin gene (responsible for [Duchenne's muscular dystrophy \(DMD\)](#)). It is 2.4 Mbp; has 80 coding regions and encodes only a 3,700 amino acid-long protein. This is one reason why it has a very high mutation rate (see below). In comparison, insulin gene is 1.43 kb with three coding regions and the final product is 51 amino acid-long.

Definitions (see [Glossary](#))

Dominant, recessive, co-dominant, incomplete dominant.

Transcription, translation [central dogma of genetics; semi-conservative replication].

Intron, exon: introns end with the dinucleotide ApG [3' splice site / acceptor] and start with the dinucleotide GpT [5' splice site / donor].

Untranslated regions (UTRs): These are the regions flanking translated part of a gene. They are transcribed (represented on mRNA) but not translated (do not exist in the peptide product). The 5' UTR is usually the initial part of exon 1 and 3' UTR is the latter half of the last exon.

[Beadle and Tatum's](#) original 1941 hypothesis predicting one gene - one enzyme had to be revised first as one gene - one polypeptide; and finally one gene - multiple polypeptides. This is because alternative splicing can create multiple polypeptide products with differing activities. Other mechanisms that create more than one product from a single gene include overlapping genes and bidirectional genes ([examples](#)).

Wobble hypothesis - degeneracy / redundancy of the genetic code [for example, arginine, leucine and serine are each encoded by six different triplets].

[Mutations](#), imprinting, penetrance (all patients with ankylosing spondylarthropathy have HLA-B27 but only 2-3% of the population have the same genetic marker which is an example of low penetrance), mosaicism (Lyonisation, X inactivation), methylation, epistasis are important concepts in understanding classic and nonclassic genetic phenomena.

Trinucleotide repeats and genetic anticipation (see [Clinical Genetics](#)).

Gene expression

a) Ubiquitous: Housekeeping genes, most metabolic enzymes, ribosomal proteins, actin, tubulin, HLA class I and beta-2 microglobulin.

b) Tissue-specific: Myoglobin, gamma-globulin, TCR, HLA class II, growth hormone and other hormones.

DNA is a negatively charged **acidic** molecule because of the phosphate groups. Each of the purine or pyrimidine **bases** is a nitrogenous base.

Terminology: **ApT** (phosphodiester - covalent bond) on the same strand vs **AT** (hydrogen bond) the base pair on different strands

Start codon: AUG codes for methionine. It does not necessarily mean that each polypeptide starts with Met because most of the time it is eliminated by post-translational modifications. Rarely the start codon is GUG and encodes valine.

Stop codons: UAA, UAG or UGA do not code for an amino acid. A nonsense mutation creates one of these codons. **Redundancy (or degeneracy) of the genetic code** also applies to the stop codon. Both AUG and UGA code for stop. Any mutation creating a triplet of one of the stop codons is called a chain-terminating or **nonsense mutation**. Other types of mutations are silent (**synonymous**) mutation [the new triplet still codes for the same amino acid due to the redundancy of genetic code], and **non-synonymous** ones: **missense** mutation and **frameshift** mutations (insertion or deletion of one or more nucleotides).

At the end of the transcription, the resultant mRNA contains leader sequence, coding region and a trailer sequence.

CpG dinucleotide islands are often located at 5' of genes. A common point mutation is a transition-type substitution between the two pyrimidines, C and T (35-50% of all point mutations). Cytosine, when linked to a guanine (CpG), is often methylated. 5-methylcytosine (C) is unstable and when deaminated yields thymine (T). CpG is therefore replaced by TpG. Such a mutation is shown as (C245T) or (245C>T), the number showing the position of the nucleotide change relative to the transcription initiation site ([Human Gene Mutation Nomenclature](#), [Mutations](#) in [Molecular Cell Biology](#)).

The number of hydrogen bonds between G and C is three, but between A and T, it is two. High GC content makes the DNA more stable and gives it a higher melting point (T_m).

[More on Gene Expression](#)

[DNA replication](#)

Steps involved in DNA replication:

1. Identification of the origin of replication (not well-characterized in mammalian cells),
2. Unwinding of double stranded DNA to provide a single-stranded DNA template by helicase and topoisomerase,
3. Formation of the replication fork,
4. Initiation of DNA synthesis and elongation (by primase and DNA polymerase) during which single-strand binding proteins prevent premature reannealing of DNA,
5. Ligation of the newly synthesized DNA segments by ligase.

DNA replication proceeds from 5' end to 3' end corresponding to N-terminal to C-terminal of the subsequent protein. Translation of the mRNA is initiated by the interaction between eukaryotic initiation factors 4F (eIF4F; mRNA cap-binding protein) and 7-methylguanosine (m7G) cap on 5' mRNA ([Translation Initiation Book Chapter](#)).

Termination of translation is achieved by recognition of stop codons by termination/release factor ([Translation Regulation Book Chapter](#)).

Germline Gene Mutation Rates

Mutation rate is expressed as the number of new mutations per locus per generation; it is estimated as the incidence of new, sporadic cases of an autosomal dominant or X-linked disease that is fully penetrant such as [achondroplasia](#). The new mutation rate ranges between 10^{-4} to 10^{-7} with a median 10^{-6} . The factors influencing the mutation rate are the gene size, mutational mechanism, presence of hotspots (methylated CpG nucleotides). Being very large ones, [Duchenne's muscular dystrophy \(DMD\)](#) and [neurofibromatosis](#) genes have very high mutation rates. The reason for very high mutation rate in [achondroplasia](#) is due to a hotspot causing the G380R (Gly308Arg) mutation (nucleotide 1138G>A) in [fibroblast growth factor receptor-3](#). Other diseases

due to high germline (de novo) mutation rate are: [Rett syndrome](#); [congenital adrenal hyperplasia](#); [Rubinstein-Taybi syndrome](#); [Marfan syndrome](#).

DNA repair

Mammalian DNA polymerase ϵ is capable of proofreading of newly synthesized DNA. Both DNA polymerase ϵ and β (corresponding to E.coli DNA polymerase II) can repair DNA. Mechanisms of DNA repair:

1. Mismatch repair: copying errors (single base mismatching or two to five base unpaired loops) can be corrected by strand cutting, exonuclease digestion and replacement. Mutations of human mismatch repair genes [MSH2](#), [PMS1](#) and [PMS2](#) are related to hereditary nonpolyposis colon cancer ([HNPCC](#)).
2. Base excision-repair: Spontaneous or induced point mutations can be corrected by base removal and replacement.
3. Nucleotide excision-repair: An approximately 30-nucleotide oligomer can be removed and replaced (cut-and-patch repair).
4. Double-strand break repair: Ionizing radiation, chemotherapy and oxidative free radicals are responsible for these breaks which can be repaired by unwinding, alignment and ligation.

Some aspect of DNA repair mechanisms is deficient in inherited diseases: [Xeroderma pigmentosum](#), [ataxia-telangiectasia](#), [Fanconi anemia](#), [Bloom syndrome](#), and [Cockayne syndrome](#).

See also [DNA Repair Mechanisms](#) in the [Molecular Biology Web Book](#).

The most frequent single base alteration is deamination of cytosine to uracil. With corrective action this results in a C to T as well as a G to A point mutation on the other strand. Cytosine, when linked to guanine (CpG), is often methylated. 5-methyl-cytosine is unstable and when deaminated yields an Uracil. This is corrected to a thymine (a C to T mutation). When this strand replicates, at the residue corresponding to this T, now there is an adenine instead of guanine (G to A mutation). In general, transition type substitutions (between C and T, or G and A) are more common than transversion type substitutions (between purine 'A/G' and pyrimidine 'T/C' nucleotides). C to T transversion type of mutation within the [MSH2](#) gene causing HNPCC is an example of this type of mutation with clinical relevance. See also [Mitochondrial DNA](#).

For more, see the [Gene Transcription](#) chapter in the [Molecular Biology Web Book](#) and [Molecular Structure of Genes and Chromosomes](#) in [Molecular Cell Biology](#).

[A review on history of human cytogenetics \(Trask BJ. Nat Rev Genet 2002\)](#)

[Genes & Chromosomes Lecture Notes](#) [Chromosome Abnormalities](#)

[Gene Expression](#) [DNA From the Beginning](#)

[Chromosome Basics](#) [Chromosomal Abnormalities Tutorial: \(1\) \(2\)](#) [Chromosome Analysis](#)

[Cytogenetics & Cell Genetics \(Journal\)](#)

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on March 8, 2004

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

GENE EXPRESSION

M.Tevfik Dorak, MD, PhD

Genes are transcribed from 5' to the 3' of the **sense strand** via RNA polymerases. It is actually the **antisense or template strand**, which is transcribed (3' to 5') and gives a strand identical to the sense strand. It is possible that a gene is encoded on the sense strand and another one on the anti-sense strand in opposite direction ([example](#)). Also possible is the overlapping genes which are frequent in viruses and plasmid/phages. See also [Michael W. Pfaffl's Page](#) for an excellent review of gene structure & expression as well as molecular genetic methods to quantitate gene expression, including microarray methods, and videos.

Regions relevant in gene expression

Enhancers: A sequence on either side of the gene (cis-acting = on the same chromosome) that stimulates a specific promoter. It is not transcribed.

Promoters: A sequence(s) in the close vicinity of the transcription initiation site 5' (upstream) to the gene. It may be a general (cis-acting) or tissue/cell-specific one (cis-, or trans-acting = on a different chromosome). Initial binding site for RNA polymerase. Transcription factors bind to the promoters and allow RNA polymerase to act. The promoter is not transcribed itself. The common promoters, TATA and CAAT boxes, are found about 30 bp and 75 bp, respectively, upstream of the transcription initiation site.

Transcription initiation (cap) site: This is where the transcription of DNA to immature (precursor) pre-mRNA (nuclear RNA or nRNA) starts. It is immediately 5' to the gene; the beginning of the first exon. This sequence adds a 7-methylated GTP (7 mG) cap to the beginning of the mRNA (to protect it against the activity of 5'-exonuclease). From here to the translation initiation site, the sequence codes for the **5'-untranslated or UT (ribosome-binding) region** and the **signal peptide**. The 5'-UT region is transcribed but not translated. It contains the site (the **leader sequence**) at which ribosomes initially bind to mRNA to start translation. The **signal sequence** is translated at the N-terminal and directs the protein to its correct cellular location (endoplasmic reticulum, Golgi apparatus, cell membrane, etc) or outside the cell through the cell membrane, and is finally removed at the final destination. The events that occur to mRNA before it leaves the nucleus are collectively called RNA processing or [post-transcriptional modification](#) (capping, polyadenylation and splicing). A large number of DNA binding proteins, collectively called transcription factors, regulate cell and tissue specificity of

gene expression through their influence on RNA polymerase II-mediated transcription. An ever increasing number of diseases are attributed to mutations in transcription factor genes ([Semenza GL, 1998](#)). In general, germline mutations in transcription factor genes result in malformation syndromes and somatic mutations involving many of the same genes contribute to carcinogenesis. Some of the better known transcription factors are: c-JUN, c-MYC, CBP, CREB, E2F1, STAT, SRY, PIT-1, ETS-1, [RUNX-2/AML-3](#), [GATA-1-6](#), [PBX2](#), [RFXAP](#) and [nuclear factor kappa-B \(NFkB\)](#). The activity of transcription factors are also regulated by coactivators and corepressors. Some of the diseases caused by mutations in transcription factors are: [Rubinstein-Taybi syndrome](#), [vitamin D-resistant rickets type IIA](#), [bare lymphocyte syndrome type II](#), [acute lymphoblastic leukemia](#), [acute myeloid leukemia](#) and [Down's syndrome-associated acute megakaryoblastic leukemia](#). Transcription factors linked to the cell cycle, [RB1](#) and [P53](#), play major roles in neoplastic development. They are involved in [retinoblastoma](#), [osteogenic sarcoma](#) and [Li-Fraumeni syndrome](#) and other familial cancer syndromes ([Levine et al, 1991](#)).

Translation initiation site (ATG): This sequence represents the beginning (N-terminal) of translated protein. [5' of DNA codes for N-terminal of a polypeptide]. It codes for a methionine but methionine is subject to post-translational elimination most of the time. Thus, each mature mRNA's first codon is for methionine (AUG) but not all polypeptides start with methionine. Translation takes place in the ribosome in the cytoplasm. According to the official Human Gene Nomenclature rules the 'A' nucleotide of the ATG is nucleotide number +1, and all other sequence variation should be numbered using this nucleotide as reference. The nucleotide 5' to +1 is numbered -1; there is no base 0 [see [Nomenclature Page](#)].

Exon-intron boundaries: Each intron starts with GpT and ends with ApG. The introns are subject to splicing out (post-transcriptional modification which also includes 5' capping and 3' polyadenylation). The highly conserved intronic 5'GT and 3'AG sequences are essential for correct splicing. Although the introns are not represented in the resultant polypeptide, they may contain some regulatory sequences. An example of this is the intron 35 of the MHC class III gene C4 (complement component 4), which contains promoter activity for the gene lying next to it CYP21A2 (21-hydroxylase).

Stop codon: One of the three codons marks the end of transcription. The triplet before the stop codon codes for the last amino acid of a polypeptide chain (C-terminal). [3' of DNA codes for C-terminal of a polypeptide.]

Untranslated Regions (UTRs): 5' UTR usually contains gene- or developmental stage-specific and common regulators of expression (motifs, boxes, response or binding elements), and 3' UTR is also involved in gene expression although it does not contain well-known transcription control sites. 3' UTR sequences (called cytoplasmic polyadenylation elements or adenylation control elements) can control the nuclear export, polyadenylation status, subcellular targeting and rates of translation and degradation of mRNA. The involvement of 3' UTR is well documented in controlling male and female gametogenesis and in early embryonic development. [Myotonic dystrophy](#) is a disease caused by the expansion of the triplet repeats in the 3' UTR of a protein kinase gene, [DMPK](#).

Polyadenylation signal: This sequence is immediately after (downstream to) the stop codon and codes for a poly-A tail which varies in length. It is in the 3' untranslated region

(histone mRNAs lack poly-A tail). Poly(A) tail is believed to stimulate translation initiation whereas its shortening triggers entry of mRNA into the decay pathway.

Position effect or tissue/cell-specific expression of genes in gene therapy depends on the effects of enhancers and promoters.

The triplets on the DNA are transcribed to codons on mRNA [in the nucleus]. After splicing out intronic sequences, the codons are read by anti-codons of tRNA to be translated to amino acids. After various post-translational modifications (which may include phosphorylation, glycosylation, etc), a protein is made. Thus, the **stages of protein synthesis** are: transcription, splicing (nuclear processing), translation and post-translational modifications (such as glycosylation, deamidation, acetylation, hydroxylation, sulfation, lipidation, methylation or phosphorylation; including removal of the N-terminal methionine in most proteins). Such posttranslational changes in the molecule may play a role in disease pathogenesis despite having no genetic mutation. An example is the replacement of beta 82 Lysine by Asparagin or Aspartic acid in hemoglobin ([Charache et al, 1977](#)). Several other posttranslational deamidation in hemoglobin molecule have been reported (see [OMIM 141900](#)). In [Celiac disease](#), the deamidation of gamma-gliadin creates an epitope that acts as the self-antigen in the initiation of this autoimmune disease ([Molberg et al, 1998](#)).

See also **paramutation** (paramutation is an allelic interaction that results in meiotically heritable changes in gene expression), [methylation](#), [genomic imprinting](#) and **allelic exclusion** ([glossary](#)). Such epigenetic changes and especially their heritability are one of the very hot debates of recent years (see *Hidden Inheritance* by G Vines in *New Scientist*, 28 Nov 1998, pp.27-30; [Epigenetics: Special Issue of Science, 2001](#)). The National Fragile X Foundation website explains the molecular basis of a well-known epigenetic disease, [fragile X syndrome](#). Common techniques to measure gene expression are Northern blotting, ribonuclease protection assay, reverse-transcription (RT) and real-time PCR. These are briefly described in the [biotechnology](#) section.

See also

[Gene Transcription](#) chapter in the [Molecular Biology Web Book](#)
[Molecular Structure of Genes and Chromosomes](#) in [Molecular Cell Biology](#)
[Gene Quantification Page by MW Pfaffl](#)

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on Sept 21, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

CLINICAL GENETICS

M. Tefvik Dorak, MD, PhD

It is correct that more than 4500 inherited genetic diseases are known (see [Mendelian Inheritance Website](#) and NCBI [Genes & Diseases Online](#)) but still only 2% of human diseases can be attributed to primarily genetic causes.

Each of us inherits hundreds of genetic mutations from our parents, as they did from their forebears. In addition, the DNA in our own cells undergoes an estimated 30 new mutations during our lifetime, either through mistakes during DNA copying or cell division or, more often, because of damage from the environment. Because most of these are in somatic cells, they are not passed on to the next generation. It is also estimated that each human being is a carrier of around five recessive lethal genes.

Examples of genetic disorders

Single-gene disorders (autosomal or sex-linked): Cystic fibrosis (in the most frequent mutation, the 1480 amino acid-long wild protein is missing a phenylalanine at position 508 and becomes 1479 amino acid-long, however, more than 30 mutations causing a defective protein have been identified), [hereditary hemochromatosis](#) (C282Y missense mutation: cysteine at position 282 is replaced by a tyrosine), [myotonic dystrophy](#) and [Huntington's chorea](#) (trinucleotide repeat polymorphism with parental origin effect: stronger when it is inherited from the father; it also causes genetic anticipation due to changes in the number of repeats), [sickle cell anaemia](#) (a point missense mutation), thalassaemia ([alpha](#), [beta](#); point 'stop' mutations), haemophilia ([A](#), [B](#)), [phenylketonuria](#)-PKU (point mutation), [Duchenne's muscular dystrophy \(DMD\)](#), [adenosine deaminase \(ADA\) deficiency](#) causing severe combined immune deficiency (SCID), [Tay-Sachs disease](#) (hexosaminidase A deficiency). In [hypermobility type of Ehlers-Danlos syndrome](#), haploinsufficiency due to a 30-kb deletion of tenascin-X (TNXB) gene is responsible for the disease. In [Cri-Du-Chat syndrome](#) (5p deletion), the genetic basis of the phenotype is haploinsufficiency for the telomerase reverse transcriptase gene ([TERT](#)), which is included in the deleted part of chromosome 5. In the rare disease [erythropoietic protoporphyria](#), haploinsufficiency for ferrochelatase ([FECH](#)) contributes to the clinical phenotype but is not the only reason for the disease expression. Dominant negative mutations are involved in [osteogenesis imperfecta type I](#) and [autosomal dominant nephrogenic diabetes insipidus](#).

Diseases with multifactorial etiology (also called complex genetic disorders): Insulin-dependent diabetes mellitus (IDDM), multiple sclerosis (MS), rheumatoid arthritis (RA), cancer, autism, and schizophrenia. These diseases show familial aggregation but not strong familial segregation.

mtDNA disorders: Leber's hereditary optic neuropathy (LHON), neurologically-associated retinitis pigmentosa (NARP), myoclonic epilepsy and ragged red-fiber disease (MERRF), maternally inherited myopathy and cardiomyopathy (MMC).

Other genetic disorders: [Chromosomal abnormalities](#), DNA repair defects (genomic instability; xeroderma pigmentosum, ataxia telangiectasia, Bloom's syndrome, Fanconi's anaemia).

Chromosomal disorders (see also [Merck Manual Chapter 261](#))

a) numerical (aneuploidy): Trisomy 21 (Down syndrome; 47, +21), Trisomy 13 (Patau syndrome; 47, +13), Trisomy 18 (Edwards syndrome; 47, +18), Monosomy X (Turner syndrome; 45, X), Klinefelter syndrome (47, XXY)

b) structural: deletion (Di George syndrome: del 22; Cri-du-chat syndrome: 5p-)

c) others: uniparental disomy (uniparental copy of chromosome 15q: [Prader-Willi syndrome](#) 'maternal' or [Angelman's syndrome](#) 'paternal', see below); Mosaicism (skewed X-chromosome inactivation: [colour blindness](#))

Mendelian segregation patterns (mode of transmission) in single gene disorders

(link to a [tutorial](#)):

In single gene disorders (as opposed to multifactorial-complex disorders), the population frequency is low, penetrance of the causative gene is high, and the contribution of environment is lower with notable exceptions of PKU and few others. Genetic counselling is important for personal-decision making involving reproductive issues. In assigning the inheritance pattern of a genetic disease, important features of pedigrees to consider are: sex ratio in affected individuals, male-to-male transmission, mother-to-son transmission (XLR) - mother-to-son/daughter transmission (mtDNA); proportion of offspring (proportion of siblings) affected and consanguinity. Situations that can confound the interpretation include: quasi-dominant inheritance (a homozygote and heterozygote (carrier) mating for a recessive disease); autosomal dominant sex-limited inheritance, *de novo* mutations (including new mutation on the second X-chromosome in a woman carrier for a X-chromosome mutation), phenocopy, skewed X-chromosome inactivation (atypical Lyonisation) and mosaicism.

Autosomal recessive (AR): Two unaffected (carrier) parents produce diseased children of both sexes (the risk is 25%); some degree of consanguinity is usually involved; there is a *horizontal* pattern in the pedigree. There does not have to be a diseased individual in the pedigree. Autosomal recessive disorders are usually common in populations with high level of inbreeding (restricted gene pool). Examples are [Tangier disease](#) in Tangier Island off the coast of Virginia, USA; many genetic disorders in Ashkenazi Jews (Tay-Sachs Disease, [Gaucher disease](#), Fanconi anaemia, Niemann-Pick Disease); [congenital adrenal hyperplasia](#) in Yupik Eskimos; and thalassaemias in Cyprus and Sardinia. Heterozygotes for the recessive disease genes (carriers) are usually clinically unaffected but some biochemical evidence for carrier status may be found: heterozygotes for cystinosis are always asymptomatic, but intracellular cystine levels are up to 10 times greater than the normal amounts (homozygotes have levels 100-fold greater than normal). Similarly, heterozygotes for 21-hydroxylase gene ([CYP21A2](#)) mutations are usually asymptomatic but their adrenal sex steroid hormone levels may be found higher than normal after stimulation (see [Figure 2](#) in an online review). AR diseases do not usually show variable expressivity within a given family.

Autosomal dominant (AD): Every patient has an affected parent; the risk of transmission of the disease to the offspring is 50%; no sex preference; two affected parents may have a healthy child (25% chance); homozygotes may have a more severe disease or may not exist (due to early, including embryonic, lethality, as in [acute intermittent porphyria](#)); normally there is no generation skipping resulting in a *vertical* pattern in the pedigree. If both parents appear to be normal for an AD disease but the offspring has it, the possibilities are as follows: biologic parents are different; incomplete penetrance in (affected) parent; phenocopy (the disorder is not genetically determined but mimicking one) or genocopy (having a phenotype caused by an AD disease but caused by a different genetic mechanism); *de novo* mutation; gonadal mosaicism for the

disease mutation in one parent. Examples of AD diseases are adult polycystic kidney disease, Huntington chorea, and von Willebrand disease. AD diseases are usually due to mutations in receptor proteins ([familial hypercholesterolaemia](#)) or structural proteins (haemoglobin C, procollagen). Expressivity may show gradual variation in successive generations.

X-linked recessive (XLR): Affected males transmit the gene (not necessarily the disease) to all daughters but not to sons (and to half of their grandsons); hemizygous males and homozygous females are affected, thus, it is more frequent in males; all affected females have an affected father and their mother is either an obligate carrier or affected. Female carriers of X-linked recessive diseases are generally asymptomatic but exceptions occur as in [hereditary hypophosphatemia with vitamin D-resistant rickets](#) and [Duchenne's muscular dystrophy \(DMD\)](#); in other words, XLR diseases are more severe in males. Transmission is usually from carrier females to sons, which results in a criss-cross pedigree. Sporadic XLR disease is due to spontaneous mutation (1/3) or carrier mother (2/3). If an XLR disease occurs in a female, possible explanations include: atypical Lyonisation, new mutation on the X-chromosome of a carrier woman; 45,X; Xp deletion; translocation involving the mutation on the X-chromosome and an autosomal chromosome (no inactivation of the translocated X-chromosome part).

X-linked dominant (XLD): All daughters but no son of affected males are affected (no male-to-male transmission); a heterozygous affected woman transmits the disease to half of her children with males and females at equal risk; on average, twice as many females as males will be affected. One reason contributing to this ratio is the ***in utero lethality in males*** carrying X-linked mutations (see below). If the sex ratio in affected offspring is ignored, the pedigree has a vertical appearance (like AD transmission).

Mitochondrial transmission: In this exceptional situation, mitochondrial diseases are transmitted only from mothers to both sexes equally (a review on [Clinical Mitochondrial Genetics](#) by Chinnery et al; [Spectrum of Mitochondrial Disease](#) by Naviaux). [Leber optic atrophy](#), [Kearns-Sayre syndrome](#), [diabetes-deafness syndrome](#), [MELAS syndrome](#), and [Wolfram syndrome](#) (mitochondrial form) are examples of mitochondrial DNA (mtDNA) disorders.

Common genetic disorders

Autosomal recessive (AR) diseases: Cystic fibrosis (CF), oculocutaneous albinism (OCA), hereditary hemochromatosis (HH), [congenital adrenal hyperplasia](#) (multiple mutations), sickle cell disease, beta thalassaemia (multiple mutations), AT, ADA deficiency / SCID. The important characteristics of AR diseases, ethnic group specificity and increased risk with consanguinity, are seen in most of these diseases.

Autosomal dominant (AD) diseases: Brachydactyly (the first Mendelian trait identified in humans), [Huntington's disease](#), [familial hypercholesterolaemia](#), familial polycystic disease, one type of Alzheimer's disease, [neurofibromatosis type 1](#) and [achondroplasia](#). [Acute intermittent porphyria](#) is an interesting autosomal dominant disease in that a heterozygous mutation causes a total enzyme deficiency.

X-linked diseases: Many conditions - including haemophilia A and B, G6PD deficiency, red-green colour blindness, hereditary myopia, night blindness and ichthyosis - are sex-linked traits in humans. The most common XLR disease is [DMD](#) (the DMD gene is 2.4 Mbp with 80 exons. Average exon size is 150 bp). Some other common XLR diseases are [haemophilia A](#), [haemophilia B](#), [fragile X syndrome](#), [non-specific X-chromosomal mental retardation \(MRXS3\)](#) and [X-linked Bruton agammaglobulinaemia \(BTK\)](#).

[Testicular feminisation](#) / androgen insensitivity syndrome is another X-linked trait that causes XY individuals to develop into phenotypic females. Prostate cancer has an X-linked form ([HPCX](#)).

X-linked dominant (XLD) diseases / traits: [Xg\(a\)](#) blood group (usually shown as X_g), [Vitamin D-resistant rickets \(X-linked\)](#), [incontinentia pigmenti](#) and [Rett syndrome](#) (the last two are X-linked dominant ([Xq28](#)) and therefore lethal in hemizygous males and only seen in females -see below for more examples-). The [Rett syndrome](#) gene (RTT) on [Xq28](#) encodes MeCP2. It appears that 99.5% of the mutations are new ones (*de novo*), therefore the likelihood for a second girl having the disease in the same family is no more than 0.5%. The MeCP2 mutation affects brain development in such important ways that boys die either before or shortly after birth and never have the chance to develop actual Rett syndrome. The severity of the syndrome in girls is a function of the percentage of cells with a normal copy of MeCP2 that are left to function after random X inactivation. If X inactivation happens to turn off the X chromosome carrying the mutated gene in a large proportion of cells, the symptoms will be mild. If, instead, a larger percentage of cells has the healthy X chromosome turned off, the onset could be earlier and the symptoms more severe.

Cytoplasmic DNA (mtDNA) disorders: Discussed above (a review on [Clinical Mitochondrial Genetics](#) by Chinnery et al; [Spectrum of Mitochondrial Disease](#) by Naviaux).

Nonclassical genetic phenomena

* Some autosomal dominant disorders may show **sex-limitation**. This happens when an affected male is infertile and the females do not express the disease. Then, the pedigree pattern will be identical to an X-linked recessive trait where males do not reproduce. [Male-limited precocious puberty](#) is such a disease. Females transmit it to half of their sons. The pedigree suggests a sex-linked inheritance.

* In some diseases, it has been documented that mutation rate is higher in males than in females: [Duchenne's muscular dystrophy \(DMD\)](#), [Lesch-Nyhan syndrome](#) and [haemophilia](#). In these diseases, increased **paternal age effect** will increase the risk of the disease in the offspring.

* When two affected parents have a child:

1. In autosomal recessive disorders, all children will be affected,
2. In autosomal dominant disorders, they may have a healthy child (if both are heterozygous, there is 25% chance of a healthy offspring).

* **Uniparental disomy (UPD)** is the presence of a pair of chromosomes originated from the same parent. The first step is usually the formation of a trisomy and then the loss of a chromosome (due to postzygotic nondisjunction) but gametic complementation (fusion of a gamete with two copies of the same chromosome with a gamete with none of the same chromosome) is also possible. The end result may be a homozygote child from a carrier parent. One particular UPD results in an interesting situation: Maternal UPD for chromosome 15 (15q11-q13) will result in [Prader-Willi syndrome](#), while paternal UPD for the same chromosome will cause [Angelman's syndrome](#). UPD does not have to occur for the whole chromosome. Due to chromosomal rearrangements, UPD can occur for a portion of the chromosome. [Beckwith-Wiedemann syndrome](#), for example, is sometimes due to partial UPD for chromosome 11p15 with paternal imprinting. Complete [hydatidiform mole](#) is a UPD, i.e., its genetic origin is completely paternal (even if the karyotype is 46,XX, which represents duplication of a haploid (23,X) sperm).

* **Premature centromere separation** or **heterochromatin repulsion** is another mechanism, which causes genetic disease: [Roberts syndrome](#) is an example.

* **Genetic anticipation** is seen mainly in autosomal dominant diseases ([Huntington's disease](#), [congenital myotonic dystrophy](#), many [cerebral ataxias](#)) but also in [Fragile X syndrome](#), and in a single autosomal recessive disease ([Friedreich's ataxia](#)). Curiously, genetic anticipation may also show sex-limitation. In [Huntington's disease](#), it only occurs when a male transmits the disease while the opposite is true for [congenital myotonic dystrophy](#) (**maternal effect**). Anticipation is due to trinucleotide repeat expansions, which may be in the coding region (Huntington's disease) or in the untranslated region (Fragile X syndrome). Opposite of anticipation is also true and is reported in [congenital myotonic dystrophy](#). This is called **reverse mutation**, **negative expansion** or **contraction** (see review by [Brook \(1993\)](#)).

* When a disease occurs first time in a family:

1. This is the first *-de novo-* mutation (in a disease for which the mutation has a very high rate as in [Rett syndrome](#); also occurs in [congenital adrenal hyperplasia](#), [achondroplasia](#) - seven-eighths of mutations are new due to methylated CG dinucleotides in [FGFR3](#)). *De novo* microdeletion or point mutations in the transcriptional coactivator [CREB-binding protein](#) (CBP) on [16p13.3](#) cause [Rubinstein-Taybi syndrome](#) with a recurrence risk of 0.1% in sibs. The mutation rate in the [NF1](#) gene is one of the highest known in humans, with approximately 50% of all [NF1](#) patients presenting with novel mutations. About one-quarter of affected individuals with [Marfan syndrome](#) have new mutations; thus, a **paternal age effect** is present in sporadic cases as in all genetic syndromes with a high rate of new mutations (mutation rate is expressed as the number of new mutations per locus per generation; it is estimated as the incidence of new, sporadic cases of an autosomal dominant or X-linked disease that is fully penetrant such as [achondroplasia](#). The new mutation rate ranges between 10^{-4} to 10^{-7} with a median 10^{-6}).
2. Both parents are healthy carriers of a recessive mutation (consanguinity or frequent mutations).
3. If the gene has been shown to be segregating in previous generations with no diseased individuals, the penetrance may have been low in earlier generations ([acute intermittent porphyria](#) is an example of low penetrance genetic disease). If it is an X-linked recessive disease and all the children so far are girls, the disease appears in the first male child.
4. It may be an adult-onset disease with genetic anticipation but the carriers of the gene in the previous generations did not live long enough to express the disease. [Huntington's disease](#) is an example of late-onset genetic disease with anticipation. A similar situation may arise due to **parent of origin effect**. Differences in gene expression according to the parent from whom the gene was derived occur in [Huntington's disease](#) and in [myotonic dystrophy](#), and might be due to a difference in methylation of the genes in the two sexes (see [Marx, 1988](#)) or unknown modifier genes acting in a sex-limited fashion.

* **Multifactorial inheritance:** This pattern applies to complex diseases (like diabetes, hypertension, schizophrenia, cancer) where both multiple genes and environmental factors play a role in the development of the disease. It has been best worked out in [pyloric stenosis](#). These disorders are presumed to result from additive effects of multiple genes with low penetrance. Individual mutations may not have any particular phenotype, but when act in concert and in the presence of the necessary environmental conditions, they may produce a disease phenotype. Under a model of multiple interacting loci, no single locus could account for more than a 5-fold increase in the risk of first-degree relatives. The disease shows increased incidence in families but with no recognizable inheritance pattern. The features of multifactorial inheritance are:

- The more severe the condition, the greater the risk to sibs,
- [Carter effect](#): the sibs or offspring of a patient in less commonly affected sex have higher susceptibility to the disease,
- If it is a rare disease, the frequency of the disease among relatives is higher,
- If more than one individual in a family is affected, recurrence risk is higher,
- The risk falls rapidly as one passes from 1st to 2nd degree relatives,

There are two models proposed to explain multifactorial inheritance: multifactorial threshold model (a combined effect of multiple genes interacting with environmental factors; i.e. several or many genes, each of small effect, combine additively with the effects of non-inherited factors) and mixed model (a multifactorial liability with the involvement of a major gene). See [Introductory Statistical Genetics](#). In multifactorial inheritance, the presence of even a major-disease causing allele is only predisposing rather than predictive.

From genotype to phenotype

Recessive traits usually result in defective enzymes / proteins (CAH, albinism, CF, SCID), dominants often alter structural, carrier or receptor proteins ([familial hypercholesterolaemia](#) (LDL receptor), haemoglobin C, procollagen, fibrillin, elastin).

Albinism: In autosomal recessive oculocutaneous albinism, melanocytes cannot produce enough melanin. In heterozygotes there is enough of it.

Sickle cell anaemia: In heterozygotes, there is enough HbA and no symptoms or signs of sickle cell disease.

Cystic fibrosis: In heterozygotes although the mutant CFTR (CF trans membrane conductance regulator) protein is produced, there is still enough wild type (not mutated) protein functioning properly. In homozygotes, the normal protein is totally missing.

In autosomal dominant diseases heterozygosity and homozygosity may affect the disease expression. In [Huntington's disease](#), heterozygotes and homozygotes have the same clinical phenotype whereas in [achondroplasia](#), homozygotes have a much more severe clinical course. Most X-linked dominant diseases cause ***in utero lethality in males*** ([Rett syndrome](#), [incontinentia pigmenti](#), [Goltz syndrome / focal dermal hypoplasia](#), [chondrodysplasia punctata type B](#), [orofaciodigital syndrome 1](#)).

In **haploinsufficiency** and **dominant-negative mutations**, a mutation in one copy of the gene may cause the disease phenotype despite having one copy still functional (see above for examples). These mutations will cause an autosomal dominant inheritance pattern. [Osteogenesis imperfecta](#) is due to dominant-negative mutations in type I collagen gene ([COL1A1](#)) mutations. Similarly, aquaporin-2 gene ([AQP2](#)) has a dominant-negative effect in causing [type II nephrogenic diabetes insipidus](#). Diseases caused by haploinsufficiency are characterized by the lack of a correlation between the severity of the genetic defect (for example, size of a deletion) and the phenotype. One example is [DiGeorge syndrome](#). Histone H2AX is a genomic caretaker that requires the function of both gene alleles for optimal protection against tumorigenesis. Therefore disruption of one allele causes genomic instability and tumour susceptibility due to haploinsufficiency ([Celeste et al, 2003](#)). Another example of haploinsufficiency is [PAX6](#) gene on 11p causing [aniridia type II / WAGR syndrome](#).

In many genetic diseases, the disease has a variable clinical spectrum from mild to severe and even lethal (variable expressivity). This phenotypic heterogeneity is usually due to involvement of different alleles (**allelic heterogeneity**) or genes (**locus heterogeneity**) in disease pathogenesis. [Gaucher disease](#), [heritable collagen disorders](#), [muscle diseases](#), [dominantly inherited ataxias](#), neurofibromatosis [type 1](#) and [type 2](#),

[congenital adrenal hyperplasia](#), and [alpha-1-antitrypsin deficiency](#) are examples of diseases showing phenotypic heterogeneity. Examples of locus heterogeneity include [hypertrophic cardiomyopathy](#) (CMH, IHSS) which may be caused by mutations in cardiac troponin I ([TNNI3](#)), cardiac troponin T2 ([TNNT2](#)), alpha-tropomyosin ([TPM1](#)), cardiac myosin binding protein C ([MYBPC3](#)), cardiac myosin heavy chain-alpha ([MYH6](#)) etc; [congenital adrenal hyperplasia](#) may be caused by mutations in [CYP21A2](#) or [CYP11B1](#). Although most complex diseases (such as [essential hypertension](#), [noninsulin-dependent diabetes mellitus](#), [schizophrenia](#), [prostate cancer](#), [familial glioma](#) and [inflammatory bowel diseases](#)) appear to be examples of locus heterogeneity because many loci are involved, this is not the case because 'combinations' of susceptibility alleles of various loci contribute to disease pathogenesis rather than different alleles of various loci causing the same disease.

Just like Beadle's one gene-one enzyme (protein) idea is no longer strictly true, one gene can also cause different diseases through different imprinting patterns depending on the parent of origin, or different mutations. Confusingly, this is also called **allelic heterogeneity**. Different mutations in cardiac sodium channel gene [SCN5A](#) cause either [long QT syndrome-3](#) (LQT3) or [Brugada syndrome](#) (both may cause sudden death in young healthy individuals). An interesting example of allelic heterogeneity concerns [KVLQT1](#) gene. While [Jervell and Lange-Nielsen syndrome](#) (JLNS) is due to homozygosity for a mutation in the [KVLQT1](#) gene, heterozygosity for a mutation in the same gene causes [long QT syndrome-1](#) (LQT1). [Dystrophin](#) gene mutations cause [Duchenne's muscular dystrophy](#), [Becker's muscular dystrophy](#) or [X-linked dilated cardiomyopathy](#). [Myosin VIIA](#) gene (MYO7A) mutations result in [autosomal recessive sensorineural deafness type 2](#) (DFNB2), [autosomal dominant nonsyndromic sensorineural deafness type 11](#) (DFNA11) and [Usher type 1B syndrome](#) (USH1B). It is interesting that different mutations in the DFNB2 gene may result in either dominant or recessive hearing loss ([Tamagawa Y et al. 1996](#)). Other examples: [fibroblast growth factor receptor 2](#) (FGFR2) mutations cause [Crouzon](#), [Pfeiffer](#) or [Apert](#) syndrome; mutations in L1 cell adhesion molecule gene ([L1CAM](#) / [CD171](#)) cause [X-linked hydrocephalus](#) or [MASA syndrome](#); peripheral myelin protein 22 ([PMP22](#)) mutations cause [Charcot-Marie-Tooth disease type 1A](#) or [hereditary neuropathy with liability to pressure palsies](#); cystic fibrosis gene ([CFTR](#)) mutations do not only cause [cystic fibrosis](#) but also [congenital bilateral aplasia of the vas deferens](#); heterozygosity for CFTR mutations is associated with ['idiopathic' chronic pancreatitis](#).

Population Genetics

How ever rare a genetic disease may be, the carrier state for the disease gene may be very common. Phenylalanine hydroxylase deficiency (classic PKU) has a frequency of 1 in 10,000 but a carrier frequency of 1 in 50. Similarly, the mutations of CYP21A2 ([congenital adrenal hyperplasia](#)) may be present in 1 in 3 Ashkenazi Jews ([Speiser PW et al. 1985](#)), [CFTR](#) mutations in 1 in 20, and the [C282Y mutation of HFE](#) in 1 in 10 Northern Europeans. This is one reason why eugenics approaches would never work in elimination of a disease. For each affected individual (for a recessive disease), there may be up to 100 carriers showing no symptoms of the disease.

Heterozygosity for an allele that causes an autosomal recessive disease in homozygotes is usually asymptomatic but detectable by laboratory tests. Heterozygosity for mutations causing [congenital adrenal hyperplasia](#), [hereditary hemochromatosis](#), [alpha-1-antitrypsin deficiency](#), [cystic fibrosis](#), [ataxia telangiectasia](#), [homocystinuria](#), and [hereditary fructose intolerance](#) results in biochemically detectable changes.

To explain the propagation of disease genes, compensating heterozygote advantage has been suggested for a few of them: [Phenylketonuria](#) (PKU) mutation ([Woolf et al, 1975](#); [Woolf, 1986](#); [Woo SL et al 1989](#)); cystic fibrosis mutation (CFTR) ([Pier GB et al, 1998](#)); [21-hydroxylase](#) mutations (CY P21A2) ([Witchel SF et al, 1997](#)); [HFE](#) mutations ([Datz et al, 1998](#)); [alpha-thalassaemia](#) mutation (HBA1) ([Flint J et al, 1986](#)); beta-E-thalassaemia ([haemoglobin E](#)) ([Chotivanich et al. 2002](#)); [sickle cell disease](#) mutations (HbF) ([Friedman & Trager, 1981](#)); and [glucose-6-phosphate dehydrogenase](#) mutations ([Luzzatto L et al, 1969](#)). The concept of compensating heterozygote advantage (or balancing selection) in this context was first proposed by [JBS Haldane in 1949](#).

An increased frequency of heterozygosity for [alpha-1-antitrypsin deficiency](#) in twins and parents of twins has been noted, which led to the suggestion that 'increased' fertility and twinning may be heterozygous advantages for antitrypsin deficiency ([Lieberman et al, 1979](#)). It is believed that the S allele of the PI gene may increase ovulation rate and enhance the success of multiple pregnancies ([Clark & Martin, 1982](#); [Boomsma et al, 1992](#)). Heterozygote advantage via lower miscarriage rate has been suggested in [phenylketonuria](#) ([Woolf et al, 1975](#); [Woolf, 1986](#)).
More at [Population Genetics Theory](#).

CLINICAL GENETICS LINKS

[Clinical Genetics Computer Resources](#)

Self Study Guides for Health Care Workers: [Virtual Children's Hospital](#) [University of South Dakota](#) [West Midlands NHS](#)

[CDC HuGe Case Studies](#) [National Society of Genetic Counselors](#) [J Genet Counselling](#)

[Genetic Counselling Tutorial](#) [Prenatal Diagnosis Tutorial](#)

[Tutorials in Human Genetics](#) [Online Human Genetics Course \(Lectures\)](#) [Clinical Genetics Educational Tools](#)

[Medical Genetics in Merck Manual](#) [Clinical Molecular Genetics](#) [Genetics in e-Medicine](#)
[Genetics Image Library](#)

[OMIM](#) [Genetic Syndromes](#) [Genes & Diseases Online](#) [Yahoo Database of Genetic Disorders](#)

[Gene Reviews](#) [Human Genetics & Medical Research](#)

[Medical Genetics at Utah University: Photos, Videos, Exams](#) [Turner Syndrome: Medscape CME](#)

[Nomenclature](#) [Cytogenetics Nomenclature](#) [Chromosomal Variation in Man Online Database](#)

[Virtual Children's Hospital-Clinical Genetics](#) [Information for Genetic Professionals](#)
[Human Genetics News](#)

[ACMG \(American College of Medical Genetics\)](#) [NCHPEG \(The National Coalition for Health Professional Education in Genetics\)](#)

[Standards and Guidelines for Clinical Genetics Laboratories](#)

[Human Genetics & Cytogenetics Links](#) [Clinical Cytogenetics Examples](#) [Cytogenetics Gallery](#)

[POSSUM: Clin Genet Diagnosis Software](#) [POSSUM Demo](#) [MENDEL for Pedigree Analysis](#)

[Every Scot to be Offered Gene Test](#)

[Genetics at GlaxoSmithKline](#) (educational material with animations)

Medical Genetics Special Feature by Discover Magazine: [Part 1](#) - [Part 2](#) - [Part 3](#)

Journals

[Journal of Medical Genetics](#) [European Journal of Human Genetics](#)

[Clinical Genetics](#) [Genetics in Medicine](#) [American Journal of Human Genetics](#)

[AJMG](#) [AJMG: Seminars in Medical Genetics](#) [AJMG: Neuropsychiatric Genetics](#)

[BMC Medical Genetics](#)

[Annual Review of Genomics and Human Genetics](#)

[Cochrane Reviews in Cystic Fibrosis and Genetics Disorders](#)

Books

[Medical Genetics at a Glance](#) [PDQ Medical Genetics](#) [Genetics PreTest](#)

[Medical Genetics by LB Jorde \(2003\)](#) [Genetic Medicine by B Childs \(2003\)](#)

[Thompson & Thompson Genetics in Medicine \(2001\)](#)

[Medical Genetics: Pearls of Wisdom by WG Sanger \(2002\)](#)

[A Comprehensive Primer on Medical Genetics by TF Thurmon \(1999\)](#)

[Emery and Rimoin's Principles and Practice of Medical Genetics \(2002\)](#)

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on March 11, 2004

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

CANCER GENETICS

M.Tevfik Dorak, MD, PhD

Cancer has a genetic component in its etiology. This is evident from increased risk in family members of cancer patients, twin studies, the presence of familial forms of cancer, experimental studies and classical/molecular genetic studies. In the majority of human cancers, there is no clear-cut mode of inheritance. This and other evidence suggest that cancer is a **multigenic** (oncogenes, tumor suppressor genes, MHC genes), **multifactorial** (radiation, chemicals, hormones, viruses, ultraviolet, diet, etc) and **multistep** (transformation, promotion, overt cancer) process. In sporadic tumors, there is no inherited genetic abnormality predisposing to tumor. Genetic susceptibility, however, occurs when a mutation is present in the heritable genetic material (germ-line). About 5-10% of cancers result from germ-line mutations and most are evident as cancer family syndromes.

Knudson proposed a two-hit model for the development of cancer in 1971 from the evidence in familial retinoblastoma cases. This simplified model suggests two separate genetic events in the development of retinoblastoma. The first one is inherited in the familial form and the time required for the second hit to occur is shorter than two hits to occur in a normal person. This is why the familial form occurs at an earlier age and is usually bilateral. In general, inherited cancer occurs earlier and is more severe, whereas, sporadic cancer has later onset.

The immune system is one of the factors influencing susceptibility to cancer. The evidence for this is that in immune deficiencies, there is an increased risk for cancer. These tend to be lymphoid tumors. Immunogenicity of cancer is suggested by the presence of tumor-specific cytotoxic T lymphocytes (CTL) and tumor infiltrating lymphocytes (TIL). Immunogenicity is impaired by the loss of MHC expression. The **immune surveillance theory** proposes that oncogenic transformation of normal cells is a frequent occurrence but the immune system clears them as they emerge. This theory has not been convincingly proven yet. The immune surveillance theory implies that homozygosity for the MHC molecules would be deleterious. Several modes of cancer treatment are based on the role played by the immune system (in vitro activation of CTLs and TILs, vaccination with tumor-derived peptides and even gene therapy using TILs).

Cancer is a combination of uncontrolled cellular proliferation and immortality (lack of apoptosis). Cell division is regulated by growth factors, their cell surface receptors, membrane tyrosine kinase, signaling molecules (GTP-binding proteins), nuclear/transcription factors (the signal transduction system) and growth regulatory

(inhibiting) factors. The genes with growth promoting activity are generally called **oncogenes** (dominant) and those with growth inhibitory activities are **tumor suppressor genes** (recessive). **Proto-oncogenes** may encode surface membrane proteins ([HER2/neu/erbb2](#)), signal transduction pathway molecules ([ras](#)) or transcription factors. Tumor suppressor genes encode cell cycle regulators, adhesion molecules ([APC](#)), DNA repair enzymes ([hMSH2](#), [hMLH1](#)) or signal transduction pathway molecules.

Genes involved in oncogenesis

The human genome contains more than 50 genes (cellular oncogenes, c-onc) that are similar to the genes carried by carcinogenic retroviruses (viral oncogenes, v-onc). The cellular counterparts of v-onc are normal cellular genes coding for proteins playing roles in normal cell growth and division. They are normally called **proto-oncogenes** but when activated they show their oncogenic effects (**oncogenes**). Examples of oncogenes include *sis* which encodes a chain of platelet derived growth factor (PDGF); *int-2* have similarities to fibroblast growth factor (FGF); [c-erb-B](#) (on chromosome 7p) which encodes a truncated form of the receptor for epidermal growth factor ([EGFR](#)); mutant *ras* genes have a reduced capacity to terminate a growth stimulating signal; *abl* and *src* have tyrosine kinase activity; *c-myb* and *c-myc* are stimulators of cell cycle; and *bcl-2* blocks apoptosis and promotes cell survival. One of the earliest phenomena in tumor formation is **genomic instability**. It is due to defects in DNA repair and cell cycle controls. This can happen by gain-of-function mutations in proto-oncogenes or loss-of-function mutations in tumor suppressor genes.

a. [Proto-oncogenes](#) may be activated to act like oncogenes by the following events:

1. Amplification: *Int-2* and *c-erb-B2* in breast cancer; *N-myc* in neuroblastoma.
2. Insertional mutagenesis: *c-myc* (nuclear DNA-binding phosphoproteins) activation by EBV in Burkitt's lymphoma (*Int3* activation by MMTV in mouse mammary tumors). *c-myc* may be activated by rearrangement, amplification or overexpression.
3. Chromosomal translocations: In Burkitt's lymphoma, *c-myc* (8q24), in 85% of follicular lymphoma, *bcl-2* (18q21), and in mantle-cell lymphoma *bcl-1* (11q13) translocate to the immunoglobulin heavy chain gene region on chromosome 14q32 (subsequently these oncogenes are over expressed); in chronic myeloid leukemia, the *bcr* gene moves next to *c-abl* [t (9;22) (q34; q11)] resulting in the expression of a fusion protein. In M3 AML (acute promyelocytic leukemia) *RAR α* / *PML* genes are rearranged by translocation t(15,17).
4. Mutations in coding sequences: Activating point mutations occur in the *ras* genes in about 30% of human cancers. *ras* gene family consists of N-*ras*, H-*ras* and K-*ras* whose products are involved in intracellular transduction of external stimulation of growth factor receptors. Mutated RAS proteins are expressed but functionally appear to have lost the ability to be inactivated. In general, mutations changing the activity of a gene may be missense, nonsense or frameshift type.

Gene expression and function can also be affected by non-DNA base pair changes, namely epigenetic effects:

5. Demethylation or hypomethylation: *bcl-2* overexpression in chronic lymphoid leukemia, *c-erb-B1* (epidermal growth factor receptor) overexpression in breast cancer.

A number of genes involved in regulation of the cell cycle, *cyclins*, may also act as oncogenes when mutated (for example, cyclin D2 in mantle cell lymphoma and cyclin D1

and E in breast cancer). On the other hand, germ-line mutation of a cyclin-dependent kinase inhibitor *CDKN2/p16* has been implicated as a tumor suppressor gene in hereditary melanoma.

b. Tumor suppressor genes (TSGs) act like recessive anti-oncogenes. Theoretically, both copies of a TSG must be lost or inactivated for oncogenesis through loss-of-heterozygosity events (unlike a single mutation in proto-oncogenes). There are, however, exceptions to this expectation. The best-known TSG is *p53* (on chromosome 17p) and mutation of a single copy of the two copies is enough for the deleterious effect. This is because mutant *p53* protein monomers are more stable than the normal *p53* proteins and can form complexes with the normal wild type *p53* acting in a dominant-negative manner to inactivate it. Therefore, one mutated copy is enough for the loss of whole *p53* function. The tumor suppressor gene *p53* is the guardian of the genome because it prevents progress through the cell cycle when there is something wrong to allow the damage or fault to be repaired. Because mutation of a single allele (single hit) is enough for neoplastic transformation, *p53* mutations are the most frequent genetic abnormalities found in human cancers (over 50% in bladder, breast, colon and lung cancers). Different mutations differentially occur in specific cancers (like the codon 249 mutation in aflatoxin-related liver cancer). Inherited (germ-line) *p53* mutation is the cause of a well-known inherited cancer syndrome ([Li-Fraumeni syndrome](#); childhood sarcomas, early-onset familial breast cancer and other neoplasms). Most familial cancers are related to defects in TSGs.

Other TSGs are the retinoblastoma (*RB1*) gene on chromosome 13q, the deleted in colorectal cancer (*DCC*) gene on chromosome 18q, and the Wilm's tumor gene (*WT1*) on chromosome 11p13. In these cases, the loss of both alleles is required for neoplastic transformation (via mutation or deletion). The *RB1* gene encodes a nuclear protein that is involved in the regulation of the cell cycle (it suppresses growth). Various viral proteins interact with the RB protein and inhibit its action (adenoviral E1A, HPV-E7 protein and SV40 large T antigen). The *DCC* gene is involved in cell adhesion and is deleted in over 70% of colorectal cancer cases. *WT1* is a transcription factor for normal kidney and gonadal development. Like most TSGs, *p53* and *RB1* are also transcription factors.

Alterations in the patterns of DNA methylation are a common genomic alteration in human cancer. Abnormal methylation of CpG islands in the promoters of TSGs (such as *RB*, von-Hippel Lindau gene and *p16*) can contribute to their functional inactivation as one of the two hits. For a recent review of DNA methylation in neoplasia, see [Rountree et al. Oncogene 2001;20:3156](#). The more conventional way of TSG inactivation is its loss due to deletion or mutation. The loss-of-heterozygosity events are genomic deletions that discard the normal copies of TSGs, or uncover the existing TSG mutations. Inherited abnormalities of TSGs are associated with familial cancer syndromes that cause a variety of cancers at an early age.

c. DNA repair genes: DNA repair involves base-excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MSH) (see [Table](#)). While oncogenes act as accelerators of growth during G1 phase of cell cycle, and suppressor genes act as stop signals during S phase, DNA repair genes are not directly involved in cell growth. Their role is repairing DNA mismatches during replication of DNA just before the chromosomes condense in G2 phase for mitosis. These are DNA damage response genes and their defects result in susceptibility to a range of tumor types (usually skin or colon cancer and hematological malignancies). Mutations in DNA mismatch repair genes represent more of a general susceptibility state than a transforming event. BRCA1 and BRCA2 are two of the DNA repair genes. Mutations in critical oncogenes and TSGs are

more likely to occur in the repair-deficient cells. One mechanism for inactivation of one of the DNA repair genes (*MLH1*) is methylation silencing. The best-known DNA repair genes *MSH2* and *MLH1* are involved in hereditary non-polyposis colon cancer (HNPCC) when mutated. DNA repair genes are referred to as 'caretakers' while the genes involved in cell cycle control are called 'gatekeepers.' [See Kinzler & Vogelstein. Nature 1997;386:761-3.] For a review of polymorphisms of DNA repair genes and associations with cancer risk see [Wood et al, 2001](#) and [Goode et al, 2002](#). See also [DNA Repair Interest Group Website](#) and [DNA Repair Mechanisms](#) in the [Molecular Biology Web Book](#), and [DNA Damage and Repair and Their Roles in Carcinogenesis](#) in [Molecular Cell Biology](#).

d. Growth factors: They regulate cellular proliferation through receptor mediated autocrine or paracrine mechanisms. Transforming Growth Factor- α ($TGF\alpha$) is one of them (ligand for epidermal growth factor receptor-EGFR). It is over expressed in 50% of invasive breast cancer cases. Insulin-like Growth Factor-1 (IGF1) is a physiologic mediator and stimulator of normal cell growth. Most breast cancers express receptors for it.

e. Other genes: Ataxia-telangiectasia (*ATM*) gene confers increased risk for leukemia and lymphoma in homozygous subjects for its mutations. *BRCA1* (probably a transcription factor) and *BRCA2* are involved in familial breast cancer susceptibility (hereditary form is no more than 5% of all breast cancers).

Genetic prediction

Having a gene may be necessary but most of the time is not sufficient for developing a disease (in the cases of polygenic diseases, the presence of an environmental component, genetic penetrance etc.). Penetrance is defined as the proportion of individuals with a specific genetic alteration who will express the associated trait. Thus, genetic tests are not always perfect predictors of health risks. The *BRCA1* mutations, for example, cause familial breast cancer in 60% of women by the age of 60 (45-90% risk for life-time). The detection of a mutation in this gene (>100kb) does not necessarily indicate an absolute risk. Examples of high penetrance are polyposis-associated colon cancer and *MEN2a*-associated medullary thyroid cancer (penetrance close to 100%). Pharmacogenetic studies deal with the polymorphisms in xenobiotic enzyme genes. One of them, *CYP1A1* is involved in the activation of polycyclic aromatic hydrocarbons (PAH). The susceptibility allele of *CYP1A1* increases the risk of lung cancer in smokers, whereas, those lacking the PAH activating allele are relatively protected from the carcinogenic effects of smoking. There is also parental imprinting effect in inherited cancer susceptibility. Paraganglionoma occurs only if the mutation on chromosome 11 is inherited from the father. In neurofibromatosis type 2, however, children of affected females show earlier and more severe symptoms than children of affected males.

Genetic anticipation refers to the younger age or increased severity of a disease in successive generations. This phenomenon is better known in some non-malignant diseases caused by unstable triplet repeats (as in [Huntington's disease](#) and [congenital myotonic dystrophy](#)) but is also rarely seen in familial leukemia and ovarian cancer. Like anticipation and imprinting, sex influence and mosaicism also cause a non-Mendelian inheritance pattern in familial cancers. Mosaicism refers to the presence of the normal (wild type) genotype as well as the abnormal (mutant) genotype in the germ-line of an individual. In this case, the parent will not have the phenotype but the offspring will. Germ-line mosaicism has been observed in retinoblastoma.

One important issue about genetic prediction of cancer is genetic heterogeneity. There may be a number of distinct genotypes associated with the same phenotype. Absence of a known genotype causing a particular cancer may not mean the lack of genetic susceptibility to that cancer. Similar to the effect of genetic heterogeneity, phenocopies also confound pedigree analysis in cancer families. Phenocopy is a trait that appears to be identical to a genetic trait but that does not have a genetic basis. Sporadic form of a tumor in a cancer family may make the interpretation of the pedigree difficult.

Genetic therapy in cancer

Antisense therapy (for HPV in cervical cancer); targeted cytotoxic treatment against fusion proteins, somatic gene therapy (using carrier vectors incorporated with a toxin, an enzyme activating the cytotoxic agents; using DNA/liposome complex containing a foreign MHC antigen; or TILs infected with retroviruses carrying TNF or IL-2); immunization with tumor-specific proteins (p53, fusion proteins) or idiotypic antibodies in B-cell malignancies.

[Cancer Genetics Web: Genes - Chromosomes - Diseases](#)

[Atlas: Chromosomes in Cancer](#) [Main chromosome anomalies in hematologic malignancies](#)

[What Makes a Cancer Cell a Cancer Cell?](#) in [Cancer Medicine Online](#)

[Cancer](#) in [Molecular Cell Biology](#)

[Cancer Genetics](#) [Clinical Cancer Genetics](#)

[NCI Division of Cancer Epidemiology and Genetics](#)

[Overview of Cancer in Merck Manual](#)

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on Nov 29, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

TRANSPLANTATION GENETICS

M.Tevfik Dorak

Various histocompatibility loci play a role in transplantation genetics. The most important one in the determination of the fate of a transplanted cell/tissue/organ is the major histocompatibility complex (MHC). The MHC exists in all vertebrates examined to date (from the *Xenopus* and shark to humans) and similar histocompatibility systems exist even in invertebrates. The MHC was first identified through tumor transplantation studies in mice by Peter Gorer in 1937. This is why it is called a histocompatibility complex but it has many more biological functions (for details, click [here](#)). George Snell, Jean Dausset and Baruj Benacerraf received the Nobel Prize in 1980 for their contributions to the discovery and understanding of the MHC in mice and humans (other noteworthy names in the HLA field are Jon van Rood, Adriana van Leeuwen, Jan Klein, Walter Bodmer, John Trowsdale and Peter Parham). The MHC is the most polymorphic expressed genetic system. This looks like a price to pay in transplantation for the other biological functions of the MHC (e.g., avoidance of inbreeding). The main function of the classical MHC antigens is peptide presentation to the immune system to help distinguishing self from non-self (even protists have systems to recognize self and non-self). It is called human leukocyte antigens (HLA) in humans and consists of three classical regions: class I (HLA-A,B,Cw), class II (HLA-DR,DQ,DP) and class III (no HLA genes). HLA matching also has some relevance in blood and blood products transfusion under special circumstances.

Histocompatibility systems:

MHC (major histocompatibility complex)

mHAg (minor histocompatibility antigen)

Sex-linked histocompatibility systems (H-Y antigen coded by SMCY)

Blood groups

MHC typing:

Cellular (mixed lymphocyte reaction for HLA-Dw, primed lymphocyte test for HLA-DP)

Serological (HLA-A,B,C, DR, DQ)

Molecular (HLA-A,B,C, DR, DQ, DP)

Other histocompatibility tests:

Mixed Lymphocyte Reaction (MLR)

Lymphocyte cross-matching

Other cellular assays to detect the precursor frequency of (cytotoxic or helper)

lymphocytes sensitized against the mismatched HLA antigen

Donor sources:

Patients themselves (autologous)

HLA-identical siblings (allogeneic)

HLA-mismatched siblings or other related/unrelated subjects

HLA-matched unrelated donors

Cadaver donors (HLA-matched or mismatched)

Cloned human embryos, tissue engineering? [see *Sci Am* April 1999 issue]

Most commonly used transplantations:

Bone Marrow Transplantation (Hematopoietic Stem Cell Transplantation): HLA matching is an absolute requirement, and even mHAg differences may result in immunological complications. Associated with graft-versus-host disease (GvHD; an attack of immunocompetent donor cells to immunosuppressed recipient cells), engraftment failure and high probability of rejection (reverse of GvHD). The first two

complications correlate with the degree of histocompatibility (so does the infection risk). Thus, its use is limited by the availability of HLA-matched donors. The alternatives are autologous BMT, cord blood/fetal liver stem cell transplantation, and xenogenic transplantation. Cord blood stem cell transplantation has been clinically evaluated and the GvHD incidence is much lower than in conventional BMT. This is due to the lower immunogenicity and lower immune capability of the umbilical cord cells. Likewise, intra uterine BMT can be attempted to exploit the immaturity of the fetus' immune system. It is also possible to use gene therapy to lower the risk of immunological complications due to HLA-mismatching (such as the addition of adenovirus E3 genes which hide the cells from the immune system "stealth technology").

Pancreatic islet cell transplantation: For the treatment of Insulin-Dependent Diabetes Mellitus.

Cornea transplantation: HLA matching is not very relevant (at least HLA-DR) mainly because of the lack of vascularization of the cornea but also because of the immunological privilege of the cornea.

Solid organ transplantations: Kidney, Liver, Heart, Lung, Pancreas, Intestine. HLA matching is not crucial but beneficial.

Internet Resources

[What is Tissue Typing \(Guy's Hospital\)](#)

[Understanding Tissue Typing \(Michigan University\)](#)

[Transplantation in Merck Manual](#)

[MHC and Transplantation Chapter in Online Immunology Textbook](#)

[MHC Diversity](#) [Structure and Function of MHC Proteins](#)

[HLA-Related Links](#)

M.Tevfik Dorak, B.A., M.D., Ph.D.

Last updated on May 3, 2002

[Back to Contents](#)

[Back to Evolution](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Biostatistics](#)

[Glossary](#) [Homepage](#)

[Back to Contents](#)

[Back to Evolution](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Biostatistics](#)

[Glossary](#) [Homepage](#)

VIRAL and BACTERIAL GENETICS

M.Tevfik Dorak

Historical landmarks in viral and bacterial genetics

1944 Avery's pneumococcal transformation experiment shows that DNA is the hereditary material
1946 Lederberg & Tatum describes bacterial conjugation using biochemical mutants
1950 Barbara McClintock finds transposable elements in Maize
1952 Hersey & Chase shows that the hereditary material of the bacteriophage is DNA; Zinder & Lederberg achieves phage-mediated transduction in Salmonella
1953 Cavalli-Sforza et al show the F factor in bacteria (Cavalli-Sforza LL, Lederberg J, Lederberg M. J Gen Microbiol 1953;8:89)
1961 Jacob & Monod describes the operon structure
1977 Sanger sequences the phage Φ X174 (identification of overlapping genes).

Definitions

A **virus** is a disease causing agent consisting of a nucleic acid molecule and protein coat. Viruses are incapable of autonomous replication and have to use a host cell's translational system. A **bacterium** is a prokaryotic cell with its own circular DNA. A **bacteriophage** is a virus that infects bacteria only. Viruses would appear to be the simplest infectious particle. The discovery of **viroids**, nucleic acid without a protein capsule and **prions**, infectious proteins, subtracts another level of complexity. Both viroids and prions can cause diseases.

Properties of Viruses

They can be morphologically variable and even complex. Their morphology is one way of classifying viruses. They contain DNA or RNA but never both. Although they have a protein coat around their hereditary material, they lack properties of cells such as membranes, ribosomes, enzymes and ATP synthesis ability. Thus, they can only proliferate by using a host cell's translational machinery. In other words, they are obligate intracellular parasites. Their replication cycle consists of attachment and entry into the cell; replication of viral nucleic acid; synthesis of viral proteins; and finally, assembly of viral components and escape from host cells. Viruses are widely used as vectors in [gene therapy](#).

Properties of Bacteriophages

They have a simple structure, which consists of their double-stranded DNA and protein coat. Only the DNA enters into the bacteria. Bacterial RNA polymerase is composed of five individual polypeptide subunits (α_2 , β , β' , σ). The σ (sigma) factor is responsible for initiating transcription by recognizing bacterial promoter DNA sequences. Some phages supply their own σ factors to instruct the bacteria to transcribe phage genes preferentially. They are virtually viruses but can only infect bacteria. Bacteriophages usually infect only one species of bacteria but there are some who can infect several species even in different genera. Their life cycle may be either lytic (virulent phage) or lysogenic (temperate phage). Their DNA may be integrated into the host chromosome and remain as a prophage. Integration is achieved by recombination between a 15 bp sequence called *att* (for attachment) in the host chromosome and an identical sequence in the phage chromosome. This recombination requires an integrase (Int) enzyme encoded by

the phage. Bacteriophages are used for DNA cloning in molecular biology. DNA fragments can be inserted into a phage and following transfection of a competent bacteria, many copies of the desired DNA fragment can be obtained.

Properties of Bacteria

A bacterium has four types of genetic material: its single (haploid), covalently closed, circular dsDNA chromosome (in a supercoiled state); a plasmid(s); a bacteriophage or prophage; and a transposon. Genetic exchange between bacteria can occur by transfection, transduction or conjugation. Conjugation involves F+ male bacterium and F- female bacterium. Bacteria are haploid, but following a gene transfer (such as conjugation), they can be partially diploid (merozygote). This may result in a double cross-over event between the circular DNA and the linear newly introduced DNA if the two copies of the DNA are related. Sexual reproduction and meiosis do not occur in bacteria but genetic recombination to increase diversity is still possible by horizontal gene transfer (see below).

While bacteria are haploid organisms, **plasmids** can be considered as additional mini-chromosomes. Plasmids can be 1 to 300 kb long and may exist as multiple, free copies in a bacterium. As a rule, small plasmids occur in multiple copies per cell (high copy number), and large plasmids have a low copy number. Plasmids cannot replicate outside a bacterium. More than one types of plasmids can co-inhabit the same bacterium. Up to 10 kb (on average 3 kb) long DNA fragments can be inserted into a plasmid. They can enter the cells in two ways: vertical (via cell division - binary fission) or horizontal transmission (bacterial gene swapping). Some plasmids may contain genes that confer an evolutionary advantage to their hosts. These can be anti-bacterial toxins, catabolic enzymes (to use unusual carbon sources), virulence factors (pathogenic toxins), enzymes to degrade toxic compounds (like polychlorinated biphenyls, pesticides) and most importantly, antibiotic resistance (conferred by R plasmids). Sometimes, they may confer resistance up to five antibiotics at the same time. Plasmids can be exchanged between unrelated bacteria. This is the reason for speedy spread of antibiotic resistance among them.

Sometimes, a bacterium also contains a prophage as an inserted DNA fragment into its chromosome and this additional genetic material may be beneficial for it. For example, a prophage of the bacterium *Corynebacterium diphtheriae* carries a gene that encodes the diphtheria toxin causing the disease. Temperate phages may exist in a bacteria in a non-replicating, latent state (prophage). Naturally, every time a bacteria divides (every 15 to 20 minutes), the prophage will also be replicated.

Transposable elements cannot exist as free particles in a bacteria. They are integrated in the bacterial genome or into the genetic material of a plasmid or a prophage. They have the ability to move between these sites using an enzyme

called transposase. Transposons may also encode proteins that are useful for bacteria (such as antibiotic or heavy metal resistance factors).

Transformation involves the uptake of DNA from the environment. Cells that are able to take up DNA are called competent. While some bacteria (*H. influenzae*, *B. subtilis*) are naturally competent owing to some surface proteins they possess, others can be made competent by various treatments (calcium chloride treatment or electroporation). This kind of transformation is an important method used in [genetic engineering](#). Bacteria can take up DNA from other bacteria in nature (potentially from genetically modified bacteria) but the fate of such DNA is usually degradation. Historically, the principal application of transformation experiments was genetic mapping studies on naturally competent bacteria (co-transformation frequencies are inversely related to map distances).

Transduction involves the transfer of a bacterial DNA by means of a phage particle. Here, a desired piece of DNA is packaged into the bacteriophage head. The bacterial DNA (which can be an entire plasmid) can be transferred to a new cell when it is infected by the phage particle. In specialized transduction, the genome of a temperate phage (such as λ) integrates as a prophage into a bacterium's chromosome usually at a specific site. In generalized transduction, it is not a specific DNA segment but whatever DNA has been loaded into the phage is transferred. For example, the μ and P1 phages of *E. coli* can achieve generalized transduction. Transduction can also be used to establish gene order and for mapping purposes (only closely spaced genes will show co-transduction). In nature, a phage may transfer parts of bacterial DNA from one bacterium to another.

In conjugation, a direct contact between a male (carrying a fertility factor, or F+) and a female (F-) bacteria results in a one-way genetic material transfer (from male-to-female). Gram-negative bacteria (like *E. coli*) use a physical bridge called (sex) pilus (encoded by a conjugative plasmid) for gene transfer in conjugation, whereas, gram-positive bacteria (like pneumococcus) use a protein called clumping factor to get together. Some phage use the pili as receptors to attach to the bacteria. During conjugation in *E. coli*, the F factor (which is a conjugative plasmid) is not lost from the donor as it is only one of the strands of the plasmid that has been transferred. Subsequent replications of the bacteria restore the double-stranded state of the plasmid. When a conjugative plasmid initiates conjugation, other plasmids can be transferred (this is called mobilization). Conjugation is the exception to the rule that bacteria reproduce asexually. Although conjugation resembles sexual reproduction, an important difference is that conjugation is a one-way process. The F factor may exist as a free plasmid or may be inserted into the bacterial genome. Some conjugative plasmids (like the F factor of *E. coli*) can achieve transfer of chromosomal genes. An *E. coli* strain that has this property is called Hfr strain (for high frequency recombination). It is important to know that chromosomal genes are transferred before the plasmid itself. If the bridge is broken during transfer, the recipient will

remain F-. Controlled conjugation experiments can be used for gene mapping. Indeed, this approach was used to show the circularity of the E. coli chromosome and to determine the location of 1900 of its genes. Transformation, transduction and conjugation are means of horizontal gene transfer in nature (see *Bacterial Gene Swapping in Nature* by RV Miller. Scientific American 1998 (January), pp.47-51).

[Viral and Bacterial Genetics](#) (Ch. 12 of the Mol & Cell Biology Course, University of Texas)

[Prokaryotic Genetics](#) (Ch. 7.01 of Biology Hypertextbook, MIT)

[Microbial Genetics](#) (Ch. 9 of Brock Biology of Microorganisms, Prentice-Hall, Inc.)

[Microbiology Videos](#)

Chapter 9 of [Online Microbiology Book: Genetic Regulatory Mechanisms in Bacteria](#)

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Aug 16, 2003

[Back to Contents](#)

[Back to Evolution](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Biostatistics](#)

[Glossary](#)

[Homepage](#)

[Back to Contents](#)

[Back to Evolution](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Biostatistics](#)

[Glossary](#)

[Homepage](#)

PLANT GENETICS

M.Tevfik DORAK

Plant taxonomy

Plant Kingdom has about 260,000 species divided into two Phyla (or divisions in plants):

1. **Bryophyta**: They lack a vascular system for the internal conduction of water, minerals and food (lower plants), and depend on direct contact with surface water. This group includes mosses, liverworts and hornworts. There is always an alternation of generations between morphologically distinct sporophyte and gametophyte. The familiar leafy plant of Bryophytes is the sexual, gamete-producing (gametophyte) generation of their life cycle.

2. Tracheophyta (vascular plants, higher plants): This group consists of plants that have a vascular system, i.e., xylem and phloem (water/mineral and food-

conducting tissues, respectively). Tracheophyte leafy plants are the asexual, spore-producing, diploid (sporophyte) generation of their life cycle. One of its Subphylum is Pteropsida consisting of the following Superclasses:

i. Filinicae: **Ferns**. They do not reproduce by seeds but by spores like the Phylum Bryophyta. Alternation of generation is typical of ferns and Bryophyta.

ii. **Gymnosperms**: Cone-bearing woody seed plants. Includes cycads, ginkgo, **conifers** (pines, cedars) and gnetophytes.

iii. **Angiosperms**: Flower plants (divides into monocots and dicots). The gymnosperms and angiosperms are collectively called Spermatophyte (seed-bearing) plants. In this group, the gametophyte (haploid) generation does not occur as an independent plant (as in ferns). The vestigial gametophytes are contained in the sporophyte tissue as a few nuclei and can only be seen by a microscope (the embryonic sac and the pollen grain). The sporophyte embryo is contained in a seed which is dispersed from the plant. The angiosperms, therefore, cannot produce asexual spores and there is no obvious alternation of generations. The haploid pollen and ovule produced by a flower are thought to contain the remains of the gametophyte generation which was typical of the ancestors of the angiosperms (up to and including ferns). Link to **Plant Evolution and Classification** in Kimball's Biology Pages.

Plant evolution

Evolution of eukaryotes from a presumed bacteria-like ancestor is one of the major events in evolutionary history. They have a distinct nucleus, organelles involved in energy metabolism (mitochondria and chloroplast), extensive internal membranes and a cytoskeleton of protein fibres and filaments. Chloroplasts (photosynthesis) in green plants and algae originated as free living bacteria related to the cyanobacteria [the chloroplastic DNA is more similar to free-living Cyanobacteria DNA than to sequences from the plants the chloroplasts reside in]. The eukaryotic mitochondria (ATP synthesis) are endosymbionts like chloroplasts. Mitochondria were acquired when aerobic Eubacteria were engulfed by anaerobic host cells. As they conferred useful functions like aerobic respiration and photosynthesis (chloroplasts), they were retained as endosymbionts. This must have happened after the nucleus was acquired by the eukaryotic lineage. The origin of eukaryotic nucleus is almost certainly autogenous and not a result of endosymbiosis. Mitochondria are believed to have originated from an ancestor of the present-day purple photosynthetic bacteria that had lost its capacity for photosynthesis (chloroplasts from an ancestral Cyanobacterium).

All land plants evolved from the green algae or Chlorophyta. In the period before the Permian (the Carboniferous), the landscape was dominated by seedless ferns and their relatives. Vascular plants first appeared in Silurian (439-409 Mya). After the Permian extinction, gymnosperms became more abundant. They

evolved seeds and pollens (encased sperm). Angiosperms evolved from gymnosperms during the early Cretaceous about 140-125 Mya. They further diversified and dispersed during the late Cretaceous (97.5-66.5 Mya). Water lilies are one of the most ancient angiosperm plants. Currently, over three fourths of all living plants are angiosperms. The angiosperms developed a close contact with insects which promoted cross-pollination and resulted in more vigorous offspring. Their generation time to reproduce is short, and their seeds can be dispersed by animals. For these reasons, the angiosperms were able to travel and disperse all around the world. The important events in the evolution of the angiosperms were the evolution of showy flowers (to attract insects and birds), the evolution of bilaterally symmetrical flowers (adaptation for specialized pollinators), and the evolution of larger and more mobile animals (to disperse fruits and seeds). See *also* **A Brief History of Life**.

Polyploidy is an important mechanism in the evolution of plants. It is a situation in which the organism has more than two ($2n$) sets of chromosomes. It can be $3n$, $4n$ or more. A high proportion (47%) of angiosperms are polyploid. It arises as a result of meiotic irregularities and gives rise to sterile progeny which can still reproduce asexually. The original South American potato is a tetraploid ($4n$). Many of the common food plants (strawberries, apples, potatoes) are polyploid as this results in larger flowers and fruits (as well as larger cells, thicker and fleshier leaves). The wheat now grown for bread (*T. aestivum*) is hexaploid ($6n = 42$ chromosomes). Polyploidy can be induced by treatment of colchicine. Triploid offspring can be produced by crossing a colchicine-induced or naturally occurring tetraploid and a diploid. Odd number polyploids are sterile because they cannot segregate chromosomes evenly into gametes in meiosis (odd numbers are not divisible by two). Sterility caused by triploidy is useful to produce seedless fruit that is easier to eat (banana) or better tasting (less bitter cucumbers). Polyploidy is a common mechanism for sympatric speciation (reproductive isolation without geographical isolation).

Plant biology

Plants are eukaryotic, multicellular organisms. A plant cell differs from an animal cell in that they have rigid cell walls composed of cellulose, chlorophyll containing plastids (chloroplasts), and are able to photosynthesize. Fungi are different because they lack chlorophyll and chloroplasts, their cell walls contain chitin. Algae were formerly thought to be plants because of their rigid cell wall and photosynthesizing ability. They are, however, currently placed in the Kingdom Protocista because of the variety of cell pigments, cell wall types and different forms and structures. See **Plant Anatomy and Physiology**.

Plant reproduction

Asexual reproduction: Potato (tubers), strawberry (runners), iris (rhizomes), gladiolus (combs).

Sexual reproduction: Because land plants are immobile, alternation of generations has evolved in some groups to allow fertilization. The plants that

possess leaves, roots, stems and flowers are sporophytes (asexually reproducing). This generation gives rise to the gametophyte generation (sexually reproducing). One spore type (microspore) develops into a male gametophyte and the other kind (megaspore) into a female gametophyte. The female gametophyte remains protected in the carpel of the flower. When fertilized, an embryo is formed (a seed) which is a young sporophyte able to form a new organism when germinated.

The female gametophyte (ovum) of a flowering plant is formed in the ovules at the centre of the flower. After meiosis, an embryo sac forms (this is the female gametophyte generation of the alternation of generations in flowering plants). This consists of the ovum and several other haploid cells. The male gametophyte is the pollen grain. Two haploid sperm are produced in each pollen grain. The pollen grain should reach to the stigma of the recipient -female- flower. Following pollination, it germinates and a pollen tube grows down into the ovule carrying the two nonmotile sperm. One of these fuses with the ovum to form a zygote while the other fuses with two further cells of the embryo sac to form a triploid nutritive tissue called endosperm. This is called double fertilization and is unique to plants. The zygote divides to give rise to two cells. One will form the embryo and the other a supportive structure (suspensor). Embryonic differentiation starts but does not proceed for long. When the development ceases, the embryo becomes packaged in a seed, specialized for dispersal.

Pollination: This can be achieved via self-pollination (autogamy) or cross-pollination (allogamy). Most plants reproduce by both self and cross-pollination. Cross-pollinating plants produce better-quality seeds and more varied (adaptable) offspring. Because of the advantages of cross-pollination, most plants have evolved mechanisms to prevent self-pollination. One of them is production of some chemicals that prevent pollen from growing on the stigma of the same flower, or from developing the pollen tubes in the style (**self-incompatibility system**). Some plants produce only one kind of (either male or female) flowers (dioecious) and some are dichogamous (the two separate sex organs develop at different times in the same flower: protandry or protogyny). Cross-pollination can be achieved by wind, insects (honey bee), bats and birds. One feature that developed as a result of insect pollination is **pollen-tube competition**. When a number of pollen grains is deposited on a stigma, each pollen grows a pollen tube to reach the ovule. Whichever reaches first, it fertilizes the ovule. The fastest growing pollen tube usually carries the best genes and results in a more vigorous offspring. Therefore, apart from avoidance of self-fertilization, there is selection for the best cross-pollinating pollen as well (sexual selection in plants). See *a/so* **Pollination Adaptations**.

Unique genetic features of plants

Ability to photosynthesize

Totipotency of plant cells

Hermaphroditism and ability to reproduce both sexually and asexually

Double fertilization
Polyploidy
Alternation of generations
Mitosis in haploid state

Agricultural biotechnology

To create plants with altered characters, gene transfer has been used extensively in recent years. In principle, it is the same procedure as used in other organisms. Plants are especially suitable for genetic modification because most plant cells are totipotent. This means that a plant can be generated from a single genetically modified cell (i.e., it would not require fertilization). The most commonly used carrier vector in plant genetic modification is the Ti plasmid of *Agrobacterium tumefaciens*. The 30 kb-long T-DNA part of this plasmid is able to integrate into plant chromosomes. This plasmid can be used to transfer up to 40 kb inserted DNA into a protoplast (a plant cell whose cell wall has been destroyed enzymatically). The engineered protoplast has the ability to act as a (fertilized) gem cell and to regenerate into a whole plant (totipotency). An alternative gene transfer method for plant cells is penetrating the cells with DNA coated gold and tungsten spheres fired from a special gun (this technique is called biolistics). The common applications of genetic modification in plants are resistance to insects, viruses, herbicides and commercial applications (such as higher yield, seedless fruits, long-lasting tomatoes). Plants can also be used as bioreactors to produce desired recombinant proteins. See also [Gene Therapy](#).

Further Reading

Flower Power by B Furlow (New Scientist, 9 Jan 1999, pp.22-26)

[Internet Directory for Botany](#) .. [Virtual Library of Botany](#) .. [Botanical Glossaries](#)

[Plant Biology Course Notes](#) .. [Plant Breeding Course Notes](#) .. [Biology of Plants Course Notes](#)

[Botany Course](#) .. [Plant Genetics Course](#) .. [Plant Reproduction \(1\)](#) .. [Plant Reproduction \(2\)](#)

[General Plant Biology Course Notes](#) .. [Plant Molecular Genetics Course Notes](#) .. [Plant Biology \(mini-essays\)](#)

[Arabidopsis thaliana Website](#) .. [Maize Genome Database](#) .. [Kew Gardens \(London\)](#)

**M.Tevfik Dorak, B.A. (Hons), M.D.,
Ph.D.**

Last updated on Dec 13, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

BIOTECHNOLOGY

M. Tefvik DORAK

An overview of [Biotechnology](#)

Tools used in biotechnology

DNA extraction: Depending on the cell characteristics, DNA extraction from animal cells differs from DNA extraction from plant or prokaryotic cells. Links to [Roche Manual](#), and [Qiagen Handbook](#) (see also [Gentra Puregene Protocols](#) for technical reports on DNA extraction). For RNA-related information, see [Ambion](#) website.

Hybridization techniques: Southern blotting, Northern blotting and in situ hybridization (including fluorescent in situ hybridisation - FISH). Hybridization techniques allow picking out the gene of interest from the mixture of DNA/RNA sequences. Hybridization only occurs between single stranded and complementary nucleic acids. The level of similarity between the probe and target determines the hybridization temperature. See the overview of [blotting techniques](#) from the [Biology Hypertextbook](#), an animation of [Southern blotting](#), and an example of [DNA fingerprinting](#).

Enzymatic modification of DNA: DNA ligase and restriction enzymes (may create sticky ends or blunt ends) are used to manipulate DNA. Most restriction enzymes recognize palindromic sequences. These are short sequences which are the same on both strands when read 5' to 3' (such as the *MspI* restriction site CCGG and that of *EcoRI* GAATTC). See the action of [EcoRI](#). (These enzymes are called restriction endonucleases or restriction enzymes because they restrict viral replication as first discovered in 1962 by Werner Arber and only recognize site-specific or restricted sequences of DNA).

Cloning into a vector: vectors can be a plasmid (pBR322, pUC including Blue Script), lambda (λ) bacteriophage, cosmid, PAC, BAC, YAC, expression vectors. The Ti plasmid is the most popular vector in agricultural biotechnology. Plasmids can accommodate up to 10 kb foreign DNA, phages up to 25 kb, cosmids up to 44 kb, YACs usually several hundred kb but up to 1.5 Mb. Gene cloning contributed to the following areas: identification of specific genes, genome mapping, production of recombinant proteins, and the creation of genetically modified organisms. Link to examples of [plasmids](#).

Gene libraries: Genomic (restriction digestion, sonication) or cDNA libraries are made to identify a gene. See the construction of a [human genomic library](#).

Polymerase Chain Reaction (PCR)

Using the thermostable DNA polymerase obtained from *Thermophilus aquaticus* (briefly *Taq*), the PCR amplifies a desired sequences millions-fold. It requires a primer pair (18-30 nucleotides) to get the DNA polymerase started, the four nucleotides (dNTPs), a template DNA and certain chemicals including magnesium chloride (as a cofactor for *Taq* polymerase). The three steps in a cycle of the PCR - denaturation (the separation of the strands at 95°C), annealing (annealing of the primer to the template at 40 - 60°C),

and elongation (the synthesis of new strands) - take less than two minutes. *Taq* polymerase extends primers at a rate of 2 - 4 kb/min at 72° C (the optimum temperature for its activity). Each cycle consisting of these three steps is repeated 20 - 40 times to get enough of the amplified segment. Annealing temperature of each primer is calculated using its base composition. For primers less than 20 base-long: $T_m = 4(G+C) + 2(A+T)$. For more accurate calculation of the T_m value, visit [IDT BioTools OligoAnalyzer](#).

The conventional PCR is able to amplify DNA sequences up to 3 kb but the newer enzymes allow amplification of DNA fragments up to 30 kb-long. Nanogram levels of template DNA (even from a single cell) is enough to obtain amplification. The more recent '[real-time PCR](#)' techniques are able to detect the sequence of interest in 20 picogram of total RNA. *Taq* polymerase has a relatively high misincorporation rate. It has been genetically modified to reduce the misincorporation rate.

See a lecture note on [PCR](#), an article on [PCR](#), an animation of [PCR](#), a technical guide to [PCR](#), and a PowerPoint presentation on [PCR](#). Roche Molecular Diagnostics website has an overview of [PCR method](#) and several [video presentations](#).

Different versions of PCR: Nested PCR (for increased sensitivity and specificity), reverse transcriptase (RT) PCR (starts with mRNA instead of genomic DNA), amplified fragment length polymorphism (AFLP) (replaced Southern blotting), overlap PCR (joins two PCR products together), inverse PCR (amplifies an unknown DNA sequence flanking a region of known sequence).

Applications of PCR

1. Diagnostic use in medical genetics, medical microbiology and molecular medicine
2. HLA typing in transplantation
3. Analysis of DNA in archival material
4. Forensic analysis (DNA fingerprinting)
5. Preparation of nucleic acid probes
6. Clone screening and mapping
7. [Gene quantitation](#)

DNA sequencing

The new technology allows direct sequencing of DNA fragments rather than trying to figure out the gene order, DNA mutations and new genes by traditional methods such as RFLP analysis, chromosomal walking or even transduction and conjugation experiments in bacteria. DNA sequencing has now reached the automated stage and is routinely used in many laboratories even for HLA typing. In automated sequencing, a single sequencing reaction is carried out in which the four ddNTPs are labeled with differently colored dyes. At the end of the reaction, the mixture is run in a polyacrylamide gel and the colored chains are detected as they migrate through the gel. The detection system identifies the terminal base from the wavelength of the fluorescence emitted upon excitation by a laser. The DNA polymerase used in a sequencing reaction is usually part of the *E. coli* polymerase known as the Klenow fragment or a genetically modified DNA polymerase from the phage T7 (Sequenase). The usual *Taq* DNA polymerase can also be used for this purpose.

See an animation of [DNA sequencing](#) (by dideoxy chain termination method).

Applications of biotechnology

1. Recombinant protein and enzyme synthesis (biopharmaceuticals)

2. Genetic modification of bacteria, animal and plant cells (genetic engineering)
3. Transgenic and knock-out animals to study gene function
4. Cloning
5. DNA fingerprinting (forensic science)
6. Biological warfare

[Recombinant DNA Chapter in Biology Hypertextbook](#)
[Recombinant DNA in Kimball's Biology](#) [Protocols in Molecular Biology](#)
[A Virtual Tour in Agricultural Biotechnology](#)
[Biotechnology Animations & Videos](#)
[Genentech: The first biotechnology company](#)

**M.Tevfik Dorak, B.A. (Hons), M.D.,
Ph.D.**

Last updated on March 9, 2004

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

GENETIC ENGINEERING

M.Tevfik Dorak

Gene therapy

Gene therapy is one of the many applications of genetic engineering. It involves introducing a new gene or modifying an existing gene (itself or its activity) in cells. It can be used to treat or prevent diseases.

There are two forms of gene therapy:

1. Germ cell therapy (unethical and not allowed in humans, easily performed in plants)
2. Somatic cell therapy (currently used to treat cystic fibrosis, severe combined immune deficiency [SCID], some tumors, etc). This kind of gene therapy can be applied to the

whole body (*in vivo* therapy) or to the cells removed from a patient (*ex vivo* therapy). In the latter, the engineered cells are returned to the patient.

Gene delivery methods:

1. Recombinant retroviral or adenoviral vector-mediated: Retroviruses can only accept up to 7 kb of introduced DNA. They only infect dividing cells. Because of this, they are used in *ex vivo* therapies. They integrate into the host genome. Thus, their effect is long lasting, but insertional mutagenesis is a potential problem. Retroviruses are used in the treatment of SCID. Adenovirus can accommodate larger genes and does not integrate into the host genome. Their effect is transient. The main problem with adenovirus is the immune response they elicit (used in the treatment of cystic fibrosis and α 1-antitrypsin deficiency).
2. Liposome-mediated transfer: DNA is encapsulated in liposomes (lipid micels). This method has no side effects but is less efficient in transferring DNA to target cells. It does not cause an immune reaction. Liposomes can be used *in vivo* and *ex vivo* and carry any size of DNA fragment.
3. Microinjection (germ cell therapy): DNA is injected directly into the nucleus of a fertilized ovum viewed under the microscope. Currently, this method is used routinely to produce transgenic animals.
4. Biolistics: This method is used in plants. It involves coating special metal spheres with DNA and firing them into the plant cell from a special gun. Biolistics is used as an alternative to Ti plasmid-mediated gene transfer. See [Plant Genetics](#).

Gene therapy in humans

In humans, genetic engineering is only attempted in somatic cells in the form of gene therapy. Technically, it is possible to modify a fertilized egg at the four-cell stage (or even clone) but there is no ethical approval for such germ cell gene therapy at present. The first gene therapy for a human disease was successfully achieved for SCID by introducing the missing gene ADA into the peripheral lymphocytes of a 4-year-old girl and returning modified lymphocytes to her (in 1990). As of January 2000, more than 350 gene therapy protocols have been approved in USA. Human gene therapy uses the same strategies to deliver the genes. This can be achieved *ex vivo* or *in vivo* (*in vivo*: viral vectors, liposomes; *ex vivo*: cells such as tumor-infiltrating lymphocytes (TNF insertion), bone marrow stem cells as in bone marrow transplantation, or peripheral lymphocytes -as in SCID (ADA insertion)- are taken out, modified and returned to the patient). Alternatively, antisense treatment can be tried to prevent the transcripts from being translated into unwanted proteins (as being tried to combat HIV infection). Even naked DNA (i.e., not inserted into a carrier virus or liposome) can be injected directly to the patients to substitute the defective gene (DNA vaccines). DNA vaccines are plasmids being inserted the desired gene. Plasmids may directly transfect (animal) cells *in vivo*. This strategy is usually used to elicit cytotoxic T cell type immune response using an antigenic product of a pathogen (e.g., HIV). A popular gene therapy method is using the 'suicide gene'. The gene in question is thymidine kinase from a herpes simplex virus (HSV-tk) and is delivered to the target cells (usually cancer cells). When ganciclovir, an otherwise harmless antiviral agent, is given to patients, HSV-tk-bearing cells convert it to a toxic substance and the cells die. This is usually used in the treatment of (brain) tumors, but also used in the treatment of GvHD following BMT. Any other gene, which would render cancer cells highly sensitive to selected drugs, can be used as the suicide gene. A logical use of gene therapy in cancer

would be either replacing the missing tumor suppressor gene or blocking the effects of an oncogene.

The tricks that can be used to treat HIV infections include using dominant negative mutations (generating inactive versions of proteins HIV needs to replicate), or delivering genes into CD4+ T lymphocytes (target cells for HIV infection) that would be transcribed to short mRNAs mimicking essential viral control mRNAs to interfere with the viral regulatory mechanisms.

Other applications of genetic engineering

Chakrabarty's bacterium was constructed by using classical genetic selection to combine genes originally located on four different plasmids onto one compound plasmid. It is used to clean up oil spills.

Proteins with industrial applications such as **Rennin**, a protein used in making cheese, can be produced by recombinant DNA technology. This technology deals with isolating a gene (or its cDNA) from its natural host and inserting it into a different (asexually reproducing) species' genome so that it can be copied every time the new host cell (usually a bacterium) replicates. Before the advent of genetic engineering it was extracted from the fourth stomach of cattle. The new technology is known as the 'cheese from microbes'.

Enzymes used in genetic engineering such as restriction endonuclease (biological scissors), DNA polymerase (for replication of DNA), reverse transcriptase (to make DNA from RNA) and ligase (to ligate two DNA fragments) are produced by recombinant DNA technology (by cloning in high copy number plasmids in bacteria). Other enzymes now routinely produced by recombinant DNA technology are rennin (an enzyme used in cheese making mentioned above), lipase (cheese making), α -amylase (beer making), bromelain (meat tenderizer, juice clarification), catalase (antioxidant in food), cellulase (alcohol and glucose production), and protease (detergents).

Proteins for medical applications such as insulin (previously extracted from pig pancreas, and since 1982 produced by recombinant DNA technology), clotting factor IX (which is lacking in hemophilia B patients and previously supplied as fresh plasma from volunteer donors), tissue plasminogen activator (t-PA, used in acute myocardial infarction), growth hormone (previously extracted from the pineal glands of cadavers), tumor necrosis factor (TNF), γ -interferon (γ -IFN) and interleukin-2 (IL-2) (TNF, γ -IFN, IL-2 are used in immunotherapy of cancers), erythropoietin (EPO, to stimulate red blood cell production) and vaccines (such as HBV, rubella) can be produced by recombinant bacteria.

Secretion of valuable (modified) proteins in milk of **transgenic animals** such as factor IX and elastase inhibitor α 1-antitrypsin (used in the treatment of emphysema) can be another source for certain protein. The gene for the desired product is usually combined with a gene coding for a milk protein that is expressed only in mammary glands. The combined gene is then inserted in fertilized eggs. These are implanted into recipient females. The desired protein can be harvested from a female's milk. Because the gene is now in all cells of the animal (including germ cells), its offspring will also have it. Transgenic and gene **knockout animals** are also used for research purposes. The first transgenic animal was a supermouse with a rat growth hormone gene (1982).

Plants: Plants most commonly used in genetic engineering are maize, tomato, potato, cotton and tobacco. The main aims are to induce tolerance to herbicides, resistance to insect pests or viral disease and to improve crop quality.

Herbicide tolerance: A gene from a soil bacterium (*A. tumefaciens*) codes for an enzyme (PAT) which inactivates the herbicide Basta. When plants are engineered to contain this gene they are not sensitive to the herbicide any more (already tried on sugar beet, tobacco and oilseed rape). This allows selective killing of weeds by herbicides. A plasmid from *A. tumefaciens* called Ti is used to integrate the gene into the plant genome (the PAT gene is inserted into the toxin gene of the plasmid).

Resistance to infection by viruses: Genes encoding antisense copies of viral genes were used with limited success to prevent viral infection. The transfer of a gene encoding a viral coat protein has been successful with an unknown mechanism.

Insect resistance: Potato plants have been engineered to contain a pea lectin gene. Lectin interferes with digestion of plants in insects but does not harm the plant. It is also possible to use a *Bacillus thuringiensis* toxin called protoxin as an insecticide. This has been tried in tomato but did not work very well because of low expression.

Quality improvement: Genetic engineering has also been used to modify plants to create genetically-modified (GM) foods (also called genetically engineered organisms 'GEO' perhaps more appropriately). As tomatoes age, they soften due to the effects of an enzyme called polygalacturonase which breaks down cell walls. Its production can be blocked by activating the antisense gene to inactivate mRNA for its gene. Thus, it is still transcribed but no translation occurs. These tomatoes can ripen on the plant and are still suitable (hard enough) for mechanical handling and transport (long lasting tomatoes = [FlavrSavr tomato](#)). The FlavrSavr tomato was the first GM food approved by the FDA to go on the market in 1994 (now discontinued). Tomato paste from genetically engineered tomatoes ([Zeneqa Tomato Paste](#); also discontinued) and oil from genetically engineered oilseed rape were the first two whole foods declared safe in the UK (in 1995) (see the link for all [Transgenic Crops](#)). Transgenic plants such as soybean and rice can be engineered to have the essential amino acids they normally lack. [Golden Rice](#) is for example genetically enhanced with beta carotene. Genetic engineering in plants has also been used to alter pigmentation in flowers, to improve nutritional quality of seeds and to obtain seedless fruits. See also [Plant Improvements: Biotechnology](#), [Transgenic Plants](#) and [Genetic Engineering in Plants](#) (National Geographic, May 2002).

Drawbacks and potential dangers of genetic engineering

1. Gene integration into the right place: Position effect or tissue/cell-specific expression of genes in gene therapy depend on the effects of enhancers and promoters.
 2. Immune response against the vector used for gene delivery
 3. Spread of the gene conferring antibiotic resistance to the vector
 4. Escape of the introduced virus or bacteria to the environment (to avoid this a modified strain of *E.coli* is used in recombinant DNA applications which cannot survive in nature but is useful in the laboratory. The laboratories using this technology are also under stringent control to avoid any contamination of the outside world)
 5. Acceleration of the evolution of resistance to the toxins or antibiotics introduced
- Modern biotechnology is achieving what classical biotechnology could not have in such a short time and across a species barrier. There is no need for generations of classical

breeding schemes any more and hybridization between different species can be achieved (like inserting a human gene into a mouse germ cell).

Cloning

Historical developments:

1. Nuclear transfer in frogs resulting in viable juveniles up to the tadpole stage in the 1950s (Elsdale et al. J Embryol Exp Morph 1960;8:437)
2. Nuclear transfer in sheep using donor nucleus from early morula stage (8-16 cell) embryos (Willadsen SM. Nature 1986;320:63)
3. Nuclear transfer in sheep using a cultured embryonic (totipotent) cell's nucleus (March 7, 1996 - Nature)
4. Cloning of Dolly the sheep by nuclear transfer from an adult's udder cell (Febr 27, 1997 - Nature) with a success rate of 0.3% (1 in 277 attempts)
5. Cloning of Polly the Sheep by transfer of a transgenic nucleus of a fetal fibroblast cell - (Dec 19, 1997 - Science) [an alternative to pronuclear injection in creation of transgenic animals]
6. Successful use of nuclear transfer in mice (Cumulina, success rate 2-2.8%) and cows in 1998 followed by cloning of many other mammals (cattle, goats, rabbits, cats, pigs, mules, horses). The most recent example is the cloned horse who was delivered by her dam twin in 2003 (see below).

Cloning is the ultimate phase in creating an animal with a desired trait. It follows the same path as selective breeding, artificial insemination, egg transplantation and in vitro fertilization. In genetics, cloning means creation of genetically identical animals (mammals) by means of nuclear transfer. In nuclear transfer, the chromosomes of an unfertilized oocyte (egg) are removed (enucleation) and a nucleus from a mature (diploid) cell is placed into the egg. Note that no fertilization takes place and it is the presence of a diploid nucleus inside an egg, which initiates cleavage divisions. It has been established that an unfertilized oocyte is a better recipient than a zygote. This is asexual reproduction of an animal, which would normally reproduce sexually. A clone resembles its (single) parent, the animal from which the donor cell was taken. This technology can be used either to create genetically identical animals or to introduce a genetic modification to the mammal. Another application is the production of undifferentiated cells which can then be induced to differentiate to any organ or cell type. The success rate is invariably very low. It may take hundreds of tries to succeed in nuclear transfer. It is, however, highest in sheep. This is probably because of the fact that transcription of the embryonic genome does not begin until the 8-16 cell stage in sheep, whereas it is the late 2-cell stage in mice.

In the previous cloning experiments, embryonic cells were used as the source of nucleus and the nucleus was transferred to a fertilized egg whose own nucleus was removed (in mice in 1980s). In the creation of Dolly, however, the nucleus of an adult-differentiated-cell was used. The key to the new procedure was synchronization of the cell cycle of the adult cell with that of the egg. The Dolly experiment showed that for successful cloning, an -undifferentiated- embryonic cell nucleus was not the only choice. Dolly was created from a mammary gland (udder) cell of a six-year old ewe. This showed that any adult cell still has the potential to start from the undifferentiated state and give rise to new differentiated cells (reversibility of differentiation). To achieve synchronization, the (donor) mammary cell is deprived of nutrients to stop its cell cycle (Go phase). The nucleus is then implanted into the egg, and an electrical current is applied to simulate fertilization which initiates egg activation which normally occurs at the time of sperm

penetration (in parthenogenesis, egg activation is achieved by physical or chemical stimuli). The egg recombined with a new nucleus begins dividing and proceeds to become an early embryo (blastocyst). This is then implanted into the uterus of an ewe (surrogate mother). The lamb that is born is a clone of the donor. At present, dedifferentiation can only be achieved by forming an embryo from the donor cell and culturing the embryo to the stage when it has a few hundred still undifferentiated cells. Then, the cells would be separated and grown in culture. Cloning, when it becomes more practical to do, has the potential to make copies of a bred line of cattle, sheep or other economically important animals without having to do artificial selection. Another application is to clone genetically modified pigs to use their organs for xenotransplantation. The modification is required to prevent immune response.

It has to be remembered that the cloned animal may not be fully [identical to the animal](#) who donated the nucleus. This is because of some epigenetic phenomena (methylation patterns) and cytoplasmic (extra-nuclear) inheritance. The low success rate in nuclear transfer may be due to a random loss of correct imprinting. Differentiated female cells, for example, contain one active X chromosome, but in early female embryos both X chromosomes are active. It is also a concern that imprinting that results in functional differences in the male and female genome may complicate the reprogramming of the genome after nuclear transfer. This may be why out of the 277 mammary-gland-cell nuclei that were fused with enucleated eggs, only 29 developed to the blastocyst stage and only one of those resulted in a live birth, Dolly, in surrogate mothers. Similarly, it took scientists 87 tries to successfully clone the first cat 'CopyCat' in 2001. Cloning of a horse in 2003 surprised many scientists because the surrogate mother was the donor of the egg. Therefore she gave birth to her own clone ([Nature](#), [New Scientist](#), [Science News](#), [BBC](#)). This time Italian researchers had to try more than 800 fusions. Most failed to develop from the cleavage stage to blastocysts and at early implantation. Interestingly, the failure rate was higher in male embryos (8 of 467 males vs 14 of 286 females, meaning three times high failure in male embryos; $P = 0.01$), which is probably the first concrete documentation of the same phenomenon known to occur in human pregnancies (see [Dorak MT et al, 2002](#) for references).

Because of the totipotency of the first four cleavage cells in the early embryo, a woman who is suffering from a mitochondrial DNA disease can still have a healthy baby by having the nucleus of her embryo implanted into a donor oocyte whose own nucleus has been removed. This is nuclear transfer (from an embryonic cell) but not cloning. Another development in recent years in reproductive physiology is the feasibility of ICSI (intracytoplasmic sperm injection) that enables sterile man to have a child. The bioethical implications of these applications are discussed elsewhere (see [Bioethics](#)). One important and usually neglected piece in discussion of the ethics of cloning is what will happen in the possible outcome of having a disabled / malformed cloned offspring. Are the scientists who have done the cloning going to be held responsible? It appears that cloned mammals usually have a collection of disorders called large offspring syndrome (LOS). Most cloned mammals so far have also developed deformities and arthritis (like Dolly). There is still a long way to go to optimize the conditions of cloning if it will ever be a common practice.

Suggestion for further reading

Wilmut I: Cloning for medicine. Scientific American 1998 (Dec):30-35.

A Student's Guide to Biotechnology - Debatable Issues. Greenwood Press, 2002.

Internet Resources

[A Beginner's Guide to Genetic Engineering](#)

[NIH Recommendations for Gene Therapy](#) [Virus Vectors & Gene Therapy](#)

[Cloning in New Scientist](#) [Cloning in Scientific American](#) [Cloning of Dolly Explained](#)
[Clone Age in BBC Science](#)

M.Tevfik Dorak, B.A. (Hons), M.D.,
Ph.D.

Last updated on Aug16, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

GENETIC ENGINEERING

M.Tevfik Dorak

Gene therapy

Gene therapy is one of the many applications of genetic engineering. It involves introducing a new gene or modifying an existing gene (itself or its activity) in cells. It can be used to treat or prevent diseases.

There are two forms of gene therapy:

1. Germ cell therapy (unethical and not allowed in humans, easily performed in plants)
2. Somatic cell therapy (currently used to treat cystic fibrosis, severe combined immune deficiency [SCID], some tumors, etc). This kind of gene therapy can be applied to the whole body (*in vivo* therapy) or to the cells removed from a patient (*ex vivo* therapy). In the latter, the engineered cells are returned to the patient.

Gene delivery methods:

1. Recombinant retroviral or adenoviral vector-mediated: Retroviruses can only accept up to 7 kb of introduced DNA. They only infect dividing cells. Because of this, they are used in *ex vivo* therapies. They integrate into the host genome. Thus, their effect is long

lasting, but insertional mutagenesis is a potential problem. Retroviruses are used in the treatment of SCID. Adenovirus can accommodate larger genes and does not integrate into the host genome. Their effect is transient. The main problem with adenovirus is the immune response they elicit (used in the treatment of cystic fibrosis and α 1-antitrypsin deficiency).

2. Liposome-mediated transfer: DNA is encapsulated in liposomes (lipid micels). This method has no side effects but is less efficient in transferring DNA to target cells. It does not cause an immune reaction. Liposomes can be used in vivo and ex vivo and carry any size of DNA fragment.

3. Microinjection (germ cell therapy): DNA is injected directly into the nucleus of a fertilized ovum viewed under the microscope. Currently, this method is used routinely to produce transgenic animals.

4. Biolistics: This method is used in plants. It involves coating special metal spheres with DNA and firing them into the plant cell from a special gun. Biolistics is used as an alternative to Ti plasmid-mediated gene transfer. See [Plant Genetics](#).

Gene therapy in humans

In humans, genetic engineering is only attempted in somatic cells in the form of gene therapy. Technically, it is possible to modify a fertilized egg at the four-cell stage (or even clone) but there is no ethical approval for such germ cell gene therapy at present. The first gene therapy for a human disease was successfully achieved for SCID by introducing the missing gene ADA into the peripheral lymphocytes of a 4-year-old girl and returning modified lymphocytes to her (in 1990). As of January 2000, more than 350 gene therapy protocols have been approved in USA.

Human gene therapy uses the same strategies to deliver the genes. This can be achieved ex vivo or in vivo (in vivo: viral vectors, liposomes; ex vivo: cells such as tumor-infiltrating lymphocytes (TNF insertion), bone marrow stem cells as in bone marrow transplantation, or peripheral lymphocytes -as in SCID (ADA insertion)- are taken out, modified and returned to the patient). Alternatively, antisense treatment can be tried to prevent the transcripts from being translated into unwanted proteins (as being tried to combat HIV infection). Even naked DNA (i.e., not inserted into a carrier virus or liposome) can be injected directly to the patients to substitute the defective gene (DNA vaccines). DNA vaccines are plasmids being inserted the desired gene. Plasmids may directly transfect (animal) cells in vivo. This strategy is usually used to elicit cytotoxic T cell type immune response using an antigenic product of a pathogen (e.g., HIV).

A popular gene therapy method is using the 'suicide gene'. The gene in question is thymidine kinase from a herpes simplex virus (HSV-tk) and is delivered to the target cells (usually cancer cells). When ganciclovir, an otherwise harmless antiviral agent, is given to patients, HSV-tk-bearing cells convert it to a toxic substance and the cells die. This is usually used in the treatment of (brain) tumors, but also used in the treatment of GvHD following BMT. Any other gene, which would render cancer cells highly sensitive to selected drugs, can be used as the suicide gene. A logical use of gene therapy in cancer would be either replacing the missing tumor suppressor gene or blocking the effects of an oncogene.

The tricks that can be used to treat HIV infections include using dominant negative mutations (generating inactive versions of proteins HIV needs to replicate), or delivering genes into CD4+ T lymphocytes (target cells for HIV infection) that would be transcribed to short mRNAs mimicking essential viral control mRNAs to interfere with the viral regulatory mechanisms.

Other applications of genetic engineering

Chakrabarty's bacterium was constructed by using classical genetic selection to combine genes originally located on four different plasmids onto one compound plasmid. It is used to clean up oil spills.

Proteins with industrial applications such as **Rennin**, a protein used in making cheese, can be produced by recombinant DNA technology. This technology deals with isolating a gene (or its cDNA) from its natural host and inserting it into a different (asexually reproducing) species' genome so that it can be copied every time the new host cell (usually a bacterium) replicates. Before the advent of genetic engineering it was extracted from the fourth stomach of cattle. The new technology is known as the 'cheese from microbes'.

Enzymes used in genetic engineering such as restriction endonuclease (biological scissors), DNA polymerase (for replication of DNA), reverse transcriptase (to make DNA from RNA) and ligase (to ligate two DNA fragments) are produced by recombinant DNA technology (by cloning in high copy number plasmids in bacteria). Other enzymes now routinely produced by recombinant DNA technology are rennin (an enzyme used in cheese making mentioned above), lipase (cheese making), α -amylase (beer making), bromelain (meat tenderizer, juice clarification), catalase (antioxidant in food), cellulase (alcohol and glucose production), and protease (detergents).

Proteins for medical applications such as insulin (previously extracted from pig pancreas, and since 1982 produced by recombinant DNA technology), clotting factor IX (which is lacking in hemophilia B patients and previously supplied as fresh plasma from volunteer donors), tissue plasminogen activator (t-PA, used in acute myocardial infarction), growth hormone (previously extracted from the pineal glands of cadavers), tumor necrosis factor (TNF), γ -interferon (γ -IFN) and interleukin-2 (IL-2) (TNF, γ -IFN, IL-2 are used in immunotherapy of cancers), erythropoietin (EPO, to stimulate red blood cell production) and vaccines (such as HBV, rubella) can be produced by recombinant bacteria.

Secretion of valuable (modified) proteins in milk of **transgenic animals** such as factor IX and elastase inhibitor α 1-antitrypsin (used in the treatment of emphysema) can be another source for certain protein. The gene for the desired product is usually combined with a gene coding for a milk protein that is expressed only in mammary glands. The combined gene is then inserted in fertilized eggs. These are implanted into recipient females. The desired protein can be harvested from a female's milk. Because the gene is now in all cells of the animal (including germ cells), its offspring will also have it. Transgenic and gene **knockout animals** are also used for research purposes. The first transgenic animal was a supermouse with a rat growth hormone gene (1982).

Plants: Plants most commonly used in genetic engineering are maize, tomato, potato, cotton and tobacco. The main aims are to induce tolerance to herbicides, resistance to insect pests or viral disease and to improve crop quality.

Herbicide tolerance: A gene from a soil bacterium (*A. tumefaciens*) codes for an enzyme (PAT) which inactivates the herbicide Basta. When plants are engineered to contain this gene they are not sensitive to the herbicide any more (already tried on sugar beet, tobacco and oilseed rape). This allows selective killing of weeds by herbicides. A plasmid from *A. tumefaciens* called Ti is used to integrate the gene into the plant genome (the PAT gene is inserted into the toxin gene of the plasmid).

Resistance to infection by viruses: Genes encoding antisense copies of viral genes were used with limited success to prevent viral infection. The transfer of a gene encoding a viral coat protein has been successful with an unknown mechanism.

Insect resistance: Potato plants have been engineered to contain a pea lectin gene. Lectin interferes with digestion of plants in insects but does not harm the plant. It is also possible to use a *Bacillus thuringiensis* toxin called protoxin as an insecticide. This has been tried in tomato but did not work very well because of low expression.

Quality improvement: Genetic engineering has also been used to modify plants to create genetically-modified (GM) foods (also called genetically engineered organisms 'GEO' perhaps more appropriately). As tomatoes age, they soften due to the effects of an enzyme called polygalacturonase which breaks down cell walls. Its production can be blocked by activating the antisense gene to inactivate mRNA for its gene. Thus, it is still transcribed but no translation occurs. These tomatoes can ripen on the plant and are still suitable (hard enough) for mechanical handling and transport (long lasting tomatoes = [FlavrSavr tomato](#)). The FlavrSavr tomato was the first GM food approved by the FDA to go on the market in 1994 (now discontinued). Tomato paste from genetically engineered tomatoes ([Zeneca Tomato Paste](#); also discontinued) and oil from genetically engineered oilseed rape were the first two whole foods declared safe in the UK (in 1995) (see the link for all [Transgenic Crops](#)). Transgenic plants such as soybean and rice can be engineered to have the essential amino acids they normally lack. [Golden Rice](#) is for example genetically enhanced with beta carotene. Genetic engineering in plants has also been used to alter pigmentation in flowers, to improve nutritional quality of seeds and to obtain seedless fruits. See also [Plant Improvements: Biotechnology](#), [Transgenic Plants](#) and [Genetic Engineering in Plants](#) (National Geographic, May 2002).

Drawbacks and potential dangers of genetic engineering

1. Gene integration into the right place: Position effect or tissue/cell-specific expression of genes in gene therapy depend on the effects of enhancers and promoters.
 2. Immune response against the vector used for gene delivery
 3. Spread of the gene conferring antibiotic resistance to the vector
 4. Escape of the introduced virus or bacteria to the environment (to avoid this a modified strain of *E.coli* is used in recombinant DNA applications which cannot survive in nature but is useful in the laboratory. The laboratories using this technology are also under stringent control to avoid any contamination of the outside world)
 5. Acceleration of the evolution of resistance to the toxins or antibiotics introduced
- Modern biotechnology is achieving what classical biotechnology could not have in such a short time and across a species barrier. There is no need for generations of classical breeding schemes any more and hybridization between different species can be achieved (like inserting a human gene into a mouse germ cell).

Cloning

Historical developments:

1. Nuclear transfer in frogs resulting in viable juveniles up to the tadpole stage in the 1950s (Elsdale et al. *J Embryol Exp Morph* 1960;8:437)
2. Nuclear transfer in sheep using donor nucleus from early morula stage (8-16 cell) embryos (Willadsen SM. *Nature* 1986;320:63)

3. Nuclear transfer in sheep using a cultured embryonic (totipotent) cell's nucleus (March 7, 1996 - Nature)
4. Cloning of Dolly the sheep by nuclear transfer from an adult's udder cell (Febr 27, 1997 - Nature) with a success rate of 0.3% (1 in 277 attempts)
5. Cloning of Polly the Sheep by transfer of a transgenic nucleus of a fetal fibroblast cell - (Dec 19, 1997 - Science) [an alternative to pronuclear injection in creation of transgenic animals]
6. Successful use of nuclear transfer in mice (Cumulina, success rate 2-2.8%) and cows in 1998 followed by cloning of many other mammals (cattle, goats, rabbits, cats, pigs, mules, horses). The most recent example is the cloned horse who was delivered by her dam twin in 2003 (see below).

Cloning is the ultimate phase in creating an animal with a desired trait. It follows the same path as selective breeding, artificial insemination, egg transplantation and in vitro fertilization. In genetics, cloning means creation of genetically identical animals (mammals) by means of nuclear transfer. In nuclear transfer, the chromosomes of an unfertilized oocyte (egg) are removed (enucleation) and a nucleus from a mature (diploid) cell is placed into the egg. Note that no fertilization takes place and it is the presence of a diploid nucleus inside an egg, which initiates cleavage divisions. It has been established that an unfertilized oocyte is a better recipient than a zygote. This is asexual reproduction of an animal, which would normally reproduce sexually. A clone resembles its (single) parent, the animal from which the donor cell was taken. This technology can be used either to create genetically identical animals or to introduce a genetic modification to the mammal. Another application is the production of undifferentiated cells which can then be induced to differentiate to any organ or cell type. The success rate is invariably very low. It may take hundreds of tries to succeed in nuclear transfer. It is, however, highest in sheep. This is probably because of the fact that transcription of the embryonic genome does not begin until the 8-16 cell stage in sheep, whereas it is the late 2-cell stage in mice.

In the previous cloning experiments, embryonic cells were used as the source of nucleus and the nucleus was transferred to a fertilized egg whose own nucleus was removed (in mice in 1980s). In the creation of Dolly, however, the nucleus of an adult-differentiated-cell was used. The key to the new procedure was synchronization of the cell cycle of the adult cell with that of the egg. The Dolly experiment showed that for successful cloning, an -undifferentiated- embryonic cell nucleus was not the only choice. Dolly was created from a mammary gland (udder) cell of a six-year old ewe. This showed that any adult cell still has the potential to start from the undifferentiated state and give rise to new differentiated cells (reversibility of differentiation). To achieve synchronization, the (donor) mammary cell is deprived of nutrients to stop its cell cycle (G₀ phase). The nucleus is then implanted into the egg, and an electrical current is applied to simulate fertilization which initiates egg activation which normally occurs at the time of sperm penetration (in parthenogenesis, egg activation is achieved by physical or chemical stimuli). The egg recombined with a new nucleus begins dividing and proceeds to become an early embryo (blastocyst). This is then implanted into the uterus of an ewe (surrogate mother). The lamb that is born is a clone of the donor. At present, dedifferentiation can only be achieved by forming an embryo from the donor cell and culturing the embryo to the stage when it has a few hundred still undifferentiated cells. Then, the cells would be separated and grown in culture. Cloning, when it becomes more practical to do, has the potential to make copies of a bred line of cattle, sheep or other economically important animals without having to do artificial selection. Another

application is to clone genetically modified pigs to use their organs for xenotransplantation. The modification is required to prevent immune response.

It has to be remembered that the cloned animal may not be fully [identical to the animal](#) who donated the nucleus. This is because of some epigenetic phenomena (methylation patterns) and cytoplasmic (extra-nuclear) inheritance. The low success rate in nuclear transfer may be due to a random loss of correct imprinting. Differentiated female cells, for example, contain one active X chromosome, but in early female embryos both X chromosomes are active. It is also a concern that imprinting that results in functional differences in the male and female genome may complicate the reprogramming of the genome after nuclear transfer. This may be why out of the 277 mammary-gland-cell nuclei that were fused with enucleated eggs, only 29 developed to the blastocyst stage and only one of those resulted in a live birth, Dolly, in surrogate mothers. Similarly, it took scientists 87 tries to successfully clone the first cat 'CopyCat' in 2001. Cloning of a horse in 2003 surprised many scientists because the surrogate mother was the donor of the egg. Therefore she gave birth to her own clone ([Nature](#), [New Scientist](#), [Science News](#), [BBC](#)). This time Italian researchers had to try more than 800 fusions. Most failed to develop from the cleavage stage to blastocysts and at early implantation. Interestingly, the failure rate was higher in male embryos (8 of 467 males vs 14 of 286 females, meaning three times high failure in male embryos; $P = 0.01$), which is probably the first concrete documentation of the same phenomenon known to occur in human pregnancies (see [Dorak MT et al, 2002](#) for references).

Because of the totipotency of the first four cleavage cells in the early embryo, a woman who is suffering from a mitochondrial DNA disease can still have a healthy baby by having the nucleus of her embryo implanted into a donor oocyte whose own nucleus has been removed. This is nuclear transfer (from an embryonic cell) but not cloning. Another development in recent years in reproductive physiology is the feasibility of ICSI (intracytoplasmic sperm injection) that enables sterile men to have a child. The bioethical implications of these applications are discussed elsewhere (see [Bioethics](#)). One important and usually neglected piece in discussion of the ethics of cloning is what will happen in the possible outcome of having a disabled / malformed cloned offspring. Are the scientists who have done the cloning going to be held responsible? It appears that cloned mammals usually have a collection of disorders called large offspring syndrome (LOS). Most cloned mammals so far have also developed deformities and arthritis (like Dolly). There is still a long way to go to optimize the conditions of cloning if it will ever be a common practice.

Suggestion for further reading

Wilmot I: Cloning for medicine. Scientific American 1998 (Dec):30-35.

A Student's Guide to Biotechnology - Debatable Issues. Greenwood Press, 2002.

Internet Resources

[A Beginner's Guide to Genetic Engineering](#)

[NIH Recommendations for Gene Therapy](#) [Virus Vectors & Gene Therapy](#)

[Cloning in New Scientist](#) [Cloning in Scientific American](#) [Cloning of Dolly Explained](#)
[Clone Age in BBC Science](#)

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Aug16, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

ETHICS AND GENETICS

M.Tevfik DORAK

Bioethics is the study of moral issues in the fields of medical treatment and research. Biomedical ethics is not a new branch. Hippocrates, the Greek physician, is not only the father of modern medicine, but also of medical ethics. About 24 centuries ago, he said physicians should not give poisons to patients and should advocate for the patient's interest. There was no consensus then either. Hippocrates opposed abortion while Plato was in favor of it. Today, every new biomedical development, such as in vitro fertilization, organ or bone marrow transplantation from a living relative, genetic modifications of all sorts bring about dilemmas and conflicting opinions. One difference today from Plato's time is that economic considerations are taken into account too.

Following the discovery of the CF gene, a rush by the public to get genetic testing done was expected but this did not happen. A survey of 20,000 people showed that people are not interested in knowing their genetic make-up unless they have a relative with a genetic disease or they are involved in a pregnancy. It appears that insurance companies and employers are more interested in such information for obvious reasons. This is where one of the greatest ethical conflicts of genetic revolution starts. In a recent review (1999), the following guidelines for genetic testing of CF mutations are drawn up: Genetic testing for CF should be offered to adults with a positive family history of CF, to partners of people with CF, to couples currently planning a pregnancy, and to couples seeking prenatal care. The panel does not recommend offering CF genetic testing to the general population or newborns. Comprehensive educational programs targeted to health care professionals and the public should be developed using input from people living with CF and their families and from people from diverse racial and ethnic groups. Additionally, genetic counseling services must be accurate and provide balanced information to afford individuals the opportunity to make autonomous

decisions. Every attempt should be made to protect individual rights, genetic and medical privacy rights, and to prevent discrimination and stigmatization. It is essential that the offering of CF carrier testing be phased in over a period to ensure that adequate education and appropriate genetic testing and counseling services are available to all persons being tested. [*Arch Intern Med* 1999 Jul 26;159(14):1529-39]

The ongoing Human Genome Project (HGP) has spared 5% of its budget for investigations into ethical, legal, and social issues of its findings. This unusual but well justified event in the history of science shows how great the implications of the forthcoming findings will be. The aim is to benefit from the findings of HGP rather than causing social disruption. Today, there are about 2,000 professional medical ethicists in the USA, coming from academic disciplines as law, medicine, philosophy, political science and theology.

Perhaps the most popularized ethical question in genetics is eugenics. In the past, numerous discussions have taken place for marriage laws, sterilization and immigration regulations in view of the principles of eugenics. The new genetic technology is likely to initiate similar discussions. In this respect, cloning and germ cell therapy are the most likely candidates to ignite the hottest debates. These two techniques are currently not allowed to be used in humans.

Currently, in some countries and in some states of the USA, legislation exists to regulate genetic testing, genetic screening, counseling, and discrimination in employment and insurance matters against individuals with genetic disorders. The California Hereditary Disorders Act of 1990 is an example. It regulates access to genetic services, confidentiality of genetic information, discrimination against affected individuals and carriers, the voluntary nature of screening programs, the reproductive rights of those at risk of passing a genetic disorder to their offspring, and professional and public education programs about genetics. This law establishes the basic elements of genetic testing: autonomy, confidentiality, privacy and equity. Ideally, all screening (including newborns) should be voluntary (informed choice), and should be done only after informed consent (very similar to the Nuremberg Code). The person should be able to choose not to proceed any more at any stage of the procedures. The results should remain confidential and anonymous, therefore should not be used to discriminate anybody on any grounds. Potential areas of conflict in clinical genetics include genetic testing in children, the distinction between research and service, and the rights of the individual versus the rights of the extended family, the doctor and the society. Think about the implications for the members of the family when somebody is discovered to have the Huntington's disease gene, or storing DNA samples from convicted criminals, volunteer blood donors, and all newborns!

In principle, there is no serious objection to somatic cell gene therapy as this is no different from the medical treatment of an individual either by medicine, surgery or transplantation. The potential problem is its use for 'enhancement', in other words, for cosmetic purposes. The opponents of this objection could easily ask about the proportion of cosmetic surgery performed in plastic surgery departments or the ratio of these to life-saving plastic surgical operations. When it comes to germ cell gene therapy, the potential use of it for eugenics creates a problem. This is because, any change in the germ cell will be passed on to the following generations forever. When it is used for medical purposes, i.e., to eliminate a disease gene and to replace it with the correct version, this is in principle acceptable but isn't this (negative) eugenics? However harsh it sounds, isn't it the diseases that have played major roles in the evolution of species? Human germ cell therapy is currently banned because of the fears of positive eugenics.

Patenting life forms and newly found genes is another hot topic brought about by the HGP. In a landmark 1980 ruling, the US Supreme Court decided that Dr Chakrabarty

(see [Gene Therapy](#)) could patent a bacterium that digests crude oil. The Court said that the intent of Congress in establishing patent law was that patents should cover anything made by human hand. Since then, hundreds of patents have been issued for genetically engineered organisms, mainly bacteria. In 1986, the US Department of Agriculture approved the sale of the first living genetically altered organism—a virus, used as a pseudorabies vaccine, from which a single gene had been removed. Since then several hundred patents have been awarded for genetically altered bacteria and plants. In 1987, the Patent Office ruled to issue patents for non-human, multicellular organisms including animals produced by genetic engineering (not by natural breeding!). The examples include genetically engineered pigs by germ cell gene therapy to have human growth hormone gene to grow up faster, a goat and sheep chimera called geep, and many transgenic mice. Humans modified genetically cannot be patented but the techniques can be. By extension, the Patent Office also issues patents for genes. Among the patent holders, an interesting one is the NIH itself. One day, it may be possible that large biotechnology companies may hold the patents for all livestock genomes.

Like all great moral issues, there will never be a permanent consensus in bioethics. Each society will reach a temporary solution that seems to make sense in their times. One should remember that in 1974, recombinant technology was banned in the USA. When five years had elapsed, it was thought to be an appropriate technique to use. Today its use for good causes is enormously popular and economically rewarding.

Ethics of Cloning

In vitro fertilization (including those using a donated oocyte), insemination with donor sperm, intracytoplasmic sperm injection (ICSI) are the recent techniques which enable individuals who would otherwise could not have a child to have one. These means of having a child have been widely accepted without any major ethical concern despite that a man with abnormal sperm and a homosexual woman can now reproduce just like anybody else. As a potential use of cloning, what if a lesbian couple wants to have a baby using one's oocyte and a nucleus from the other? Cloning creates a clone of one parent (the source of the nucleus), but not a shared descendant of both the father and mother (except the contribution of the mitochondrial genome by the female). It can be predicted that in some cases public opinion for cloning may be favorable. For example, if the male partner is sterile, it may be acceptable for this couple to have a baby through cloning. The mother would still contribute with her mitochondrial genome, intrauterine influences and subsequent nurture. From now on, the technical barrier has been overcome and it is the moral barrier that tops the agenda in cloning research. Would cloning be used to create second-class citizens or would it revive slavery? If so, should it still be banned considering the fact that these have been achieved without using high technology anyway? There are also objections to human cloning in terms of the social prejudice such children will have to face. But, is it going to be any different from what already happens to children of mixed-race couples? For therapeutic abortion, on the other hand, there are worries that an embryo is being 'killed' to treat somebody. Is it really a case that an unimplanted conceptus (that can only be called a pre-embryo) can be seen as a living subject? There is a philosophical point of view that creating human life for the sole purpose of preparing therapeutic material would not conform to the dignity of life principle. In the UK, currently, the use of human eggs is illegal if the intent is to create an embryo even only for cell replacement (therapeutic cloning). In most countries, including USA, legislation does not exist to stop therapeutic cloning. It is then simply a matter of professional ethics. Cloning allows a woman with a mitochondrial DNA-linked disease to have a healthy baby that would be impossible otherwise. What

would be the public's reaction to such an attempt? It is clear that there will be medically justified uses of cloning in humans, what is not clear is that if any license is issued for any application of cloning, who is going to draw the line for further applications? Cloning is such a technique that shortcuts the safeguards imposed by sexual reproduction. Even a sterile person can now have a child. Patients with cancer are routinely offered storage of their gametes before they are treated by chemotherapy or radiotherapy after which they would usually become sterile. With cloning, they can have a child anytime. Since cancer has a genetic component, are we not going to keep these genes in the population at higher than ever frequencies by doing so? If eugenics is wrong, is this, the opposite of eugenics, right?

Eugenics

The success of artificial breeders in improving the inherited characters of domesticated animals, cultivated plants even microorganisms raised the issue whether the course of the human evolution can also be changed. Eugenics is the false science of improving the quality of the human species through selective breeding. The word eugenics comes from the Greek for good genes. Any policy that is thought by advocates to stimulate the prevalence of 'good genes' is considered eugenic in its effect. Its origin goes back to earlier times. Plato's Republic describes a society in which there is a continuous selection to improve humans through selective breeding. In modern times, the establishment of social Darwinism paved the way for eugenic movements. Modern eugenics relies on the idea that careful planning through selective breeding is the key to improve society. Eugenics supplies a biological or genetic interpretation to its means and aims. If it is a particular race that is to be targeted, the eugenicist will first offer a so-called scientific basis for such a plan. This usually consists of statistical 'evidence' that the race in question is less capable of achievement, more prone to anti-social behavior, or responsible for a prevalent social problem. Most importantly and most of the time wrongly, the eugenicist will insist that this 'inferiority' has genetic basis. In 1900, with the birth of modern genetics together with the belief that humans are the superior species, the interest in improving the human race led to the eugenics movement. There are two basic types of eugenic action:

1. Negative eugenics emphasizes the restriction on reproduction of unfit types. The idea is to improve the human species by identifying individuals and couples at risk of maintaining and spreading inferior genes and to prevent such persons from reproducing.

2. Positive eugenics encourages the reproduction of 'high quality' individuals. Very often, however, the identification of 'good' hereditary human traits is a subjective and even political matter.

Many organizations devoted to eugenic purposes arose around the world, but the movement was especially strong in England, the United States, and Germany between 1910 and 1940. From the beginning, the movement was closely associated with a sense of white Anglo-Saxon superiority. Sir Francis Galton (a cousin of Charles Darwin) is the founder of the English eugenics movement. He coined the term eugenics in 'Inquiries into Human Faculty' in 1883 and continued to advocate his ideas until his death in 1911. He had been drawn to the study of human heredity and eugenics by his curiosity by the hereditary genius in his own family. Galton, who was primarily a statistician, founded a eugenics laboratory and established a research scholarship of eugenics at University College, London in 1904. In his will, he provided funds for a chair of eugenics at University College, London University. The fellowship and later the chair at University College were both occupied by Karl Pearson, a brilliant mathematician who helped to

create the science of biometry (the statistical aspects of biology). In his book, *Hereditary Genius* (1869), Galton proposed that a system of arranged marriages between selected men and women would produce a gifted race. In another book, *Natural Inheritance* (1889), Galton developed statistical methods to the study of man. He was the first to recognize the value of the study of twins for research in heredity. Interestingly, Galton's eugenics movement did not gain wide acceptance, because of the lack of scientific and technical foundation. Moving on the same path after Galton, Pearson felt that the high birth rate of the poor was a threat to civilization. Pearson became the Galton Professor of eugenics at University College in 1911. He shares the blame for the discredit later brought on eugenics in the United States and for making possible the dreadful misuse of the word eugenics in Adolf Hitler's propaganda. The English Eugenics Society, founded by Galton in 1907 as the Eugenics Education Society, opposed Pearson's views but was unable to stop the growing racial discrimination of that time. The Eugenics Society (England) is now known as the Galton Institute.

In the United States, eugenics exerted considerable influence on popular opinion and was reflected in some state and federal legislation. The American Eugenics Society was founded in 1926 by men who believed that the white race was superior to other races. They even thought that the 'Nordic' white was superior to other whites. They thought of races as discrete groups. They did not know that all races are mixtures of many types, the distribution of genes among the races varying in proportions rather than in kind. The American Eugenics Society promoted the idea that the upper classes had superior hereditary qualities that justified their being the ruling class. The result of these activities was the passage of the Immigration Act of 1924 (the Johnson Act), which limited quota immigrants to about 150,000 annually. It was a coalition of eugenicists and some big-business interests who pushed through the Johnson Act which limited immigration into the United States from eastern European and Mediterranean countries. Later, it became clear that the material the eugenicists had presented to congressional hearings had little scientific foundation.

Another consequence was the sterilization laws. Between 1907 and 1943, 30 states in the USA passed sterilization laws. By 1935, sterilization laws had also been passed in Denmark, Switzerland, Germany, Norway, and Sweden. Most of these laws provided for the voluntary or compulsory sterilization of insane, mentally retarded, epileptic, criminal and sexually deviant people. In California, sterilizations averaged more than 350 cases per year, with a total of 9,931 by 1935. Laws were also passed restricting marriages between members of various racial groups. The American Society survives and flourishes to the present day, although, since a name change in 1973, it has been known as the Society for the Study of Social Biology. The new name is not believed to reflect an alteration in its goals. Unquestionably the greatest abuse of the concept of eugenics took place in Hitler's Germany, when as a rationale for producing a 'master race', the Nazis murdered millions of people considered to have inferior genes. Eugenicists were embarrassed by Hitler. After the war, they instituted various strategies to cover up the collaboration that had existed between German, American, and English eugenicists. For example, they adopted a policy of 'crypto-eugenics' and founded cover organizations like the Population Council and the International Planned Parenthood Federation to carry out their aims.

After the German Nazis used eugenics against Jews, Gypsies, the insane, and homosexuals, the assumptions of eugenicists came under sharp criticism which led to the discreditation of eugenics. Recent developments in the diagnosis and treatment of genetic defects have stirred up a eugenics debate within the wider context of medical ethics. Since the 1950s there has been a renewed interest in eugenics. Because certain diseases are now known to be genetically transmitted, many couples choose to undergo

genetic screening, in which they learn the chances that their offspring might be affected. The practice of modern genetic counseling is in a way a eugenic activity, in that it attempts to prevent the conception or birth of individuals with most serious forms of maldevelopment who would be burdens to themselves and to their families. This form of negative eugenics identifies individuals and couples at risk of perpetuating genes that lead to heritable diseases and disorders. It is, however, important that information on these risks is given to couples so that they can make informed and personal decisions about reproduction without societal pressure.

Counterbalancing this trend, however, has been medical progress that enables victims of many genetic diseases to live fairly normal lives and even to reproduce. Genetic surgery, in which harmful genes are altered by direct manipulation, is also being studied. It could obviate eugenic arguments for restricting reproduction among those who carry harmful genes. Such conflicting innovations have complicated the controversy surrounding eugenics. Furthermore, the concept of eugenics tends to ignore the sizable role that environment plays in the establishment of human characteristics. Suggestions for expanding eugenics programs, which range from the creation of sperm banks for the genetically superior to the potential cloning of human beings, have met with vigorous resistance from the public, which often views such programs as unwarranted interference with nature or as opportunities for abuse by authoritarian regimes. Thus, the use of eugenics as happens in modern genetics today is generally acceptable but the potential for the use of the same principle for racist purposes still disconcerts many societies. From this point of view, the situation is similar to the ethical approaches to cloning. Its potential contributions to human health are unquestionable but the possibility exists that once the method is perfected, it can fall into wrong hands.

Further reading

Journal of Medical Ethics, April 1999 issue (devoted to the New Genetics and Ethics)
Nature, 16 Oct 1997, pp.658-663 (Briefing: Bioethics) and 16 Dec 1999, pp.743-746
(The Future of Cloning)

Classic Cases in Medical Ethics: Accounts of Cases That Have Shaped Medical Ethics (Pence & Pence, 1999)
Ethics of Research With Human Subjects: Selected Policies & Resources
(Sugerman, 1998)

Internet Resources

Center for Bioethics at the University of Pennsylvania School of Medicine
The Genetic Revolution: Ethical Issues **Genetics & Ethics**
Public Perspectives on Human Cloning **Eugenics Watch**
Belmont Report on Ethical Principles of Research on Human Subjects
Access BioExcellence Ethics Links **Ethics in Medicine (Washington University)**
NIH BioEthics Guide **NIH Ethics & Genetics**
NIH Office of Human Subjects Research **Guidance On The Research Use Of**
Stored Samples Or Data
NCI Regulations on Research with Human Subjects **Use Of Human Subjects In**
Research
Online IRB Training **NIH Human Participant Protections Education (Online)**
Online Bioethics Course

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Nov 24, 2003

[Back to Contents](#) [Back to Evolution](#) [Back to HLA](#) [Back to MHC](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#)
[Homepage](#)

BASIC POPULATION GENETICS

M.Tevfik Dorak, M.D., Ph.D.

G.H. Hardy (the English mathematician) and W. Weinberg (the German physician) independently worked out the mathematical basis of population genetics in 1908. Their formula predicts the expected genotype frequencies using the allele frequencies in a diploid Mendelian population. They were concerned with questions like "what happens to the frequencies of alleles in a population over time?" and "would you expect to see alleles disappear or become more frequent over time?"

Hardy and Weinberg showed in the following manner that if the population is very large and random mating is taking place, allele frequencies remain unchanged (or in equilibrium) over time unless some other factors intervene. If the frequencies of allele A and a (of a biallelic locus) are p and q, then $(p + q) = 1$. This means $(p + q)^2 = 1$ too. It is also correct that $(p + q)^2 = p^2 + 2pq + q^2 = 1$. In this formula, p^2 corresponds to

the frequency of homozygous genotype AA, q^2 to aa, and $2pq$ to Aa. Since 'AA, Aa, aa' are the three possible genotypes for a biallelic locus, the sum of their frequencies should be 1. In summary, Hardy-Weinberg formula shows that:

$$p^2 + 2pq + q^2 = 1$$

AA Aa aa

If the observed frequencies do not show a significant difference from these expected frequencies, the population is said to be in Hardy-Weinberg equilibrium (HWE). If not, there is a violation of the following assumptions of the formula, and the population is not in HWE.

The assumptions of HWE

1. Population size is effectively infinite,
2. Mating is random in the population (the most common deviation results from inbreeding),
3. Males and females have similar allele frequencies,
4. There are no mutations and migrations affecting the allele frequencies in the population,
5. The genotypes have equal fitness, i.e., there is no selection.

The Hardy-Weinberg law suggests that as long as the assumptions are valid, allele and genotype frequencies will not change in a population in successive generations. Thus, any **deviation from HWE** may indicate:

1. Small population size results in random sampling errors and unpredictable genotype frequencies (a real population's size is always finite and the frequency of an allele may fluctuate from generation to generation due to chance events),
2. Assortative mating which may be positive (increases homozygosity; self-fertilization is an extreme example) or negative (increases heterozygosity), or inbreeding which increases homozygosity in the whole genome without changing the allele frequencies. Rare-male mating advantage also tends to increase the frequency of the rare allele and heterozygosity for it (in reality, random mating does not occur all the time),
3. A very high mutation rate in the population (typical mutation rates are $< 10^{-5}$ per generation) or massive migration from a genotypically different population interfering with the allele frequencies,
4. Selection of one or a combination of genotypes (selection may be negative or positive). Another reason would be unequal transmission ratio of alternative alleles from parents to offspring (as in mouse *t*-haplotypes).

The implications of the HWE

1. The allele frequencies remain constant from generation to generation. This means that hereditary mechanism itself does not change allele frequencies. It is possible for one or more assumptions of the equilibrium to be violated and still not produce deviations from the expected frequencies that are large enough to be detected by the goodness of fit test,

2. When an allele is rare, there are many more heterozygotes than homozygotes for it. Thus, rare alleles will be impossible to eliminate even if there is selection against homozygosity for them,
3. For populations in HWE, the proportion of heterozygotes is maximal when allele frequencies are equal ($p = q = 0.50$),
4. An application of HWE is that when the frequency of an autosomal recessive disease (e.g., sickle cell disease, hereditary hemochromatosis, congenital adrenal hyperplasia) is known in a population and unless there is reason to believe HWE does not hold in that population, the gene frequency of the disease gene can be calculated (for an example visit the [Cancer Genetics](#) website and choose Topics).

It has to be remembered that when HWE is tested, mathematical thinking is necessary. When the population is found in equilibrium, it does not necessarily mean that all assumptions are valid since there may be counterbalancing forces. Similarly, a significant deviance may be due to sampling errors (including **Wahlund effect**, see below and [Glossary](#)), misclassification of genotypes, measuring two or more systems as a single system, failure to detect rare alleles and the inclusion of non-existent alleles. The Hardy-Weinberg laws rarely holds true in nature (otherwise evolution would not occur). Organisms are subject to mutations, selective forces and they move about, or the allele frequencies may be different in males and females. The gene frequencies are constantly changing in a population, but the effects of these processes can be assessed by using the Hardy-Weinberg law as the starting point.

The direction of departure of observed from expected frequency cannot be used to infer the type of selection acting on the locus even if it is known that selection is acting. If selection is operating, the frequency of each genotype in the next generation will be determined by its relative fitness (W). Relative fitness is a measure of the relative contribution that a genotype makes to the next generation. It can be measured in terms of the intensity of selection (s), where $W = 1 - s$ [$0 \leq s \leq 1$]. The frequencies of each genotype after selection will be $p^2 W_{AA}$, $2pq W_{Aa}$, and $q^2 W_{aa}$. The highest fitness is always 1 and the others are estimated proportional to this. For example, in the case of heterozygote advantage (or overdominance), the fitness of the heterozygous genotype (Aa) is 1, and the fitnesses of the homozygous genotypes negatively selected are $W_{AA} = 1 - s_{AA}$ and $W_{aa} = 1 - s_{aa}$. It can be shown mathematically that only in this case a stable polymorphism is possible. Other selection forms, underdominance and directional selection, result in unstable polymorphisms. The weighted average of the fitnesses of all genotypes is the mean fitness. It is important that genetic fitness is determined by both fertility and viability. This means that diseases that are fatal to the bearer but do not reduce the number of progeny are not genetic lethals and do not have reduced fitness (like the adult onset genetic diseases: Huntington's chorea, hereditary hemochromatosis). The detection of selection is not easy because the impact on changes in allele frequency occur very slowly and selective forces are not static (may even vary in one generation as in antagonistic pleiotropy).

All discussions presented so far concerns a simple biallelic locus. In real life, however, there are many loci which are multiallelic, and interacting with each other as well as with the environmental factors. The Hardy-Weinberg principle is equally applicable to multiallelic loci but the mathematics is slightly more complicated. For multigenic and multifactorial traits, which are mathematically continuous as opposed to discrete, more complex techniques of quantitative genetics are required.

In a final note on the practical use of HWE, it has to be emphasised that **its violation** in daily life is most frequently due to genotyping errors. Allelic misassignments, as frequently happens when PCR-SSP method is used, sometimes due to allelic dropout are the most frequent causes of the Hardy-Weinberg disequilibrium. When this is observed, the genotyping protocol should be reviewed. In a case-control association study, it is of paramount importance that the control group is in HWE to rule out any technical errors. The violation of HWE in the case group, however, may be due to a real association.

Some concepts relevant to HWE

Wahlund effect: Reduction in observed heterozygosity (increased homozygosity) because of considering pooled discrete subpopulations that do not interbreed as a single randomly mating unit. When all subpopulations have the same gene frequencies, no variance among subpopulations exists, and no Wahlund effect occurs ($F_{ST}=0$). **Isolate breaking** is the phenomenon that the average homozygosity temporarily increases when subpopulations make contact and interbreed (this is due to decrease in homozygotes). It is the opposite of Wahlund effect.

F statistics: The F statistics in population genetics has nothing to do the F statistics evaluating differences in variances. Here **F** stands for fixation index, fixation being increased homozygosity resulting from inbreeding. Population subdivision results in the loss of genetic variation (measured by heterozygosity) within subpopulations due to their being small populations and genetic drift acting within each one of them. This means that population subdivision would result in decreased heterozygosity relative to that expected heterozygosity under random mating as if the whole population was a single breeding unit. Wright developed three fixation indices to evaluate population subdivision: F_{IS} (interindividual), F_{ST} (subpopulations), F_{IT} (total population).

F_{IS} is a measure of the deviation of genotypic frequencies from panmictic frequencies in terms of heterozygous deficiency or excess. It is what is known as the **inbreeding coefficient (f)**, which is conventionally defined as the probability that two alleles in an individual are identical by descent (autozygous). The technical description is the correlation of uniting gametes relative to gametes drawn at random from within a subpopulation (**I**ndividual within the **S**ubpopulation) averaged over subpopulations. It is calculated in a single population as $F_{IS} = 1 - (H_{OBS} / H_{EXP})$ [equal to $(H_{EXP} - H_{OBS}) / H_{EXP}$] where H_{OBS} is the observed heterozygosity and H_{EXP} is the expected heterozygosity calculated on the assumption of random mating. It shows the degree to which heterozygosity is reduced below the expectation. The value of F_{IS} ranges between -1 and +1. Negative F_{IS} values indicate heterozygote excess (outbreeding) and positive values indicate heterozygote deficiency (inbreeding) compared with HWE expectations.

F_{ST} measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. F_{ST} is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called **coancestry coefficient (θ)** [Weir & Cockerham, 1984] or '**Fixation index**' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (**S**ubpopulation within the **T**otal population). It is calculated as using the subpopulation (average) heterozygosity and total population expected heterozygosity. F_{ST} is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the population) and 1 = complete isolation (extreme subdivision). F_{ST} values up to 0.05 indicate negligible genetic differentiation whereas >0.25 means very great genetic differentiation within the population analyzed. F_{ST} is usually calculated for different genes, then averaged across all loci, and all populations. F_{ST} can also be used to estimate gene flow: $0.25 (1 - F_{ST}) / F_{ST}$. This highly versatile parameter is even used as a genetic distance measure between two populations instead of a fixation index among many populations (see [Weir BS, Genetic Data Analysis II, 1996](#); and [Kalinowski ST, 2002](#)).

F_{IT} is rarely used. It is the overall inbreeding coefficient (F) of an individual relative to the total population (**I**ndividual within the **T**otal population).

Detecting Selection Using DNA Polymorphism Data

Several methods have been designed to use DNA polymorphism data (sequences and allele frequencies) to obtain information on past selection events. Most commonly, the ratio of non-synonymous (replacement) to synonymous (silent) substitutions (d_N/d_S ratio; see below) is used as evidence for overdominant selection (balancing selection) of which one form is heterozygote advantage. Classic example of this is the mammalian **MHC** system genes and other **compatibility systems** in other organisms: the self-incompatibility system of the plants, fungal mating types and invertebrate allorecognition systems. In all these genes, a very high number of alleles is also noted. This can be interpreted as an indicator of some form of balancing (diversifying) selection. In the case of neutral polymorphism, one common allele and a few rare alleles are expected. The frequency distribution of alleles is also informative. Large number of

alleles showing a relatively even distribution is against neutrality expectations and suggestive of diversifying selection.

Most tests detect selection by rejecting neutrality assumption (observed data is deviate significantly from what is expected under neutrality). This deviation, however, may also be due to other factors such as changes in population size or genetic drift. The original neutrality test was [Ewens-Watterson homozygosity test](#) of neutrality (see [Glossary](#)) based on the comparison of observed homozygosity and predicted value calculated by [Ewens's sampling formula](#) which uses the number of alleles and sample size. This test is not very powerful.

Other commonly used statistical tests of neutrality are Tajima's D (theta), Fu & Li's D, D* and F. Tajima's test ([Tajima F, Genetics 1989](#)) is based on the fact that under the neutral model estimates of the number of segregating/polymorphic sites and of the average number of nucleotide differences are correlated. If the value of D is too large or too small, the neutral 'null' hypothesis is rejected. [DnaSP](#) calculates the D and its confidence limits (two-tailed test). Tajima did not base this test on coalescent but Fu and Li's tests ([Fu & Li, Genetics 1993](#)) are directly based on coalescent. The tests statistics D and F require data from intraspecific polymorphism and from an outgroup (a sequence from a related species), and D* and F* only require intraspecific data. [DnaSP](#) uses the critical values obtained by [Fu & Li, Genetics 1993](#) to determine the statistical significance of D, F, D* and F* test statistics. [DnaSP](#) can also conduct the Fs test statistic ([Fu YX, Genetics 1997](#)). The results of this group of tests (Tajima's D and Fu & Li's tests) based on allelic variation and/or level of variability may not clearly distinguish between selection and demographic alternatives (bottleneck, population subdivision) but this problem only applies to the analysis of a single locus (demographic changes affect all loci whereas selection is expected to be locus-specific which are distinguishable if multiple loci are analysed). Tests for multiple loci include the **HKA test** described by Hudson, Kreitman and Aguade ([Genetics, 1987](#)). This test is based on the idea that in the absence of selection, the expected number of polymorphic (segregating) sites within species and the expected number of 'fixed' differences between species (divergence) are both proportional to the mutation rate, and the ratio of them should be the same for all loci. Variation in the ratio of divergence to polymorphism among loci suggests selection.

A different group of neutrality tests that are not sensitive to demographic changes include McDonald-Kreitman test ([McDonald JH & Kreitman M, Nature 1991](#)) and d_N/d_S ratio test. McDonald-Kreitman test compares the ratio of the number of nonsynonymous to synonymous 'polymorphisms' within species to that ratio of the number of nonsynonymous to synonymous 'fixed' differences between species in a 2x2 table (see a worked example [here](#)). The most direct method of showing the presence of positive selection is to compare the number of nonsynonymous (d_N) to the number of synonymous (d_S) substitutions in a locus. A high (>1) value of (d_N/d_S) substitutions suggest fixation of nonsynonymous mutations with a higher probability than neutral (synonymous) ones. Statistical properties of this test are given by [Goldman N & Yang Z, Mol Biol Evol, 1994](#) and by [Muse SV & Gaut BS, Mol Biol Evol, 1994](#). The d_N/d_S ratio tests take into account of transition/transversion rate bias and codon usage bias.

For other tests and software to perform these statistics, see [DNA Sequence Polymorphism, DnaSP](#). See also: [Statistical Tests of Neutrality \(Lecture Note by P Beerli\)](#); [Statistical Tests of Neutrality of Mutations against Excess of Recent Mutations \(Rare Allels\)](#); [Statistical Tests of Neutrality of Mutations against an Excess of Old Mutations or a Reduction of Young Mutations](#); [Estimation of theta](#); Properties of statistical tests of neutrality for DNA polymorphism data, [Simonsen et al Genetics 1995;141:413-29](#). Review of statistical tests of selective neutrality on genomic data, [Nielsen R, Heredity 2001](#); and a [Lecture Note by Gil McVean](#).

Linkage disequilibrium (LD)

The tendency for two 'alleles' to be present on the same chromosome (positive LD), or not to segregate together (negative LD). As a result, specific alleles at two different loci are found together more or less than expected by chance. The same situation may exist for more than two alleles. Its magnitude is expressed as the delta (Δ) value and corresponds to the difference between the expected and the observed haplotype frequency (see [Measures of LD by Devlin & Risch, 1995](#) for further details). It can have positive or negative values. LD is decreased by recombination. Thus, it decreases every generation of random mating unless some process opposing the approach to linkage 'equilibrium'. Permanent LD may

result from natural selection if some gametic combinations confer higher fitness than other combinations. For more on LD, see [Statistical Analysis in HLA and Disease Association Studies](#). Link to a lecture on [linkage disequilibrium](#); [online LD analysis](#). Software to perform LD analysis: [Genetic Data Analysis, EH, 2LD, MLD, PopGene, Arlequin 2000](#), and [Online Easy LD](#). Please note that LD has nothing to do HWE and should not be confused with it (see [Possible Misunderstandings in Genetics](#)).

Genetic distance (GD)

Genetic distance is a measurement of genetic relatedness of samples of populations (whereas genetic diversity represents diversity within a population). The estimate is based on the number of allelic substitutions per locus that have occurred during the separate evolution of two populations. (See lecture notes on [Genetic Distances, Estimating Genetic Distance](#); and [GeneDist: Online Calculator of Genetic Distance](#). The software [Arlequin](#), [PHYLIP](#), [GDA](#), [PopGene](#), [Populations](#) and [SGS](#) are suitable to calculate population-to-population genetic distance from allele frequencies.

Genetic Distance can be computed on freeware PHYLIP. Most components of PHYLIP are available [on the web](#). One component of the package [GENDIST](#) estimates genetic distance from allele frequencies using one of the three methods: Nei's, Cavalli-Sforza's or Reynold's (see papers by [Cavalli-Sforza & Edwards, 1967](#), [Nei et al, 1983](#), [Nei M, 1996](#) and [lecture note \(1\)](#) and [\(2\)](#) for more information on these methods). GENDIST can be run [online](#) using default options ([Nei's genetic distance](#)) to obtain genetic distance matrix data. The PHYLIP program [CONTML](#) estimates phylogenies from gene frequency data by maximum likelihood under a model in which all divergence is due to genetic drift in the absence of new mutations (Cavalli-Sforza's method) and draws a tree. The program is also available [on the web](#) and runs with default options. If new mutations are contributing to allele frequency changes, Nei's method should be selected on GENDIST to estimate genetic distances first. Then a tree can be obtained using one of the following components of PHYLIP: [NEIGHBOR](#) also draws a phylogenetic tree using the genetic distance matrix data (from GENDIST). It uses either [Nei and Saitou's](#) (1987) "[Neighbor Joining \(NJ\) Method](#)," or the [UPGMA](#) (unweighted pair group method with arithmetic mean; average linkage clustering) method ([Sneath & Sokal, 1973](#)). [Neighbor Joining](#) is a distance matrix method producing an unrooted tree without the assumption of a clock (the evolutionary rate does not have to be the same in all lineages). Major assumption of UPGMA is equal rate of evolution along all branches (which is frequently unrealistic). NEIGHBOR can be run [online](#). Other components of PHYLIP that draw phylogenetic trees from genetic distance matrix data are [FITCH](#) / [online](#) (Fitch-Margoliash method with no assumption of equal evolutionary rate) and [KITSCH](#) / [online](#) (employs Fitch-Margoliash and Least Squares methods with the assumption that all tip species are contemporaneous, and that there is an evolutionary clock -in effect, a molecular clock). ([Mathematical formulae of various genetic distance measures.](#)) Another freeware [PopGene](#) calculates Nei's genetic distance and creates a tree using UPGMA method from genotypes. For genetic distance calculation on Excel, try freeware [GenAIEx](#) by Peakall & Smouse.

Because of different assumptions they are based on the NJ and UPGMA methods may construct dendrograms with totally different topologies. For an example of this and a review of main differences between the two methods, see [Nei & Roychoudhury, 1993 \(free full-text\)](#). Both methods use distance matrices (also Fitch-Margoliash and Minimal Evolution methods are distance methods). The principle difference between NJ and UPGMA is that NJ does not assume an equal evolutionary rate for each lineage. Since the constant rate of evolution does not hold for human populations, NJ seems to be the better method. For the genetic loci subject to natural selection, the evolutionary rate is not the same for each population and therefore UPGMA should be avoided for the analysis of such loci (including the HLA genes). The leading group in HLA-based genetic distance analysis led by [Arnaiz-Villena](#) proposes that the most appropriate genetic distance measure for the HLA system is the DA value first described by [Nei et al, 1983](#). Unlike UPGMA, NJ produces an unrooted tree. To find the root of the tree, one can add an outgroup. The point in the tree where the edge to the outgroup joins is the best possible estimate for the root position. One persistent problem with tree construction is the lack of statistical assessment of the phylogenetic tree presented. This is best done with widely available bootstrap analysis originally described in Felsenstein J:

Evolution 1985;39:783-791 (available through [JSTOR](#) if you have access) and [Efron et al. 1996](#); and reviewed in [Nei M, 1996](#)). For a discussion of statistical tests of molecular phylogenies, see [Li & Gouy, 1990](#) and [Nei M, 1996](#). For the topology to be statistically significant the bootstrap value for each cluster should reach at least 70% whereas 50% overestimates accuracy of the tree. Bootstrap tests should be done with at least 1000 (preferably more) replications.

Nei noted that some genes are more suitable than others in phylogenetic inference and that most tree-building methods tend to produce the same topology whether the topology is correct or not [Nei M, 1996](#). He also added that sometimes adding one more species/population would change the whole tree for unknown reasons. An example of this has been provided in a study of human populations with genetic distances [Nei & Roychoudhury, 1993](#). The properties of most popular genetic distance measures have been reviewed ([Kalinowski, 2002](#)). Whichever is used, large sample sizes are required when populations are relatively genetically similar, and loci with more alleles produce better estimates of genetic distance. However, in a simulation study, Nei et al concluded that more than 30 loci should be used for making phylogenetic trees ([Nei et al, 1983](#)). There seems to be a consensus that estimated trees are nearly always erroneous (i.e., the topological arrangement will be wrong) if the number of loci is less than 30 ([Nei M, 1996](#); Jorde LB. Human genetic distance studies. Ann Rev Anthropol 1985;14:343-73; available through [JSTOR](#) if you have access). If populations are closely related even 100 loci may be necessary for an accurate estimation of the relationships by genetic distance methods. Cavalli-Sforza et al have noted important correlations between the genetic trees and linguistics evolutionary trees with the exceptions for New Guinea, Australia and South America ([Cavalli-Sforza et al. 1994](#)).

Especially for the HLA genes, phylogenetic trees can be constructed by using the Nei's DA genetic distance values and NJ method with bootstrap tests on [DISPAN](#). [Correspondence analysis](#), a supplementary analysis to genetic distances and dendrograms, displays a global view of the relationships among populations ([Greenacre MJ, 1984](#); [Greenacre & Blasius, 1994](#); [Blasius & Greenacre, 1998](#)). This type of analysis tends to give results similar to those of dendrograms as expected from theory ([Cavalli-Sforza & Piazza, 1975](#)), and is more informative and accurate than dendrograms especially when there is considerable genetic exchange between close geographic neighbors ([Cavalli-Sforza et al. 1994](#)). In their enormous effort to work out the genetic relationships among human populations, Cavalli-Sforza et al concluded that two-dimensional scatter plots obtained by correspondence analysis frequently resemble geographic maps of the populations with some distortions ([Cavalli-Sforza et al. 1994](#)). Using the same allele frequencies that are used in phylogenetic tree construction, [correspondence analysis](#) using allele frequencies can be performed on the [ViSta \(v7.0\)](#), [VST](#), [SAS](#) but most conveniently on Multi Variate Statistical Package [MVSP](#). Link to [a tutorial on correspondence analysis](#).

Internet Links

[History of Population Genetics and Evolution in A History of Genetics](#) by AH Sturtevant

[ASHI 2001 Biostatistics and Population Genetics Workshop Notes](#)

[Microsatellites and Genetic Distance \(Primer on Genetic Distance\)](#)

[HWE in Kimball's Biology Pages](#) [Online HWE Test](#) [Online GD Calculation](#) [Online Easy LD](#)

[Population Genetics Simulations](#) [Molecular Evolution / Computational Pop Genet Course](#)

[Lectures on Population Genetics \(1\) & \(2\) & \(3\) & \(4\) & \(5\)](#)

[Statistical Genetics Websites](#)

[Freeware Population Genetic Data Analysis Software \(List of Features\):](#)

[Arlequin 2000](#) [PopGene](#) [GDA](#) [Genetix](#) [GenePop](#) [GeneStrut](#) [SGS](#) [TFPGA](#) [MVSP](#)
[PHYLIP\(Online\)](#) [DISPAN](#) [ViSta](#) [GenAlEx](#) [CLUMP](#) [TDT](#) [HAPLOTYPER](#) [PHASE v2.0](#)
[EasyLD](#)

[MSA \(for microsatellite data\)](#) [POPULATIONS](#) [WINPOP v2.0](#) [QUANTO](#)

[GSF: Genetic Software Forum](#) [Partition for Online Bayesian Analysis](#)

[Comprehensive List of Genetic Analysis Software \(1\) \(2\) \(3\)](#)

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated March 19, 2004

[Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#)
[Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

ELEMENTARY EVOLUTIONARY BIOLOGY

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

[Introduction](#)

[A Brief History of Life](#) [Human Evolution](#)

[Basic Population Genetics](#) [Frequency Dependent Selection](#)

[Sexual Selection](#) [Evolution of Sexual Reproduction](#)

[Molecular Clock](#) [Mitochondrial DNA](#) [Mimicry](#)

[Speciation](#) [Cladistics](#)

[Major Histocompatibility Complex](#)

[Links to Evolution-Related Sites](#) [Glossary](#)

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

INTRODUCTION TO EVOLUTIONARY BIOLOGY

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

These are not meant to be comprehensive notes but just brief ones on basic evolutionary biology at very introductory level. This page is intended to be the basis for the following discussions of more specialized topics.

Contents

1. Major events in evolutionary history
2. Basic concepts in evolutionary biology
 Natural Selection, Genetic Drift, Mutation
3. Death & extinction
4. Some rules and theories in evolutionary biology

Major events in evolutionary history

Evolution of eukaryotes from a prokaryotic ancestral cell
Photosynthesis

Evolution of lungs in amphibians and then land animals

Evolution of amnion, allantois and shelled egg (the conquest of dry land by vertebrates)

Evolution of feathers and wings leading to the evolution of flight

Upright walking and tool using
Artistic inclination and thinking

Basic concepts in evolutionary biology

Evolution is the result of the simultaneous occurrence of multiple causes including (stochastic) chance phenomena and the more deterministic selective phenomena. In the production of variation, chance dominates, while selection itself operates largely by necessity. Any property that evolves must be controlled by a gene that can vary. Evolution is more than merely a change in gene frequencies. It also includes the origin of the variation.

Four processes account for most of the changes in allele frequencies in population

Those that increase genetic variation:

1. Mutation - recombination
2. Migration (gene flow)

Those that decrease genetic variation:

3. Natural selection (stabilizing selection)
4. Random genetic drift (3 & 4 are the most important ones)
- (5). Meiotic drive

All but genetic drift are deterministic ones.

There are two evolutionary dimensions: time (adaptedness) and space (speciations and multiplication of lineages).

Evolution is not progress. It is about adaptation to surroundings. This does not necessarily mean improvement over a period of time. Evolution cannot be equated with progress. However, no progress can take place without it.

Probably all traits are a product of genetic and environmental effects. In general, the more genes are involved, the more continuous is the variation. According to the Fundamental Theorem of Natural Selection, the rate of adaptive change is proportional to the amount of genetic variation in the population. The rate of increase in fitness is equal to the genic variance in the population [RA Fisher 1930].

Evolution can occur without morphological change, and morphological change can occur without evolution. Sibling species are genetically distinct but may be very similar morphologically. Gross evolution tends to be irreversible [Dollo's law]. It would be a mistake, however, to consider evolution totally irreversible. Atavism (the reappearance of certain characters typical of remote ancestors) argues against this.

Only heritable changes contribute to evolution (not environmentally induced phenotypical changes). Behavior evolves through natural selection similar to the evolution of morphological characters.

Darwin's basic principle: Evolution is due to genetic variation and natural selection acting on heritable characters. Charles Darwin recognized natural selection as the mechanism

of evolution in 1838, but did not publish the *Origin of Species* until 1859. Alfred Russel Wallace reached the same conclusion independently in 1858.

Darwin's five major theories

1. The organisms steadily evolve over time (evolution theory)
2. Different kinds of organisms descended from a common ancestor (common descent theory)
3. Species multiply over time (speciation theory)
4. Evolution takes place through the gradual change of populations (gradualism theory)
5. The mechanism of evolution is the competition among vast numbers of unique individuals for limited resources under selective pressures, which leads to differences in survival and reproduction (natural selection theory)

Weaknesses of Darwin's natural selection theory

1. Blending inheritance was favored rather than discrete Mendelian genes (which was unknown at the time)
2. No knowledge of Mendel genetics was available
3. Phyletic gradualism was favored as the type of speciation
4. Fecundity was not emphasized in the description of fitness
5. Sexual selection: Sexually selected characters were seen as ornaments but they may be advertising genuine male qualities (Hamilton & Zuk, 1982)

Evidence used by Darwin for his natural selection theory

1. Biogeography: Distinct features of cosmopolitan species and the presence of endemic species (Darwin's finches: of the 14 finch species of the Galapagos islands, 13 are endemic)
2. Morphology and embryology: Homologous structures among related species; similarities in the embryos of related species
3. Paleontology: Gradual change in the fossil record, evident extinctions
4. Taxonomy and systematics: Morphological similarities among related taxa

In evolution, it is not the survival of the individual that matters but of the offspring of that individual (relevant in altruism).

Natural selection is one mechanism responsible for evolutionary change. It does not act on species or on populations, but on [the phenotypes not genotypes of] individuals.

Genes mutate, individuals are selected, populations evolve. Natural selection is the non-random survival of randomly varying hereditary characteristics. Evolution acts on the portion of variation which is controlled genetically. Natural selection favors traits that enhance reproduction.

No gene is ever directly exposed to selection, but only in the context of an entire genotype, and a gene may have different selective values in different genotypes (i.e., the gene is not the target of selection). An individual is favored by selection owing to the overall quality of its genotype. This is a confusing issue in modern evolutionary biology. This arises from the fact that in classical Darwinian school, it is believed that the unit of selection is the individual but in neo-Darwinism (the New Synthesis), it is the gene. The classical belief that the unit of selection is the individual is true within any one generation, but there is no continuity in individuals in a sexually reproducing species, the only continuity is in the continuation of copies of alleles. This is why the New Synthesis considers selection to act on particular alleles in relation to their average contribution to all the individuals that carry copies of them.

Selection types

1. Stabilizing: Most adaptive character is preserved as long as the adaptive peak does not change. A typical example of stabilizing selection was presented by H. Bumpus in 1899. He measured the wings of house sparrows (*Passer domesticus*) killed in a storm in New York. He found that those with markedly long or short wings were more frequently killed. Stabilizing selection does not allow new variations to emerge. Once a character is optimized, natural selection keeps it as it is like keeping the number of fingers, egg size and number, mate choice adaptations, seasonal timing of migrations, birth weight in humans etc. stable. Both extremes in variation are selected against and eliminated. Aristotle's description of wild animals and plants, written 2,500 years ago, are still accurate today as natural selection must have been preventing their further evolution [from a stable state]. Thus, natural selection cannot be equated to evolution as sometimes it prevents evolution, but it is a major mechanism involved in evolution.

2. Directional: Strong selection favors one of the extreme phenotypes. This type of selection decreases variation (sexual selection of male characters). Antibiotic resistance by bacteria and insecticide resistance by insects are other examples. By favouring those who are resistant, the variation in the population decreases, and the population eventually consists of only resistant individuals. A common form of directional selection causes character displacement when two species compete. When there are two species of finches on an island, each will evolve to have a different size beak (small and large) whereas either species alone has an intermediate beak size elsewhere.

3. Disruptive selection: Both extreme phenotypes are preferred over the intermediate one. This selection occurs as a result of the heterozygous being at a disadvantage, thus the two homozygotes are selected (underdominance). In a bi-allelic polymorphism, if one of the alleles is remarkably rare than the other, the rare allele may be lost. Selection for two different colours in North American lacewings but not for intermediate color is an example. This happens because the two extreme color patterns provide the best camouflage in the two different niches but the intermediate one does not offer any protection. Similarly, the African swallow tail butterfly produces two distinct morphs, both of which mimic distasteful butterflies of other species with aposematic coloration (Batesian mimicry). This type of selection tends to increase or maintain the diversity in the population. It might even cause one species to evolve into two.

Evolution requires genetic variation. Mutation is a change in a gene (variation). Natural selection operates on this variation and the population evolves. Natural selection is the only mechanism and the driving force of adaptive evolution. The most common action of natural selection is to remove unfit variants as they arise via mutation (most mutations are lost due to drift). Selection only distinguishes between existing variants, does not create new ones. Without selection, genetic diversity would be low and primarily controlled by two parameters: how frequently new alleles arise due to mutation (or immigration) and how frequently alleles are lost due to genetic drift, which is dependent on the size of the population.

Constraints on natural selection

1. Natural selection does not induce variability. The genetic variation needed for the most adaptable phenotype may not be available or forthcoming,
2. The different components of the phenotype are dependent of one another and none of them can respond to selection without interacting with the others,

3. Most genes are pleiotropic and most components of the phenotype are polygenic characters,
4. Capacity of organisms for non-genetic modifications: the more plastic the phenotype is (owing to developmental flexibility), the more this reduces the force of adverse selection pressures,
5. Much of the differential survival and reproduction in a population is still the result of chance. This also limits the power of natural selection.

Natural selection is often confused with evolution. The original definition of evolution was descent with modification. This includes both the origin as well as the spread of new variants or traits. Evolution may thus occur as a result of natural selection, genetic drift, or both, as long as there is a continual supply of new variation (such as mutation, recombination, gene flow). Natural selection does not necessarily give rise to evolution. Unlike evolution, natural selection is a non-historical process that depends upon only current ecological and genetic conditions. Evolution depends not only on these current conditions but also upon their entire history. Natural selection deals with frequency changes brought about by differences in ecology among heritable phenotypes; evolution includes this as well as random effects and the origin of these variants.

Natural selection does not have any foresight. It only allows organisms to adapt to their current environment. Selection merely favors beneficial genetic changes when they occur by chance. Neither mutations arise as an adaptive response to the environment, but may prove fortuitously to be adaptive after they arise. Unstable environments drive and induce evolution. Environment does not cause change, it causes the need for change, which is recognized and acted upon by the organism. Think about why all trees in a rain forest are tall. What happened to the short ones?

Most major evolutionary changes occur by the gradual accumulation of minor mutations, accompanied by very gradual phenotypic transitions.

Evidence for evolution by natural selection in contemporary populations:

1. The resistance of the house fly (*Musca domestica*) to DDT first reported in 1947,
2. The change in the frequencies of differently colored peppered moths with industrial revolution in England,
3. Establishment of new HLA alleles in isolated and inbred populations where most subjects would be homozygous and new alleles increasing heterozygosity rate would be favored.

Random Genetic Drift: Populations do not exactly reproduce their genetic constitutions in successive generations. There is a random/chance component of gene-frequency change. In other words, only a fraction of all possible zygotes become mature adults and not all alleles available in parental gene pool are transmitted to the offspring. If a pair of parents have one child, not all of their alleles will have passed on to the next generation. In a large population, the random nature of the process will average out. But, in a small population the effect could be rapid and significant. Random genetic drift is a change in the allele frequencies in a population that cannot be ascribed to the action of any selective process. It is a binomial sampling error of the gene pool.

In principle, any individual may by chance meet with an accident or fail to meet a mate, and so fail to make a contribution to the next generation, irrespective of how well it is adapted. Therefore, its genes will not be represented in the next generation. This is not important in a large population. In small populations, such chance happenings can have

important effects on gene frequencies and the population may drift away from the adaptive peak. Genetic drift can be an aid as well as a hindrance to adaptation.

Like natural selection, drift also decreases genetic variation. There are, however, mechanisms that replace variation depleted by selection and drift (mutation -the most important one-, recombination, gene flow). Genetic drift and natural selection are the two most important mechanisms of evolution. Their relative importance depends on estimated population sizes. Drift is much more important in small populations and in those who breeds in demes. In principle, genetic drift acts on a smaller time scale and natural selection in the long-term.

Genetic bottleneck: Sudden and remarkable reduction in the population size due to natural disasters, disease, or predation (Northern elephant seal population at the end of the last century; American buffalo population in the beginning of this century, and cheetahs). The small number of survivors may represent only a fraction of the original polymorphism present in the population. When the population grows to large numbers, genetic variation may be limited. The cheetah appears to have gone through a population bottleneck, probably at the end of the pleistocene. Because of the limited variation in the population, in the mid-1980s, an outbreak of feline infectious peritonitis caused 50-60% mortality rate showing the importance of genetic diversity in maintaining healthy populations.

Genetic drift caused by bottlenecking may have been important in the early evolution of human populations when calamities decimated tribes. The unselected small group of (lucky) survivors is unlikely to be representative of the original population in its genetic make-up (this is an example of evolution by luck but not by fitness).

When a population has been founded by a few or even by a single gravid female, this population cannot contain more than a fraction of the total genetic variability of the parent population. This is called founder effect. When a population is started by one or a small group of individuals randomly separated from the parent population, chance may dictate that the allele frequencies in the newly founded population will be quite different from those of the parent population. Many species on islands (*Drosophila* of Hawaii, Darwin's finches at Galapagos) display the consequences of founder effects.

In any population, some proportion of loci is fixed at a selectively unfavorable allele because the intensity of selection is insufficient to overcome the random drift to fixation. Drift is intensified as selection pressures increase. This is because, strong selection decreases the effective population size.

If the genetic variation observed in populations is inconsequential to survival and reproduction (ie, neutral), the drift will be the main determinant. If the gene substitutions affect fitness, natural selection is the main driving force.

Mutation provides the new and different genetic raw material for evolution. It is the ultimate source of all new genetic variation. Mutations are frequently reversible, being able to mutate back to normal. Some genes are prone to recurring, often similar, mutations. Advantageous gene mutations are retained in a population, while deleterious ones tend to be eliminated because its inheritors are not as viable.

The rate of mutation per base per replication is approximately 10^{-9} (10^{-5} to 10^{-6} per gene per generation). Chromosomal mutations are more common. For example, reciprocal translocations occur at a rate of 10^{-4} to 10^{-3} per gamete per generation. Most mutations are thought to be neutral with regards to fitness. A change in environment can cause a neutral allele to have a selective value. In the short term, evolution can run on stored variation for which mutation is the ultimate source.

Pseudogenes evolve much faster than their working counterparts (the same applies to introns). Mutations in them do not get incorporated into proteins, so they are not subject to selection. Also, silent nucleotide sites (that can be changed without changing the sequence of the protein) are expected to be more polymorphic than replacement nucleotide sites within a population and show more differences between populations. This is because silent changes are not subject to selection.

As the neutral theory predicts, the rate of evolution is greater in functionally less constrained molecules. Proteins with vital functions cannot tolerate mutations as they would interfere with the viability of the organism.

Li and Graur (1991) calculated the rate of evolution for silent vs. replacement sites in humans and rodents. Silent sites evolved at an average rate of 4.61 nucleotide substitution per 10^9 years. Replacement sites evolved much slower at an average rate of 0.85 nucleotide substitution per 10^9 years.

Most neutral alleles are lost soon after they appear. Alleles are added to the gene pool by mutation at the same rate they are lost to drift. For neutral alleles that do fix, it takes an average of $4N_e$ generations (N_e = effective population size) to do so. When a new allele has a positive selective value, s , the expected time to fixation is less than $4N_e$ generations and is, approximately, $(2/s)\log_e(2N_e)$ generations (Ayala FJ, 1994).

Deleterious mutants are selected against but remain at low frequency in the gene pool. In diploids, a deleterious recessive mutant may increase in frequency due to drift (unopposed by selective forces). Deleterious alleles also remain in populations at a low frequency due to a balance between recurrent mutations and selection (the mutation load). It is estimated that each of us, on average, carries three to five recessive lethal alleles.

Most new mutants, even beneficial ones, are lost due to drift (but may recur many times). An allele that conferred a one percent increase in fitness has only a two percent chance of fixing. A beneficial mutant may be lost several times, but eventually it will arise and stick in a population. Even deleterious mutations may recur. Directional selection depletes genetic variation at the selected locus as the fitter allele sweeps to fixation. Sequences linked to a selected allele also increase in frequency due to hitchhiking. Eventually, recombination will bring the two loci to linkage equilibrium (their association will be random).

Recombination within a gene can form a new allele. Recombination is a mechanism of evolution because it adds new alleles and combinations of alleles to the gene pool.

Genetic drift, gene flow and the breeding structure of a species should, in principle, affect all loci in a similar fashion.

Neutralism: According to the neutral theory of molecular evolution, the great majority of evolutionary changes at the molecular level do not result from natural selection but, rather, from random fixation of selectively neutral or near-neutral mutants through random genetic drift. It assumes that only a minute fraction of molecular changes are adaptive and most mutations are selectively neutral (neither advantageous nor non-advantageous). As a result, polymorphisms are maintained by the balance between mutational input and random extinction. The frequency of synonymous base changes in a population is a matter of genetic drift not natural selection. Molecular changes that are less likely to be subject to natural selection occur more rapidly in evolution. Thus, neutral evolution occurs at higher rate. Initially, a wide variety of observations seemed to be consistent with the neutral theory. Eventually, however, several lines of evidence have emerged arguing against it.

Inbreeding leads to an increase in homozygosity at all loci because the breeding pairs are initially genetically more similar to one other than would be the case if a pair of individuals had been taken at random from the population. Inbreeding distributes genes from the heterozygous to homozygous state without changing the allele frequencies. **Outbreeding** does not eliminate but preserves deleterious alleles in heterozygous state (masking effect). Self-fertilization, the most extreme case of inbreeding, most definitely eliminates them.

Balancing selection involves opposing selection forces. A balanced equilibrium results when two alleles selected against in the homozygous state are retained because of the superiority of heterozygotes (heterozygote advantage or overdominance). This is why a recessive deleterious allele will never be eliminated by selection as it will be maintained in heterozygous form. Selection can maintain a polymorphism when the heterozygote is fitter than either homozygote.

In many mammals, including humans, more than 50% of zygotes are male and, for reasons that are poorly understood, this proportion gradually falls between conception and birth. The primary sex ratio is estimated to be at least 120 (males):100 (females) at conception in humans [McMillen MM. Science 1979;204:89]. This is evidence for prenatal selection and it appears that an MHC-mediated, male-specific selection based on heterozygous advantage is operating [Dorak MT; unpublished].

The alleles that increase in frequency relative to others are said to be fitter, so the change in these relative frequencies measures neo-Darwinian fitness. Fit does not necessarily mean biggest, fastest or strongest. Evolutionary fitness refers to reproductive success and survival.

Mendel's ideas: Heterozygous parents produce equal quantities of gametes containing the contrasting alleles; genes of different characters behave independently as they are assorted into gametes; genes are non-blending and very stable (he did not use the word gene but meant it). Identical looking individuals may be genetically different since part of the genetic variety is masked by dominance.

Mendel's first law (law of segregation): The two alleles received one from each parent are segregated in gamete formation, so that each gamete receives one or the other with equal probability.

Mendel's second law (law of recombination): Two characters determined by two unlinked genes are recombined at random in gametic formation, so that they segregate independently of each other, each according to the first law. (Note that recombination here is not used to mean crossing-over in meiosis.)

In diploid organisms, the extent to which an allele spreads or recedes in a population depends upon which alleles it becomes associated with in heterozygotes. For example, a recessive deleterious gene will be protected from selection in heterozygous associations with advantageous dominants. Alternatively, selection against a deleterious dominant will lead to the elimination of advantageous recessives when they are associated with it in the heterozygotes.

Similarities of the proteins of coding region DNAs of two species may be due to convergent evolution or similarity by descent. But, similarities in non-coding region DNAs can only be similarity by descent and it means that these two species have only recently diverged.

The largest amount of human genetic diversity is being preserved in African genomes. Evolutionarily speaking, Africans thus have a larger allelic pool to draw on for both fitness and survival. [See Disotell TR: Sex-specific contributions to genome variation. *Curr Biol* 1999;9:R29 and Olerup et al. HLA-DR and -DQ gene polymorphism in West Africans is twice as extensive as in North European Caucasians, *PNAS* 1991;88:8480.]

Sexual reproduction: (see also **Evolution of Sexual Reproduction**)

1. The production of gametes by meiosis
2. The recombination of genes by crossing-over
3. The random allocation of homologous chromosomes to each meiotic product
4. The production of new individuals by syngamy (fusion of two gametes; usually from two separate individuals except in self-fertilization). Sexual reproduction generates diversity within populations (evolutionary plasticity) whereas parthenogenesis limits diversity.

Sexual reproduction exposes a new array of genotypes to the environment at each generation, while keeping its basic elements, the alleles, and their respective frequencies about the same. As a result, populations of sexually reproducing organisms enjoy adaptability in the face of a changing environment far beyond the reach of the asexual species. Sexual reproduction may have evolved because of the benefit of having two copies of a gene. This is more adaptive in terms of DNA repair and elimination of a deleterious mutation.

Death: Natural selection cannot prevent senescence and death because it cannot eliminate certain kinds of alleles. The alleles that exert their deleterious effects after an organism has ceased reproducing cannot be eliminated from the gene pool (like Huntington's chorea, Alzheimer's disease, late-onset malignancies). Death is an effect of natural selection but is not a character favored by it. Most popular theories of senescence are antagonistic pleiotropy and mutation accumulation theories.

Extinction: Extinction is the ultimate fate of all species. More than 99.9% of all evolutionary lines that once existed on earth have become extinct. The average life span of a species in the paleontological record is 4 million years (Raup DM 1994). The

Permian extinction (250 mya) was the largest extinction in history. It is estimated that 96% of all species (50% of all families) met their end. Often, the appearance of a new species is instrumental in the extinction of another species. All other things being equal, the risk of extinction is higher for small -peripheral- populations, populations of species of short-lived individuals than for populations of species of long-lived individuals; for species with a low intrinsic rate of increase, and for those populations whose environment varies greatly. Very little is known about the actual causes of extinction or about how species extinction relates to population extinction. Most extinction is probably due to several factors acting together, not just a single cause. Species do not become extinct because they fail to adapt. Extinction occurs when their habitat is removed or changed to a state where there will not be enough individuals with an adapted genotype to maintain the species. Extinction via predation by humans is an example. Surely, there was no animal with a genotype adapted to protection from predation by firearms and not enough time was given to them to adapt. Currently, human alteration of the ecosphere is causing a global mass extinction. Extinction is a normal part of evolution, and overall, it has taken place almost as often as speciation.

Some laws and principles in evolutionary biology

Allen's Rule: Within species of warm-blooded animals (birds + mammals) those populations living in colder environments will tend to have shorter appendages than populations in warmer areas.

Allometry Equation: Most lines of relative growth conform to $y=bx^a$ where y and x are the two variates being compared, b and a are constants. The value of a , the allometric exponent, is 1 if the growth is isometric; allometry is said to be positive when $a>1$ and negative when $a<1$.

Biejernik's Principle (of microbial ecology): Everything is everywhere; the environment selects.

Bergmann's Rule: Northern races of mammals and birds tend to be larger than Southern races of the same species.

Coefficient of Relatedness: $r=n(0.5)^L$ where n is the alternative routes between the related individuals along which a particular allele can be inherited; L is the number of meiosis or generation links.

Cope's 'law of the unspecialized': The evolutionary novelties associated with new major taxa are more likely to originate from a generalized member of an ancestral taxon rather than a specialized member.

Cope's Rule: Animals tend to get larger during the course of their phyletic evolution.

There is a gradient of increasing species diversity from high latitudes to the tropics (see New Scientist, 4 April 1998, p.32).

Two or more similar species will not be found inhabiting the same locality unless they differ in their ecological requirements, for example in their food or their breeding habits, in their predators or their diseases.

Fisher's Fundamental Theorem: The rate of increase in fitness is equal to the additive genetic variance in fitness. This means that if there is a lot of variation in the population the value of S_w will be large.

Fisher's Theorem of the Sex Ratio: In a population where individuals mate at random, the rarity of either sex will automatically set up selection pressure favoring production of the rarer sex. Once the rare sex is favored, the sex ratio gradually moves back toward equality.

Galton's Regression Law: Individuals differing from the average character of the population produce offspring which, on the average, differ to a lesser degree but in the same direction from the average as their parents.

Gause's Rule (competitive exclusion principle): Two species cannot live the same way in the same place at the same time (ecologically identical species cannot coexist in the same habitat). This is only possible through evolution of niche differentiation (difference in beak size, root depths, etc.).

Haeckel's Law (the infamous biogenetic law): Ontogeny recapitulates phylogeny, i.e., an embryo repeats in its development the evolutionary history of its species as it passes through stages in which it resembles its remote ancestors. (Embryos, however, do not pass through the adult stages of their ancestors; ontogeny does not recapitulate phylogeny. Rather, ontogeny repeats some embryonic features of ancestors are present in embryonic development (L. Wolpert: *The Triumph of Embryo*. Oxford University Press, 1991)).

Hamilton's Altruism Theory: If selection favored the evolution of altruistic acts between parents and offspring, then similar behavior might occur between other close relatives possessing the same altruistic genes which were identical by descent. In other words, individual may behave altruistically not only to their own immediate offspring but to others such as siblings, grandchildren and cousins (as happens in the bee society).

Hamilton's Rule (theory of kin selection): In an altruistic act, if the donor sustains cost C , and the receiver gains a benefit B as a result of the altruism, then an allele that promotes an altruistic act in the donor will spread in the population if $B/C > 1/r$ or $rB - C > 0$ (where r is the relatedness coefficient).

Hardy-Weinberg Law: In an infinitely large population, gene and genotype frequencies remain stable as long as there is no selection, mutation, or migration.

When there is no selection, mutation, migration (gene flow) in a pan-mictic population in infinite size, the genotype frequencies will remain constant in this population. For a bi-allelic locus where the gene frequencies are p and q :

$$p^2 + 2pq + q^2 = 1 \quad (\text{more on HWE})$$

Selection Coefficient (s): $s = 1 - W$ where W is relative fitness. This coefficient represents the relative penalty incurred by selection to those genotypes that are less fit than others. When the genotype is the one most strongly favored by selection its s value is 0.

Heritability: the proportion of the total phenotypic variance that is attributable to genetic causes:

$$h^2 = \text{genetic variance} / \text{total phenotypic variance}$$

Natural selection tends to reduce heritability because strong (directional or stabilizing) selection leads to reduced variation.

Lyon hypothesis: The proposition by Mary F Lyon that random inactivation of one X chromosome in the somatic cells of mammalian females is responsible for dosage compensation and mosaicism.

Muller's Ratchet: The continual decrease in fitness due to accumulation of (usually deleterious) mutations without compensating mutations and recombination in an asexual lineage (HJ Muller, 1964). Recombination (sexual reproduction) is much more common than mutation, so it can take care of mutations as they arise. This is one of the reasons why sex is believed to have evolved.

Protein clock hypothesis: The idea that amino acid replacements occur at a constant rate in a given protein family (ribosomal proteins, cytochromes, etc) and the degree of divergence between two species can be used to estimate the time elapsed since their divergence.

Selection Differential (S) and Response to Selection (R): Following a change in the environment, in the parental (first) generation, the mean value for the character among those individuals that survive to reproduce differs from the mean value for the whole population by a value of (S). In the second, offspring generation, the mean value for the character differs from that in the parental population by a value of R which is smaller than S. Thus, strong selection of this kind (directional) leads to reduced variability in the population.

van Baer's Rule: The general features of a large group of animals appear earlier in the embryo than the special features.

Further reading

- What Evolution Is* by Ernst Mayr (Basic Books, 2001) " A MUST "**
- The Origins of Life* by J Maynard-Smith & E Szathmary (OUP, 1999)**
- Evolution: An Introduction* by SC Stearns & RF Hoekstra (OUP, 2000)**
- Symbiotic Planet : A New Look at Evolution* by L Margulis (Perseus, 2000)**
- Biology, Evolution, and Human Nature* by TH Goldsmith & WF Zimmerman (John Wiley & Sons, 2000)**
- The Book of Life* by SJ Gould (Norton, 2001)**
- The Way of the Cell* by FM Harold (OUP, 2001)**

Internet Links

Evolution-Related Links

[PBS Evolution](#) [BBC Education: Evolution](#)

[M.Tevfik Dorak, BA \(Hons\), MD, PhD](#)

Last updated March 26, 2003

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population](#)
[Genetics](#) [Homepage](#)

A Brief History of Life

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

It is believed that formation of the solar system started 4.6 billion years ago and life was probably present on earth 4 billion years ago. The main source of energy was heat (originated from accretion, radioactive decay and meteorite impacts). This heat must have caused partial melting within the Earth resulting in separation of molten metal core and silicate mantle and subsequently release of heat and associated gases from the Earth's interior (outgassing). The lack of oxygen would have meant the lack of ozone layer. Thus, short wave solar uv light would not have been absorbed. It is more likely that late outgassing occurred producing a slightly more reducing than today's atmosphere dominated by carbon monoxide (CO) and nitrogen as N₂ but also included sulphur dioxide (SO₂), hydrogen (H₂), ammonia (NH₃) and methane (CH₄). There is little doubt that the Earth possessed a hydrosphere (early oceans) 3.8 billion years ago. The evidence for this comes from the ancient rocks at Isua in Greenland. This anoxic ocean would be rich in soluble reduced (ferrous) iron. It is possible that in solutions containing reduced iron, CO₂ can be reduced to formaldehyde (HCHO) by irradiation with uv light. This implies that reduced iron may have played an important role as a source of reducing power in prebiotic chemistry. The primitive atmosphere contained a lot of CO₂ which is a well-known greenhouse gas. In short, the environmental conditions under which life arose on the primitive Earth were very different from those of today.

Among the theories on the origin of life (App.1), the chemical theory originally proposed by Oparin and Haldane and subsequently modified in 1980s is the most plausible one. Any theory that is put forward should offer an explanation to the creation of the basics of life: replication and metabolism. Evidence from extraterrestrial environments, interstellar molecules and from other planets of the solar system show that chemical evolution is able to produce organic molecules.

It is conceivable that the building blocks of life, amino acids and nucleic acid bases, can be formed from the normal chemistry of carbon, hydrogen, oxygen, nitrogen and sulphur. The pre-biotic chemistry would get enough energy to catalyze the reactions from the Sun (short wave UV light), electric discharges, and heat (decay of radioactive elements, meteorite impacts and volcanic activity).

The essence of the chemical theory is that life arose on Earth from organic molecules produced abiogenically. It is believed that the appearance of life followed a period of anaerobic chemical evolution and first protocells, organized replicating systems and finally true cells evolved. This is the Oparin-Haldane theory first proposed in 1920s. According to this scenario, organic molecules accumulated in the oceans (primeval soup) and protocells and primitive cells obtained energy using these chemicals (i.e., they were anaerobic heterotrophs). Once the most likely substrates for prebiotic chemistry, hydrogen cyanide (HCN) and formaldehyde (HCHO), were formed they would have to be concentrated to react further. It has been proposed that clay particles provided surfaces where HCN, HCHO and other molecules could be adsorbed and concentrated. If the chemicals are available in enough quantities, more complex molecules can be produced from HCN and HCHO in the presence of energy. For example HCHO can proceed to form sugars (including riboses of nucleic acids DNA and RNA) from glycoaldehyde using uv light or electrical discharges (formose reaction). Similarly, HCN can form amino acids under mildly reducing conditions using the same energy sources. Heating and cooling of these amino acids would generate protein-like protenoids. Despite these advances made to explain the pre-biotic chemistry of proteins, it has not yet been possible to come up with an idea to explain the formation of nucleic acids in the primitive Earth.

An alternative theory assumes that it was not the nucleic acids as in modern life forms but something else in primitive life that acted as the replicating system. Cairns-Smith proposed that electrically charged surfaces of clay particles may have been the primitive replicating templates. It is thought that the distribution of electrical charges on the clay surface acted as the genetic information and ionic interactions with newly formed surface layers were the basis of transmission of the information. The emergence of this idea coincided with other evidence suggesting that the first type of metabolism that arose was autotrophic. The energy may have been provided by the oxidation of ferrous iron in iron-rich clays. It was also suggested that a primitive kind of photosynthesis arose when iron-rich clays were exposed to UV radiation. The most recent and comprehensive theory holds that the primitive ocean was too dilute for all these reactions to occur and it must have all happened in the atmosphere. There is no experimental support for this theory yet.

The common ancestor of all life probably used RNA (ribosyme) as its genetic material. RNA can have catalytic activity similar to that of protein enzymes so RNA can combine both a genetic (informational) role and a functional (enzymatic) role. RNA is the precursor of DNA and plays a role in several key

cellular processes, including DNA synthesis, translation and protein synthesis. The general idea is that RNA was once the genetic material of early cells or protocells and DNA evolved later. How RNA first arose is still a mystery. Reverse transcription, whereby DNA is formed from RNA is an obvious route by which DNA could have arisen eventually to replace RNA. Translation of RNA into proteins must have evolved before switching to the DNA. Several theories have been put forward to explain the evolution of translation.

Whatever may have happened before its emergence, the first organisms which have left living descendants were simple unicellular prokaryotes. These cells lack a true nucleus (eu-karyon), mitochondria or chloroplasts. They probably first emerged 4 billion years ago and dominated the Earth for the next 2 billion years (the age of prokaryotes). The universal ancestor cell followed three distinct pathways:

1. Evolution of Archaeobacteria (prokaryotes; thermophilic Sulphobacteria, halophilic, methanogenic)
2. Evolution of Eubacteria (prokaryotes; ordinary bacteria and Cyanobacteria - blue-green algae-)
3. Evolution of nuclear line to form Eukaryotes (and eventually protists, fungi, plants, animals) (nearly all are aerobic)

The great majority of modern eukaryotes arose through the association of two or more different prokaryotes (a host anaerobic prokaryote engulfing an aerobic one as its mitochondria or chloroplast). Molecular phylogenetic studies using the rRNA sequencing data (since it has a very slow evolutionary rate it is suitable for the analysis of organisms with a long evolutionary history), showed that Sulphobacteria evolved least from the universal ancestor cell. Archaeobacteria as a whole group diverged less than either Eukaryotes or Eubacteria from the universal ancestor and remain closest to it. There are also other features of cells that suggest the same, for example, their membrane lipids are ether linked whereas eubacteria and eukaryotes have ester linked membrane lipids. There is no evidence to suggest that Archaeobacteria are the ancestor of Eubacteria and Eukaryotes, therefore, it is generally believed that the ancestral life form gave rise to three groups.

One great transformation in the evolution of early life forms was the shift from anoxic to oxygen-containing atmosphere resulted from the oxygen released by photosynthetic Cyanobacteria. Therefore, Cyanobacteria are photo-autotrophs (see App.2). They use light as a source of energy, and CO₂ as a source of carbon (photosynthesis). Heterotrophs, on the other hand, use organic molecules synthesized outside their body as a source of energy and carbon (aerobic respiration). In all life forms, energy is obtained from high energy electrons during cyclic or non-cyclic electric transport schemes during which the energy is

transferred to and stored into ATP and low energy electrons are released at the end. An exception to this general scheme is the fermenters who store their energy in the form of proton motive force (PMF) as well as ATP. In some photo-autotrophs, electrons are energised through the absorption of light by a special light-sensitive pigment (bacteriorhodopsin) and then transferred to other electron carriers.

In chemo-autotrophs, high energy electrons are obtained from an external, inorganic substance. Sulphobacteria use H_2S as the energy source and the electron acceptor for the low energy electrons is oxygen which is reduced to water. Some species use substances such as nitrate and reduce it to nitrite. Autotrophs can also assimilate carbon from a simple inorganic molecule into complex organic molecules. The most widespread way of doing this is the Calvin cycle which fixes C from simple molecules like CO_2 into their organic molecules. The Calvin cycle is found in many photosynthetic Eubacteria and all green Eukaryotes.

Heterotrophs have variation in the terminal electron acceptor. It is oxygen in aerobic respiration and sulphate or nitrate in anaerobic respiration. When there is no electron acceptor, fermentation under anaerobic conditions occurs resulting in formation of lactic acid. This may be the only system in some prokaryotes and is an emergency system in others.

The main feature of Archaeobacteria is that they live in extreme conditions. Sulphobacteria live in extreme heat and in environments rich in sulphur. Sulphobacteria are anaerobic chemo-autotrophs. If the hot spring scenario is correct, the Sulphobacteria should closely resemble the ancestral prokaryotes. Methanogens are obligate anaerobes, and Halobacteria grow in very salty environments (usually aerobic heterotrophs). The Calvin cycle is generally absent in Archaeobacteria.

Eubacteria are more diverged than Archaeobacteria. They use several types of energy metabolism. In this group, purple bacteria and Gram(+) bacteria are chemo-autotrophs and Cyanobacteria are oxygenic photo-autotroph, many are heterotrophic with either anaerobic or aerobic respiration. Chlamydia have no energy metabolism at all. They depend on their hosts as parasites for eukaryotic cells. By evolving oxygen respiration 3.5 billion years ago, Cyanobacteria enabled the eukaryotic lineage to become aerobic (not immediately though. The atmosphere became rich in oxygen only 2 billion years ago as iron acted as oxygen sink). When oxygen was plentiful in the atmosphere, it was difficult for anaerobic cells to survive. Some retreated to inhospitable and relatively anaerobic environments and some acquired aerobic bacteria as endosymbionts (see below). The phylogenetic relationships suggest that aerobic heterotrophs (like all aerobic animals) and chemo-autotrophs (gram+ bacteria) evolved from early photo-autotrophs. This implies that using light as an energy source may be just as ancient as fermenting organic compounds. As an oxygenic photo-

autotroph, Cyanobacteria use two molecules of water as the electron donor and oxidize it to oxygen while releasing four electrons. By releasing oxygen, they changed the environmental conditions (ozone formation) and enabled the evolution of aerobic respiration.

The autotrophic nature of early life on Earth has been suggested by stromatolite (characteristic structures formed by Cyanobacteria) fossils on Precambrian rocks and carbon isotope ratios confirming that autotrophs fixing carbon via the Calvin cycle must have existed for 3.5 billion years. Bacteria are the only life forms found in the rocks for a long time (3.5 to 2.1 billion years ago). Eukaryotes were numerous 1.9 to 2.1 billion years ago and fungi-like things appeared about 0.9 billion years ago.

Evolution of eukaryotes from a presumed bacteria-like ancestor is one of the major events in evolutionary history. They have a distinct nucleus, organelles involved in energy metabolism (mitochondria and chloroplast), extensive internal membranes and a cytoskeleton of protein fibers and filaments. Chloroplasts (photosynthesis) in green plants and algae originated as free living bacteria related to the cyanobacteria [the chloroplastic DNA is more similar to free-living Cyanobacteria DNA than to sequences from the plants the chloroplasts reside in]. The eukaryotic mitochondria (ATP synthesis) are endosymbionts like chloroplasts. Mitochondria were acquired when aerobic Eubacteria were engulfed by anaerobic host cells. As they conferred useful functions like aerobic respiration and photosynthesis (chloroplasts), they were retained as endosymbionts. This must have happened after the nucleus was acquired by the eukaryotic lineage. The origin of eukaryotic nucleus is almost certainly autogenous and not a result of endosymbiosis. Mitochondria are believed to have originated from an ancestor of the present-day purple photosynthetic bacteria that had lost its capacity for photosynthesis (chloroplasts from an ancestral Cyanobacterium).

Animals start appearing prior to the Cambrian, about 600 million years ago. The earliest known animals occur in the widespread late Precambrian Ediacaran fauna (c. 620-550 Ma). The Cambrian explosion (radiation), one of the most important events in the history of life, began about 540 Ma at the start of the Cambrian period. This is the largest radiation ever recorded. All the phyla of animals (except the subphylum vertebrates) appeared around the Cambrian. This explosion may be due to the availability of oxygen in large amounts enabling large animals to evolve. The oldest known vertebrates (armored-jawless fish) appeared first in Ordovician (510-439 Ma). The second part of Carboniferous saw the evolution of the first reptiles, a group that evolved from amphibians and lived entirely on land. About 380 million years ago, during the Devonian period, a group of fishes evolved limbs and began to leave the water. All tetrapods (amphibians, reptiles, birds and mammals) evolved as a result of this move. During the Permian period (290-245 Ma), the Earth's land areas became welded into a single landmass (Pangaea). Many forms of marine animals

disappeared and the reptiles spread rapidly. The land was first colonized by lung-bearing fish about 250 Ma and then by amphibians. The Mesozoic era is known as the age of reptiles from which mammals and later birds evolved. The first mammals appeared in the beginning of Mesozoic, during Triassic. Mammals had an adaptive radiation in Paleocene following the mass extinction at late Cretaceous. First anthropoid appeared in Miocene about 20 Ma. The Pliocene epoch is the climax of the age of mammals. The Pleistocene is marked by an abundance of large mammals.

All land plants evolved from the green algae or Chlorophyta. In the period before the Permian (the Carboniferous), the landscape was dominated by seedless ferns and their relatives. Vascular plants first appeared in Silurian (439-409 Mya). After the Permian extinction, gymnosperms became more abundant. They evolved seeds and pollens (encased sperm). The angiosperms evolved from the gymnosperms during the early Cretaceous about 140-125 Mya. They further diversified and dispersed during the late Cretaceous (97.5-66.5 Mya). Currently, over three fourths of all living plants are angiosperms. The angiosperms developed a close contact with insects which promoted cross-pollination and resulted in more vigorous offspring. Their generation time to reproduce is short, and their seeds can be dispersed by animals. For these reasons, the angiosperms were able to travel and disperse all around the world. The important events in the evolution of the angiosperms were the evolution of showy flowers (to attract insects and birds), the evolution of bilaterally symmetrical flowers (adaptation for specialized pollinators), and the evolution of larger and more mobile animals (to disperse fruits and seeds).

Appendix 1. Theories on the origin of life:

1. Divine creation
2. Spontaneous generation
3. Extraterrestrial theories (panspermia)
4. Chemical theory

Appendix 2. Metabolism in prokaryotes:

1. Autotrophs
 - a. photo-autotrophs (photosynthetic)
 - i. anoxygenic
 - ii. oxygenic (only in Cyanobacteria)
 - b. chemo-autotrophs (Sulphobacteria)
2. Heterotrophs (they make their food by oxidation of nitrogen, sulphur or other elements; they are uncommon)
 - a. aerobic respiration (most Halobacteria)
 - b. anaerobic respiration (Methanogens)

Appendix 3. Possibilities on the first living cell:

1. Photo-autotroph: As an anaerobic but oxygenic photosynthetic prokaryote Cyanobacteria is the strongest candidate. It makes its own food and this is fueled by light. The fossil evidence (stromatolite and oncolite) for their existence and carbon isotope ratio suggesting oxygenic photosynthesis took place by 3.5 billion years ago support this idea.

or

2. Chemo-autotroph: If the hot-spring scenario is correct, it would have to be Sulphobacteria. This anaerobic thermophilic bacteria uses H_2S as electron donor. It does not use the Calvin cycle.

Whatever it was, the first living cell replicated by a mechanism not based on nucleic acids. There is not enough evidence to support the suggestions that either heterotrophs or fermenters may have been the first living cells.

see also <http://www.ucmp.berkeley.edu/allife/threedomains.html>

[Transcripts of the BBC series 'Life on Earth'](#)

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated June 8, 2001

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population Genetics](#) [Homepage](#)

[Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#) [Homepage](#)

BASIC POPULATION GENETICS

M.Tevfik Dorak, M.D., Ph.D.

G.H. Hardy (the English mathematician) and W. Weinberg (the German physician) independently worked out the mathematical basis of population genetics in 1908. Their formula predicts the expected genotype frequencies using the allele frequencies in a diploid Mendelian population. They were concerned with

questions like "what happens to the frequencies of alleles in a population over time?" and "would you expect to see alleles disappear or become more frequent over time?"

Hardy and Weinberg showed in the following manner that if the population is very large and random mating is taking place, allele frequencies remain unchanged (or in equilibrium) over time unless some other factors intervene. If the frequencies of allele A and a (of a biallelic locus) are p and q, then $(p + q) = 1$. This means $(p + q)^2 = 1$ too. It is also correct that $(p + q)^2 = p^2 + 2pq + q^2 = 1$. In this formula, p^2 corresponds to the frequency of homozygous genotype AA, q^2 to aa, and $2pq$ to Aa. Since 'AA, Aa, aa' are the three possible genotypes for a biallelic locus, the sum of their frequencies should be 1. In summary, Hardy-Weinberg formula shows that:

$$p^2 + 2pq + q^2 = 1$$

AA Aa aa

If the observed frequencies do not show a significant difference from these expected frequencies, the population is said to be in Hardy-Weinberg equilibrium (HWE). If not, there is a violation of the following assumptions of the formula, and the population is not in HWE.

The assumptions of HWE

1. Population size is effectively infinite,
2. Mating is random in the population (the most common deviation results from inbreeding),
3. Males and females have similar allele frequencies,
4. There are no mutations and migrations affecting the allele frequencies in the population,
5. The genotypes have equal fitness, i.e., there is no selection.

The Hardy-Weinberg law suggests that as long as the assumptions are valid, allele and genotype frequencies will not change in a population in successive generations. Thus, any **deviation from HWE** may indicate:

1. Small population size results in random sampling errors and unpredictable genotype frequencies (a real population's size is always finite and the frequency of an allele may fluctuate from generation to generation due to chance events),
2. Assortative mating which may be positive (increases homozygosity; self-fertilization is an extreme example) or negative (increases heterozygosity), or inbreeding which increases homozygosity in the whole genome without changing the allele frequencies. Rare-male mating advantage also tends to increase the frequency of the rare allele and heterozygosity for it (in reality, random mating does not occur all the time),
3. A very high mutation rate in the population (typical mutation rates are $< 10^{-5}$ per generation) or massive migration from a genotypically different population interfering with the allele frequencies,
4. Selection of one or a combination of genotypes (selection may be negative or positive). Another reason would be unequal transmission ratio of alternative alleles from parents to offspring (as in mouse *t*-haplotypes).

The implications of the HWE

1. The allele frequencies remain constant from generation to generation. This means that hereditary mechanism itself does not change allele frequencies. It is possible for one or more assumptions of the equilibrium to be violated and still not produce deviations from the expected frequencies that are large enough to be detected by the goodness of fit test,
2. When an allele is rare, there are many more heterozygotes than homozygotes for it. Thus, rare alleles will be impossible to eliminate even if there is selection against homozygosity for them,
3. For populations in HWE, the proportion of heterozygotes is maximal when allele frequencies are equal ($p = q = 0.50$),
4. An application of HWE is that when the frequency of an autosomal recessive disease (e.g., sickle cell disease, hereditary hemochromatosis, congenital adrenal hyperplasia) is known in a population and unless there is reason to believe HWE does not hold in that population, the gene frequency of the disease gene can be calculated (for an example visit the [Cancer Genetics](#) website and choose Topics).

It has to be remembered that when HWE is tested, mathematical thinking is necessary. When the population is found in equilibrium, it does not necessarily mean that all assumptions are valid since there may be counterbalancing forces. Similarly, a significant deviance may be due to sampling errors (including **Wahlund effect**, see below and [Glossary](#)), misclassification of genotypes, measuring two or more systems as a single system, failure to detect rare alleles and the inclusion of non-existent alleles. The Hardy-Weinberg laws rarely holds true in nature (otherwise evolution would not occur). Organisms are subject to mutations, selective forces and they move about, or the allele frequencies may be different in males and females. The gene frequencies are constantly changing in a population, but the effects of these processes can be assessed by using the Hardy-Weinberg law as the starting point.

The direction of departure of observed from expected frequency cannot be used to infer the type of selection acting on the locus even if it is known that selection is acting. If selection is operating, the frequency of each genotype in the next generation will be determined by its relative fitness (W). Relative fitness is a measure of the relative contribution that a genotype makes to the next generation. It can be measured in terms of the intensity of selection (s), where $W = 1 - s$ [$0 \leq s \leq 1$]. The frequencies of each genotype after selection will be $p^2 W_{AA}$, $2pq W_{Aa}$, and $q^2 W_{aa}$. The highest fitness is always 1 and the others are estimated proportional to this. For example, in the case of heterozygote advantage (or overdominance), the fitness of the heterozygous genotype (Aa) is 1, and the fitnesses of the homozygous genotypes negatively selected are $W_{AA} = 1 - s_{AA}$ and $W_{aa} = 1 - s_{aa}$. It can be shown mathematically that only in this case a stable polymorphism is possible. Other selection forms, underdominance and directional selection, result in unstable polymorphisms. The weighted average of the fitnesses of all genotypes is the mean fitness. It is important that genetic fitness is determined by both fertility and viability. This means that diseases that are fatal to the bearer but do not reduce the number of progeny are not genetic lethals and do not have reduced fitness (like the adult onset genetic diseases: Huntington's chorea, hereditary hemochromatosis). The detection of selection is not easy because the impact on changes in allele frequency occur very slowly and selective forces are not static (may even vary in one generation as in antagonistic pleiotropy).

All discussions presented so far concerns a simple biallelic locus. In real life, however, there are many loci which are multiallelic, and interacting with each other as well as with the environmental factors. The Hardy-Weinberg principle is equally applicable to multiallelic loci but the mathematics is slightly more complicated. For multigenic and multifactorial traits, which are mathematically continuous as opposed to discrete, more complex techniques of quantitative genetics are required.

In a final note on the practical use of HWE, it has to be emphasised that **its violation** in daily life is most frequently due to genotyping errors. Allelic misassignments, as frequently happens when PCR-SSP method is used, sometimes due to allelic dropout are the most frequent causes of the Hardy-Weinberg disequilibrium. When this is observed, the genotyping protocol should be reviewed. In a case-control

association study, it is of paramount importance that the control group is in HWE to rule out any technical errors. The violation of HWE in the case group, however, may be due to a real association.

Some concepts relevant to HWE

Wahlund effect: Reduction in observed heterozygosity (increased homozygosity) because of considering pooled discrete subpopulations that do not interbreed as a single randomly mating unit. When all subpopulations have the same gene frequencies, no variance among subpopulations exists, and no Wahlund effect occurs ($F_{ST}=0$). **Isolate breaking** is the phenomenon that the average homozygosity temporarily increases when subpopulations make contact and interbreed (this is due to decrease in homozygotes). It is the opposite of Wahlund effect.

F statistics: The F statistics in population genetics has nothing to do the F statistics evaluating differences in variances. Here **F** stands for fixation index, fixation being increased homozygosity resulting from inbreeding. Population subdivision results in the loss of genetic variation (measured by heterozygosity) within subpopulations due to their being small populations and genetic drift acting within each one of them. This means that population subdivision would result in decreased heterozygosity relative to that expected heterozygosity under random mating as if the whole population was a single breeding unit. Wright developed three fixation indices to evaluate population subdivision: F_{IS} (interindividual), F_{ST} (subpopulations), F_{IT} (total population).

F_{IS} is a measure of the deviation of genotypic frequencies from panmictic frequencies in terms of heterozygous deficiency or excess. It is what is known as the **inbreeding coefficient (f)**, which is conventionally defined as the probability that two alleles in an individual are identical by descent (autozygous). The technical description is the correlation of uniting gametes relative to gametes drawn at random from within a subpopulation (Individual within the Subpopulation) averaged over subpopulations. It is calculated in a single population as $F_{IS} = 1 - (H_{OBS} / H_{EXP})$ [equal to $(H_{EXP} - H_{OBS}) / H_{EXP}$] where H_{OBS} is the observed heterozygosity and H_{EXP} is the expected heterozygosity calculated on the assumption of random mating. It shows the degree to which heterozygosity is reduced below the expectation. The value of F_{IS} ranges between -1 and +1. Negative F_{IS} values indicate heterozygote excess (outbreeding) and positive values indicate heterozygote deficiency (inbreeding) compared with HWE expectations.

F_{ST} measures the effect of population subdivision, which is the reduction in heterozygosity in a subpopulation due to genetic drift. F_{ST} is the most inclusive measure of population substructure and is most useful for examining the overall genetic divergence among subpopulations. It is also called **coancestry coefficient (θ)** [Weir & Cockerham, 1984] or '**Fixation index**' and is defined as correlation of gametes within subpopulations relative to gametes drawn at random from the entire population (Subpopulation within the Total population). It is calculated as using the subpopulation (average) heterozygosity and total population expected heterozygosity. F_{ST} is always positive; it ranges between 0 = panmixis (no subdivision, random mating occurring, no genetic divergence within the population) and 1 = complete isolation (extreme subdivision). F_{ST} values up to 0.05 indicate negligible genetic differentiation whereas >0.25 means very great genetic differentiation within the population analyzed. F_{ST} is usually calculated for different genes, then averaged across all loci, and all populations. F_{ST} can also be used to estimate gene flow: $0.25 (1 - F_{ST}) / F_{ST}$. This highly versatile parameter is even used as a genetic distance measure between two populations instead of a fixation index among many populations (see [Weir BS, Genetic Data Analysis II, 1996](#); and [Kalinowski ST, 2002](#)).

F_{IT} is rarely used. It is the overall inbreeding coefficient (F) of an individual relative to the total population (Individual within the Total population).

Detecting Selection Using DNA Polymorphism Data

Several methods have been designed to use DNA polymorphism data (sequences and allele frequencies) to obtain information on past selection events. Most commonly, the ratio of non-synonymous (replacement) to synonymous (silent) substitutions (d_N/d_S ratio; see below) is used as evidence for overdominant selection (balancing selection) of which one form is heterozygote advantage. Classic example of this is the

mammalian **MHC** system genes and other **compatibility systems** in other organisms: the self-incompatibility system of the plants, fungal mating types and invertebrate allorecognition systems. In all these genes, a very high number of alleles is also noted. This can be interpreted as an indicator of some form of balancing (diversifying) selection. In the case of neutral polymorphism, one common allele and a few rare alleles are expected. The frequency distribution of alleles is also informative. Large number of alleles showing a relatively even distribution is against neutrality expectations and suggestive of diversifying selection.

Most tests detect selection by rejecting neutrality assumption (observed data is deviate significantly from what is expected under neutrality). This deviation, however, may also be due to other factors such as changes in population size or genetic drift. The original neutrality test was **Ewens-Watterson homozygosity test** of neutrality (see **Glossary**) based on the comparison of observed homozygosity and predicted value calculated by **Ewens's sampling formula** which uses the number of alleles and sample size. This test is not very powerful.

Other commonly used statistical tests of neutrality are Tajima's D (theta), Fu & Li's D, D* and F. Tajima's test (**Tajima F, Genetics 1989**) is based on the fact that under the neutral model estimates of the number of segregating/polymorphic sites and of the average number of nucleotide differences are correlated. If the value of D is too large or too small, the neutral 'null' hypothesis is rejected. **DnaSP** calculates the D and its confidence limits (two-tailed test). Tajima did not base this test on coalescent but Fu and Li's tests (**Fu & Li, Genetics 1993**) are directly based on coalescent. The tests statistics D and F require data from intraspecific polymorphism and from an outgroup (a sequence from a related species), and D* and F* only require intraspecific data. **DnaSP** uses the critical values obtained by **Fu & Li, Genetics 1993** to determine the statistical significance of D, F, D* and F* test statistics. **DnaSP** can also conduct the Fs test statistic (**Fu YX, Genetics 1997**). The results of this group of tests (Tajima's D and Fu & Li's tests) based on allelic variation and/or level of variability may not clearly distinguish between selection and demographic alternatives (bottleneck, population subdivision) but this problem only applies to the analysis of a single locus (demographic changes affect all loci whereas selection is expected to be locus-specific which are distinguishable if multiple loci are analysed). Tests for multiple loci include the **HKA test** described by Hudson, Kreitman and Aguade (**Genetics, 1987**). This test is based on the idea that in the absence of selection, the expected number of polymorphic (segregating) sites within species and the expected number of 'fixed' differences between species (divergence) are both proportional to the mutation rate, and the ratio of them should be the same for all loci. Variation in the ratio of divergence to polymorphism among loci suggests selection.

A different group of neutrality tests that are not sensitive to demographic changes include McDonald-Kreitman test (**McDonald JH & Kreitman M, Nature 1991**) and d_N/d_S ratio test. McDonald-Kreitman test compares the ratio of the number of nonsynonymous to synonymous 'polymorphisms' within species to that ratio of the number of nonsynonymous to synonymous 'fixed' differences between species in a 2x2 table (see a worked example [here](#)). The most direct method of showing the presence of positive selection is to compare the number of nonsynonymous (d_N) to the number of synonymous (d_S) substitutions in a locus. A high (>1) value of (d_N/d_S) substitutions suggest fixation of nonsynonymous mutations with a higher probability than neutral (synonymous) ones. Statistical properties of this test are given by **Goldman N & Yang Z, Mol Biol Evol, 1994** and by **Muse SV & Gaut BS, Mol Biol Evol, 1994**. The d_N/d_S ratio tests take into account of transition/transversion rate bias and codon usage bias.

For other tests and software to perform these statistics, see **DNA Sequence Polymorphism, DnaSP**. See also: **Statistical Tests of Neutrality (Lecture Note by P Beerli)**; **Statistical Tests of Neutrality of Mutations against Excess of Recent Mutations (Rare Allels)**; **Statistical Tests of Neutrality of Mutations against an Excess of Old Mutations or a Reduction of Young Mutations**; **Estimation of theta**; Properties of statistical tests of neutrality for DNA polymorphism data, **Simonsen et al Genetics 1995;141:413-29**. Review of statistical tests of selective neutrality on genomic data, **Nielsen R, Heredity 2001**; and a **Lecture Note by Gil McVean**.

Linkage disequilibrium (LD)

The tendency for two 'alleles' to be present on the same chromosome (positive LD), or not to segregate together (negative LD). As a result, specific alleles at two different loci are found together more or less

than expected by chance. The same situation may exist for more than two alleles. Its magnitude is expressed as the delta (Δ) value and corresponds to the difference between the expected and the observed haplotype frequency (see [Measures of LD by Devlin & Risch, 1995](#) for further details). It can have positive or negative values. LD is decreased by recombination. Thus, it decreases every generation of random mating unless some process opposing the approach to linkage 'equilibrium'. Permanent LD may result from natural selection if some gametic combinations confer higher fitness than other combinations. For more on LD, see [Statistical Analysis in HLA and Disease Association Studies](#). Link to a lecture on [linkage disequilibrium](#); [online LD analysis](#). Software to perform LD analysis: [Genetic Data Analysis, EH, 2LD, MLD, PopGene, Arlequin 2000](#), and [Online Easy LD](#). Please note that LD has nothing to do HWE and should not be confused with it (see [Possible Misunderstandings in Genetics](#)).

Genetic distance (GD)

Genetic distance is a measurement of genetic relatedness of samples of populations (whereas genetic diversity represents diversity within a population). The estimate is based on the number of allelic substitutions per locus that have occurred during the separate evolution of two populations. (See lecture notes on [Genetic Distances](#), [Estimating Genetic Distance](#); and [GeneDist: Online Calculator of Genetic Distance](#). The software [Arlequin](#), [PHYLIP](#), [GDA](#), [PopGene](#), [Populations](#) and [SGS](#) are suitable to calculate population-to-population genetic distance from allele frequencies.

Genetic Distance can be computed on freeware PHYLIP. Most components of PHYLIP are available [on the web](#). One component of the package [GENDIST](#) estimates genetic distance from allele frequencies using one of the three methods: Nei's, Cavalli-Sforza's or Reynold's (see papers by [Cavalli-Sforza & Edwards, 1967](#), [Nei et al, 1983](#), [Nei M, 1996](#) and [lecture note \(1\)](#) and [\(2\)](#) for more information on these methods). GENDIST can be run [online](#) using default options ([Nei's genetic distance](#)) to obtain genetic distance matrix data. The PHYLIP program [CONTML](#) estimates phylogenies from gene frequency data by maximum likelihood under a model in which all divergence is due to genetic drift in the absence of new mutations (Cavalli-Sforza's method) and draws a tree. The program is also available [on the web](#) and runs with default options. If new mutations are contributing to allele frequency changes, Nei's method should be selected on GENDIST to estimate genetic distances first. Then a tree can be obtained using one of the following components of PHYLIP: [NEIGHBOR](#) also draws a phylogenetic tree using the genetic distance matrix data (from GENDIST). It uses either [Nei and Saitou's](#) (1987) "[Neighbor Joining \(NJ\) Method](#)," or the [UPGMA](#) (unweighted pair group method with arithmetic mean; average linkage clustering) method ([Sneath & Sokal, 1973](#)). [Neighbor Joining](#) is a distance matrix method producing an unrooted tree without the assumption of a clock (the evolutionary rate does not have to be the same in all lineages). Major assumption of UPGMA is equal rate of evolution along all branches (which is frequently unrealistic). NEIGHBOR can be run [online](#). Other components of PHYLIP that draw phylogenetic trees from genetic distance matrix data are [FITCH](#) / [online](#) (Fitch-Margoliash method with no assumption of equal evolutionary rate) and [KITSCH](#) / [online](#) (employs Fitch-Margoliash and Least Squares methods with the assumption that all tip species are contemporaneous, and that there is an evolutionary clock -in effect, a molecular clock). ([Mathematical formulae of various genetic distance measures.](#)) Another freeware [PopGene](#) calculates Nei's genetic distance and creates a tree using UPGMA method from genotypes. For genetic distance calculation on Excel, try freeware [GenAIEx](#) by Peakall & Smouse.

Because of different assumptions they are based on the NJ and UPGMA methods may construct dendrograms with totally different topologies. For an example of this and a review of main differences between the two methods, see [Nei & Roychoudhury, 1993 \(free full-text\)](#). Both methods use distance matrices (also Fitch-Margoliash and Minimal Evolution methods are distance methods). The principle difference between NJ and UPGMA is that NJ does not assume an equal evolutionary rate for each lineage. Since the constant rate of evolution does not hold for human populations, NJ seems to be the better method. For the genetic loci subject to natural selection, the evolutionary rate is not the same for each population and therefore UPGMA should be avoided for the analysis of such loci (including the HLA genes). The

leading group in HLA-based genetic distance analysis led by [Armaiz-Villena](#) proposes that the most appropriate genetic distance measure for the HLA system is the DA value first described by [Nei et al., 1983](#). Unlike UPGMA, NJ produces an unrooted tree. To find the root of the tree, one can add an outgroup. The point in the tree where the edge to the outgroup joins is the best possible estimate for the root position. One persistent problem with tree construction is the lack of statistical assessment of the phylogenetic tree presented. This is best done with widely available bootstrap analysis originally described in Felsenstein J: Evolution 1985;39:783-791 (available through [JSTOR](#) if you have access) and [Efron et al. 1996](#); and reviewed in [Nei M., 1996](#)). For a discussion of statistical tests of molecular phylogenies, see [Li & Gouy, 1990](#) and [Nei M., 1996](#). For the topology to be statistically significant the bootstrap value for each cluster should reach at least 70% whereas 50% overestimates accuracy of the tree. Bootstrap tests should be done with at least 1000 (preferably more) replications.

Nei noted that some genes are more suitable than others in phylogenetic inference and that most tree-building methods tend to produce the same topology whether the topology is correct or not [Nei M., 1996](#). He also added that sometimes adding one more species/population would change the whole tree for unknown reasons. An example of this has been provided in a study of human populations with genetic distances [Nei & Roychoudhury, 1993](#). The properties of most popular genetic distance measures have been reviewed ([Kalinowski, 2002](#)). Whichever is used, large sample sizes are required when populations are relatively genetically similar, and loci with more alleles produce better estimates of genetic distance. However, in a simulation study, Nei et al concluded that more than 30 loci should be used for making phylogenetic trees ([Nei et al., 1983](#)). There seems to be a consensus that estimated trees are nearly always erroneous (i.e., the topological arrangement will be wrong) if the number of loci is less than 30 ([Nei M., 1996](#); Jorde LB. Human genetic distance studies. Ann Rev Anthropol 1985;14:343-73; available through [JSTOR](#) if you have access). If populations are closely related even 100 loci may be necessary for an accurate estimation of the relationships by genetic distance methods. Cavalli-Sforza et al have noted important correlations between the genetic trees and linguistics evolutionary trees with the exceptions for New Guinea, Australia and South America ([Cavalli-Sforza et al. 1994](#)).

Especially for the HLA genes, phylogenetic trees can be constructed by using the Nei's DA genetic distance values and NJ method with bootstrap tests on [DISPAN. Correspondence analysis](#), a supplementary analysis to genetic distances and dendrograms, displays a global view of the relationships among populations ([Greenacre MJ, 1984](#); [Greenacre & Blasius, 1994](#); [Blasius & Greenacre, 1998](#)). This type of analysis tends to give results similar to those of dendrograms as expected from theory ([Cavalli-Sforza & Piazza, 1975](#)), and is more informative and accurate than dendrograms especially when there is considerable genetic exchange between close geographic neighbors ([Cavalli-Sforza et al. 1994](#)). In their enormous effort to work out the genetic relationships among human populations, Cavalli-Sforza et al concluded that two-dimensional scatter plots obtained by correspondence analysis frequently resemble geographic maps of the populations with some distortions ([Cavalli-Sforza et al. 1994](#)). Using the same allele frequencies that are used in phylogenetic tree construction, [correspondence analysis](#) using allele frequencies can be performed on the [ViSta \(v7.0\)](#), [VST](#), [SAS](#) but most conveniently on Multi Variate Statistical Package [MVSP](#). Link to [a tutorial on correspondence analysis](#).

Internet Links

[History of Population Genetics and Evolution in A History of Genetics](#) by AH Sturtevant

[ASHI 2001 Biostatistics and Population Genetics Workshop Notes](#)

[Microsatellites and Genetic Distance \(Primer on Genetic Distance\)](#)

[HWE in Kimball's Biology Pages](#) [Online HWE Test](#) [Online GD Calculation](#) [Online Easy LD](#)

[Population Genetics Simulations](#) [Molecular Evolution / Computational Pop Genet Course](#)

[Lectures on Population Genetics \(1\) & \(2\) & \(3\) & \(4\) & \(5\)](#)

[Statistical Genetics Websites](#)

[Freeware Population Genetic Data Analysis Software \(List of Features\):](#)

[Arlequin 2000](#) [PopGene](#) [GDA](#) [Genetix](#) [GenePop](#) [GeneStrut](#) [SGS](#) [TFPGA](#) [MVSP](#)

[PHYLIP\(Online\)](#) [DISPAN](#) [ViSta](#) [GenAIEx](#) [CLUMP](#) [TDT](#) [HAPLOTYPER](#) [PHASE v2.0](#)
[EasyLD](#)

[MSA](#) (for microsatellite data) [POPULATIONS](#) [WINPOP v2.0](#) [QUANTO](#)

[GSF: Genetic Software Forum](#) [Partition for Online Bayesian Analysis](#)

[Comprehensive List of Genetic Analysis Software \(1\) \(2\) \(3\)](#)

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated March 19, 2004

[Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#) [Back to HLA](#) [Back to MHC](#) [Glossary](#)
[Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

Frequency Dependent Selection

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

An evolutionary process where the fitness of a phenotype is dependent on the relative frequency of other phenotypes in the population is called frequency dependent selection. In positive frequency dependent selection, the fitness of a phenotype increases as it becomes more common. In negative frequency dependent selection, the fitness of a phenotype decreases as it becomes more common. In other words, less frequent phenotypes have higher fitnesses than common ones.

Haldane was the first to recognize that host-parasite relationships can lead to a cyclical coevolutionary arms race at the molecular level. Natural selection would favor those parasite genotypes that can best evade host resistance mechanisms of the most common host genotypes, because they would be capable of persisting in a large proportion of the host population. This disproportionate parasite burden would then lower the fitness of individuals bearing the common allele, thereby favoring rare alleles. As rare alleles increase in frequency the whole process is repeated leading to a stable polymorphism of host and parasite genotypes, with newly arising host alleles often being favored because pathogens have not had time to adapt to these new variants. Thus, rare alleles are favored when infrequent and disfavored as they become common. As this process would not allow fixation of an allele, a stable polymorphism is maintained. Frequency dependent selection is a dynamic process while the other form of balancing selection -heterozygous advantage- is a stable one (see below).

There are several examples of (negative) frequency dependent selection:

1. Maintenance of a 50:50 sex ratio: If one sex becomes more common, some of its members will not be able to mate whereas all members of the less frequent sex will mate. This will result in higher fitness for the rare sex and the sex ratio will return to the balance.
2. Maintenance of polymorphism in prey species: Predators commonly prefer the commoner type of a prey. A prey type that is rare will thus tend to be at an advantage because it suffers lower predation than commoner prey types. This maintains the polymorphism among preys.
3. The rare male effect: Females of some polymorphic species prefer to mate with males who belong to a rarer phenotype. When this type gets more common,

however, its fitness decreases as the other (now rarer) type will be preferred. Thus both types remain in the population.

4. Competition among males for females (the caller-satellite system of green frogs): Frogs choose the mating strategy (being a caller or satellite) whichever seems to be more advantageous each type. The mating success of each type depends on the frequency of the type relative to the frequency of the other type (if they are all callers or all satellites the mating success will be very low for all of them).

5. Competition for resources: Similar to the above example, the relative frequencies of alternative competitive strategies may be determined by frequency dependent selection. The more common one type becomes, the lower is the average fitness of individuals of that type and the higher is the average fitness of an alternative type.

6. In prey species mimicking a poisonous species (Batesian mimicry), this is only advantageous if the mimic is less common than the model. In other words, the model should be much commoner. The fitness of the mimics is negatively frequency dependent. If their frequency increases and especially exceeds that of the model, the advantage disappears.

Another form of selection, overdominance or heterozygous advantage where heterozygous genotype is selected over either of the homozygous genotypes, also has a (negative) frequency dependent component because rare alleles are disproportionately found in heterozygous genotypes, whereas common alleles are disproportionately found in homozygous genotypes. Thus rare alleles always appear to have an advantage. Heterozygous advantage results in a stable polymorphism, hence the alternative name balancing selection.

An interesting article suggesting frequency-dependent selection of **Left-Handedness in Humans**

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated June 8, 2001

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population](#)
[Genetics](#) [Homepage](#)

SEXUAL SELECTION

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Sexual selection is selection for characters that enhance mating success. Darwin was impressed by the fact that qualities of sexual attractiveness were often the reverse of qualities leading to individual survival. He thought that gaining a higher chance to win mates was worth the risk conferred by such characters. Bright colors, long tails, plumes, antlers and horns threaten the survival of the animal but they also give them an advantage in fighting other males or attracting females. To Darwin, sexually selected characters were of no use other than being attractive to the females. He described sexual selection as selection in relation to sex. Wallace, however, thought that all those characters were more than ornaments with some utilitarian quality which females benefited for choosing. In his view, the ornamentations are used to advertise genuine quality as only the healthiest males can afford doing so; mating with them will generate more and healthier offspring.

Sexual selection leads to the evolution of characters that are apparently maladaptive. In view of the modern view of (inclusive) fitness, however, natural selection and sexual selection are not distinct and opposing processes. Superficially, sexual selection favors elaborate sexually dimorphic characters and natural selection opposes it because of their interference with survival. When fitness is seen as reproductive success rather than simply survival, sexual selection promotes morphological characters that would enhance reproductive success, thus, this is no different from what natural selection does. As long as the sexually selected characters do not prevent the individual attaining breeding age, sexual selection increases reproductive success of the individual whatever the overall survival cost is.

Sexual selection causes sexual dimorphism since it usually concerns males. This could lead to misidentification of individual specimens, particularly fossils. Natural selection in relation to sex may cause very rapid evolutionary change (runaway selection; see below). It may even lead to the formation of new species. It is also possible that sexual selection could have caused extinction of some species. The logic behind the fact that sexual selection mainly concerns males is simple. Males produce an enormous number of small sperms without investing too much energy, whereas, females produce a few large and nutritious eggs and they are very costly to them. If a male mates with several females, he increases the number of his progeny, but for females it makes no difference. A female gains only a little from having a large number of copulations with different males. Therefore, males in general are promiscuous but females are choosy because they want their few progeny to be of highest possible quality. There is also another reason. In species where females invest more in offspring, females will be the choosier sex. Females, then, want to increase the quality of their progeny by mating with particular males rather than others. All other things being equal, males are under strong selection to mate with as many females as possible. The selection forces on females act to be selective in terms of which males father their progeny. In summary, eggs represent a limited resource for which males must compete. There is usually a trade-off between mating behavior and parental care.

Competition among males to obtain mates caused the evolution of intrasexual and intersexual selection. Pre-copulatory intrasexual selection (male-male competition) works through male aggression to claim the disputed females and intersexual selection (female choice) promoted the development of sexually attractive conspicuous characters but no fighting. In some mating systems male aggression and female choice for male characters are both involved (post-copulatory counterparts of these mechanisms are sperm competition and cryptic female choice, see below). Elephant seals are polygamous; pregnant females arrive on the breeding ground and

gather together, giving birth in September; within three weeks they are ready for mating again. In October, a single male, which can be three times larger than the female, will claim a harem of as many as a hundred females, and will fight any other male that approaches his territory, proclaiming his ownership with a loud cry. Only fully mature males have the size and strength to keep such a harem. When a male approaches a female, the female cries and starts a fight among males to be copulated by the strongest of them (Le Boeuf BJ, 1972). This is an example of sexual selection enticed by the females. A female elephant uses the same tactic to make sure she is mating with the strongest male in the group. Wood frogs (*Rana sylvatica*) exhibit explosive, synchronous breeding on a spring night. They gather in thousands at ponds where males compete intensively for females and even attempt to dislodge the amplexing male. Females also show choosiness by physically dislodging weakly amplexing males from their backs, therefore showing their preference for larger males. Red deer is another polygamous species in which a few dominant males mate with most of the females. It has to be noted that the harem masters among elephant seals and red deer have very much shorter reproductive lifespans than the females they defend. In elks, during rutting, the larger males are preferred by females through the correlation between the body size and pitch of the voice (the larger the body, the lower the pitch).

Sexual selection in species where females are heterogametic (e.g., having two different sex chromosomes 'ZW' such as birds, butterflies and moths) occurs mainly by the elaboration of ornate male secondary sexual characteristics, whereas in species where females are homogametic (e.g., having two copies of the same sex chromosome 'XX' such as mammals) sexual selection results predominantly in inter-male rivalry and the evolution of traits such as horns, antlers and large body size (all of which are testosterone-dependent). Another important factor determining the pressure of sexual selection is that males in monogamous species are subject to weaker sexual selection than males in polygynous species.

Since for biological reasons females are the choosy sex, it would be beneficial for them if there were some physical indicators of a quality of the genes of a male. Such externally visible cues to good genes would help them to make their choice. One such character would be the age of male since old age itself is a direct marker for survival ability. The selection of males with larger song repertoire by some birds may be due to the correlation of song repertoire with age. Other physical features used for selection by females may indeed be markers for male quality. A female Satin Bower Bird is, for example, attracted to a bower built by a male and is rich in rare blue objects. The males constantly fight and raid each other's bowers, so that a well-kept bower indicates the owner's superiority (Borgia G, 1985).

There are three main theories currently being considered in the explanation of female choice [another grouping consists of the Fisherian run-away, viability indicator, sensory exploitation and antagonistic seduction models (Jennions & Petrie, 2000)]:

1. Run-away Selection Theory (RA Fisher, 1958): According to Fisher, if a majority of females prefer a particular kind of male, other females would be favored if they mate with the same kind of males because their sons will be attractive to many females. Every individual will tend to inherit its mother's genes for preferring its father, and its father's genes for the qualities preferred. These two (groups of) genes will then segregate together and under certain circumstances, due to positive feedback, may lead to runaway selection of more and more exaggeration of the quality preferred. This would continue until the disadvantage, in terms of male survival, exceeds the reproductive advantage for males. Eventually, all the males in the population will end up with a tail length at the optimum point. When all males come to have the same trait, there would be no genetic advantage to a female in choosing one rather than the other. Because of this major problem with this theory, more elaborate forms of it have been developed.

2. 'Good Genes' Theories

a) Strategic Choice Handicap Theory (Zahavi A, J Theor Biol 1975 & 1977): This theory tries to explain the evolution of conspicuous characters from a different point of view. It suggests that the

physical characters that are operational in sexual selection have evolved as handicaps to mark the stronger males who have survived despite the handicaps they bear. There are two kinds of handicaps: conditional or strategic choice handicaps reducing fitness (such as long tail) and revealing handicaps which confer no reduction in fitness (such as the bright coloring). The assumption that the tails of the peacock, birds of paradise and other birds, and the antlers of deer are 'handicaps' that these traits (when inherited by the offspring) will penalize them is not very convincing. The terms 'advertisements' or 'status symbols' may describe these traits better.

b) Revealing Signal Theory (Hamilton & Zuk, 1982): This theory proposes that only males resistant to parasites would be able to display conspicuous features -not necessarily costly to them- to attract females. Thus, the ornaments (such as long tails, inflated throat pouches or bright plumage) simply reveal the state of health without damaging it so they constitute revealing handicaps. Female choice for males with the best-developed sexual characters would result in offspring that are likely to inherit genetically-determined resistance to parasites from their father. This theory is in line with Wallace's opinion about sexually selected characters and offers a possible solution to two serious problems in the theory of sexual selection: (1) choosing such a male is adaptive for females, (2) heritable genetic variation in male characters may be maintained despite the strong directional selection on these characters which should theoretically be short-lived. Since most parasites have shorter generation times than their hosts, the host has to adapt to new varieties constantly which would keep the selection going. Perhaps, as Wallace advocated, ornamentations are advertising genuine quality. The essence of the good genes theory is that an ornament can be sustained only by genuinely healthy males in good condition. Female sticklebacks, for example, use male coloration in mate choice and avoid parasitized males (Milinski & Bakker, 1990). A bright ruddy complexion in Uakari Monkeys signals good health and resistance to monkey malaria; pale faced ones are sickly and have no sex appeal. It has been shown that the offspring of peacocks with more elaborate trains enjoy improved growth and survival (Petrie M, 1994). Similarly, in male great tits (*Parus major*), heritable variation in a plumage is an indicator of viability (Norris K, 1993).

Recently, a set of data supporting the Hamilton & Zuk model was presented from the Molecular Population Biology Laboratory of Lund University (see also Dustin Penn's Work). Pheasants are polygynous and cocks defend a harem with several hens. There is a marked difference in plumage between the males and only the hen cares for the chicks. Spurs are one of the most variable ornaments of male pheasants and they found that males with longer spurs survived better than short spurred ones and this was not a by-product of male-male competition. This observation indicates that spur length reflects male quality and that females can use this trait to choose among potential fathers and pick up only those with good genetic traits to mate with in order to increase the quality of their young. They found that offspring from long spurred males had an increased survival. In a sample of 110 male pheasants, there was a significant difference in spur length between different MHC genotypes. The MHC exists in all vertebrate animals and is the genetic complex involved in immunocompetence and disease resistance. They also found fewer than expected MHC homozygous individuals and different survival of different MHC genotypes (von Schantz T *et al*, 1989, 1994, 1996). These are the first data that directly support the "good genes" hypothesis predicting that females discriminate among males on the basis of secondary sexual characters in order to pass on genes for disease resistance which will improve fitness in their offspring. The same Laboratory also reported that male song repertoire in great reed warblers correlates with post-fledging survival of offspring and, females choose (older) males with large song repertoires, resulting in extra-pair fertilizations with neighboring males (Hasselquist D *et al*, 1996).

It is well established that in vertebrates testosterone has an immunosuppressive effect, and thus males have greater susceptibility to infections (including parasitic ones). Folstad & Karter (1992) suggested that males exhibiting well-developed secondary sexual ornaments, such as the comb on a rooster or antlers on a deer, might possess genes that offset the immunosuppression that often accompanies the production of testosterone-dependent ornaments. It follows that only males with immune systems robust enough to withstand the compromise associated with high levels of testosterone can maintain elaborate secondary sex characters. Females choosing to

mate with such males then have offspring that inherit not only the attractiveness of their father, but his ability to resist pathogens while maintaining elaborate ornaments. This hypothesis also suggests that the ornaments are markers of good genes. Folstad and Karter refer to the ornament as an immunocompetence handicap. For example, male red jungle fowl (*Gallus gallus*) with larger combs have higher testosterone levels but fewer lymphocytes (less immunocompetent) than less-ornamented individuals, suggesting that maintenance of the ornament causes a compromise in immunocompetence (Zuk M et al. 1995). In this model, the good genes are proposed to be immune response genes and the same MHC genes fit into this model well. Recently, Dichtkoff *et al* (2001) have provided evidence for the role of the MHC in the development of antlers in white-tailed deer. They showed correlations between a heterozygous MHC genotype and antler development, body mass and testosterone levels. Thus, the MHC may be involved in sexual selection as the genetic origin of the 'good genes' affecting the development of secondary sexual characters and the quality of the immune response both in intra and intersexual competition.

3. Sensory Bias or Passive Attraction Theory (Parker GA, 1982): This is the simplest and a very convincing theory. It suggests that the sexually selected characters (such as loud sounds, bright plumage, long tails) are simply more conspicuous and more likely to attract the attention of a female from a distance than less intense signals. The evolution of such signals, therefore, may have nothing to do with handicaps or male quality. They may just be the result of selection on males to be noticed by females. Female Natterjack toads, for example, simply choose the louder male. This idea is consistent with a well-known phenomenon by ethologists that many animals respond to supernormal stimuli: bigger, louder, and faster. Females who choose males with conspicuous, easy to recognize signals denoting which species they belong to, mate successfully with little cost. This easy choice behavior would be favored, even if the female gains no additional advantage through the attractiveness of her sons (Fisher's theory) or genetically high quality children of both sexes (good genes theories). Such easy choice signals of males may simply be enabling females to find a mate with minimal cost.

The overall conclusion is that although sexual selection is widely referred to in the literature as if it were a process distinct from natural selection, it should be regarded -as Darwin did- as a form of natural selection. Sexual selection is not some extra force in opposition to natural selection. It is about selection in relation to sex which is governed by the same processes involved in the evolution of characters that are not related to sex. It is basically natural selection acting differently on the two sexes.

The less appreciated form of sexual selection that has gained popularity lately is the **cryptic female choice**. This is a postcopulatory process of sperm selection probably for the genetically most divergent ones (another inbreeding avoidance mechanism). The selection for sperm may start as early as in the vagina, may involve the process called sperm capacitation in the uterus and the long journey of the sperm along the reproductive tract. Finally, the ovum itself may have mechanisms to distinguish between sperms. The overall effect is selective fertilization. This is particularly important in animals where forced copulations (as in yellow dung fly, reptiles and ducks) or extra-matings (as in zebra finch, mice) take place. Cryptic female choice forms the post-copulatory continuation of female mate choice whereas sperm competition is the continuation of male-to-male competition. Cryptic female choice consists of many other mechanisms including selective abortion. Pre-copulatory sexual selection results in non-random mating and post-copulatory selection in selective fertilization/implantation/abortion. Most interestingly, similar mechanisms also exist in plants: pollen competition/selection, differential pollen tube growth, selective fertilization, selective fruit and seed abortion.

The above-mentioned (pre-copulatory) female enticed sexual selection in seals has its correspondent in post-copulatory phase of selection. Female feral fowl prefer dominant males but if they are sexually coerced by subordinate males and their manipulation of dominant males fails, they differentially eject the sperm from subdominant males (Pizzari & Birkhead, 2000). This example shows that sexual selection continues after copulation and post-copulatory mechanisms

exist to fix the mistakes made in earlier phases. In fact, pre-copulatory mechanisms may not even exist for several reasons (males cannot reliably signal their quality or females cannot detect signals). In these situations, females are only left with the post-copulatory mechanisms they can use. By mating polyandrously, sexual selection occurs *in copula* (Jennions & Petrie, 2000). Post-copulatory mechanisms of female choice may be more successful at detecting genetic compatibility because males cannot disguise their identity as easily (Zeh & Zeh, 1997; Newoomer et al, 1999). Genetic compatibility is detected via signals on sperm, and sperm-soma or egg-sperm interactions but avoidance of genetic incompatibility is not the only benefit of post-copulatory sexual selection (Evans & Magurran, 2000).

Sexual selection and speciation: Sexual selection affects the process of mating and the slightest change to the process of mating, if it prevents interbreeding, means a new species exists (via reproductive isolation mechanisms).

Evidence for sexual selection: Preference of prominently colored males of the Guppy (*Poecilia reticulata*) by females correlates with the presence of more colored male fish (Houde & Endler, 1990). The number of colorful fish is high when there is not much predation (sexual selection), but low when there are predators in the habitat (natural selection). Color patterns of natural populations of guppies are, therefore, a compromise between sexual selection and predation avoidance. They are relatively more conspicuous to guppies at the times and places of courtship and relatively less conspicuous at the times and places of maximum predator risk (Endler JA 1991).

The great crested newt provides an example of a correlation between a sexually selected character and general condition. Newts with higher crests during breeding time are in better conditions. Therefore, selection of males in terms of their crest height is selection for healthy ones.

The male fiddler crab *E. albus* has one greatly enlarged claw with which he displays to females; the female fiddler crabs mate preferentially with males with larger claws. In this case this exaggerated male character is positively advantageous in survival terms as it provides enhanced protection against predation by birds.

In seaweed flies, the male character preferred by females is the adult size. The adult size has large additive genetic variance in males, but not in females. The benefit of this selection is that virtually all the variance in male size is attributable to a chromosomal inversion system which is also a major determinant of larval viability, and male size is a reliable indicator of offspring survival for females (Wilcockson RW et al, 1995).

Courtship behavior

Courtship behavior widely varies among species. Cream-like hawk moth's courtship behavior consists of aphrodisiac pheromones, love songs and aerobatic flights. Firefly uses light to attract mates. Fairy tern courtship begins with the male's flying around the female offering her a mackerel. During the courtship, males keep offering females more fish as gifts. Probably most bizarre behavior during courtship is the one observed in mantis. The female chomps off the male's head during copulation and as a result of this, copulation continues more effectively as the brain has an inhibiting effect on male's sexual performance.

Mating systems

A. Monogamy: mating occurs with one mate only. 90% of birds are monogamous. This is because their young require parental care (see below).

B. Polygamy: mating occurs with many mates

1. polygyny: a male mates with several females, occurs especially when males have the control of resources or females. Creates strong selection for larger and stronger males (elephant seals).

2. polyandry: a female mates with several males. This is a very rare mating system. It occurs when suitable breeding sites are scarce and nests are subject to heavy predation (American Jacana or Lilly trotter).

Mating system is determined by several factors

1. Quantity of resources: rich environments tend to favor polygyny, impoverished environments monogamy.
2. Distribution of resources: patchy distribution of resources favors polygyny.
3. Predation: monogamy is favored as a couple would better defend their territory.
4. Other determinants of social behavior: living in groups may alter reproductive behavior.
5. Female availability in time: synchronous breeding among females tends to favor polygyny.
6. Requirements of the young: depending on the development level of the newborn and the level of care it would require, males may be monogamous or polygynous.

In many species, females mate with more than one male as an insurance in case the first male is infertile. Multiple paternity also offers opportunities to increase the genetic diversity and the survival rate of the offspring. A general rule is that the smaller the proportion of males that mate, the more intense will be the selection for larger male size (10% in elephant seals, hence the huge size).

Suggested further reading

Lecture Notes on Sexual Selection (University of British Columbia), Albany University by JL Brown and UCSC by B Sinervo

A Lecture on Sexual Selection and the Biology of Beauty by AP Moller

An Article Entitled "How Females Choose Their Mates" (Dugatkin & Godin, Sci Am, April 1998)

MB Andersson. Sexual Selection. Princeton University Press, 1994

TR Birkhead. Promiscuity: An Evolutionary History of Sperm Competition and Sexual Conflict. Harvard University Press, 2000

(see also the article by the same author: Hidden choices of females. Natural History (US), 2000;109(9):66-71

WG Eberhard & C Cordero. Sexual selection by cryptic female choice on male seminal products - a new bridge between sexual selection and reproductive physiology.

Trends in Ecology and Evolution 1995;10:493

J-G J Godin & LA Dugatkin. Female mating preference for bold males in the guppy, *Poecilia reticulata*. Proc Natl Acad Sci 1996;93:10262-7

TH Goldsmith & WF Zimmerman: Biology, Evolution, and Human Nature. John Wiley & Sons, 2000

JL Gould & CG Gould. Sexual Selection: Mate Choice and Courtship in Nature. Scientific American Library, 1997

J Sparks. Battle of the Sexes. BBC, 1999

MF Willson. Sexual selection in plants and animals. Trends in Ecology and Evolution 1990;5:210

A Zahavi & A Zahavi. The Handicap Principle. OUP, 1997

Human Pheromones and Sexual Selection

The Mating Game at Discovery.com

Encyclopedia Britannica Articles on Sexual Selection: (1) (2) (*subscribers only*)

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on March 27, 2003

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population Genetics](#) [Homepage](#)

Evolution of Sexual Reproduction

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Asexual reproduction is still used by some organisms but in general failed to pass the test of natural selection. Sexual reproduction is the favored way of reproducing for many organisms. In sexual reproduction, new combinations of genes can be assembled on the same chromosomes through recombination. Independent assortment during meiosis, which changes combinations of chromosomes, generates endless genetic diversity. This variation enables a species to overcome novel environmental changes by fast adaptive change. In asexual reproduction, however, natural selection has to wait for some sort of mutation or change due to drift to take place, to act on. Sexual reproduction can also put two beneficial mutations together (although there is always a possibility to break a favorable combination too), or eliminate a deleterious one. The phenomenon that a favorable combination of genes may be broken during meiosis is called the 'cost of recombination'. Overall, groups reproducing sexually can evolve more quickly than those do not, because the combination of beneficial mutations will occur more quickly and deleterious mutations will accumulate more slowly. This is why in eukaryotic multicellular life forms sexual reproduction is a rule. Even in bacteria, which reproduce asexually, there are mechanisms allowing gene transfers between organisms (see conjugation in [Viral and Bacterial Genetics](#)). It is believed that sexual reproduction evolved as early as 2.5 to 3.5 billion years ago (Bernstein H *et al.*, Am Nat 1981). Sexual reproduction is beneficial particularly when environment changes (including changes in the number and genetic constitution of parasites), and when offspring are dispersed widely to end up in different places from their parents. This is why aphids produce winged offspring when reproduced sexually, and wingless ones when reproduced asexually. Similarly, common grass grows asexually (locally) but produces seeds by sexual reproduction to travel away, so that in the new environments, there will be diversity for best adaptation. A host genotype successful against parasites may not be so in the next generation as the rate of evolution in parasites is so fast. The only way in which an animal makes sure that its descendants will be able to deal with different parasites is to reproduce sexually. This is currently the most widely favored theory on the evolution of sex. An alternative theory suggests that it may have evolved as an adaptation for competition with other species.

The evidence suggests that no major group of organisms (except one group of invertebrate rotifers) has evolved and diversified over long periods using parthenogenesis (development of eggs without fertilization 'virgin birth' as in aphids or Daphnia (water flea), or apomixis in plants) alone. When, however, the cost of sex (this is the cost of males, mating and recombination) outweighs the benefits of sex, a species may switch back to asexual reproduction. This is what is believed to have happened to dandelions. Interestingly, some species have adapted to have both modes of reproduction depending on the environmental conditions (see the [link](#) for the Encyclopaedia Britannica article at the end). Aphids reproduce parthenogenetically in the spring and summer when food supply is plenty. This enables them to reproduce very rapidly (and produce wingless offspring). When they face the long winter with no food around, they switch to sexual reproduction and deposit their fertilized large eggs on plants (alternation of generations). These eggs start a new generation of asexually reproducing (winged) females next spring. Similarly, plants may also have alternation in generations with different modes of reproduction. In some plants, one generation consists of diploid plants (sporophytes). Some of their cells undergo meiosis and produce haploid spores. These spores develop directly into haploid plants (gametophytes). Haploid gametophytes produce gametes by mitosis which will fuse to form a diploid zygote that develops into a diploid sporophyte plant. Most algae, fern, mosses and some vascular plants go through these separate phases in their life cycle.

The two important features of sexual reproduction are meiosis (production of haploid gametes to fuse with a different haploid gamete and crossing-over during this process) and syngamy (fusion of two haploid gametes to produce a diploid zygote). These features result in production of offspring significantly varied and rather different from the parents. In asexual reproduction, however, the parent produces progeny that are exact genetic replicas of themselves. Despite the expectation that sex should produce greater genetic diversity, asexual species may have very high levels of heterozygosity for unknown reasons. Asexual reproduction is generally favored when a species lives in a stable environment. The exception of sexual reproduction resulting in haploid offspring is what is seen in social insects such as wasps and bees. In the situation called 'haplodiploidy', male offspring are haploid as they result from unfertilized eggs, and females are diploid -the product of fertilization of an egg by a sperm-. This creates a situation that sisters are genetically 75% identical to each other as they all get the same haploid set from their fathers and either haploid set from their mothers. This has important implications in altruism (see Hamilton's rule in the [introduction](#)).

One consequence of sexual reproduction is anisogamy (the existence of two kinds of gamete, eggs and sperm) which is the basis of morphological and behavioral differences between the sexes (see [sexual selection](#)).

Modes of reproduction

Mitotic parthenogenesis
Sexual parthenogenesis
Self-fertilizing hermaphroditism [simultaneous or sequential]
Sex with polyembryony
Inbreeding sex
Outbreeding sex

Inbreeding: Despite sexual reproduction, inbreeding lowers offspring fitness. This is called inbreeding depression. This occurs because of the presence of deleterious recessive mutations in populations. Inbreeding depression is due to homozygous offspring expressing deleterious alleles. In humans, for example, it is estimated that each individual carries three to five recessive lethal genes. In humans, >40% of progeny born to full sib matings either die before reproductive age or suffer severe disabilities (May RM, 1979). One theoretically possible advantage of inbreeding is that it reduces the cost of recombination. Mechanisms evolved to prevent inbreeding include differential dispersal by two sexes, similarly for plants, differential seed dispersal by the two sexes and the self-incompatibility (SI) system, and **MHC**-disassortative mating preferences (**Penn DJ & Potts WK**, 1999). The most extreme form of inbreeding, self-fertilization, however, still occurs in many plants. This usually occurs in monoecious plants in which both male and female flowers exist on the same plant. A mechanism evolved to prevent inbreeding in plants is dioecious plants which have only one kind of flower (male or female) on each plant. Other mechanisms have evolved to prevent self-fertilization in monoecious plants: differences in flowering times of male and female flowers, in hermaphrodites with perfect flowers, male and female parts of the flower are physically separated, and the self-incompatibility system which reduces the viability of self-fertilized embryos.

Self-incompatibility, is a genetically controlled mechanism to prevent inbreeding in plants. The SI loci prevent matings with self and also reduce matings with close kin; when pollen and maternal tissue share an incompatibility allele, fertilization is prevented, thereby enforcing disassortative matings (Haring V, 1990). The secreted glycoproteins encoded by the SI loci are envisaged to interact with a pollen component to cause arrest of pollen tube growth. This finding that some plants avoid inbreeding through disassortative mating preferences controlled by a highly polymorphic self-incompatibility system provides a striking precedent for the evolution of MHC-disassortative mating to avoid inbreeding in vertebrates (Jordan WC & Bruford MW, 1998; **Penn DJ & Potts WK**, 1999). There seems to be no mechanism to counteract outbreeding depression.

Inbreeding leads to an increase in homozygosity at all loci because the breeding pairs are initially genetically more similar to one other than would be the case if a pair of individuals had been taken at random from the population. Inbreeding distributes genes from the heterozygous to homozygous state. Thus, homozygosity increases without any change in allele frequencies.

Although inbreeding may be harmful to a species, sometimes outbreeding may also cause depression. It appears that there is an optimum amount of inbreeding and outbreeding. Very close relatives (including self) and totally unrelated individuals are unsuitable as mates, but medium to close relatives should be preferred. In plants, fertilization by donors of plants about 10m distant from them generally yields higher numbers of seeds than self-fertilization, fertilization with near-neighbors, or fertilization with plants further away than 10m.

Sexual reproduction in plants: The advantages of cross-pollination are such that plants have evolved elaborate mechanisms to prevent self-pollination and to have their pollen carried to distant plants. Many plants avoid self-pollination by producing chemicals (products of the SI system) that prevent pollen from growing on the stigma of the same flower, or from developing pollen tubes in the style. Other plants, such as date palms, some orchard trees, and stinging nettles, have become dioecious, producing only male (staminate) flowers on some plants and female (pistillate) flowers on others. Some plants are dichogamous—that is, the pistil ripens before or after the stigma in the same flower becomes receptive. See also [Plant Genetics](#).

Summary

Advantages of sex: slower rate of reproduction but faster evolution, lower extinction rates, fast removal of deleterious mutations and better adaptation to host-parasite arms race.

Disadvantage of sex: cost of recombination, cost of mating (competition, no mating, wrong mating), cost of males.

Asexual reproduction: generates genetically identical progeny, conservation of harmful mutations (but also favorable genotypes), new genotypes are generated only through mutation.

A course on [Sexual Evolution](#)

The abstract of a review on [Advantages of Sexual Reproduction](#)

A full-text book by H-R Gregorius on [Mating Systems](#)

Matt Ridley's book: [The Red Queen: Sex and the Evolution of Human Nature](#)

PBS Programs: [Nature of Sex](#) [Evolution: Sex](#)

Encyclopedia Britannica article on [Sexual Reproduction](#) (*subscribers only*)

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated on March 26, 2003

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population](#)
[Genetics](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population](#)
[Genetics](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

Molecular Clock

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

The controversial hypothesis of molecular clock (MC) is a consequence of the neutral theory of evolution. It holds that in any given DNA sequence, mutations accumulate at an approximately constant rate as long as the DNA sequence retains its original functions. The difference between the sequences of a DNA segment (or protein) in two species would then be proportional to the time since the species diverged from a common ancestor (coalescence time). This time may be measured in arbitrary units and then it can be calibrated in millions of years for any given gene if the fossil record of that species happens to be rich. MCs do not behave metronomically, i.e., neutral mutations do not yield literally constant rates of molecular change but are expected to yield constant average rates of change over a long period of time (known as 'stochastically constant' rates). The concept of MC fits well with Kimura's neutrality theory because the rate of neutral evolution in genetic sequences is equal to the mutation rate of neutral alleles. Even where selection operates, an averaging of selection coefficients over long periods could also provide a MC for times of divergence. Undeniably, different DNA sequences (or proteins) or even different parts of the same gene (or protein) evolve at markedly different rates. Different functional constraints on structure and varying intensities of selectional forces as well as generation time (the rate of meiotic production of gametes) are the major sources of variation in evolutionary rates. In general, rRNA evolves slowly and mtDNA rapidly. Among other proteins, fibrinopeptide changes relatively rapidly and is useful for studying closely related species but cytochrome c is useful for studying the whole period of life on earth. Fast mutating sequences cannot be used to go back in evolutionary time because the mutations effectively randomize the sequences. It is also possible that reverse mutations will occur and the comparisons will be very difficult. Introns and pseudogenes evolve rapidly and nondegenerate sites in protein coding sequences (exons) slowly. Molecular differences between species are used to infer phylogenetic relationships. Molecular evolution from living fossils provides an example that constant rate of molecular evolution occurs independent of morphological evolution.

A MC must be calibrated first and this requires a reliable fossil record. Only after this calibration, a MC can be used for phylogenetic inference. A conventional calibration for

evolutionary rate of animal mtDNA is about 2% sequence divergence per million years between pairs of lineages separated for less than 10 million years (or 20×10^{-9} substitutions per site per year). Beyond 15-20 million years, mtDNA sequence divergence begins to plateau, presumably as the genome becomes saturated with substitutions at variable sites. 16S rRNA gene, however, has an evolutionary rate of 1% sequence divergence per 50 million years. Although mean evolutionary rates in the nuclear genome may vary among taxa, they do so in a consistent fashion. For example, molecular evolution appears slower in primates than in rodents and especially slow in hominoids. Using the highly conserved protein cytochrome c sequences, it has been found that the two most divergent species by far are two species of ascomycetes (fungus) that separated twice as long ago as the ancestors of mammals separated from the ancestors of insects.

To test a MC, a relative rate test that does not depend on absolute divergence times is used. Each test requires at least two related species (A and B) and an outside reference species (C) known to have branched off prior to the separation of A and B. Theoretically, the distances (A to C) and (B to C) estimated by the MC should be similar. If there is a statistically significant difference, then this particular MC is unreliable for this taxa. Many relative rate tests have been conducted for molecular data from different species. Uniform average rates of DNA clocks in birds, rats, mice and hamsters have been established. The relative rate test does show up some discrepancies in molecular rates in many cases. This is not surprising as many allele frequencies are clearly modulated by natural selection. The main advantage of today's favorite phylogenetic analysis method, cladistics, is that it does not rely on MCs.

Microsatellite loci have properties that make them suitable for dating evolutionary events. The mutation rate is so high (5.6×10^{-4} in dinucleotide microsatellites) that a reasonable number of mutational events can occur within the short time of modern human evolution. For CA dinucleotide microsatellites, the average estimate for mutation rate is about 1/5000 per generation. The application of the microsatellite data to human population studies gave an estimate of 156 (95 to 290) thousand years for separation of Africans and non-Africans.

Since most molecular systems evolve at heterogeneous rates across most taxa, if precise clocks exist, they are local rather than universal. MCs keep far from perfect time, but in certain cases (where they have been tested and approved), they provide invaluable information in phylogenetic studies.

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated Jan 3, 2003

[Back to HLA](#)

[Back to MHC](#)

[Back to Genetics
Homepage](#)

[Back to Biostatistics](#)

[Glossary](#)

Mitochondrial DNA

M. Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Mitochondrial DNA (mtDNA) of higher animals is a circular molecule of some 16,000 bases. It corresponds to chloroplastic DNA of plants both of which are collectively known as cytoplasmic DNA. The mtDNA has no repetitive DNA, spacers, or introns. It encodes 13 mRNAs, 22 tRNAs and 2 rRNAs. mtDNA is usually the only type of DNA to survive in ancient bone specimens because of its abundance; 500-1000 copies per cell instead of only two copies of most nuclear DNA. Unlike nuclear DNA which gets mixed around each generation (this is each meiosis), the only alteration to mtDNA is an accidental change caused by mutation, copying errors, or other accidents, i.e., it does not recombine. mtDNA is maternally inherited. All the copies in an individual are usually identical but populations may be highly polymorphic. The lack of polymorphism within individuals suggests that at some point in the germ-line, the effective number of copies must have been small.

The rate of base substitution in mtDNA is much higher than nuclear DNA. An estimate of the initial rate of sequence divergence is 20×10^{-9} per site per year per evolutionary line (this is 2% sequence divergence per million years between pairs of lineages; 10 times faster than the highest rates in nuclear DNA). This estimate applies only to the first 10 million years of species separation after which mtDNA sequence divergence begins to plateau as many bases are conserved and the genome becomes saturated with substitutions at variable sites. This high rate of substitution makes mtDNA particularly valuable in studying the relationships in recently diverged lineages.

The fact that the mtDNA is inherited only through the female line without crossing-over provides unique information to phylogenetic studies as it preserves information about ancestry. The only source of sequence variation is mutation at a well-worked out stochastically constant rate so that divergence times (coalescence times) can be estimated.

Among human populations, restriction maps of mtDNA reveal rather little sign of geographical structuring. This suggests that existing human populations migrated from a common center relatively recently. Based on the divergence rate of 20×10^{-9} , the mean time of divergence of human races is of the order of 50,000 years. mtDNA data have also been used to estimate effective population size (N_e ; those contributing to the next generation) in the past. The total divergence time elapsed

since a single ancestral sequence (coalescent) gave rise to today's variants is $4xN_e$ generations (time for a neutral mutation to get fixed). mtDNA is haploid and maternally inherited; hence, the mean coalescence is $2xN_f$, where N_f is the number of mothers. When the total divergence is estimated within the existing human populations, based on the sequence divergence rate of $20x10^{-9}$ per year, it has been concluded that the single common ancestor for all the variant sequences in today's human populations existed 10 to 20 thousand generations ago (200 to 400 thousand years, assuming 20 years as generation time). This means that N_f must have been about 5 to 10 thousand.

It is important to understand the claim that all existing human mitochondria are probably derived from a single female living less than half a million years ago does not mean that our ancestral lineage was ever reduced to a single pair or that only one female contributed to our nuclear genome. Surely most of the females within the effective population at the time contributed to our nuclear genome but the female all our mtDNAs trace back lived 200 to 400 thousand years ago. More technically, the extant mtDNA alleles coalesce to a single ancestral molecule extant at that time. It is a mathematical certainty that each gene will coalesce into one ancestor, the others just have not been able to make it today (similar to the extinction of surnames). Each one of our 40,000 genes can be traced back to its own ancestral gene. It is important to remember that only those females who have daughters have the chance to pass on their mtDNA to following generations. A similar calculation has been made for HLA-DRB1 genes. All extant DRB1 alleles seem to have derived from one ancestor lived more than 65 million years ago.

One study compared a long stretch of mtDNA in humans in different primate species. This study by Horai et al. (**J Mol Evol 1992;35:32-43**) showed that humans are closer to chimpanzee than gorillas in terms of their mtDNA genealogy. When the divergence of mtDNA is compared among populations, the initial mtDNA split was found between Africans and others, followed by progressively younger calibrated ages for specific Asian, Australian, New Guinea, European, and native American mtDNA types. This pattern has been thought to support the Out of Africa model in the origin of modern humans. This estimate is based on the concept that greater sequence variation in mtDNA on a continent is a sign of greater longevity. African populations are the oldest because they harbor the greatest mtDNA variation.

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated June 8, 2001

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population Genetics](#) [Homepage](#)

MIMICRY

M. Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

[Mirror Site](#) (no pop-up ads)

Mimicry is the resemblance of one organism (mimic) to another (model) such that these two organisms are confused by a third organism (receiver). The model and mimic are not usually taxonomically related. In molecular mimicry, pathogenic organism (or a parasite) mimics a molecule of the host so that it escapes recognition as foreign (a kind of aggressive mimicry, see below). An evolving mimicry takes advantage of previously evolved communication signals and responses between organisms (for example, between a predator and a warningly colored prey). To be successful and beneficial to the mimic, the model should be an abundant species whose noxious characteristics have left a lasting impression on predator.

Batesian mimicry: First described by the British naturalist Henry Walter Bates in 1852. He found two unrelated but similarly marked families of Brazilian forest butterflies one of which (model) was poisonous to the birds and the other palatable ones (mimic) survived because of the resemblance to the poisonous ones. They usually mimic the aposematic coloration of the model species. In this kind of mimicry, the mimicking organism has evolved some features of a poisonous organism but is not poisonous itself. This is essentially equivalent to camouflage. Batesian mimicry is particularly common among insects. The mimicry by grasshoppers of poisonous tiger beetle is another example from the insect world. Theoretically, selection only favors the mimic if it is less common than the model. The fitness of the mimics is negatively frequency-dependent.

Mullerian mimicry: The German zoologist Fritz Muller proposed an explanation to Bates's paradox in 1878. Bates had observed a resemblance among several unrelated butterflies all of which were inedible. This paradoxical observation puzzled him. Muller realized that the explanation might lie in the advantage to one inedible species in having a predator learn from another. Once the predator has learned to avoid the particular conspicuous warning coloration with which it had its initial contact, it would then avoid all other similarly patterned species, edible or inedible. Maximum protection is gained by Mullerian mimics when all individuals have the same signal (signal standardization). The number of individuals sacrificed in educating the predators is spread over all of the species sharing the same warning pattern (called mimicry rings). This tendency of inedible and noxious species to evolve to have the same or similar warning signals is called Mullerian mimicry. One example is the black and yellow striped bodies of social wasps, solitary digger wasps and the caterpillars of the cinnabar moths. Mullerian mimicry could be considered not to be true mimicry because the receiver is not actually deceived and it is not obvious which organism is the model and which one is the mimic.

Aggressive mimicry: The organism mimics a signal that is attractive or deceptive to its prey. The examples are the egg mimicry by cuckoos and praying mantis mimicking flowers and vegetation to attract insects (a wolf in sheep's clothing). Another example is

that cuckoo bees lay their eggs in the nests of humblebees, which they closely resemble. Host mimicry by parasites, in which the host is both the model and receiver, is an extension of aggressive mimicry. Most examples occur in birds and between viruses and their hosts including humans.

A lecture on mimicry from [University College of London](#)

Examples of [Mimicry in Sea Animals](#)

An article by Lev-Yadun on [Aposematic Coloring in Plants](#)

A high school activity on [Mimicry](#) with a list of examples

[Female mimicry in garter snakes](#) by Mason & Crews, 1985 & by [Shine et al, 2001](#)

[Motion camouflage](#) by Dragon Flies (New Scientist)

Encyclopedia Britannica article on [mimicry](#) (*subscribers only*)

[M.Tevfik Dorak](#), BA (Hons), MD, PhD

<http://www.openlink.org/dorak>

Last updated March 27, 2003

[Back to Evolution](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Genetics](#)

[Back to Biostatistics](#)

[Glossary](#) [Homepage](#)

[Back to HLA](#)

[Back to MHC](#)

[Back to Genetics](#)

[Back to Evolution](#)

[Biostatistics](#)

[Population](#)

[Genetics](#) [Homepage](#)

Speciation

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Species: A population of organisms interbreeding only with each other. Subspecies are genetically diverged groups of a species. In taxonomy, the term race is used interchangeably with subspecies. Species ranks below family and genus. In practice, taxonomists identify species on the basis of morphological characteristics with the help of more-detailed genetic and chromosomal analyses if necessary.

Different species concepts

Biological species concept (Dobzhansky, 1937; Mayr, 1940): Species are groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups. Speciation is thus seen in terms

of the evolution of isolating mechanisms and is said to be complete when reproductive barriers are sufficient to prevent gene flow between the two new species. The problem is that capacity to interbreed cannot always be tested nor can the potential for interbreeding. For asexually reproducing organisms and fossils, this concept does not apply.

Recognition species concept (Paterson, 1985): A species is defined as an inclusive population of individual biparental organisms which share a common fertilization system. Speciation thus occurs when a different fertilization system evolves. This definition can sometimes be applied to fossil species. In a single habitat in USA, up to 40 cricket species live together. Each one of them, however, has a different song sung by the male and recognized by the female which altogether make up the mate recognition system. The crucial event in the formation of a new species is the evolution of a new mate recognition system. The advantage of this concept is that mate recognition systems can be observed directly whereas interbreeding (the biological species concept) may have to be inferred indirectly.

Biological and recognition species concepts only apply to species that reproduce sexually and neither allows for the existence of hybrids between species.

Evolutionary species concept (Simpson, 1951): A species consists of all individuals that share a common evolutionary history. It is not always clear what constitutes a common evolutionary history, and chronospecies (species occurred with gradual evolution along the same line) are part of the same lineage but still different species. The predecessor of a chronospecies is said to have gone pseudo extinction. It is applicable to both living and fossil species and also to both sexual and asexual species. It fails to say anything about how speciation occurs.

Description of species

Allopatric: If they occupy different ranges

Sympatric: If they coexist in the same habitat

Parapatric: If their ranges are adjacent and if they have a zone of contact

Speciation: The formation of new species may involve transformation of one species into another (anagenesis) or splitting up of one species into a number of others (cladogenesis). Speciation is generally described as multiplication of species by the division of one species into two or more separate species, thus leaves out the anagenetic speciation. Speciation is the direct result of changes in the gene pool. Isolation (with subsequent reduction in gene flow) and disruptive or diverging selection result in speciation. The common factor in all mechanisms of speciation is a reduction in gene flow between two populations. This starts the

divergence and speciation eventually occurs. Speciation is not necessarily adaptive. Gene flow in allopatric models is reduced by geographical separation, whereas gene flow in sympatric or parapatric models is reduced by other means.

The relative importance of adaptation in speciation is controversial. Chance, however, plays some part in speciation as in emergence of allopatry. When divergences (eventually leading to speciation) between populations are not adaptive, they are attributed to genetic drift (some Hawaiian *Drosophila* species and the North American flowering plant *Clark lingual*).

Allopatric speciation: This is the most common pattern and it takes place when populations become geographically separated. Progressive divergence as a result of physical separation leads to speciation. The debate is whether adaptation or chance (genetic drift) plays a major role in divergence in allopatry (in small population it is more likely to be drift). Speciation of *Drosophila* in Hawaii following volcanic eruptions is an example. When two species get together after a period of allopatric divergence (secondary contact): (1) they have already speciated and cannot interbreed; (2) their hybrids have lowered fitness and natural selection rapidly acts to develop reproductive isolation mechanisms; (3) they interbreed successfully and mix again as single species. If one of the split populations is very small a major role will be played by the founder population in the development of divergence (drift will be fast, inbreeding will lead to increased homozygosity, ecological shift will occur rapidly, intraspecific competition will be low and the population will have a flush phase). In such small population faster divergence results in a rapid budding kind of speciation (*peripatric speciation* or founder effect speciation). However, in small populations likelihood of extinction is very high before any allele is fixed and many alleles may be lost as a result of inbreeding.

Sympatric speciation: Results from disruptive selection for alternative adaptive models (disruptive selection= selection for the extreme phenotypes instead of the intermediate ones). Changes in host, food or habitat preference, resource partitioning may start sympatric speciation. Habitat isolation within a sympatric species would stop interbreeding and may lead to speciation. It is interesting that closely related sympatric insect species usually use different host plants while closely related allopatric species use identical or similar plants. A common example of sympatric speciation from the same species would involve their colonizing different trees to lay their eggs. The European mosquito *Anopheles* group consists of six morphologically indistinguishable species. They are isolated reproductively as they breed in different habitats. Some breed in brackish water, some in running fresh water and some in stagnant fresh water. Therefore, they never meet to breed. If this happens for subpopulations of a species, speciation may follow.

In North American species of lacewings (genus *Chrysopa*), speciation may have been initiated by disruptive selection on genetic variation in color, favoring

homozygotes in their respective habitats, where they are protected against predators. The intermediate ones do not have a cryptic color and are eliminated quickly (in fact, *C.downesi* and *C.carnea* have different breeding times). In this example of sympatric model, speciation is initiated by disruptive selection operating on a single freely interbreeding population which is followed by physical separation (only after the genetic divergence has already begun). In allopatric model, geographical separation always occurs before genetic divergence (and initiates it).

Another sympatric model in which speciation does not require habitat isolation is competitive speciation which results from slight differences in resource utilization. This intraspecific competition can lead to the establishment of a stable polymorphism in sympatry even in a homogeneous environment. The end result is the division of a single gene pool into two or more adaptive types.

Parapatric speciation: In this model, geographical separation is not complete and two diverging populations share a boundary with no barrier to dispersal across it. This is also a result of disruptive selection. An ancestral species spreads over a spatially variable area and this leads to geographical separation by primary contact area. Geographical differentiation leads to the formation of a cline, which acts as a barrier to gene flow so that further divergence can take place in the populations on each side of the cline. Hybrids between two parapatric populations are less fit and assortative mating is favored by natural selection. The reduction of gene flow through a hybrid zone will depend on the dispersal distances, selection against hybrids and selection against pure types either in the hybrid zone or on the wrong side of the hybrid zone. Indeed, in nature, when a species covers a large geographical area, individuals at the extreme ends of its distribution can be very different.

Reproductive isolation

After subsequent sympatry (secondary contact), initially slight differences in mate recognition traits are exaggerated by selection in favor of pre-zygotic isolation - through assortative mating- (reinforcement theory). Pre-zygotic isolating mechanisms due to mate recognition may evolve during allopatry. This is called recognition in allopatry hypothesis by Maynard Smith. For sympatric species, pre-zygotic isolation -through natural selection- evolves more rapidly between species who produce unfit hybrids.

Isolating mechanisms prevent gene flow between sympatric species. These may be pre-zygotic preventing the formation of hybrids or post-zygotic preventing the reproduction of hybrids. Pre-zygotic isolating mechanisms are the result of natural selection favoring isolation to prevent the waste of reproductive effort. Genetic drift may also play a role in the development of reproductive isolation (see below).

Some pre-zygotic (post-mating) isolating mechanisms:

1. Allopatric separation
2. Ecological or habitat isolation (sympatric) (Anopheles group)
3. Different flowering, pollination (in plants) or mating season (temporal isolation) (Pinus radiata and P. muricata are sympatric but shed their pollens at different times. Hybrids are rare and less vigorous. The American toads Bufo americanus and B. fowleri have different breeding times and do not mate. Their habitat preferences are different too.
4. Ethological (or sexual) isolation: The sexual attraction between males and females is reduced or absent. Differences in courtship patterns in Drosophila species in Hawaii is an example. Sexual selection and assortative mating result in ethological isolation. In general, when strong, positive assortative mating maintains the polymorphism, the different types may become genetically distinct, and these may eventually become true species.

Slight changes caused by genetic drift could lead to rapid divergence in reproductive morphology and sexual behavior between two populations as a result of runaway selection. Changes in only a few genes controlling a male sexual trait and female preference for that trait could lead to pre-mating reproductive isolation between two populations and hence to speciation without any adaptive value.

5. Isolation by different pollinators: The two species attract different kinds of insects etc. as pollinators and their gametes never get together.
6. Mechanical isolation: The reproductive organs of the sexes are not anatomically identical and this impedes reproduction. Lack of pollen tube growth down style of a different plant species is an example.
7. Gametic isolation: Gamete transfer takes place but fertilization does not occur. Many species of Drosophila show an insemination reaction as a result of which sperm is killed in the vagina.
8. A genetic change in some members of the population (like chromosomal reorganizations such as polyploidy)

Some post-zygotic (post-mating) isolating mechanisms:

1. F1 hybrids inviable: Fertilization occurs but embryonic development does not. In crosses between sheep and goats, the embryos die early in their development.

2. F1 hybrids infertile: The hybrids occur but do not produce functioning gametes. The classic example is the cross between female horse and male ass, the mule.

3. Hybrid breakdown: F1 hybrids are viable and fertile, but F2, backcross or later-generation hybrids are inviable or infertile.

Reproductive character displacement is a post-mating mechanism imposed by natural selection to prevent hybrid formation when they are inviable or infertile. It involves increased differentiation between the reproductive systems of two species living in sympatry. This phenomenon has nothing to do with speciation because it involves already differentiated species.

Endemic species

Speciation on isolated islands is an example of allopatric speciation. These areas provide a unique combination of empty habitats, novel environments, geographical isolation and sometimes lack of predators. In these new habitats, selection pressures will be strong and invading groups will evolve quickly, producing new species as a by-product. The ancestors of the faunas and floras of oceanic islands are believed to have arrived by long-distance dispersal mechanisms. Once the organisms have arrived in these isolated places, in the absence of competitors, they would diversify and speciate as they occupy specialized niches. The only constraint would be intraspecific competition. This kind of rapid speciation is driven by natural selection rather than genetic drift and is called adaptive radiation. In these small founder populations, genetic drift would also play a role at neutral loci in addition to strong selective forces acting on other loci.

1. Darwin's Finches of Galapagos Islands: Of the 14 species of finches living on the islands today, 13 are endemic. Finches are poor fliers and colonization from the mainland could not have been frequent. On the other hand the marine birds are strong fliers and only two of the 13 marine birds are endemic. The differences in beak shape and size of finches correlate with their feeding habits.

2. Cichlid fishes of lake Victoria: There are 170 species of the cichlid genus *Haplochromis* alone in Lake Victoria believed to have originated from a single ancestral species or a group of species. None but one lives elsewhere. The difference in dental patterns correlate with their feeding habits. Reproductive isolation between species is very marked, and male coloration may be a major factor in species recognition during mating. It is believed that they diverged each time water levels fell in the lake isolating peripheral population from the main body of the lake. Sexual selection seems to have involved in the speciation process.

3. Hawaiian Drosophilids: More than a quarter of the known species of the family *Drosophila* live in the Hawaii archipelago. Out of probably more than 800 species

in Hawaii, 95% are endemic. All of the endemic species may be descendants of a single gravid female arrived at the oldest island Kauai about six million years ago. The habitat on these volcanic islands is very patchy and have been repeatedly split up by lava flows. It is clear that founder events have occurred frequently in these patches leading to new species. Despite the huge morphological diversity, there is little genetic diversity among the endemic species. It appears that courtship behavior has driven speciation in Hawaiian islands. Even closely related sympatric species have differences in their reproductive behavior. This is an example of sexual selection causing speciation probably due to the effects of genetic drift in small founder population.

(Hawaiian Honey Birds)

4. Silversword alliance: In Hawaii, 95% of the native plant species are endemic.
5. Tristania (endemic snails of the Tristan de Cunha archipelago): All six species of snails from the genus Tristania are different from the rest of the snails on other Atlantic islands.
6. Lemurs in Madagascar and Comoro islands.

Suggested reading: Coyne JA: Genetics and speciation. Nature 1992;355:511-5.

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated June 8, 2001

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Biostatistics](#) [Population](#)
[Genetics](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

CLADISTICS

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Definitions

clade: A group of organisms evolved from a common ancestor (from Greek *klados* meaning branch). An ancestor and all its descendants are called a clade. Clade is synonym of monophyletic group.

cladistics: A method of classification of animals and plants on the basis of those shared derived characteristics that were not present in their distant ancestors (synapomorphies) which are assumed to indicate common ancestry. It uses strict monophyly as the only criterion for grouping related species. In cladistic taxonomy, evolution is seen as a process of progressive bifurcations of lineages. Every species, therefore, has a sister species whether recognizable or not, and this pair derived from an ancestral species. Two taxa are more closely related to each other (sister groups) if they share a more recent common ancestor relative to a third. Anagenetic component of evolutionary change is ignored in cladistics. Cladistics is accepted as the best method available for phylogenetic analysis because it recognizes and employs evolutionary theory.

cladogram: The output as a branching diagram from a (cladistic) phylogenetic analysis postulating relationship of different taxa. All cladograms are hypothetical and none can be proved correct for sure. Among the alternative cladograms the one which is best supported by the character/sequence data is the most representative one.

cladogenesis: The evolutionary division (bifurcation) of lineages causing a proliferation of species.

anagenesis: Descent with modification within any given single lineage (change without speciation). Anagenetic changes create grade groupings (as opposed to clade groupings).

grades: A grouping of taxa which show similar modifications with respect to their ancestors. Grade taxa may be polyphyletic, paraphyletic or monophyletic. A paraphyletic taxon is the lowest grade in a phylogenetic tree.

phylogeny: Evolutionary relationship that can be studied in three ways:

1. Phenetics (Mayr): Method of classification in which taxa are grouped together with other taxa that they most closely resemble phenotypically. It is based on overall similarity. It accepts all monophyletic, paraphyletic and polyphyletic groups. An important issue in the phenetic method is the question of measuring phenetic similarity (it is not objective). If molecular sequence data is used in phenetic analysis, stochastically constant rates of molecular evolution (the molecular clock theory) must be assumed if the phenogram is to be equated with a phylogeny. Pheneticists usually make no assumptions to distinguish sources of resemblance. Any similarity, whether symplesiomorphy or synapomorphy, is taken into account.

2. Evolutionary Systematics (an eclectic approach, Simpson). This is a synthesis of phenetics and cladistic principles. There is no single underlying method of analysis. It defines groups by homologies (although without distinguishing between primitive and derived homologies) and ignores analogies. (This method uses both paraphyletic and monophyletic groups but excludes polyphyletic groups.)

3. Cladistics (Hennig): Inferred recency of common ancestry. The members of a group in a cladistic classification share a more recent common ancestor with one another than with the members of any other group. Only monophyletic groups are used in this analysis. Cladists attempt to distinguish between symplesiomorphic and synapomorphic similarity and to identify clades on the basis of synapomorphs only. A cladogram shows only the hierarchical arrangement of increasingly inclusive synapomorphies observed in a set of compared taxa. A cladogram is not precisely equivalent to a phylogeny, because none of the compared taxa is recognized as an ancestor. The nodes can be treated as representing unknown ancestors.

plesiomorphy: A primitive (ancestral) character state (the original state of the characteristic). This character is present in the ancestor or in the outgroup.

Criteria suggesting primitiveness of a character:

1. Presence in fossils (paleontological antiquity),
2. Commonness in an array of taxa,
3. Early appearance in ontogeny,
4. Presence in an outgroup.

apomorphy: A derived or newly evolved (specialized, advanced) character state (the changed state of the characteristics). For example, presence of hair is a primitive character state for all mammals, whereas the hairlessness of whales is a unique derived state (autapomorphy) for one subclade within the Mammalia.

symplesiomorphy: A primitive character state shared by two or more taxa.

synapomorphy: A derived character state shared by two or more taxa and held to reflect their common ancestry. This is the key in inferring relationships to common ancestry. It is not just the presence of shared characteristics that is important (as in phenetics) but also the presence of shared derived characteristics. Synapomorphies can be identified by the study of developmental patterns (ontogeny) or outgroup comparison.

autapomorphy: A derived character state unique to one taxon. A species with an autapomorphy is unlikely to be the ancestor of the other species in the clade on the grounds of parsimony. It is more parsimonious if it is the further evolved species (rather than the ancestor). The absence of autapomorphies allows the possibility of a species being directly ancestral to another.

monophyletic group: A set of taxa containing a common ancestor and all of its common descendants.

paraphyletic group: A paraphyletic group can be recognized from a cladogram as a group representing a primitive grade grouping and including some but not all of the descendants of a common ancestor. A paraphyletic taxon is united by common possession of shared primitive characters (symplesiomorphy) but not the derived (specialized) characters, thus, those included in the paraphyletic group have changed little from their ancestor. In other words, a paraphyletic group lacks the derived character states of the other descendants. The taxa included in a paraphyletic group are those that have continued to resemble the ancestor; the excluded taxa have evolved rapidly and no longer resemble their ancestor. Among the lizards, crocodiles and birds, the reptiles (lizard+crocodile) are paraphyletic to birds. In determination of the evolutionary relationships, paraphyletic groups are best avoided as they convey no information. Paraphyletic grouping implies a closer relationship among the taxa within the group, but the descendants left out may be more closely related to a taxon in the paraphyletic group.

polyphyletic group: A set of taxa descended from different ancestors. The ultimate common ancestor of all taxa in the group is not a member of the group. A polyphyletic taxon assembles species with independently evolved similarities derived from separate ancestors. Birds and bats evolved to have wings but are not descendants of a common ancestor, thus, they form a polyphyletic group. Such taxa are rejected from modern systems of classification except phenetics.

sister species: taxa stemming from the same node in a phylogeny. They may be morphologically identical but will be genetically divergent.

homology: Similarity by common ancestry; a character shared by a set of species which was present in their common ancestor (a similarity in a character or sequence between two species may be due to homology 'evolutionary relationship' or analogy 'convergent evolution'). For a homology to become established, characters must also occur in the same topographical position within an organism and also agree with other characters (congruence) about relationship of taxa. If the structure has been modified through descent in one lineage (anagenesis), it may be difficult to establish homology.

Homologies are divided into two groups: derived homologies are those unique to a particular group of species and their ancestors; ancestral homologies are found in the ancestor of a group of species and some, but not all, of its descendants. Recognition of homology becomes increasingly difficult with increased time since divergence.

analogy: Characters showing similarity due to convergence (as opposed to shared ancestry). Endothermy in birds and mammals is believed to be analogous.

homoplasy: The component of overall similarity due to convergence from unrelated ancestors. Repeated events in the evolution of a taxon that result in possession by two or more species of a similar or identical trait that has not derived from a common ancestor; A homoplastic character may arise from convergence, parallel evolution, or evolutionary reversal. Homoplasies create noise in cladistic analysis unless overcome by a large number of informative synapomorphies.

parallel evolution: The evolution of similar or identical features independently in related lineages, thought to be based on similar modifications of the same developmental pathways.

outgroup: A related taxon retaining many primitive characters and is believed to have a common ancestor more distantly in the past than the taxa being classified. Baboon or macaque are an outgroup to humans, chimpanzees and gorillas because they split off from the common ancestor of humans, chimpanzees and gorillas earlier. (A taxon that diverged from a group of other taxa before they diverged from each other.) The outgroup taxon is the least related one to the others in a group of taxa. Outgroup comparison of homologous sequences in cladistic analysis of molecular data is used to determine the primitive-derived polarity of alternative bases at a given position. A prominent danger in outgroup comparison of molecular data arises from the probability that some mutations (e.g., base substitutions) may be repeated and others reversed. This may be brought about by the appearance of plesiomorphy in lineages which in fact is doubly apomorphic. Presence in an outgroup is the best indicator for a character to be primitive.

coalescent (coalesce=unite into one whole): The approximated ancestral allele/haplotype/ sequence in a cladogram from which all other alleles in the sample may have descended (the ancestral sequence). If a population remains intact, the average coalescence time is the substitution time for neutral alleles.

Principles of Cladistics

There are three basic assumptions in cladistics:

1. Any group of organisms are related by descent from a common ancestor,
2. There is a bifurcating pattern of cladogenesis,
3. Change in characteristics occurs in lineages over time. This is the most important and the least controversial assumption.

Hierarchical order in nature is manifested in the distribution of characters shared among organisms. Common ancestry is worked out from the number of synapomorphies among taxa. For any three taxa, two taxa are more closely related to each other if they share a more recent common ancestor relative to a third.

In a phylogenetic tree, it is assumed that all branches are dichotomous (not polytomic). A single ancestral lineage split to give rise to two descendants. It is also assumed that lineages split but never rejoin lineages (reticulation). These are unrealistic assumptions of cladistics which are violated in real life.

Methods in phylogenetic analysis

1. Analysis of character states: synapomorphies are the basis for cladistics. (A supposed synapomorphy should not be the result of independent evolutionary development!) The

hypothesis of a fossil species being a direct ancestor may be rejected if it possesses autapomorphies.

2. Outgroup comparisons

3. Embryology: ontogeny recapitulates phylogeny, development proceeds from general to particular (systematic hierarchy).

4. Molecular sequence data: For a group of closely related organisms, one needs a rapidly evolving molecule that varies sufficiently within the group. Mitochondrial DNA (mtDNA) is a good example. It is abundant in the cell, therefore, easier to get hold of from some fossils; its inheritance is strictly maternal, its total sequence is known and short; and it does not cross-over (Y chromosomes do not cross-over either). For a more distantly related group, a slowly evolving molecule such as rRNA would be better. The more similar the sequence, the more recent we infer that the duplication event has occurred. The selection of unambiguously aligned DNA segment is essential to ensure homology in sequence comparisons. Chance matching in non-homologous alignments would confuse the cladistic analysis.

5. Area cladograms

Methods to construct an evolutionary tree (phylogenetic inference)

1. Maximum parsimony: The selection of the simplest phylogenetic tree requiring the least number of substitutions from among all possible phylogenetic trees as the most likely to be the true phylogenetic tree.

2. Maximum likelihood: For this method, protein sequences are much more reliable than the DNA sequences.

3. Neighbor-joining: A simplified version of the minimal evolution method. An evolutionary distance is computed for all pairs of sequences, and a phylogenetic tree is constructed from pairwise distances by using the smallest distances by inferring a bifurcating tree.

How to construct cladograms

1. Choose the taxa which are evolutionarily related,

2. Determine the characters,

3. Determine the polarity of characters (original or derived) by:

i. outgroup comparisons (primitive character),

ii. presence in fossils (primitive character),

iii. early appearance in ontogeny (primitive character),

iv. commonness in an array of taxa (primitive character).

4. Group taxa by synapomorphies not symplesiomorphies,

5. Work out conflicts that arise by some clearly stated method like parsimony (minimizing the number of conflicts),

6. Build the cladogram following these rules:

i. all taxa go to the endpoints of the cladogram, not at nodes (splitting points). The two taxa split from the node are called sister taxa,

ii. all nodes must have a list synapomorphies which are common to all taxa above the node (unless the character is later modified),

iii. all synapomorphies appear on the cladogram only once unless the character state was derived separately by evolutionary parallelism.

M.Tevfik Dorak, BA (Hons), MD, PhD

Last updated July 26, 2002

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Biostatistics](#) [Glossary](#)
[Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

Major Histocompatibility Complex

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

[Compatibility Systems in Nature](#)

[Origin of the MHC and Its Polymorphism](#) [Evolutionary Histories of HLA-DRB Haplotypes](#)

[MHC: Structure and Function](#) [MHC and Leukemia](#)

[Glossary](#) [HLA-Related Links](#)

[Infection & Immunity Reviews](#)

M.Tevfik Dorak

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

Compatibility Systems in Nature

M.Tevfik Dorak, MD, PhD

Self and non-self recognition is a common requirement for all living organisms. Most taxa have their own systems to identify self and non-self. Nature makes use of these

systems for several purposes. One common use is to avoid inbreeding by identifying individuals, cells or gametes as different from self (self-incompatibility as in mate choice, selective fertilization). Also in adults, the same discrimination finds its use in co-operation between individuals or cells (self-compatibility as in kin recognition, colony formation, nuclear fusion, dual recognition, transplantation). It is not infrequent that the same system is involved in both levels of recognition in the same species.

Bacteria, protozoa, fungi and plants have well studied compatibility systems. In the animal kingdom, invertebrates have an allorecognition system which is primarily involved in immunological recognition but also in fertilization. The other major division in the animal kingdom, the vertebrates, invariably has the MHC. Recent evidence suggests that inbreeding avoidance is one of the less appreciated functions of the MHC. Here, a review of the compatibility systems in nature will be followed by the involvement of the MHC in reproductive phenomena. The similarities between the MHC and other compatibility systems will be emphasized with the view that the MHC is primarily an inbreeding avoidance system and its role in histocompatibility is a secondary one¹. The best known compatibility systems are the protozoan pheromone system^{2,3}, the fungal (in)compatibility systems⁴⁻⁸, angiosperm (flowering plants) self-incompatibility system^{6,9,10}, and the invertebrate allorecognition systems¹¹⁻¹⁴. All of these systems are primarily involved in prevention of matings between genetically similar individuals to avoid the harmful effects of inbreeding. As a result of the most polymorphic compatibility system, the fungus *Schizophyllum commune* (the edible mushroom) has the maximal (98.8%) rate of outbreeding in nature. Even in bacteria in which reproduction is asexual (by binary fission) as a rule, there are mechanisms to increase genetic diversity via horizontal gene transfer (conjugation) (see [Microbial Genetics](#)). The green alga *Spirogyra* and the Zygomycetes phylum of the fungal kingdom use the mechanism for the same purpose.

Fungal compatibility systems

The fungal incompatibility system regulates both sexual reproduction and somatic compatibility. The first major review on this subject was presented by Raper in 1966¹⁵. More recent reviews cover the mating types and pheromone systems in Basidiomycetes and Ascomycetes^{5,8,16-20}. Between the two phyla, Basidiomycetes species may have thousands of (tetrapolar) mating types as well as a pheromone system. The somatic compatibility system which regulates self/nonself recognition during vegetative growth in filamentous fungi has also been extensively reviewed^{5,21,22}.

In fungi, both asexual and sexual reproductions are observed. In sexually reproducing fungi, there is no distinction between male and female structures but there is a genetically determined difference among individual fungi. This is due to the mating types. Individuals of the same mating type cannot mate with one another. The nuclei of most fungi are haploid except when a zygote is formed in sexual reproduction. The diploid zygotes undergo meiosis, producing haploid nuclei that will be integrated into the spores. When haploid fungal spores germinate, their nuclei divide mitotically to produce hyphae (the structural unit of a fungus in its vegetative phase or mycelium). These haploid hyphae in filamentous fungi may be in a dikaryotic stage ($n+n$) which is different from haploid (n) or diploid ($2n$) state. The co-existence of two different nuclei (heterokaryon or dikaryon) in the same cell is regulated by the somatic/vegetative/heterokaryon compatibility system.

The studies on the compatibility systems in Basidiomycetes¹⁹ include those on the smut fungus *Ustilago maydis*²³⁻²⁶, *U. hordei*⁸ and the edible members (mushrooms) *Schizophyllum commune*⁷ and *Coprinus cinereus*²⁷. The members of the other phylum Ascomycetes^{17,18} studied are the unicellular yeasts *Saccharomyces cerevisiae*²⁸⁻³⁰ and

Schizosaccharomyces pombe^{17; 19}; the filamentous ascomycetes Neurospora crassa³¹⁻³⁴, Podospora anserina^{35; 36} and Cochliobolus heterostrophus; and Aspergillus nidulans³⁷⁻³⁹. Unique features of these groups are that Basidiomycetes have more complex mating type systems consisting of much more polymorphic mating types and a pheromone/pheromone receptor system, and (filamentous) Ascomycetes have a somatic (heterokaryon) incompatibility system in addition to the sexual mating types. The heterokaryon compatibility system (het loci) regulates the heterokaryon formation in filamentous fungi. Filamentous fungi are capable of hyphal fusion to form dikaryotic heterokaryons during their vegetative growth but the formation of this new composite mycelium is under the control of het loci as well as one of the mating type loci^{5; 21; 40-42}. The coexpression of the antagonistic het alleles triggers a lethal reaction and prevents the formation of viable heterokaryons. In Neurospora crassa, allelic differences at any one of at least 11 het loci trigger an incompatibility response (thus, the heterokaryon is homozygous at all het loci). The number of the het loci in other Ascomycetes is 17 in P. anserina, eight in A. nidulans and at least five in Cryphonectria parasitica (the chestnut blight ascomycete)^{5; 21}. The heterokaryon formation and the role of similarity in this process is very similar to the control of colony formation and fusion in invertebrates¹³, kin recognition in mice^{43; 44}, and transplant acceptance in the animal kingdom⁴⁵⁻⁴⁹. Heterokaryosis is the first step in sexual reproduction of non-filamentous fungi such as Basidiomycetes, but here diversity is favored and this is regulated by the mating types¹⁹. Allelic incompatibility in the het loci does not generally affect sexual function; strains with numerous het differences can mate.

In N. crassa, which is heterothallic (self-incompatible), strains of opposite mating type, A and a, must interact to give the series of events resulting in sexual reproduction: fruiting body formation, meiosis, and the generation of dormant ascospores. While the mating type sequences must be of the opposite kind for mating to occur in the sexual cycle, two strains of opposite mating type cannot form a stable heterokaryon during vegetative growth. In haploid heterothallic species, the genome only contains one of the A or a mating type loci. The genus Neurospora also includes homothallic (self-compatible) species. Those carry a single haploid nucleus and are able to form fruiting bodies, undergo meiosis, and produce new haploid spores. One such species, N. terricola, contains one copy each of the A and the a sequences within each haploid genome³⁴. Homothallism in these species is not due to mating-type switching, as it is in Saccharomyces cerevisiae³². In S. cerevisiae, the genome contains both mating type loci and switching between them is possible⁸. Each haploid genome contains both the a and α genes and normally one of them is transcribed with its interaction with the MAT locus while the other one is silent. Sometimes, a new copy of the silent mating type gene is made, the other one is removed from the MAT locus and the new type is transcribed.

During (sexual) conjugation in S. cerevisiae, two cells of opposite mating type (MAT α and MAT a) fuse to form a diploid zygote. Conjugation requires that each cell locates an appropriate mating partner. This is achieved by pheromones and pheromone receptors. In MAT-a cells, both production of a-pheromone and response to α -pheromone are necessary for successful 'courtship'⁵⁰. Unlike the pheromone system of Basidiomycetes, in the yeast, the pheromone system is under the control of the mating types (not independent).

The function of the mating types is obviously avoidance of inbreeding with no homozygosity allowed and consequently, absolute heterozygote advantage. This helps to increase genetic diversity of survival chances of the species. On the other hand, the function of the heterokaryon compatibility system is not clear. While having two nuclei offers the advantages of diploidy (for example, masking recessive deleterious

genes), why they have to be the same genetic type is not known. One of the hypotheses is that the horizontal transfer of cytoplasmic genetic elements is reduced between incompatible strains and this protects strains of natural populations against invasion by harmful cytoplasmic genetic elements (stable RNA, mitochondria and plasmids). The prevention of horizontal gene transfer, however, is not absolute and the het loci differ in their efficiency in this process^{38; 40; 51}.

Mating types in fungi:

BIPOLAR MATING TYPES:

Zygomycetes: (+) and (-)

All heterothallic ascomycetes have single-locus, two-allele mating systems:

S.cerevisiae: a (a1, a2) and α (α 1, α 2) [type switching is possible]

N.crassa: A (mtA-1) and a (mta-1) [idiomorphic]

U.hordei: a and b (linked)

TETRAPOLAR MATING TYPES:

U.maydis: diallelic pheromone system a (a1, a2) and multiallelic b locus (homeodomain transcription factor)

S.commune and *C.cinereus*: multiallelic A (transcription factor) and multiallelic B (pheromone and pheromone receptor)

Plant self-incompatibility system

Mechanisms that prevent self-pollination are of crucial importance for maintaining genetic diversity within flowering plant (angiosperm) populations. This is because the flowers often have male and female organs within close proximity on the same plant and not infrequently on the same flower. Self-incompatibility, is a genetically controlled mechanism to reject its own pollen. For a classical treatment of this subject, the reader is referred to the monograph by de Nettancourt⁵². More recent reviews have dealt with the nature, molecular and population genetics of this system^{9; 10; 20; 53-57}. See also [Plant Genetics](#).

Some flowers have developed mechanical barriers for their own pollen to prevent them from reaching the female organ (pistil) in the same flower or plant. Some plants have timing differences between their male and female flowerings. The self-incompatibility systems creating a topological barrier (due to different morphologies of their flowers) are called heteromorphic self-incompatibility systems^{10; 52; 53}. The homomorphic self-incompatibility (SI) involves the rejection of self-pollen and was first recognized by Darwin. Over half of the flowering plants have flowers with similar shape and this type of self-incompatibility^{52; 58}. The homomorphic type is further classified into gametophytic and sporophytic types. In the former pollen's own SI type is perceived by the stigma and should not match either of the plant's SI alleles for successful fertilization. In the more interesting sporophytic type, the two alleles of pollen's parent are recognized by the stigma and there should be no matching combination between the two alleles of the stigma and two alleles of the plant from which the pollen has derived to avoid self-rejection. The gametophytic type is more common (found in 60 families of angiosperms) than the sporophytic type (found in six families)¹⁰. The two types are not related and evolved independently. The gametophytic type has been studied in Papaveraceae (poppies), Poaceae, Rosaceae, Scrophulariaceae, and Solanaceae (including tobacco, potato and tomato). The sporophytic type has only been studied in Brassicaceae (including cabbage and mustard, for example *Arabidopsis thaliana*). Despite being very

common among angiosperms, the SI system in different families have different origins, in other words, they evolved independently several times^{56; 59}. Self-incompatible (heterothallic) plants necessarily produce offspring that are heterozygous at the S locus which in general, contains 30-50 alleles^{53; 60}. The alleles of the S locus confer genetic identity (S haplotype specificity) on the pollen and stigma of self-incompatible plants. The S locus of the sporophytic type has two genes encoding two proteins expressed on the stigma surface. These are a transmembrane S receptor protein kinase (SRK) and S locus glycoprotein (SLG) which has RNase activity⁶¹. It is the SRK gene product which determines the S haplotype specificity of the stigma but the SI response is stronger if SLG of the same haplotype is also expressed.⁵⁷ The corresponding protein on the pollen surface has recently been identified as a member of the pollen coat protein family (SCR)^{62; 63}. When a self pollen reaches the stigma on the same flower or plant, a self-rejection reaction takes place. The biochemical mechanism of self-rejection involves the cytotoxic action of the RNase activity^{64; 65}. The end result is the prevention of pollen tube growth. In the gametophytic type, the same is achieved by a single glycoprotein with an RNase activity⁶⁶. Just like the fungal mating types, the plant self-incompatibility system provides an example of balancing selection in the maintenance of their alleles^{22; 60; 67}. It is easy to imagine how this works. Any new allele would have selective advantage since a pollen with this allele will always be accepted by the stigma until this allele reaches a remarkable frequency in the population. Once it has been established, the frequency will still be maintained through heterozygote advantage. Since this occurs for any new allele created as a result of mutations, balancing selection results in extreme polymorphism detected in these compatibility systems including the vertebrate MHC^{8; 60; 68; 69}. The resulting highly diverged alleles will also have very long evolutionary life times and their existence will cover the life times of several successive species (transspecies polymorphism). The evidence suggesting this pattern of polymorphism is greater sequence similarity of alleles between species than similarities within species. This is again typical of all these compatibility systems^{22; 67; 70-76}. Another piece of evidence for the action of balancing selection on the self-incompatibility alleles is the clustering of non-synonymous mutations in hypervariable regions (HVRs) rather than a homogeneous distribution⁶⁷. Under a strictly neutral model, there would be no such heterogeneity in the distribution of substitution rates. The continuous stretches of non-polymorphic sequences in different alleles suggest that these segments are functionally constrained and any non-synonymous substitution in these parts would be deleterious and subject to purifying selection. These segments may still have a rate of high synonymous base substitutions. On the other hand, if new alleles are favored as in heterozygote advantage, non-synonymous substitutions will be concentrated on the segments encoding the allelic specificity (HVRs). This is exactly what happens in the fungal het loci²², plant SI^{67; 72} and vertebrate MHC alleles⁷⁷⁻⁸¹.

Invertebrate compatibility systems

Although the MHC multigene family is restricted to vertebrates, histocompatibility loci are also found in invertebrates where they appear to have an analogous role in the regulation of mating systems^{11; 82}. A histocompatibility system of immunorecognition is postulated to have originated in multicellular invertebrates probably beginning with coelenterates (corals)^{45; 46; 83}. The best studied invertebrate compatibility system is that of the colonial tunicates^{11; 13; 14; 84; 85}. The best known species in this group is *Botryllus schlosseri* and its compatibility system is called fusion/histocompatibility (Fu/HC). Allorecognition in *Botryllus* is principally controlled by this single Mendelian locus, with a

large number of codominantly expressed alleles⁸⁵. The number of alleles is estimated to be 30-200¹³.

Colonial tunicates are complex marine invertebrates (in fact protochordates) that undergo a variety of histocompatibility reactions in their intraspecific competition for feeding surfaces. By means of these reactions colonies fuse with kin, extend domination over a feeding surface, while isolating unrelated conspecifics. A Botryllus colony is composed of numerous units which are embedded within the translucent-gelatinous matrix, the tunic. Each hermaphroditic member possesses male and female gonads. Following fusion with nonidentical kin sharing 1 or more Fu/HC allele(s), the fused pair expands both chimeric partners via an asexual budding process, further extending domination over a feeding surface. However, at some later time point an intense set of histoincompatibility reactions occurs between fused kin, resulting in the destruction of all individuals of one of the genotypes, ending the chimeric state¹³.

Apart from prevention of fusion with non-kin^{11;84}, the Fu/HC also affects self-fertilization by sperm-egg incompatibility. Eggs resist fertilization by sperm from the same colony represented by its Fu/HC allele. This interaction results in selective fertilization by sperm bearing a different Fu/HC allele^{11;82}. This situation in hermaphroditic invertebrates is very similar to what happens in fungi and plants.

Similar to the situation with the other compatibility systems and the MHC, there is yet no evidence for a common ancestor for the invertebrate compatibility systems and the vertebrate MHC^{13;14}. This shows that the widespread existence of these systems is not a co-incidence due to a common ancestor but suggests a biological requirement to have a system to promote outbreeding. The best evidence for that is that in plants, the self-incompatibility system arose independently more than one times. It appears, however, that the main function of all these systems is to enforce heterozygosity by acting at the earliest phase of sexual reproduction. There is still a possibility that the vertebrate histocompatibility genes evolved from gametic self-nonsel recognition systems which prevent self-fertilization in hermaphroditic organisms¹¹. This idea was first put forward by the Nobel Laureate immunologists FM Burnet⁸⁶.

Vertebrate MHC and compatibility

The MHC also prevents inbreeding through its influence on mate choice in mice^{87;88} and humans^{89;90}; and on reproductive processes in rats⁹¹, mice^{92;93} and humans^{94;95}. The reproductive mechanisms are varied and range from selective fertilization to selective abortion. A major common feature of the compatibility systems is that they favor genetic dissimilarity between mates and the gametes (mate choice, selective fertilization); but similarity in co-operation (kin recognition, dual recognition, transplant matching)^{1;96}. All these functions are based on the provision of a phenotype for the genetic identity of the individual by the MHC: either cell surface molecules or chemosensory signals. See also Non-Pathogen-Based Selection in [Origin of the MHC and Its Polymorphism](#).

References

1. Jones JS, Partridge L: Tissue rejection: the price for sexual acceptance. *Nature* 304:484, 1983
2. Weiss MS, Anderson DH, Raffioni S, Bradshaw RA, Orteni C, Lupoini P, Eisenberg D: A cooperative model for receptor recognition and cell adhesion: evidence from the molecular packing in the 1.6-A crystal structure of the pheromone Er-1 from the ciliated protozoan *Euplotes raikovi*. *Proceedings of the National Academy of Sciences USA* 92:10172, 1995
3. Vallesi A, Giuli G, Bradshaw RA, Lupoini P: Autocrine mitogenic activity of pheromones produced by the protozoan ciliate *Euplotes raikovi*. *Nature* 376:522, 1995
4. Metzberg RL: The role of similarity and difference in fungal mating. *Genetics* 125:457, 1990

5. Begueret J, Turcq B, Clave C: Vegetative incompatibility in filamentous fungi: het genes begin to talk. *Trends in Genetics* 10:441, 1994
6. Hiscock SJ, Kues U, Dickinson HG: Molecular mechanisms of self-incompatibility in flowering plants and fungi - different means to the same end. *Trends in Cell Biology* 6:421, 1996
7. Wendland J, Vaillancourt LJ, Hegner J, Lengeler KB, Laddison KJ, Specht CA, Raper CA, Kothe E: The mating-type locus B alpha 1 of *Schizophyllum commune* contains a pheromone receptor gene and putative pheromone genes. *EMBO Journal* 14:5271, 1995
8. Kothe E: Tetrapolar fungal mating types: sexes by the thousands. *FEMS Microbiological Reviews* 18:65, 1996
9. Haring V, Gray JE, McClure BA, Anderson MA, Clarke AE: Self-incompatibility: a self-recognition system in plants [Review]. *Science* 250:937, 1990
10. Kao TH, McCubbin AG: How flowering plants discriminate between self and non-self pollen to prevent inbreeding. *Proceedings of the National Academy of Sciences USA* 93:12059, 1996
11. Scofield VL, Schlumpberger JM, West LA, Weissman IL: Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* 295:499, 1982
12. Grosberg RK: The evolution of allorecognition specificity in clonal invertebrates. *Quarterly Review of Biology* 63:377, 1988
13. Weissman IL, Saito Y, Rinkevich B: Allorecognition histocompatibility in a protochordate species: is the relationship to MHC somatic or structural? *Immunological Reviews* 113:227, 1990
14. Magor BG, De Tomaso A, Rinkevich B, Weissman IL: Allorecognition in colonial tunicates: protection against predatory cell lineages? *Immunological Reviews* 167:69, 1992
15. Raper JR: *Genetics of Sexuality in Higher Fungi*. New York, Ronald Press, 1966
16. Bolker M, Kahmann R: Sexual pheromones and mating responses in fungi. *Plant Cell* 5:1461, 1993
17. Kronstad JW, Staben C: Mating type in filamentous fungi. *Annual Review of Genetics* 31:245, 1997
18. Coppin E, Debuchy R, Arnaise S, Picard M: Mating types and sexual development in filamentous ascomycetes. *Microbiology and Molecular Biology Reviews* 61:411, 1997
19. Casselton LA, Olesnick NS: Molecular genetics of mating recognition in basidiomycete fungi. *Microbiology and Molecular Biology Reviews* 62:55, 1998
20. Hiscock SJ, Kues U: Cellular and molecular mechanisms of sexual incompatibility in plants and fungi. *International Review of Cytology* 193:165, 1999
21. Glass NL, Kuldau GA: Mating type and vegetative incompatibility in filamentous ascomycetes. *Annual Review of Phytopathology* 30:201, 1992
22. Wu J, Saupé SJ, Glass NL: Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences USA* 95:12398, 1998
23. Urban M, Kahmann R, Bolker M: Identification of the pheromone response element in *Ustilago maydis*. *Molecular & General Genetics* 251:31, 1996
24. Urban M, Kahmann R, Bolker M: The biallelic a mating type locus of *Ustilago maydis*: remnants of an additional pheromone gene indicate evolution from a multiallelic ancestor. *Molecular & General Genetics* 250:414, 1996
25. Kahmann R, Romeis T, Bolker M, Kumper J: Control of mating and development in *Ustilago maydis*. *Current Opinion in Genetics & Development* 5:559, 1995
26. Gillissen B, Bergemann J, Sandmann C, Schroeder B, Bolker M, Kahmann R: A two-component regulatory system for self/non-self recognition in *Ustilago maydis*. *Cell* 68:647, 1992
27. O'Shea SF, Chaure PT, Halsall JR, Olesnick NS, Leibbrandt A, Connerton IF, Casselton LA: A large pheromone and receptor gene complex determines multiple B mating type specificities in *Coprinus cinereus*. *Genetics* 148:1081, 1998
28. Weissman JD, Singer DS: Striking similarities between the regulatory mechanisms governing yeast mating-type genes and mammalian major histocompatibility complex genes. *Molecular & Cellular Biology* 11:4228, 1991
29. Naider F, Gounarides J, Xue CB, Bargiota E, Becker JM: Studies on the yeast alpha-mating factor: a model for mammalian peptide hormones. *Biopolymers* 32:335, 1992
30. Brizzio V, Gammie AE, Nijbroek G, Michaelis S, Rose MD: Cell fusion during yeast mating requires high levels of a-factor mating pheromone. *Journal of Cell Biology* 135:1727, 1996

31. Glass NL, Grotelueschen J, Metzberg RL: *Neurospora crassa* A mating-type region. *Proceedings of the National Academy of Sciences USA* 87:4912, 1990
32. Glass NL, Vollmer SJ, Staben C, Grotelueschen J, Metzberg RL, Yanofsky C: DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar. *Science* 241:570, 1988
33. Arganoza MT, Ohmberger J, Min J, Akins RA: Suppressor mutants of *Neurospora crassa* that tolerate allelic differences at single or at multiple heterokaryon incompatibility loci. *Genetics* 137:731, 1994
34. Metzberg RL, Glass NL: Mating type and mating strategies in *Neurospora*. *Bioessays* 12:53, 1990
35. Coustou V, Deleu C, Saupe S, Begueret J: The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proceedings of the National Academy of Sciences USA* 94:9773, 1997
36. Barreau C, Iskandar M, Loubradou G, Levallois V, Begueret J: The mod-A suppressor of nonallelic heterokaryon incompatibility in *Podospora anserina* encodes a proline-rich polypeptide involved in female organ formation. *Genetics* 149:915, 1998
37. Butcher AC: The relationship between sexual outcrossing and heterokaryon incompatibility in *Aspergillus nidulans*. *Heredity (Edinburgh)* 23:443, 1968
38. Coenen A, Debets F, Hoekstra R: Additive action of partial heterokaryon incompatibility (partial-het) genes in *Aspergillus nidulans*. *Current Genetics* 26:233, 1994
39. van Diepeningen AD, Debets AJ, Hoekstra RF: Heterokaryon incompatibility blocks virus transfer among natural isolates of black *Aspergilli*. *Current Genetics* 32:209, 1997
40. Debets F, Yang X, Griffiths AJ: Vegetative incompatibility in *Neurospora*: its effect on horizontal transfer of mitochondrial plasmids and senescence in natural populations. *Current Genetics* 26:113, 1994
41. Hartl DL, Dempster ER, Brown SW: Adaptive significance of vegetative incompatibility in *Neurospora crassa*. *Genetics* 81:553, 1975
42. Saupe SJ, Glass NL: Allelic specificity at the het-c heterokaryon incompatibility locus of *Neurospora crassa* is determined by a highly variable domain. *Genetics* 146:1299, 1997
43. Manning CJ, Wakeland EK, Potts WK: Communal nesting patterns in mice implicate MHC genes in kin recognition. *Nature* 360:581, 1992
44. Brown JL, Eklund A: Kin recognition and the major histocompatibility complex: an integrative review. *American Naturalist* 143:435, 1994
45. Hildemann WH, Johnson IS, Jokiel PL: Immunocompetence in the lowest metazoan phylum: transplantation immunity in sponges. *Science* 204:420, 1979
46. Hildemann WH, Linthicum DS, Vann DC: Immunoincompatibility reactions in corals (Coelenterata). *Advances in Experimental Medicine & Biology* 64:105, 1975
47. Hildemann WH: Transplantation immunity in fishes: Agnatha, Chondrichthyes and Osteichthyes. *Transplantation Proceedings* 2:253, 1970
48. Gorer PA: The genetic and antigenic basis of tumour transplantation. *Journal of Pathology & Bacteriology* 44:691, 1937
49. Snell GD: Studies in histocompatibility. *Science* 213:172, 1981
50. Jackson CL, Hartwell LH: Courtship in *Saccharomyces cerevisiae*: an early cell-cell interaction during mating. *Molecular & Cellular Biology* 10:2202, 1990
51. Caten CE: Vegetative incompatibility and cytoplasmic infection in fungi. *Journal of General Microbiology* 72:221, 1972
52. de Nettancourt D: *Incompatibility in Angiosperms*. Berlin, Springer-Verlag, 1977
53. Ebert PR, Anderson MA, Bernatzky R, Altschuler M, Clarke AE: Genetic polymorphism of self-incompatibility in flowering plants. *Cell* 56:255, 1989
54. Thompson RD, Kirch HH: The S locus of flowering plants: when self-rejection is self-interest. *Trends in Genetics* 8:381, 1992
55. Sims TL: Genetic regulation of self-incompatibility. *Critical Reviews in Plant Sciences* 12:129, 1993
56. Matton DP, Nass N, Clarke AE, Newbigin E: Self-incompatibility: How plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences USA* 91:1992, 1994
57. Kao TH, McCubbin AG: A social stigma. *Nature* 403:840, 2000

58. East EM: The distribution of self-sterility in flowering plants. *Proceedings of the American Philosophical Society* 82:449, 1940
59. Mau SL, Anderson MA, Heisler M, Haring V, McClure BA, Clarke AE: Molecular and evolutionary aspects of self-incompatibility in flowering plants. *Symposia of the Society for Experimental Biology* 45:245, 1991
60. Richman AD, Uyenoyama MK, Kohn JR: Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* 273:1212, 1996
61. Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K: The S receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature* 403:913, 2000
62. Schopfer CR, Nasrallah ME, Nasrallah JB: The male determinant of self-incompatibility in *Brassica*. *Science* 286:1697, 1999
63. Takayama S, Shiba H, Iwano M, et al.: The pollen determinant of self-incompatibility in *brassica campestris*. *Proceedings of the National Academy of Sciences USA* 97:1920, 2000
64. Clarke AE, Newbigin E: Molecular aspects of self-incompatibility in flowering plants. *Annual Review of Genetics* 27:257, 1993
65. Charlesworth D: Plant self-incompatibility. The key to specificity. *Current Biology* 4:545, 1994
66. McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE: Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature* 342:955, 1989
67. Clark AG, Kao TH: Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Sciences USA* 88:9823, 1991
68. Alberts SC, Ober C: Genetic variability of the MHC: a review of non-pathogen-mediated selective mechanisms. *Yearbook of Physical Anthropology* 36:71, 1993
69. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK: The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology* 17:179, 1997
70. Dwyer KG, Balent MA, Nasrallah JB, Nasrallah ME: DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. *Plant Molecular Biology* 16:481, 1991
71. Vekemans X, Slatkin M: Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* 137:1157, 1994
72. Ioerger TR, Clark AG, Kao TH: Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences USA* 87:9732, 1990
73. Uyenoyama MK: A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics* 139:975, 1995
74. Rivers BA, Bernatzky R, Robinson SJ, Jahnen-Dechent W: Molecular diversity at the self-incompatibility locus is a salient feature in natural populations of wild tomato (*Lycopersicon peruvianum*). *Molecular & General Genetics* 238:419, 1993
75. Boyes DC, Nasrallah ME, Vrebalov J, Nasrallah JB: The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. *Plant Cell* 9:237, 1997
76. Klein J: Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* 19:155, 1987
77. Hughes AL, Nei M: Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167, 1988
78. Hedrick PW, Whittam TS, Parham P: Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. *Proceedings of the National Academy of Sciences USA* 88:5897, 1991
79. Hughes AL, Hughes MK, Watkins DI: Contrasting roles of interallelic recombination at the HLA-A and HLA-B loci. *Genetics* 133:669, 1993
80. Hughes AL, Hughes MK: Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics* 42:233, 1995
81. Hughes AL, Yeager M: Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics* 32:415, 1998
82. Oka H: Colony specificity in compound Ascidians, in Yukova M (ed): *Profiles of Japanese Science & Scientists*, Tokyo, Kodansha, 1970, p 195

83. Hildemann WH, Raison RL, Cheung G, Hull CJ, Akaka L, Okamoto J: Immunological specificity and memory in a scleractinian coral. *Nature* 270:219, 1977
84. Grosberg RK, Quinn JF: The genetic control and consequences of kin recognition by the larvae of a colonial marine invertebrate. *Nature* 322:456, 1976
85. De Tomaso A, Saito Y, Ishizuka KJ, Palmeri KJ, Weissman IL: Mapping the genome of a model protochordate. I. A low resolution genetic map encompassing the fusion/histocompatibility (Fu/HC) locus of *Botryllus schlosseri*. *Genetics* 149:277, 1998
86. Burnet FM: "Self-recognition" in colonial marine forms and flowering plants in relation to the evolution of immunity. *Nature* 232:230, 1971
87. Williams JR, Lenington S: Factors modulating preferences of female house mice for males differing in t-complex genotype: role of t-complex genotype, genetic background, and estrous condition of females. *Behavior Genetics* 23:51, 1993
88. Potts WK, Manning CJ, Wakeland EK: Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* 352:619, 1991
89. Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu DD, Elias S: HLA and mate choice in humans. *American Journal of Human Genetics* 61:497, 1997
90. Wedekind C, Seebeck T, Bettens F, Paepke AJ: MHC-dependent mate preferences in humans. *Proceedings of the Royal Society of London - Series B: Biological Sciences* 260:245, 1995
91. Palm J: Maternal-fetal histoincompatibility in rats: an escape from adversity. *Cancer Research* 34:2061, 1974
92. Hamilton MS, Hellstrom I: Selection for histoincompatible progeny in mice. *Biology of Reproduction* 19:267, 1978
93. Wedekind C, Chapuisat M, Macas E, Rulicke T: Non-random fertilization in mice correlates with the MHC and something else. *Heredity* 77:400, 1996
94. Ober C, Elias S, Kostyu DD, Hauck WW: Decreased fecundability in Hutterite couples sharing HLA-DR. *American Journal of Human Genetics* 50:6, 1992
95. Jin K, Ho HN, Speed TP, Gill TJ: Reproductive failure and the major histocompatibility complex. *American Journal of Human Genetics* 56:1456, 1995
96. Brown JL: Some paradoxical goals of cells and organisms: the role of the MHC, in Pfaff DW (ed): *Ethical Questions in Brain and Behavior: Problems and Opportunities*, New York, Springer-Verlag, 1983, p 111

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Nov 29, 2003

[Back to HLA](#)

[Back to MHC](#)

[Back to Genetics](#)

[Back to Evolution](#)

[Back to Biostatistics](#)

[Glossary](#) [Homepage](#)

Origin of the MHC

M. Tevfik DORAK, MD, PhD

There is not a definite candidate for the primordial MHC gene. According to one hypothesis the class II MHC evolved first¹, whereas another hypothesis holds that the class I MHC originated first as a result of a recombination between an immunoglobulin-like C-domain and the peptide-binding domain of an HSP70 heat-shock protein². A phylogenetic analysis supports a relationship between the class II MHC alpha chain and beta 2-microglobulin and between the class II MHC beta-chain and the class I alpha chain¹. Most evidence supports the hypothesis that the ancestral MHC molecule had a class II-like structure and it gave rise to the class I molecule^{1; 3; 4}. This still does not explain the nature of the very first MHC gene. Ohno suggested that plasma membrane cell adhesion proteins (N-CAM) that were involved in ontogenic organogenesis since time immemorial were the ultimate ancestor of the adaptive immune system. N-CAM of the chicken for neuronal organogenesis possesses four beta 2-microglobulin-like domains and it is this domain from which the adaptive immune system originated⁵. Hughes and Nei calculated the divergence time for class II A and B genes and found 446 to 521 million years depending on the method⁶. This fits with the observation that amphibians, which diverged 370 million years ago, have both class II A and B genes⁷. The presence of all class I, III, and II genes in the amphibian *Xenopus* suggests that the physical linkage of these MHC region may be as old as 370 million years or more⁸. MHC loci do not always exist in a single tightly linked cluster as they do in mammals, but can be found in two (e.g., chicken⁹) or multiple (e.g., zebrafish¹⁰) clusters.

The innate immune system is the only defensive system in invertebrates. It also exists in vertebrates but the main immune defense in vertebrates is the adaptive immune system with its components MHC, TCR and Ig genes (and enzymes with recombinase activity such as RAG1). These main components of the adaptive immune system are missing not only in invertebrates but also in primitive 'jaw less' vertebrates^{11; 12}. The adaptive immune system seems to have evolved from the most primitive 'jaw ed' vertebrates, i.e., cartilaginous fish (sharks, rays) upwards. There is no molecular evidence to suggest whether vertebrate immune systems (and particularly the MHC molecules) are evolutionarily related to invertebrate allorecognition systems, and the functional evidence can be interpreted either way. The MHC itself does not exist in the jaw less fish^{4; 12}. MHC class I and class II molecules do exist in the cartilaginous fish (sharks)^{13; 14}. The overall data suggest that the MHC evolved following duplication of housekeeping genes sometime soon after the evolutionary divergence of the agnathans (jaw less vertebrates)^{15; 16}. It is thought that duplication allowed one copy of the genes to preserve their housekeeping functions and the other copy to diversify.

There is clearly an MHC in amphibians and birds with many characteristics like the MHC of mammals (a single genetic region encoding polymorphic class I and class II molecules) and evidence for polymorphic class I and class II molecules in reptiles. However, many details differ from the mammals, and it is not clear whether these reflect

historical accident or selection for different lifestyles or environment. For example, the adult frog *Xenopus* has a vigorous immune system with many similarities to mammals, a ubiquitous class I molecule, but a much wider class II tissue distribution than human, mouse and chicken. The *Xenopus* tadpole has a much more restricted immune response, no cell surface class I molecules and a mammalian class II distribution. The axolotl (an amphibian - salamander) has a very poor immune response (as though there are no helper T cells), a wide class II distribution and, for most animals, no cell surface class I molecule. It would be enlightening to understand both the mechanisms for the regulation of the MHC molecules during ontogeny and the consequences for the immune system and survival of the animals. These animals also differ markedly in the level of MHC polymorphism. Another difference from mammals is the presence of previously uncharacterized molecules. In *Xenopus* and reptiles, there are two populations of class I alpha chain on the surface of erythrocytes, those in association with β_2m and those in association with a disulphide-linked homodimer⁷.

MHC class I genes do not show orthologous (i.e., homologous by descent from a common ancestral locus) relationship between mammals of different orders; whereas orthologous relationships have been found among mammalian class II loci⁶. The HLA-C locus has been found only in humans, gorilla and chimpanzees but not in monkeys¹⁷. The class II gene regions seem to have arisen prior to the divergence of the orders of placental mammals. The most ancient polymorphic class II locus appears to be HLA-DQA1¹⁸. The polymorphism of this locus also correlates to the MHC class II supertypal groupings^{19;20}. There is evidence that MHC genes are subject to a birth-and-death process²¹. New genes are created by repeated gene duplication and some duplicate genes are maintained in the genome for a long time but others are deleted or become non-functional by deleterious mutations. This concept disagrees with the earlier idea that the MHC diversity and evolution are governed by concerted evolution of the multigene families of major histocompatibility complex (MHC) genes and immunoglobulin (Ig) genes. The alleles seem to have a fast turnover rate. The lack of correspondence between the human and chimpanzee alleles suggests that 5 million years of separation have been sufficient to reconfigure MHC alleles. This means that the alleles are constantly undergoing modifications during their transspecies evolution²². In Amerindian population studies, the emergence of new recombinant HLA-B alleles is accompanied by the loss of founding alleles²³.

The origin of diversity of MHC alleles

The major histocompatibility complex (MHC) loci are known to be highly polymorphic in humans, mice and certain other mammals, with heterozygosity as high as 80-90%. Four different hypotheses have been considered to explain this high degree of polymorphism: (1) a high mutation rate, (2) gene conversion or interlocus genetic exchange, (3) overdominant (balancing) selection and (4) frequency-dependent selection.

The distribution of the pattern of sequence polymorphism in human and mouse class I genes provides evidence for four co-ordinate factors that contribute to the origin and sustenance of abundant allele diversity that characterizes the MHC in the species. These include: (a) a gradual accumulation of spontaneous mutational substitution over evolutionary time but not an unusually high mutation rate; (b) selection against mutational divergence in regions of the class I molecule involved in T cell receptor interaction and also in certain regions that interact with common features of antigens; (c)

positive selection pressure in favor of persistence of polymorphism and heterozygosity at the antigen recognition site; and (d) periodic intragenic (interallelic) and more rarely, intergenic, recombination within the class I genes.

It has to be emphasized that the evolutionary interplay between mutation and recombination varies with MHC locus, and even for subregions of the same gene^{22; 24; 25}. For example, phylogenetic inferences based on the exon 2 region of HLA-DRB loci are complicated by selection and recombination (gene conversion). Noncoding region analysis may help clarify patterns of allele evolution usually with contrasting results to those obtained from coding region analyzes²⁶. The main source for the variability in the HLA gene sequences is point mutation but the mutation rate is by no means higher in the MHC than elsewhere in the genome^{27; 28}. Because of transspecies polymorphism, accumulation of point mutations over millions of years results in extensive polymorphism. In contrast, gene conversions have produced at least 80 new class I alleles since the separation of the Homo lineage and the rate of conversion is much higher than that of point mutation²⁹.

Mechanisms maintaining the extreme polymorphism of the MHC

1. *Pathogen-driven selection* favors genetic diversity of the MHC through both heterozygote advantage (overdominance) and frequency-dependent selection³⁰. Selection is thought to favor rare MHC genotypes, since pathogens are more likely to have developed mechanisms to evade the MHC-dependent immunity encoded by common MHC genotypes. Six molecular models of pathogen-driven selection have been presented³¹:

A. Pathogen Evasion Models

- Escape of a single T-cell clone recognition

- Escape into holes in the T-cell repertoire produced by T-cells anergized by pathogen variants

- Escape into holes in the T-cell repertoire induced by self-tolerance (molecular mimicry)

- Escape of MHC presentation

B. Host-Pathogen Interactions:

- Heterozygote advantage

- Pathogens bearing allo-MHC antigens

MHC associations with specific infectious diseases have been difficult to demonstrate. The best known ones are Marek's disease in chickens³², parasitic infestations in Soay sheep³³, and malaria in humans³⁴. Since most infectious agents have multiple epitopes the MHC has to deal with³⁵, this is not surprising. Rather than resistance of specific heterozygous genotypes to specific agents, it is more likely that a promiscuous heterozygous advantage is operating. This is to say that all heterozygotes are favored over all homozygotes as proposed by Flaherty³⁶. Only two examples of heterozygote advantage in human infectious diseases have been reported to date: one for a specific genotype in HIV infection³⁷ and another promiscuous heterozygote advantage in HBV infection^{38; 39}.

2. Non-pathogen driven mechanisms

a. Selection through inbreeding depression acts indirectly by favoring MHC-based disassortative mating (mating preferences)^{40;41}. Here, the MHC is exploited to discriminate against genetic similarity at highly polymorphic loci to avoid inbreeding. MHC-based disassortative matings would produce heterozygous progeny at least at the MHC which would result in increased fitness. Progeny derived from MHC-dissimilar parents would enjoy increased fitness because of reduced levels of inbreeding depression and increased resistance to infectious diseases thanks to increased MHC heterozygosity. The basis of this mechanism is that vertebrate species can detect MHC genotype by smell⁴²⁻⁴⁴. Since sharing highly polymorphic genetic markers is predictive of kinship, avoiding mating with animals that have a similar MHC genotype will reduce the likelihood of matings with relatives (inbreeding). Kinship recognition through MHC-linked chemosensory identity has been documented⁴⁵⁻⁴⁷. Thus, as the most polymorphic system, the MHC contributes to the genetic identity of an individual at the highest resolution and this is expressed as chemosensory identity. The self-incompatibility system in plants evolved to assure disassortative matings to avoid inbreeding^{48;49}, and the vertebrate MHC may be doing the same. Behavioral observations and genetic typing indicated that female mice often left their territories and mated with extraterritorial males with MHC-dissimilar haplotypes⁴⁰. A surprising 50% of offspring born in enclosures were from extra-pair copulations. Disassortative mating with respect to self-incompatibility system in plants and the MHC in vertebrates results in the rare-male-effect, and consequently, frequency-dependent selection⁵⁰. This selection contributes to the high levels of genetic polymorphism observed at the MHC loci.

b. Reproductive mechanisms: Fetuses which are unlike their mothers have increased chance to survive⁵¹. This is achieved through mate selection^{40;41;52}, selective fertilization⁵³⁻⁵⁶ and selective abortion^{51;57;58}. The mechanisms involved are unknown but the plant self-incompatibility system^{59;60} and the invertebrate allorecognition system^{61;62} provide good examples of selective fertilization.

There is no human study examined the deficit for MHC homozygosity in new borns, but there are studies in mice⁵⁸ and rats^{54;57;63}. In one of the earliest studies and its continuation, Palm found that depending on the MHC type, new born rats might have deficits for homozygosity which appeared as increased heterozygosity. He repeatedly showed that this only occurs in new born males^{57;63;64}. Also in mice, it has been noted that when deficit for homozygosity for an MHC type occurs, this concerns males⁵⁸. Another mouse study found excess heterozygosity at a different histocompatibility locus, H-3, only in males for certain combinations⁶⁵. The mouse *t*-complex, in which the MHC is embedded, contains recessive embryonic lethal genes⁶⁶. While most fetus homozygous for a particular recessive *t*-lethal die, some who bear two different *t*-lethals may survive till birth. In the group of *t6/tw5* heterozygotes, sex affects lethality and a deficit of males among live births was observed in two independent experiments^{67;68}. It was also shown that this deficit was not due to sex-reversal⁶⁸. These findings also suggest the higher sensitivity of males to prenatal lethality. Thus, maternal-fetal interactions result in heterozygote advantage for MHC haplotypes as a non-pathogen-mediated selection but with a gender effect.

Evidence for Selection on MHC alleles

1. One important feature of the MHC genes is that (as in the plant self-incompatibility and fungal compatibility system alleles⁶⁹⁻⁷¹) the ratio of non-synonymous (replacement) to synonymous (silent) substitutions (dn/ds ratio) is very high in the codons encoding the antigen recognition site of polymorphic class II molecules compared to other codons^{24; 25; 72; 73}. This pattern is evidence that the polymorphism at the antigen recognition sites is maintained by overdominant selection of which the most common form is heterozygote advantage. This kind of selection has been noted for all expressed DRB genes including DRB3 and DRB4^{25; 73; 74}. By contrast, in the case of monomorphic class II loci, no such enhancement of the rate of non-synonymous substitution was observed. This feature and the others such as (1) an extremely large number of alleles; (2) ancient allelic lineages that pre-date contemporary species (trans-species evolution) and; (3) extremely high sequence divergence of alleles make the MHC a unique system in the whole genome. These features are only shared by the self-incompatibility system of the plants^{69; 70; 75-78}, fungal mating types^{71; 79-81} and invertebrate allorecognition systems^{62; 82-84}.

2. The expected number of alleles under neutrality is far lower than the number of MHC alleles observed in natural population indicating some form of balancing (diversifying) selection has been acting on them⁸⁵⁻⁸⁷. In the case of neutral polymorphism, one common allele and a few rare alleles are expected. Only under large effective population sizes (10^5) and high mutation rates (10^{-4}) does the number of selectively neutral alleles maintained approach observed numbers. For a subdivided population over a large range of migration rates, it appears that the number of self-incompatibility alleles (or MHC-alleles) observed can provide a rough estimate of the total number of individuals in the population but it underestimates the neutral effective size of the subdivided population⁷⁷.

3. The large number of alleles showing a relatively even distribution is against neutrality expectations and indicates that diversifying, and not simply directional, selection operates in contemporary populations.

4. The observed deficiency of homozygotes in some human^{88-90; 90-93} and animal populations^{40; 57; 63; 94-96} indicates that selection favors heterozygotes, presumably because of heightened immune responsiveness. When the amino acid heterozygosities per site for HLA-A and -B loci were determined, for the 54 amino acid sites thought to have functional importance, the average heterozygosity per site was 0.301. Sixteen positions have heterozygosities greater than 0.5 at one or both loci and the frequencies of amino acids at a given position are very even, resulting in nearly the maximum heterozygosity possible. Furthermore, the high heterozygosity is concentrated in the peptide-interacting sites, whereas the sites that interact with the T-cell receptor have lower heterozygosity. Overall, these results indicate the importance of some form of balancing selection operating at HLA loci, maybe even at the individual amino acid level⁹⁷. A recent review points out that deficiencies of homozygotes may be overlooked if functional homozygotes are misclassified as heterozygotes⁵². Most human MHC alleles belong to only a few supertypes based on similarities in their peptide-binding properties⁹⁸. Thus, classifying individuals as "heterozygotes" based on high-resolution typing of alleles may fail to detect true (functional) homozygotes.

5. The observed linkage disequilibrium among tightly linked MHC genes suggests that the strength of selection is uneven within the MHC⁹⁹.

6. Studies in West Africa showed that resistance against malaria is HLA-B53 associated and this is the reason for an increased frequency of B53 in that area. The selection differential for HLA-B*5301 is estimated to be 0.028³⁴. In a Soay sheep population, the variation within the MHC is associated with juvenile survival and parasite resistance³³.

Maintenance of deleterious MHC genes has also puzzled researchers. In an attempt to explain why MHC haplotypes that predispose individuals to autoimmune diseases are common in contemporary populations, Apanius et al⁹⁹ suggested that they confer some benefit such as resistance to infectious diseases that outweighs the deleterious effects from autoimmunity. This would be especially pertinent if the benefits were expressed early in life, while the cost due to autoimmunity is paid late in life (i.e., post-reproductive period) when selection against the haplotype would be weaker. This model is based on antagonistic pleiotropy, which is one of the theories proposed to explain the senescence^{100; 101}. A related explanation for the maintenance of autoimmune-predisposing MHC haplotypes is that these alleles protect against initial infection, but the pathogen triggers autoimmunity through molecular mimicry¹⁰² or other factors⁹⁹.

Please note that this essay will be updated in 2004. More recent articles not covered in this review include the following:

1. Kulski JK et al. **Comparative genomic analysis of the MHC.** [Immunol Rev 2002;190:95-122](#)
2. Flajnik MF & Kasahara M. **Comparative genomics of the MHC.** [Immunity 2001;15:351-62](#)
3. Meyer D & Thomson G. **How selection shapes variation of the human MHC.** [Ann Hum Genet. 2001;65:1-26](#)

References

1. Hughes AL, Nei M: Evolutionary relationships of the classes of major histocompatibility complex genes. *Immunogenetics* 37:337, 1993
2. Flajnik MF, Canel C, Kramer J, Kasahara M: Which came first, MHC class I or class II? *Immunogenetics* 33:295, 1991
3. Lawlor DA, Zemmour J, Ennis PD, Parham P: Evolution of class-I MHC genes and proteins: from natural selection to thymic selection. *Annual Review of Immunology* 8:23, 1990
4. Klein J, O'HUigin C: Composite origin of major histocompatibility complex genes [Review]. *Current Opinion in Genetics & Development* 3:923, 1993
5. Ohno S: The ancestor of the adaptive immune system was the CAM system for organogenesis. *Experimental & Clinical Immunogenetics* 4:181, 1987
6. Hughes AL, Nei M: Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Molecular Biology & Evolution* 7:491, 1990
7. Kaufman J, Skjoedt K, Salomonsen J: The MHC molecules of nonmammalian vertebrates. *Immunological Reviews* 113:83, 1990
8. Nonaka M, Namikawa C, Kato Y, Sasaki M, Salter-Cid L, Flajnik MF: Major histocompatibility complex gene mapping in the amphibian *Xenopus* implies a primordial organization. *Proceedings of the National Academy of Sciences USA* 94:5789, 1997
9. Miller MM, Goto R, Bernot A, Zoorob R, Auffray C, Bumstead N, Briles WE: Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. *Proceedings of the National Academy of Sciences USA* 91:4397, 1994
10. Bingulac-Popovic J, Figueroa F, Sato A, Talbot WS, Johnson SL, Gates M, Postlethwait JH, Klein J: Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. *Immunogenetics* 46:129, 1997

11. Klein J, Sato A: Birth of the major histocompatibility complex. *Scandinavian Journal of Immunology* 47:199, 1998
12. Matsunaga T, Rahman A: What brought the adaptive immune system to vertebrates? *Immunological Reviews* 166:177, 1998
13. Kasahara M, Vazquez M, Sato K, McKinney EC, Flajnik MF: Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the cartilaginous fish. *Proceedings of the National Academy of Sciences USA* 89:6688, 1992
14. Okamura K, Ototake M, Nakanishi T, Kurosawa Y, Hashimoto K: The most primitive vertebrates with jaws possess highly polymorphic MHC class I genes comparable to those of humans. *Immunity* 7:777, 1997
15. Kasahara M, Hayashi M, Tanaka K, Inoko H, Sugaya K, Ikemura T, Ishibashi T: Chromosomal localization of the proteasome Z subunit gene reveals an ancient chromosomal duplication involving the major histocompatibility complex. *Proceedings of the National Academy of Sciences USA* 93:9096, 1996
16. Kasahara M, Nakaya J, Satta Y, Takahata N: Chromosomal duplication and the emergence of the adaptive immune system. *Trends in Genetics* 13:90, 1997
17. Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI: The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *Journal of Immunology* 156:4656, 1996
18. Gyllensten UB, Erlich HA: Ancient roots for polymorphism at the HLA-DQ alpha locus in primates. *Proceedings of the National Academy of Sciences USA* 86:9986, 1989
19. Moriuchi J, Moriuchi T, Silver J: Nucleotide sequence of an HLA-DQ alpha chain derived from a DRw9 cell line: genetic and evolutionary implications. *Proceedings of the National Academy of Sciences USA* 82:3420, 1985
20. Karr RW, Gregersen PK, Obata F, Goldberg D, Maccari J, Alber C, Silver J: Analysis of DR beta and DQ beta chain cDNA clones from a DR7 haplotype. *Journal of Immunology* 137:2886, 1986
21. Nei M, Gu X, Sitnikova T: Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences USA* 94:7799, 1997
22. Parham P, Ohta T: Population biology of antigen presentation by MHC class I molecules. *Science* 272:67, 1996
23. Parham P, Arnett KL, Adams EJ, Little AM, Tees K, Barber LD, Marsh SG, Ohta T, Markow T, Petzl-Erler ML: Episodic evolution and turnover of HLA-B in the indigenous human populations of the Americas. *Tissue Antigens* 50:219, 1997
24. Hughes AL, Nei M: Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335:167, 1988
25. Hughes AL, Nei M: Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences USA* 86:958, 1989
26. Hickson RE, Cann RL: Mhc allelic diversity and modern human origins. *Journal of Molecular Evolution* 45:589, 1997
27. Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature* 335:268, 1988
28. Parham P, Adams EJ, Arnett KL: The origins of HLA-A,B,C polymorphism. [Review]. *Immunological Reviews* 143:141, 1995
29. Zangenberg G, Huang MM, Arnheim N, Erlich H: New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genetics* 10:407, 1995
30. Potts WK, Wakeland EK: Evolution of MHC genetic diversity: a tale of incest, pestilence and sexual preference. *Trends in Genetics* 9:408, 1993
31. Potts WK, Slev PR: Pathogen-based models favoring MHC genetic diversity [Review]. *Immunological Reviews* 143:181, 1995

32. Briles WE, Briles RW, Taffs RE: Resistance to a malignant disease in chickens is mapped to a subregion of major histocompatibility (B) complex. *Science* 219:977, 1983
33. Paterson S, Wilson K, Pemberton JM: Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population. *Proceedings of the National Academy of Sciences USA* 95:3714, 1998
34. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM: Common west African HLA antigens are associated with protection from severe malaria. *Nature* 352:595, 1991
35. Fienberg SE: The analysis of multidimensional contingency tables. *Ecology* 51:419, 1970
36. Flaherty L: Major histocompatibility complex polymorphism: a nonimmune theory for selection. *Human Immunology* 21:3, 1988
37. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ: HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* 283:1748, 1999
38. Thursz MR, Thomas HC, Greenwood BM, Hill AV: Heterozygote advantage for HLA class-II type in hepatitis B virus infection [letter]. *Nature Genetics* 17:11, 1997
39. Thio CL, Carrington M, Marti D, et al.: Class II HLA alleles and hepatitis B virus persistence in African Americans. *Journal of Infectious Diseases* 179:1004, 1999
40. Potts WK, Manning CJ, Wakeland EK: Mating patterns in seminatural populations of mice influenced by MHC genotype. *Nature* 352:619, 1991
41. Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu DD, Elias S: HLA and mate choice in humans. *American Journal of Human Genetics* 61:497, 1997
42. Boyse EA, Beauchamp GK, Yamazaki K: Critical review: the sensory perception of genotypic polymorphism of the major histocompatibility complex and other genes: some physiological and phylogenetic implications. *Human Immunology* 6:177, 1983
43. Brown RE, Singh PB, Roser B: The major histocompatibility complex and the chemosensory recognition of individuality in rats. *Physiology & Behavior* 40:65, 1987
44. Singh PB, Brown RE, Roser B: MHC antigens in urine as olfactory recognition cues. *Nature* 327:161, 1987
45. Manning CJ, Wakeland EK, Potts WK: Communal nesting patterns in mice implicate MHC genes in kin recognition. *Nature* 360:581, 1992
46. Yamazaki K, Yamaguchi M, Baranoski L, Bard J, Boyse EA, Thomas L: Recognition among mice. Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. *Journal of Experimental Medicine* 150:755, 1979
47. Brown JL, Eklund A: Kin recognition and the major histocompatibility complex: an integrative review. *American Naturalist* 143:435, 1994
48. Matton DP, Nass N, Clarke AE, Newbigin E: Self-incompatibility: How plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences USA* 91:1992, 1994
49. Kao TH, McCubbin AG: How flowering plants discriminate between self and non-self pollen to prevent inbreeding. *Proceedings of the National Academy of Sciences USA* 93:12059, 1996
50. Partridge L: The rare-male effect: what is its evolutionary significance? *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 319:525, 1988
51. Clarke B, Kirby DR: Maintenance of histocompatibility polymorphisms. *Nature* 211:999, 1966
52. Penn DJ, Potts WK: The evolution of mating preferences and major histocompatibility complex genes. *American Naturalist* 153:145, 1999
53. Ho HN, Yang YS, Hsieh RP, Lin HR, Chen SU, Chen HF, Huang SC, Lee TY, Gill TJI: Sharing of human leukocyte antigens in couples with unexplained infertility affects the

- success of in vitro fertilization and tubal embryo transfer. *American Journal of Obstetrics & Gynecology* 170:63, 1994
54. Michie D, Anderson NF: A strong selective effect associated with a histocompatibility gene in the rat. *Annals of the New York Academy of Sciences* 129:88, 1966
 55. Kirby DR: The egg and immunology. *Proceedings of the Royal Society of Medicine* 63:59, 1970
 56. Wedekind C, Chapuisat M, Macas E, Rulicke T: Non-random fertilization in mice correlates with the MHC and something else. *Heredity* 77:400, 1996
 57. Palm J: Maternal-fetal histoincompatibility in rats: an escape from adversity. *Cancer Research* 34:2061, 1974
 58. Hamilton MS, Hellstrom I: Selection for histoincompatible progeny in mice. *Biology of Reproduction* 19:267, 1978
 59. Haring V, Gray JE, McClure BA, Anderson MA, Clarke AE: Self-incompatibility: a self-recognition system in plants [Review]. *Science* 250:937, 1990
 60. Thompson RD, Kirch HH: The S locus of flowering plants: when self-rejection is self-interest. *Trends in Genetics* 8:381, 1992
 61. Oka H: Colony specificity in compound Ascidians, in Yukova M (ed): *Profiles of Japanese Science & Scientists*, Tokyo, Kodansha, 1970, p 195
 62. Scofield VL, Schlumpberger JM, West LA, Weissman IL: Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* 295:499, 1982
 63. Palm J: Association of maternal genotype and excess heterozygosity for Ag-B histocompatibility antigens among male rats. *Transplantation Proceedings* 1:82, 1969
 64. Palm J: Maternal-fetal interactions and histocompatibility antigen polymorphisms. *Transplantation Proceedings* 2:162, 1970
 65. Hull P: Maternal-foetal incompatibility associated with the H-3 locus in the mouse. *Heredity (Edinburgh)* 24:203, 1969
 66. Dorak MT, Burnett AK: Major histocompatibility complex, t-complex, and leukemia [Review]. *Cancer Causes & Control* 3:273, 1992
 67. Bechtol KB: Lethality of heterozygotes between t-haplotype complementation groups of mouse: sex-related effect on lethality of t6/tw5 heterozygotes. *Genetical Research* 39:79, 1982
 68. King TR: Partial complementation by murine t haplotypes: deficit of males among t6/tw5 double heterozygotes and correlation with transmission-ratio distortion. *Genetical Research* 57:55, 1991
 69. Ioerger TR, Clark AG, Kao TH: Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences USA* 87:9732, 1990
 70. Clark AG, Kao TH: Excess nonsynonymous substitution of shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proceedings of the National Academy of Sciences USA* 88:9823, 1991
 71. Wu J, Saupe SJ, Glass NL: Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences USA* 95:12398, 1998
 72. Hughes AL, Hughes MK, Howell CY, Nei M: Natural selection at the class II major histocompatibility complex loci of mammals. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences* 346:359, 1994
 73. Hughes AL: Evolution of the HLA complex, in Jackson M, Strachan T, Dover G (eds): *Human Genome Evolution*, Oxford, Bios Scientific Publishers, 1996, p 73
 74. Klein J, O'hUigin C: Class II B Mhc motifs in an evolutionary perspective [Review]. *Immunological Reviews* 143:89, 1995

75. Dwyer KG, Balent MA, Nasrallah JB, Nasrallah ME: DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. *Plant Molecular Biology* 16:481, 1991
76. Charlesworth D, Awadalla P: Flowering plant self-incompatibility: the molecular population genetics of *Brassica* S-loci. *Heredity* 81 (Pt 1):1, 1998
77. Schierup MH: The number of self-incompatibility alleles in a finite, subdivided population. *Genetics* 149:1153, 1998
78. Boyes DC, Nasrallah ME, Vrebalov J, Nasrallah JB: The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. *Plant Cell* 9:237, 1997
79. Rivers BA, Bernatzky R, Robinson SJ, Jahnen-Dechent W: Molecular diversity at the self-incompatibility locus is a salient feature in natural populations of wild tomato (*Lycopersicon peruvianum*). *Molecular & General Genetics* 238:419, 1993
80. Hartl DL, Dempster ER, Brown SW: Adaptive significance of vegetative incompatibility in *Neurospora crassa*. *Genetics* 81:553, 1975
81. Saupe SJ, Glass NL: Allelic specificity at the het-c heterokaryon incompatibility locus of *Neurospora crassa* is determined by a highly variable domain. *Genetics* 146:1299, 1997
82. Grosberg RK: The evolution of allorecognition specificity in clonal invertebrates. *Quarterly Review of Biology* 63:377, 1988
83. Weissman IL, Saito Y, Rinkevich B: Allorecognition histocompatibility in a protochordate species: is the relationship to MHC somatic or structural? *Immunological Reviews* 113:227, 1990
84. Magor BG, De Tomaso A, Rinkevich B, Weissman IL: Allorecognition in colonial tunicates: protection against predatory cell lineages? *Immunological Reviews* 167:69, 1992
85. Hedrick PW, Thomson G: Evidence for balancing selection at HLA. *Genetics* 104:449, 1983
86. Klitz W, Thomson G, Baur MP: Contrasting evolutionary histories among tightly linked HLA loci. *American Journal of Human Genetics* 39:340, 1986
87. Potts WK, Wakeland EK: Evolution of diversity at the major histocompatibility complex. *Trends in Ecology and Evolution* 5:181, 1990
88. Degos L, Colombani J, Chaventre A, Bengtson B, Jacquard A: Selective pressure on HLA polymorphism. *Nature* 249:62, 1974
89. Kostyu DD, Dawson DV, Elias S, Ober C: Deficit of HLA homozygotes in a Caucasian isolate. *Human Immunology* 37:135, 1993
90. Black FL, Hedrick PW: Strong balancing selection at HLA loci: evidence from segregation in South Amerindian families. *Proceedings of the National Academy of Sciences USA* 94:12452, 1997
91. Hedrick PW: Evolution at HLA: possible explanations for the deficiency of homozygotes in two populations. *Human Heredity* 40:213, 1990
92. Markow T, Hedrick PW, Zuerlein K, Danilovs J, Martin J, Vyvial T, Armstrong C: HLA polymorphism in the Havasupai: evidence for balancing selection. *American Journal of Human Genetics* 53:943, 1993
93. Black FL, Salzano FM: Evidence for heterosis in the HLA system. *American Journal of Human Genetics* 33:894, 1981
94. von Schantz T, Wittzell H, Goransson G, Grahn M, Persson K: MHC genotype and male ornamentation: genetic evidence for the Hamilton-Zuk model. *Proceedings of the Royal Society of London - Series B: Biological Sciences* 263:265, 1996
95. Duncan WR, Wakeland EK, Klein J: Heterozygosity of H-2 loci in wild mice. *Nature* 281:603, 1979

96. Ritte U, Neufeld E, O'hUigin C, Figueroa F, Klein J: Origins of H-2 polymorphism in the house mouse. II. Characterization of a model population and evidence for heterozygous advantage. Immunogenetics 34:164, 1991
97. Hedrick PW, Whittam TS, Parham P: Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. Proceedings of the National Academy of Sciences USA 88:5897, 1991
98. Sidney J, Grey HM, Kubo RT, Sette A: Practical, biochemical and evolutionary implications of the discovery of HLA class I supermotifs. Immunology Today 17:261, 1996
99. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK: The nature of selection on the major histocompatibility complex. Critical Reviews in Immunology 17:179, 1997
100. Curtsinger JW, Fukui HH, Khazaali AA, Kirscher A, Pletcher SD, Promislow DE, Tatar M: Genetic variation and aging. Annual Review of Genetics 29:553, 1995
101. Albin RL: Antagonistic pleiotropy, mutation accumulation, and human genetic disease. Genetica 91:279, 1993
102. Hall R: Molecular mimicry. Advances in Parasitology 34:81, 1994.

M.Tevfik Dorak, B.A. (Hons), M.D., Ph.D.

Last updated on Nov 29, 2003

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

Evolutionary histories of HLA-DRB haplotypes

M.Tevfik Dorak

Phylogeny and evolutionary history of any gene group can be worked out by determining nucleotide substitutions to construct phylogenetic trees and to

estimate divergence times; or a set of insertions (mostly Alu elements, L1 repeats, retroviral insertions and others) that reveal the order of splitting of duplicated genes; or the presence of homologous genes in different taxa with known evolutionary histories mainly from their fossil records. When these data are applied to human HLA-DRB genes, depending on the method used or in the case of sequencing studies, depending on the part of the gene used, different scenarios emerge. There are, however, some consistent findings. The ancestor of the human DRB genes appears to have been HLA-DRB1*04-like [1; 2]. The ancestor of the DRB1*03 cluster (DRB1*03, DRB1*15, DRB3) is believed to have diverged from the ancestral HLA-DRB1*04 lineage [2; 3]. Both DRB1*04-like ancestor and the ancestor of the DRB1*03 cluster have been estimated to be older than 85 my [1]. This estimate derives from the fact that DRB1*04 alleles are found in prosimian species (note that the mouse lineage separated 75-80 Mya). It is possible that DRB2 (on DR52 haplotypes), DRB4 (on DR53 haplotypes and most closely resembles DRB2), and DRB6 (on DR51 haplotypes) might be the diverged copies of a single ancestral DRB gene [4-6]. The common ancestry of DR51 and DR52 haplotypes is also suggested by a common ERV9 LTR insertion about 40-60 million years old [5]. Phylogenetic analyses of introns 1, 4, 5 of HLA-DRB genes suggest that the present-day HLA-DR haplotypes were derived from three principal ancestral haplotypes: DRB1-DRB2 (DR52 group), DRB1-DRB5 (DR51 group), and DRB1-DRB7 (DR53 group) [2; 3]. Therefore, the main allelic lineages appear to be: an ancestral DRB1*04 (DR53 group) gene (the exon 2 motif: 9EQVKH13) giving rise to DRB1*03 lineage (DR52 group) (the exon 2 motif: 9EYSTS13), and DR51 group (represented by DRB1*15) diverging from the DR52 group later on.

The DRB4 gene may have arisen 46 Mya by a deletion from the DRB1 and DRB2 genes [2]. The DRB1 and DRB4 genes of the DR53 haplotypes have distinct evolutionary histories [7]. The DRB9 locus is about 58 million years old [8] and the pseudogenes DRB7 and DRB8 arose after DRB9 [9]. The remaining HLA-DR haplotypes, the DR1/10 and DR8 groups, probably evolved from the DR51 and DR52 haplotypes, respectively, after more recent deletion events [5]. The DRB1 genes in the DR1 and DR51 haplotypes contain similar ERV9 LTR elements located at the same positions as in the expressed DRB genes in the DR8 and DR52 haplotypes [5; 6]. These ERV9 LTRs are missing at these positions in the DR53 haplotypes. All DR53 haplotypes carry the DRB1, DRB4, DRB7, DRB8 and DRB9 genes. Of these, DRB4, DRB7 and DRB8 are exclusive to this group. The two DRB pseudogenes DRB7 and DRB8 in the DR53 haplotypes do not cluster with other DRB genes. The DRB8 gene also exists in the gorilla and seems to have appeared 18-26 Mya [10]. The DR10 haplotype has a composite origin: a mixture of segments from DR1, DR3 and independent parts can be detected [11]. Interestingly, its HV3 epitope sequence is shared by HLA-DR53 and -DR1 [12; 13].

Since DR51 (incl. DR1/10) and DR52 (incl. DR8) haplotypes seem to share a common ancestry, it is possible to divide all HLA-DR haplotypes into two

evolutionarily related groups: DR53 group and non-DR53 group as direct descendants of the two primordial DRB genes, i.e., HLA-DRB1*04 and HLA-DRB1*03, respectively. The exclusive features of the DR53 haplotypes are their unique DRB gene composition, the lack of the ERV9 LTRs in introns 4 and 5 but the presence of distinct Alu repeats [6], and the 110-160 kb extra DNA content (irrespective of the DRB1 type) [14; 15]. The other main branch is characterized by the ERV9 LTR insertions at identical positions in the intron 5 of the expressed DRB genes (DRB1*01, DRB1*15, DRB1*0301, DRB3*0101, DRB1*08021) [5]. In summary, two main, evolutionarily old branches of DR haplotypes exist in the human population. The DR53 haplotypic group represents one main branch. The second branch consists of the other DR haplotypes [5; 6].

M.Tevfik Dorak, MD, PhD

<http://www.openlink.org/dorak>

Last updated on Nov 29, 2003

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

Major Histocompatibility Complex

M.Tevfik Dorak, MD, PhD

The major histocompatibility complex (MHC) is a set of genes with immunological and non-immunological functions and present in all vertebrates studied so far^{1,2}. It was discovered during transplantation studies in mice (as the H-2 complex) by Peter Gorer in the Lister Institute in London in 1937 who later collaborated with George Snell of the Jackson Laboratories in Ben Harbor^{3,4}. Jean Dausset described the first human MHC antigen Mac (HLA-A2)⁵ followed by the discovery of 4a and 4b (HLA-Bw 4 and -Bw 6) by the Leiden group led by Jon van Rood^{6,7}. The function of the MHC can be described as pleiotropic, i.e., multiple unrelated ones⁸⁻¹¹. It is best known with its role in histocompatibility¹² and in immune regulation¹³⁻¹⁷ with many other functions not much appreciated yet^{2,18-22}. The main function of the main MHC molecules is peptide binding and presentation of them to T lymphocytes. Among the non-immune functions, the

noteworthy ones are interactions with other receptors on the cell surface^{23;24}, in particular with transferrin receptor (TfR), epidermal growth factor²⁵ and various hormone receptors²⁶⁻²⁸, and signal transduction²⁹.

In nature, different taxa of multicellular organisms have unrelated compatibility systems such as the protozoan pheromone system^{30;31}, the fungal compatibility system³²⁻³⁵, angiosperm (flowering plants) self-incompatibility system^{33;36}, and the invertebrate allorecognition systems³⁷⁻³⁹. All of these systems are primarily involved in prevention of matings between genetically similar individuals to avoid the harmful effects of inbreeding. The MHC also prevents inbreeding through its influence on mate choice in mice^{40;41} and humans^{42;43}, and on reproductive processes in rats⁴⁴, mice^{45;46} and humans^{47;48}. The reproductive mechanisms are varied and range from selective fertilization to selective abortion. A major common feature of the compatibility systems is that they favor genetic dissimilarity between mates and the gametes (mate choice, selective fertilization); but similarity in co-operation (kin recognition, dual recognition, transplant matching)^{49;50}. All these functions are based on the provision of a phenotype for the genetic identity of the individual by the MHC: either cell surface molecules or chemosensory signals.

MHC structure

The MHC in humans is called Human Leukocyte Antigens (HLA). It is located on chromosome 6p21.31 and covers a region of about 3.6 Mbp depending on the haplotype^{1;51}. The longest haplotype is the HLA-DR53 group haplotypes because of the 110-160 kb extra DNA in their DR/DQ region⁵²⁻⁵⁶. The HLA complex is divided into three regions: class I, II, and III regions as first proposed by Jan Klein in 1977⁵⁷. The telomeric region to the classical HLA complex is now called the class Ib region; and there has also been a suggestion for a class IV region located at the telomeric end of the class III region². The classical HLA antigens encoded in each region are HLA-A, -B, and -C in the class I region, and HLA-DR, -DQ and -DP in the class II region. All class I genes are between 3 and 6 kb, whereas, class II genes are 4-11 kb long⁵⁸. The 1998 Nomenclature Committee recognized more HLA genes all of which are in the class I and Ib regions: HLA-E, -F, -G, -H, -J, -K and -L⁵⁹. Among those, only HLA-E, -F and -G are expressed⁶⁰. The massive sequencing project of a human MHC haplotype has just been completed and the map positions of all of these genes are known⁵¹. The class III region has the highest gene density but some of the genes are not involved in the immune system^{2;61}. Among the genes which are of interest, HSP70, TNF, C4A, C4B, C2, BF and CYP21 should be mentioned. The HSP70 genes encode cytosolic molecular chaperons and might have donated to the PBR region to the ancestor MHC gene⁶². It has also been proposed that HSP70 may be the functional forerunners of MHC molecules because of their peptide binding and presenting abilities⁶³. By presenting intracellular contents of a cancer cells to the immune system, HSP70 behaves like a tumor rejection antigen⁶⁴⁻⁶⁸ similar to the other molecular chaperons, calreticulin and grp94/gp96⁶⁷⁻⁶⁹. An important feature of HSP70 alleles makes this locus a useful one in disease association studies. They show strong linkage disequilibrium (LD) with HLA-DR alleles⁷⁰⁻⁷². TNF(A) and TNFB (LTA) genes encode cachectin and lymphotoxin- α molecules, respectively^{2;73}. C2, C4A and C4B are the genes for some of the complement proteins, whereas, BF codes for factor B which is also involved in immune response^{74;75}. CYP21 is the gene for 21-hydroxylase which is an important enzyme in corticosteroid metabolism. Its complete deficiency causes congenital adrenal hyperplasia which was the first disease identified

to be the result of a structural change in an HLA-linked gene⁷⁶. Other genes of interest in the class III region are the human homologue of the mouse mammary tumor integration site *Int-3*, *NOTCH4*, and the homologue of a homeobox gene similar to *PBX1* involved in t(1;19) translocation in pre-B cell ALL encoded on chromosome 1q23, *PBX2* (or *HOX12*)^{2,77-79}.

Classical genetics

A highly relevant feature of the MHC antigens is their co-dominant expression. Since both alleles contribute to the phenotype equally, it is important to investigate the genotypes in disease association studies rather than the alleles on their own. If susceptibility to a disease is a recessive trait, allelic association studies may not yield a positive result. Also important is the fact that the MHC is inherited *en bloc* as a haplotype with the exception of the rare recombinational events. Recombination occurs at 1-3% frequency mostly at the HLA-A or HLA-DP ends, i.e., in 100 meioses the haplotype will be broken and reconstituted in one to three of them. The large segment from HLA-B to HLA-DQB is almost always inherited as a whole. This also has important implications in disease associations. A haplotypal association is usually stronger and more meaningful than an allelic association.

The co-dominant expression and haplotypal transmission have an important consequence: within a family, HLA-identical sibling frequency should be 25% according to Mendelian expectations. This has been, however, found to be higher than that in leukemia⁸⁰⁻⁸⁴. This would suggest preferential transmission of leukemia-associated HLA haplotypes⁸⁵. The fact that HLA-identical sibling frequency is higher than 25% in leukemic families should not be confused with the overall chance of having an HLA-identical sibling which is correlated with the family size (equal to $[1 - (0.75)^n]$ where n is the number of siblings). This probability may go up to 55% in areas where families are traditionally large⁸⁶.

Despite the enormous number of alleles at each expressed loci, the number of haplotypes observed in populations is much smaller than theoretical expectations. This is to say that certain alleles tend to occur together on the same haplotype rather than randomly segregating together. This is called linkage disequilibrium (LD) and quantitated by a Δ value^{87,88}.

The public specificities, also called supertypes and sometimes wrongly broad specificities, group a number of private specificities. In the HLA class I region, all HLA-B private specificities are grouped into two supertypal families: HLA-Bw4 and -Bw6. In recent years, the nature of HLA-B supertypes has been better understood. They are not encoded by a different gene. The antigens HLA-Bw4 and -Bw6 reside on a unique epitope on each HLA-B molecule and are distinctly different from the epitopes that determine the HLA-B specificity. Each HLA-B molecule expresses either the Bw4 or Bw6 supertype (residues 74 to 83 of the α 1 helix) in addition to a (private) HLA-B specificity. The amino acid residues 80 IALR 83 represent the Bw4 specificity, whereas, 80 NLRG 83 represent Bw6 (Ref 89).

Likewise, HLA-DR alleles are also associated with supertypes. However, the HLA-DR supertypes are not allelic with each other⁹⁰. They are encoded by separate genes (HLA-DRB3, -B4, -B5) and are distinct molecules (HLA-DR52, -DR53, -DR51, respectively). Only one or none of these genes occurs on a haplotype.

The private specificities in each supertypal family is as follows:

DR51 (DRB5): DR2 (DRB1*15/16)

DR52 (DRB3): DR3 (17/18; DRB1*03), DR5 (DRB1*11/12), DR6 (DRB1*13/14)

DR53 (DRB4): DR4 (DRB1*04), DR7 (DRB1*07), DR9 (DRB1*09)

Although all HLA-DR4 / 7 / 9 haplotypes carry the structural gene HLA-DRB4, not all of them express the HLA-DR53 molecule⁹¹. The non-expression, however, is restricted to the HLA-(B57) : DR7 (Dw 11): DQ9 haplotype⁹² due to a G to A substitution in the acceptor splice site at the 3' end of the first intron, changing the normal AG dinucleotide to AA^{93;94}. In fact, the null allele of the HLA-DRB4 gene is expressed but it is an aberrant protein⁹⁵. An exception has been reported as an unexpected expression of HLA-DR53 in a DR7 (Dw 11) : DQ9 - positive leukemia patient⁹⁶. A difference between HLA-B and -DR supertypes is that not all DR alleles are associated with a supertype. These are HLA-DR1, -DR8 and -DR10. Thus, no supertypal gene is present on these haplotypes.

An interesting group of MHC haplotypes is the ancestral or extended haplotypes (also called supratypes). These are specific HLA-B, -DR, BF, C2, C4A and C4B combinations in significant linkage disequilibrium in chromosomes of unrelated individuals. They extend from HLA-B to DR and have been conserved *en bloc*⁹⁷⁻¹⁰¹. In some Caucasian populations, the extended haplotypes constitute 25-30% of all MHC haplotypes and together with recombinants between any two of them, they account for almost 75% of unselected haplotypes^{97;98;100}. Particular extended haplotypes are identical by descent. The evidence for this is that in one study, 22 of 26 unrelated extended-haplotype-matched subjects had similar mixed lymphocyte reactivity to HLA-identical siblings⁹⁹. Matching for extended haplotypes significantly improves survival in kidney transplantation¹⁰². In Caucasians, there are 10 to 12 common extended haplotypes that show significant linkage disequilibrium. They are relatively population-specific^{101;102} and are believed to represent the original MHC haplotypes of our ancestors which are still segregating unchanged. They are easily recognized from their characteristic class III polymorphisms called complotypes^{100;102-104}. Disease associations with extended haplotypes are generally stronger than allelic associations¹⁰⁰. The best examples of extended haplotype associations are those with rheumatoid arthritis¹⁰⁵, multiple sclerosis¹⁰⁶, insulin-dependent diabetes mellitus^{100;107;108}, and systemic lupus erythematosus¹⁰⁹.

Polymorphism

One of the main characteristics of the MHC is its extreme polymorphism. Among the expressed loci, the MHC has the greatest degree of polymorphism in the human genome. The numbers of alleles recognized at the classical loci by December 1998 are presented in Table 1 (for the latest number of alleles, follow the link at the end).

Table 1. Number of alleles at the classical HLA loci

Locus	DNA-level Alleles	Serological Equivalents
HLA-A	119	40
HLA-B	245	88

HLA-C	74	9
HLA-DRB1	201	80
HLA-DQB1	39	7
HLA-DPB1	84	(-)

Data from Refs 59,91,110

This is at such a degree that it is theoretically possible for each human to possess a different set of MHC alleles. This feature of the MHC is shared by other compatibility systems in different taxa (such as the fungal mating types, invertebrate allorecognition system and plant self-incompatibility system). It is, however, important to recognise that within the allelic polymorphism at the DNA level which seems endless, there are ancient lineages which predate speciation and maintain themselves in closely related species. This is the basis of the trans-species polymorphism theory proposed by Jan Klein and has found widespread support¹¹¹. Allelic lineages may be shared by related species, such as human and apes^{112;113} or even human and mice¹¹⁴, having been present in their common ancestor. However, when primate and human HLA alleles are compared, there is no identical (private) class I allele in great apes and humans despite the similarities in polymorphic motifs^{113;115-120}. The only similarity is that the human class I supertypes Bw4 and Bw6 are cross-reactive with chimpanzees and gorillas (and even with rhesus monkeys)^{113;117;121-123}. This conservation throughout hominoid evolution is attributed to the functional importance of these two epitopes in CTL immunomodulation⁸⁹ and NK cell function¹²⁴⁻¹²⁶.

Similarly, in the HLA-DRB loci, while no private specificity has an equivalent in another species, the DRB3 / 4 / 5 loci seem to have remained as they are in all primates or even in rhesus monkeys^{127;128}. Interestingly, these loci encode the class II supertypes HLA-DR52, -53, and -51. The most ancient polymorphic class II locus appears to be HLA-DQA1^{129;130}. The polymorphism of this locus also correlates to the MHC class II supertypal groupings^{131;132}.

The haplotypal structure, phylogenetic analysis and sequence comparisons agree on the presence of five major haplotypal groups in the HLA class II region^{59;128;133-135}. These are HLA-DR1, HLA-DR51, HLA-DR52, HLA-DR8, and HLA-DR53. It appears that the oldest lineages are HLA-DRB1*04 (represented by the exon 2 motif 9 EQVKH 13) and HLA-DRB1*03 (the motif 9 EYSTS 13)¹³⁶. The analysis of intron sequences suggest that HLA-DRB1*03 first diverged from HLA-DRB1*04 more than 85 million years ago and later gave rise to HLA-DRB1*15 (Refs 133,137,138).

MHC class I and class II supertypes are biologically as functional as private specificities. MHC restriction of peptide presentation has been shown for HLA-Bw4 and -Bw6^{139;140}, HLA-DR52^{141;142} and HLA-DR53¹⁴³⁻¹⁴⁷. HLA-Bw4 appears to be more immunogenic than HLA-Bw6 judged by the antibody response in the case of mismatching in transplantation¹⁴⁸. Similar to the cross-reactive HLA-B supertypes between humans and chimpanzees¹¹⁷, HLA-DR supertypes are cross-reactive with chimpanzee¹⁴² and even with mouse class II supertypes¹⁴⁹. Particularly striking is the cross-reactions between HLA-DR53 and H-2E^k^{145;150}. Furthermore, HLA-DR53 has its own peptide binding motif

^{151;152}. The most abundant peptide eluted from the DR53 molecule is derived from an intracellular protein, calreticulin, which is involved in MHC class I biosynthesis, heat shock response and tumor rejection ^{69;153-155}.

References

1. Trowsdale J. "Both man & bird & beast": comparative organization of MHC genes [Review]. **Immunogenetics** 1995; **41**: 1-17.
2. Gruen JR, Weissman SM. Evolving views of the major histocompatibility complex. **Blood** 1997; **90**: 4252-4265.
3. Gorer PA. The genetic and antigenic basis of tumour transplantation. **Journal of Pathology & Bacteriology** 1937; **44**: 691-697.
4. Gorer PA, Lyman S, Snell GD. Studies on the genetic and antigenic basis of tumour transplantation. Linkage between a histocompatibility gene and "fused" in mice. **Proceedings of the Royal Society of London - Series B: Biological Sciences** 1948; **135**: 499-505.
5. Dausset J. Iso-leuco-antico-rps. **Acta Haematologica** 1959; **20**: 156-156.
6. van Rood, J. J. Leukocyte Grouping: A Method and Its Application. PhD dissertation, 1962. University of Leiden, Netherlands.
7. van Rood JJ, van Leeuwen A. Leukocyte grouping. A method and its application. **Journal of Clinical Investigation** 1963; **42**: 1382-1390.
8. Bodmer WF. Evolutionary significance of the HLA system [Review]. **Nature** 1972; **237**: 139-145.
9. Meruelo D, Edidin M. The biological function of the major histocompatibility complex: hypotheses [Review]. **Contemporary Topics in Immunobiology** 1980; **9**: 231-253.
10. Jonker M, Balner H. The major histocompatibility complex: a key to a better understanding of evolution. **Transplantation Proceedings** 1980; **12**: 575-581.
11. Dausset J. The major histocompatibility complex in man. **Science** 1981; **213**: 1469-1474.
12. Snell GD. Studies in histocompatibility. **Science** 1981; **213**: 172-178.
13. Benacerraf B, McDevitt HO. Histocompatibility-linked immune response genes. **Science** 1972; **175**: 273-279.
14. Zinkernagel RM, Doherty PC. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. **Nature** 1974; **248**: 701-702.
15. Benacerraf B. Role of MHC gene products in immune regulation. **Science** 1981; **212**: 1229-1238.
16. Doherty PC, Zinkernagel RM. A biological role for the major histocompatibility antigens. **Lancet** 1975; **1**: 1406-1409.
17. Zinkernagel RM. Cellular immune recognition and the biological role of major transplantation antigens. **Scandinavian Journal of Immunology** 1997; **46**: 421-436.
18. Bonner JJ. Major histocompatibility complex influences reproductive efficiency: evolutionary implications. **Journal of Craniofacial Genetics and Developmental Biology** 1986; **Suppl 2**: 5-11.
19. Lerner SP, Finch CE. The major histocompatibility complex and reproductive functions [Review]. **Endocrine Reviews** 1991; **12**: 78-90.
20. Powis SH, Geraghty DE. What is the MHC? **Immunology Today** 1995; **16**: 466-468.
21. Zavazava N, Eggert F. MHC and behavior. **Immunology Today** 1997; **18**: 8-10.
22. Penn DJ, Potts WK. The evolution of mating preferences and major histocompatibility complex genes. **American Naturalist** 1999; **153**: 145-164.
23. Svejgaard A, Ryder LP. Interaction of HLA molecules with non-immunological ligands as an explanation of HLA and disease association. **Lancet** 1976; **ii**: 547-549.
24. Edidin M. Function by association? MHC antigens and membrane receptor complexes. **Immunology Today** 1988; **9**: 218-219.
25. Schreiber AB, Schlessinger J, Edidin M. Interaction between major histocompatibility complex antigens and epidermal growth factor receptors on human cells. **Journal of Cell Biology** 1984; **98**: 725-731.
26. Phillips ML, Moule ML, Delovitch TL, Yip CC. Class I histocompatibility antigens and insulin receptors: evidence for interactions. **Proceedings of the National Academy of Sciences USA** 1986; **83**: 3474-3478.
27. Solano AR, Cremaschi G, Sánchez ML, Borda E, Stein-Borda L, Podestà EJ. Molecular and biological interaction between major histocompatibility complex class I antigens and luteinizing

- hormone receptors or beta-adrenergic receptors triggers cellular response in mice. **Proceedings of the National Academy of Sciences USA** 1988; **85**: 5087-5091.
28. Verland S, Simonsen M, Gammeltoft S, Allen H, Flavell RA, Olsson L. Specific molecular interaction between the insulin receptor and a D product of MHC class I. **Journal of Immunology** 1989; **143**: 945-951.
29. Schafer PH, Pierce SK, Jardetzky TS. The structure of MHC class II: a role for dimer of dimers. **Seminars in Immunology** 1995; **7**: 389-398.
30. Weiss MS, Anderson DH, Raffioni S, et al. A cooperative model for receptor recognition and cell adhesion: evidence from the molecular packing in the 1.6-A crystal structure of the pheromone Er-1 from the ciliated protozoan *Euplotes raikovi*. **Proceedings of the National Academy of Sciences USA** 1995; **92**: 10172-10176.
31. Vallesi A, Giuli G, Bradshaw RA, Luporini P. Autocrine mitogenic activity of pheromones produced by the protozoan ciliate *Euplotes raikovi*. **Nature** 1995; **376**: 522-524.
32. Metzberg RL. The role of similarity and difference in fungal mating. **Genetics** 1990; **125**: 457-462.
33. Hiscock SJ, Kues U, Dickinson HG. Molecular mechanisms of self-incompatibility in flowering plants and fungi - different means to the same end. **Trends in Cell Biology** 1996; **6**: 421-428.
34. Wendland J, Vaillancourt LJ, Hegner J, et al. The mating-type locus B alpha 1 of *Schizophyllum commune* contains a pheromone receptor gene and putative pheromone genes. **EMBO Journal** 1995; **14**: 5271-5278.
35. Kothe E. Tetrapolar fungal mating types: sexes by the thousands. **FEMS Microbiological Reviews** 1996; **18**: 65-87.
36. Haring V, Gray JE, McClure BA, Anderson MA, Clarke AE. Self-incompatibility: a self-recognition system in plants [Review]. **Science** 1990; **250**: 937-941.
37. Scofield VL, Schlumpberger JM, West LA, Weissman IL. Protochordate allorecognition is controlled by a MHC-like gene system. **Nature** 1982; **295**: 499-502.
38. Grosberg RK. The evolution of allorecognition specificity in donal invertebrates. **Quarterly Review of Biology** 1988; **63**: 377-412.
39. Weissman IL, Saito Y, Rinkevich B. Allorecognition histocompatibility in a protochordate species: is the relationship to MHC somatic or structural? **Immunological Reviews** 1990; **113**: 227-241.
40. Williams JR, Lenington S. Factors modulating preferences of female house mice for males differing in t-complex genotype: role of t-complex genotype, genetic background, and estrous condition of females. **Behavior Genetics** 1993; **23**: 51-58.
41. Potts WK, Manning CJ, Wakeland EK. Mating patterns in seminatural populations of mice influenced by MHC genotype. **Nature** 1991; **352**: 619-621.
42. Ober C, Weitkamp LR, Cox N, Dytch H, Kostyu DD, Elias S. HLA and mate choice in humans. **American Journal of Human Genetics** 1997; **61**: 497-504.
43. Wedekind C, Seebeck T, Bettens F, Paepke AJ. MHC-dependent mate preferences in humans. **Proceedings of the Royal Society of London - Series B: Biological Sciences** 1995; **260**: 245-249.
44. Palm J. Maternal-Fetal histoincompatibility in rats: an escape from adversity. **Cancer Research** 1974; **34**: 2061-2065.
45. Hamilton MS, Hellstrom I. Selection for histoincompatible progeny in mice. **Biology of Reproduction** 1978; **19**: 267-270.
46. Wedekind C, Chapuisat M, Macas E, Rulicke T. Non-random fertilization in mice correlates with the MHC and something else. **Heredity** 1996; **77**: 400-409.
47. Ober C, Elias S, Kostyu DD, Hauck WW. Decreased fecundability in Hutterite couples sharing HLA-DR. **American Journal of Human Genetics** 1992; **50**: 6-14.
48. Jin K, Ho HN, Speed TP, Gill TJ. Reproductive failure and the major histocompatibility complex. **American Journal of Human Genetics** 1995; **56**: 1456-1467.
49. Brown JL. Some paradoxical goals of cells and organisms: the role of the MHC. In: Pfaff DW, ed. *Ethical Questions in Brain and Behavior: Problems and Opportunities*, New York: Springer-Verlag, 1983: 111-124.
50. Jones JS, Partridge L. Tissue rejection: the price for sexual acceptance. **Nature** 1983; **304**: 484-485.

51. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. **Nature** 1999; **401**: 921-923.
52. Dunham I, Sargent CA, Dawkins RL, Campbell RD. An analysis of variation in the long-range genomic organization of the human major histocompatibility complex class II region by pulsed-field gel electrophoresis. **Genomics** 1989; **5**: 787-796.
53. Tokunaga K, Saueracker G, Kay PH, Christiansen FT, Anand R, Dawkins RL. Extensive deletions and insertions in different MHC supratypes detected by pulsed field gel electrophoresis. **Journal of Experimental Medicine** 1988; **168**: 933-940.
54. Inoko H, Ando A, Kawai J, Trowsdale J, Tsuji K. Mapping of the HLA-D region by pulsed-field gel electrophoresis: size variation in subregion intervals. In: Silver J, ed. *Molecular Biology of HLA Class II Antigens*, Florida: CRC Press, 1990: 1-17.
55. Niven MJ, Hitman GA, Pearce H, Marshall B, Sachs JA. Large haplotype-specific differences in inter-genic distances in human MHC shown by pulsed field electrophoresis mapping of healthy and type 1 diabetic subjects. **Tissue Antigens** 1990; **36**: 19-24.
56. Kendall E, Todd JA, Campbell RD. Molecular analysis of the MHC class II region in DR4, DR7, and DR9 haplotypes. **Immunogenetics** 1991; **34**: 349-357.
57. Klein J. Evolution and function of the major histocompatibility complex: facts and speculations. In: Gotze D, ed. *The Major Histocompatibility System in Man and Animals*, New York: Springer-Verlag, 1976: 339-378.
58. Browning M, McMichael A. *MHC and HLA: Genes, Molecules and Function*. Oxford: Bios Scientific Publishers, 1996;
59. Bodmer JG, Marsh SG, Albert ED, et al. Nomenclature for the factors of the HLA system, 1998. **European Journal of Immunogenetics** 1999; **26**: 81-116.
60. Le Bouteiller P. HLA class I chromosomal region, genes, and products: facts and questions. **Critical Reviews in Immunology** 1994; **14**: 89-129.
61. Aguado B, Milner CM, Campbell RD. Genes of the MHC class III region and the functions of the proteins they encode. In: Browning M, McMichael A, eds. *HLA and MHC: genes, molecules and function*, Oxford: Bios Scientific Publishers, 1996: 39-76.
62. Flajnik MF, Canel C, Kramer J, Kasahara M. Which came first, MHC class I or class II? **Immunogenetics** 1991; **33**: 295-300.
63. Srivastava PK, Heike M. Tumor-specific immunogenicity of stress-induced proteins: convergence of two evolutionary pathways of antigen presentation? **Seminars in Immunology** 1991; **3**: 57-64.
64. Srivastava PK. Peptide-binding heat shock proteins in the endoplasmic reticulum: role in immune response to cancer and in antigen presentation. **Advances in Cancer Research** 1993; **62**: 153-177.
65. Srivastava PK, Menoret A, Basu S, Binder RJ, McQuade KL. Heat shock proteins come of age: primitive functions acquire new roles in an adaptive world. **Immunity** 1998; **8**: 657-665.
66. Blachere NE, Li Z, Chandawarkar RY, et al. Heat shock protein-peptide complexes, reconstituted in vitro, elicit peptide-specific cytotoxic T lymphocyte response and tumor immunity. **Journal of Experimental Medicine** 1997; **186**: 1315-1322.
67. Suto R, Srivastava PK. A mechanism for the specific immunogenicity of heat shock protein-chaperoned peptides. **Science** 1995; **269**: 1585-1588.
68. Ishii T, Udono H, Yamano T, et al. Isolation of MHC class I-restricted tumor antigen peptide and its precursors associated with heat shock proteins hsp70, hsp90, and gp96. **Journal of Immunology** 1999; **162**: 1303-1309.
69. Basu S, Srivastava PK. Calreticulin, a peptide-binding chaperone of the endoplasmic reticulum, elicits tumor- and peptide-specific immunity. **Journal of Experimental Medicine** 1999; **189**: 797-802.
70. Partanen J, Milner C, Campbell RD, Maki M, Lipsanen V, Koskimies S. HLA-linked heat-shock protein 70 (HSP70-2) gene polymorphism and celiac disease. **Tissue Antigens** 1993; **41**: 15-19.
71. Cascino I, D'Alfonso S, Cappello N, et al. Gametic association of HSP70-1 promoter region alleles and their inclusion in extended HLA haplotypes. **Tissue Antigens** 1993; **42**: 62-66.
72. Cascino I, Sorrentino R, Tosi R. Strong genetic association between HLA-DR3 and a polymorphic variation in the regulatory region of the HSP70-1 gene. **Immunogenetics** 1993; **37**: 177-182.

73. Webb GC, Chaplin DD. Genetic variability at the human tumor necrosis factor loci. **Journal of Immunology** 1990; **145**: 1278-1285.
74. Campbell RD, Carroll MC, Porter RR. The molecular genetics of components of complement [Review]. **Advances in Immunology** 1986; **38**: 203-244.
75. Campbell RD, Dunham I, Sargent CA. Molecular mapping of the HLA-linked complement genes and the RCA linkage group [Review]. **Experimental & Clinical Immunogenetics** 1988; **5**: 81-98.
76. Levine LS, Zachmann M, New MI, et al. Genetic mapping of the 21-hydroxylase-deficiency gene within the HLA linkage group. **New England Journal of Medicine** 1978; **299**: 911-915.
77. Sugaya K, Fukagawa T, Matsumoto K, et al. Three genes in the human MHC class III region near the junction with the class II: gene for receptor of advanced glycosylation end products, PBX2 homeobox gene and a notch homolog, human counterpart of mouse mammary tumor gene int-3. **Genomics** 1994; **23**: 408-419.
78. Sugaya K, Sasanuma S, Nohata J, et al. Gene organization of human NOTCH4 and (CTG)_n polymorphism in this human counterpart gene of mouse proto-oncogene Int3. **Gene** 1997; **189**: 235-244.
79. Lu Q, Wright DD, Kamps MP. Fusion with E2A converts the Pbx1 homeodomain protein into a constitutive transcriptional activator in human leukemias carrying the t(1;19) translocation. **Molecular & Cellular Biology** 1994; **14**: 3938-3948.
80. Chan KW, Pollack MS, Braun D, Jr., O'Reilly RJ, Dupont B. Distribution of HLA genotypes in families of patients with acute leukemia. Implications for transplantation. **Transplantation** 1982; **33**: 613-615.
81. De Moor P, Louwagie A. Distribution of HLA genotypes in sibs of patients with acute leukaemia. **Scandinavian Journal of Haematology** 1985; **34**: 68-70.
82. De Moor P. Distribution of HLA genotypes in sibs of patients with acute lymphoblastic leukemia. **European Journal of Haematology** 1989; **42**: 317-318.
83. Carpentier NA, Jeannet M. Increased HLA-DR compatibility between patients with acute myeloid leukemia and their parents: implication for bone marrow transplantation. **Transplantation Proceedings** 1987; **19**: 2644-2645.
84. Dorak MT, Chalmers EA, Gaffney D, et al. Human major histocompatibility complex contains several leukemia susceptibility genes. **Leukemia & Lymphoma** 1994; **12**: 211-222.
85. Dorak MT, Burnett AK. Major histocompatibility complex, t-complex, and leukemia [Review]. **Cancer Causes & Control** 1992; **3**: 273-282.
86. O'Riordan J, Finch A, Lawlor E, McCann SR. Probability of finding a compatible sibling donor for bone marrow transplantation in Ireland. **Bone Marrow Transplantation** 1992; **9**: 27-30.
87. Mattiuz PL, Ihde D, Piazza A, Ceppellini R, Bodmer WF. New approaches to the population genetic and segregation analysis of the HL-A system. In: Terasaki P, ed. *Histocompatibility Testing 1970*, Copenhagen: Munksgaard, 1971: 193-205.
88. Begovich AB, McClure GR, Suraj VC, et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. **Journal of Immunology** 1992; **148**: 249-258.
89. Nossner E, Goldberg JE, Naftzger C, Lyu SC, Clayberger C, Krensky AM. HLA-derived peptides which inhibit T cell function bind to members of the heat-shock protein 70 family. **Journal of Experimental Medicine** 1996; **183**: 339-348.
90. Gorski J, Rollini P, Mach B. Structural comparison of the genes of two HLA-DR supertypic groups: the loci encoding DRw52 and DRw53 are not truly allelic. **Immunogenetics** 1987; **25**: 397-402.
91. Schreuder GM, Hurlley CK, Marsh SG, et al. The HLA dictionary 1999: A summary of HLA-A, -B, -C, -DRB1/3/4/5, -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR, and -DQ antigens. **Human Immunology** 1999; **60**: 1157-1181.
92. Knowles RW, Flomenberg N, Horibe K, Winchester R, Radka SF, Dupont B. Complexity of the supertypic HLA-DRw53 specificity: two distinct epitopes differentially expressed on one or all of the DR beta-chains depending on the HLA-DR allotype. **Journal of Immunology** 1986; **137**: 2618-2626.
93. Sutton VR, Kienzie BK, Knowles RW. An altered splice site is found in the DRB4 gene that is not expressed in HLA-DR7, Dw11 individuals. **Immunogenetics** 1989; **29**: 317-322.
94. Leen MP, Gorski J. DRB4 promoter polymorphism in DR7 individuals: correlation with DRB4 pre-mRNA and mRNA levels. **Immunogenetics** 1997; **45**: 371-378.

95. Sutton VR, Knowles RW. An aberrant DRB4 null gene transcript is found that could encode a novel HLA-DR beta chain. **Immunogenetics** 1990; **31**: 112-117.
96. Lardy NM, van der Horst AR, van de Weerd MJ, de Waal LP, Bontrop RE. HLA-DRB4 gene encoded HLA-DR53 specificity segregating with the HLA-DR7, -DQ9 haplotype: unusual association. **Human Immunology** 1998; **59**: 115-118.
97. Alper CA, Awdeh Z, Yunis EJ. Conserved, extended MHC haplotypes [Review]. **Experimental & Clinical Immunogenetics** 1992; **9**: 58-71.
98. Degli-Esposti MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL. Ancestral haplotypes: conserved population MHC haplotypes. **Human Immunology** 1992; **34**: 242-252.
99. Awdeh ZL, Alper CA, Eynon E, Alosco SM, Stein R, Yunis EJ. Unrelated individuals matched for MHC extended haplotypes and HLA-identical siblings show comparable responses in mixed lymphocyte culture. **Lancet** 1985; **2**: 853-856.
100. Alper CA, Awdeh ZL, Yunis EJ. Complotypes, extended haplotypes, male segregation distortion, and disease markers. **Human Immunology** 1986; **15**: 366-373.
101. Gaudieri S, Leelayuwat C, Tay GK, Townsend DC, Dawkins RL. The major histocompatibility complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. **Journal of Molecular Evolution** 1997; **45**: 17-23.
102. Wilton AN, Christiansen FT, Dawkins RL. Supratype matching improves renal transplant survival. **Transplantation Proceedings** 1985; **17**: 2211-2216.
103. Christiansen FT, Witt CS, Dawkins RL. Questions in marrow matching: the implications of ancestral haplotypes for routine practice. **Bone Marrow Transplantation** 1991; **8**: 83-86.
104. Degli-Esposti MA, Leelayuwat C, Daly LN, et al. Updated characterization of ancestral haplotypes using the Fourth Asia-Oceania Histocompatibility Workshop panel. **Human Immunology** 1995; **44**: 12-18.
105. Fraser PA, Stern S, Larson MG, et al. HLA extended haplotypes in childhood and adult onset HLA-DR4-associated arthropathies. **Tissue Antigens** 1990; **35**: 56-59.
106. Hauser SL, Fleischnick E, Weiner HL, et al. Extended major histocompatibility complex haplotypes in patients with multiple sclerosis. **Neurology** 1989; **39**: 275-277.
107. Raum D, Awdeh Z, Yunis EJ, Alper CA, Gabbay KH. Extended major histocompatibility complex haplotypes in type I diabetes mellitus. **Journal of Clinical Investigation** 1984; **74**: 449-454.
108. Christiansen FT, Saueracker GC, Leaver AL, Tokunaga K, Cameron PU, Dawkins RL. Characterization of MHC ancestral haplotypes associated with insulin-dependent diabetes mellitus: evidence for involvement of non-HLA genes. **Journal of Immunogenetics** 1990; **17**: 379-386.
109. Welch TR, Beischel LS, Balakrishnan K, Quinlan M, West CD. Major histocompatibility complex extended haplotypes in systemic lupus erythematosus. **Disease Markers** 1988; **6**: 247-255.
110. Marsh SG, Parham P, Barber LD. The HLA Facts Book San Diego: Academic Press, 2000.
111. Klein J, Satta Y, O'Huigin C, Takahata N. The molecular descent of the major histocompatibility complex [Review]. **Annual Review of Immunology** 1993; **11**: 269-295.
112. Kupfermann H, Mayer WE, O'Huigin C, Klein D, Klein J. Shared polymorphism between gorilla and human major histocompatibility complex DRB loci. **Human Immunology** 1992; **34**: 267-278.
113. Lawlor DA, Warren E, Taylor P, Parham P. Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. **Journal of Experimental Medicine** 1991; **174**: 1491-1509.
114. Lundberg AS, McDevitt HO. Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. **Proceedings of the National Academy of Sciences USA** 1992; **89**: 6545-6549.
115. Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G, Klein J. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. **EMBO Journal** 1988; **7**: 2765-2774.
116. Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. **Nature** 1988; **335**: 268-271.
117. Lawlor DA, Warren E, Ward FE, Parham P. Comparison of class I MHC alleles in humans and apes [Review]. **Immunological Reviews** 1990; **113**: 147-185.
118. Erlich HA, Gyllenstein UB. Shared epitopes among HLA class II alleles: gene conversion, common ancestry and balancing selection. **Immunology Today** 1991; **12**: 411-414.

119. Boyson JE, Shufflebotham C, Cadauid LF, et al. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. **Journal of Immunology** 1996; **156**: 4656-4665.
120. Parham P, Lawlor DA, Lomen CE, Ennis PD. Diversity and diversification of HLA-A,B,C alleles. **Journal of Immunology** 1989; **142** : 3937-3950.
121. Metzgar RS, Ward FE, Seigler HF. Study of the HL-A system in chimpanzees. In: Dausset J, Colombani J, eds. *Histocompatibility Testing 1972*, Copenhagen: Munksgaard, 1972: 55-61.
122. Bright S, Balner H. The antigens 4a and 4b in rhesus monkeys and stump-tailed macaques. **Tissue Antigens** 1976; **8**: 261-271.
123. Balner H. The major histocompatibility complex of primates: evolutionary aspects and comparative histogenetics. **Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences** 1981; **292**: 109-119.
124. Gumperz JE, Litwin V, Phillips JH, Lanier LL, Parham P. The Bw4 public epitope of HLA-B molecules confers reactivity with natural killer cell clones that express NKB1, a putative HLA receptor. **Journal of Experimental Medicine** 1995; **181**: 1133-1144.
125. Cella M, Longo A, Ferrara GB, Strominger JL, Colonna M. NK3-specific natural killer cells are selectively inhibited by Bw4-positive HLA alleles with isoleucine 80. **Journal of Experimental Medicine** 1994; **180**: 1235-1242.
126. Lanier LL, Phillips JH. Inhibitory MHC class I receptors on NK cells and T cells. **Immunology Today** 1996; **17**: 86-91.
127. Sliereendregt BL, van Noort JT, Bakas RM, Otting N, Jonker M, Bontrop RE. Evolutionary stability of transspecies major histocompatibility complex class II DRB lineages in humans and rhesus monkeys. **Human Immunology** 1992; **35**: 29-39.
128. Hickson RE, Cann RL. Mhc allelic diversity and modern human origins. **Journal of Molecular Evolution** 1997; **45**: 589-598.
129. Gyllensten UB, Erlich HA. Ancient roots for polymorphism at the HLA-DQ alpha locus in primates. **Proceedings of the National Academy of Sciences USA** 1989; **86**: 9986-9990.
130. Bergstrom T, Gyllensten U. Evolution of Mhc class II polymorphism: the rise and fall of class II gene function in primates [Review]. **Immunological Reviews** 1995; **143**: 13-31.
131. Moriuchi J, Moriuchi T, Silver J. Nucleotide sequence of an HLA-DQ alpha chain derived from a DRw9 cell line: genetic and evolutionary implications. **Proceedings of the National Academy of Sciences USA** 1985; **82**: 3420-3424.
132. Karr RW, Gregersen PK, Obata F, et al. Analysis of DR beta and DQ beta chain cDNA clones from a DR7 haplotype. **Journal of Immunology** 1986; **137**: 2886-2890.
133. Satta Y, Mayer WE, Klein J. HLA-DRB intron 1 sequences: implications for the evolution of HLA-DRB genes and haplotypes. **Human Immunology** 1996; **51**: 1-12.
134. Svensson AC, Setterblad N, Pihlgren U, Rask L, Andersson G. Evolutionary relationship between human major histocompatibility complex HLA-DR haplotypes. **Immunogenetics** 1996; **43**: 304-314.
135. Gongora R, Figueroa F, Klein J. The HLA-DRB9 gene and the origin of HLA-DR haplotypes. **Human Immunology** 1996; **51**: 23-31.
136. Klein J, O'Huigin C. Class II B Mhc motifs in an evolutionary perspective [Review]. **Immunological Reviews** 1995; **143**: 89-111.
137. Figueroa F, O'Huigin C, Tichy H, Klein J. The origin of the primate Mhc-DRB genes and allelic lineages as deduced from the study of prosimians. **Journal of Immunology** 1994; **152**: 4455-4465.
138. Satta Y, Mayer WE, Klein J. Evolutionary relationship of HLA-DRB genes inferred from intron sequences. **Journal of Molecular Evolution** 1996; **42**: 648-657.
139. Clayberger C, Rosen M, Parham P, Krensky AM. Recognition of an HLA public determinant (Bw4) by human allogeneic cytotoxic T lymphocytes. **Journal of Immunology** 1990; **144**: 4172-4176.
140. Arnett KL, Huang W, Valiante NM, Barber LD, Parham P. The Bw4/Bw6 difference between HLA-B*0802 and HLA-B*0801 changes the peptides endogenously bound and the stimulation of alloreactive T cells. **Immunogenetics** 1998; **48**: 56-61.
141. Decary F, L'Abbe D, Tremblay L, Chartrand P. The immune response to the HPA-1a antigen: association with HLA-DRw52a. **Transfusion Medicine** 1991; **1**: 55-62.

142. Bontrop RE, Elferink DG, Otting N, Jonker M, de Vries RR. Major histocompatibility complex class II-restricted antigen presentation across a species barrier: conservation of restriction determinants in evolution. **Journal of Experimental Medicine** 1990; **172**: 53-59.
143. Qvigstad E, Gaudernack G, Thorsby E. Antigen-specific T cell clones restricted by DR, DRw53 (MT), or DP (SB) class II HLA molecules. Inhibition studies with monodonal HLA-specific antibodies. **Human Immunology** 1984; **11**: 207-217.
144. Paulsen G, Qvigstad E, Gaudernack G, Rask L, Winchester R, Thorsby E. Identification, at the genomic level, of an HLA-DR restriction element for cloned antigen-specific T4 cells. **Journal of Experimental Medicine** 1985; **161**: 1569-1574.
145. Waters SJ, Winchester RJ, Nagase F, Thorbecke GJ, Bona CA. Antigen presentation by murine and human cells to a murine T-cell hybridoma: demonstration of a restriction element associated with a major histocompatibility complex class II determinant(s) shared by both species. **Proceedings of the National Academy of Sciences USA** 1984; **81**: 7559-7563.
146. Mustafa AS, Deggerdal A, Lundin KE, Meloen RM, Shinnick TM, Oftung F. An HLA-DRw53-restricted T-cell epitope from a novel Mycobacterium leprae protein antigen important to the human memory T-cell repertoire against M. leprae. **Infection & Immunity** 1994; **62**: 5595-5602.
147. Mustafa AS. Identification of mycobacterial peptide epitopes recognized by CD4+ T cells in association with multiple major histocompatibility complex class II molecules. **Nutrition** 1995; **11**: 657-660.
148. Fuller TC, Fuller A. The humoral immune response against an HLA class I allodeterminant cor

[Back to HLA](#) [Back to MHC](#) [Back to Genetics](#) [Back to Evolution](#) [Back to Biostatistics](#)
[Glossary](#) [Homepage](#)

MHC and Leukemia

M. Tefvik Dorak, MD, PhD

The first attempt to correlate a polymorphic system with disease susceptibility was made following the discovery of red blood cell groups. On the assumption that such polymorphisms must be maintained by selection, a search for associations between blood groups and specific diseases was initiated. The first such association was found between blood group A and stomach cancer in 1953¹. With the discovery of the H-2 and HLA systems and typing methods for them, it was inevitable that the MHC would be the subject of extensive disease association studies.

Animal experiments

Mice: Initial studies on susceptibility to mouse leukemia were done without much knowledge of the H-2 complex. Gross showed that when cell-free filtrates from AKR or C58 mice (inbred strains with a very high incidence of spontaneous leukemia) were injected to newborn mice of C3H/Bi or C57BR/cd strains (with low natural leukemia incidence), leukemia developed, whereas, the adult mice were resistant^{2,3}. Interestingly, these four strains (AKR, C58, C57BR, C3H) all had the same MHC haplotype, H-2^{k 4,5}. In 1964, Lilly *et al*, convincingly showed the role of the H-2^k haplotype on the

development of leukemia by inoculating leukemic extracts into mice from segregating generations with different H-2 types⁶. The H-2^k haplotype caused increased susceptibility to both 'virus-induced' and 'spontaneous' leukemia but only in homozygous form. The following studies of Lilly and others continued to show that spontaneous leukemia development was influenced by H-2^k as much as virus-induced leukaemia^{7,8}. In Lilly's studies, the H-2^k homozygotes had a high incidence of spontaneous leukaemia at about the same age (7-11 months) as the peak incidence period in AKR mice, whereas, their heterozygous (H-2^{b/k}) littermates showed a somewhat lower leukaemia incidence occurring gradually in their life span, generally later than 11 months⁷. Other mouse studies revealed that the MHC also influences several other cancers either spontaneously occurring⁹, virally-induced¹⁰ or chemically-induced¹¹.

The influence of homozygosity for the H-2^k haplotype has been repeatedly confirmed in many other studies^{8,12-16}, and one of the several leukaemia susceptibility loci has been mapped to the MHC class II region¹⁵⁻¹⁸. Heterozygosity for the H-2^k haplotype has no effect on leukaemogenesis in mice. Reviews of earlier studies concluded that the studies where the H-2^b haplotype had been included, it seemed to favour resistance to virally-induced oncogenesis including leukaemogenesis¹⁹⁻²¹. The resistance is not absolute and shows itself as later onset rather than no occurrence. Similar to Lilly's studies on mouse leukaemia, a more recent lymphoma model shows this point clearly. In a virally-induced lymphoma model in mice, 64% of animals of susceptible strains (H-2I-A^{k/k}) developed T-cell lymphoma with a mean latency period of 37 weeks. In resistant strains (H-2^{b/b} or H-2^{b/k}), only 14% of the mice developed lymphomas with a longer latency of 57 weeks. The mouse studies concluded that the contribution of the MHC to the development of leukaemia is secondary and mainly on the age at onset in a polygenic and multifactorial context^{12,22,23}. It is also well-established that the interaction of the MHC is not on transformation of a neoplastic clone but occurs during the preleukaemic phase^{10,22}. Although various mechanisms have been postulated for this effect, at present, there is only evidence to support an immune response related one¹³⁻¹⁸. Similar conclusions have been drawn also from other animal models²⁴⁻²⁶. Besides the H-2 type, female gender has been identified as another factor in determination of an effective immune response for murine leukaemia viruses¹⁴.

Rat: Susceptibility to the growth of transplantable gliosarcoma in genetically inbred rats has a correlation with the MHC type²⁷. Curiously, however, this effect is a dominant one. The presence of antibodies to histocompatibility antigens which were those of the strain in which the tumour originated predicts tumour regression which suggests an immune mechanism in operation. More recent studies showed striking correlations between chemical tumour induction and the growth and reproduction complex (grc) of the rat MHC²⁸⁻³². The grc was discovered in 1979 by Gill and Kunz³³ as an MHC-linked gene complex affecting growth and development. It is homologous to the Q/TL region of the mouse, and homozygosity for a 70 kb deletion in this region are associated with embryonic death, developmental defects, and decreased resistance to cancer^{34,35}. The mechanism of increased susceptibility to cancer is unknown. If a similar region exists in humans, it would be at the telomeric end of the class I region^{36,37}.

Cattle: Bovine leukaemia virus (BLV), a retrovirus, is associated with enzootic bovine leukosis, which is the most common neoplastic disease of cattle. Its infection appears as persistent lymphocytosis in one-third of cases and then some cases (1-5% of the total infected) proceed to lymphosarcoma³⁸. The first genetic factor in susceptibility to BLV infection was identified to be the bovine MHC (BoLA). In an early study, particular antigens, namely the class I antigens W6 and Eu28R, were found to be more frequent in cattle with persistent lymphocytosis³⁹. Later molecular studies mapped the susceptibility and resistance to BLV-induced persistent lymphocytosis to the (non-expressed) BoLA-DRB2 and -DRB3 loci in the class II region and showed that the previous class I associations were due to linkage disequilibrium^{38;40;41}. Resistance conferred by (DQA*3A-DQB*3A)-DRB2*2A-DRB3.2*11, *23 or *28 is dominant and almost absolute, whereas, susceptibility is associated with homozygosity for the haplotypes DRB2*1C-DRB3.2*16, *22, or *24; or (BoLA-A14-DQA*12-DQB*12)-DRB2*3A-DRB3.2*8 and not absolute. The map position of susceptibility or resistance has been further refined and located at the amino acid positions 70 and 71 in the second exon of BoLA-DRB3.2 molecule²⁴. All resistance alleles have the amino acids Glu⁷⁰ (E) and Arg⁷¹ (R) in these positions, and the susceptibility alleles do not have either of them. Since these amino acids are involved in peptide binding, it has been concluded that the cellular immune response is important in preventing *in vivo* spread of BLV infection. Indeed, DRB3 or a closely linked gene appear to control the number of BLV-infected peripheral B cells *in vivo*³⁸. The animals with the susceptibility genotypes tend to have a larger number of BLV-infected B lymphocytes. A four amino acid homology between susceptibility associated DRB3 alleles (residues 75 to 78) and BLV *pol* protein led to the speculation that molecular mimicry could be involved in susceptibility²⁴.

Chicken: The MHC of the chicken is traditionally called the B complex and completely sequenced⁴². The class I region (B-F) has been extensively studied because of its association with susceptibility to Marek's disease (MD) which is the most prominent naturally occurring disease of chickens. MD is an herpes virus-induced lymphoproliferative disease. Genetic resistance to Marek's disease was mapped to the B complex after a lucky recombination event⁴³. That first observation showed that the haplotype B21 was resistance and B19 was susceptibility markers. Following studies mostly agreed with the finding that B21 was the resistance haplotype⁴⁴ and also showed that the B2 haplotype had the same effect^{45;46}. Viremia levels at day 5 and 6 postinfection were found to be lower in chickens bearing the B2 haplotype compared to those with the susceptibility haplotypes⁴⁵. The most significant influence of the MHC on Rous sarcomas is on their regression or progression⁴⁷⁻⁵⁰. The interesting feature of the B-F association with this virus-induced malignancy is the dominant nature of the resistance and the interaction of the MHC genotypes with gender, males being the worst affected one in susceptibility⁵¹ and mortality⁵². With the availability of the complete sequence of the chicken B complex, an interesting possibility arose that the presence of the putative NK cell receptor gene(s) may be the reason for the MHC-linked susceptibility to Marek's disease⁴².

Similarly, Rous Sarcoma Virus (RSV) -induced malignancy is also associated with certain B complex genotypes⁴⁹. The B2 haplotype is again the resistance haplotype and

B5 is the susceptibility marker^{50;53;54} although other associations were also reported^{55;56}. The metastatic behaviour of tumours also shows a correlation with the MHC type with a gender effect. Metastasis is less frequent in chickens with the B2 haplotype and this genotype interacts with the sex of the chicken, again maleness being the unfavourable of the two⁵³. The consistent pattern continues for RSV-induced tumours that susceptibility genotypes are homozygous⁵⁷ and resistance appears to be dominant⁵⁶. The mechanism of the MHC association with RSV-induced tumour development and its clinical course is almost certainly immunological^{58;59}, and a possibility of molecular mimicry between a B5 antigen and the virus has been suggested to be the reason for susceptibility^{25;60}.

Avian Leukosis Virus (ALV) causes erythroblastosis in chickens and this malignancy is also associated with the B complex^{50;61}. It is important that the same B2 haplotype is the resistance marker against this virus, although another study found the same effect associated with B21⁶¹. This means that all three virus-induced malignancies of chickens, Marek's disease, Rous sarcoma and erythroblastosis share the same MHC-associated susceptibility (B5, B15) and resistance genotypes (B2, B21)^{50;54;59}. This unusual situation is similar to the H-2^b-associated protection from malignancies in mice and strongly suggests the involvement of an immune response component in this phenomenon.

Rabbits: Regression or malignant conversion to squamous cancer of skin warts induced by papillomavirus show strong correlations with MHC class II alleles by RFLP analysis in rabbits²⁶. This finding suggests an immune system mediated influence of the MHC on virus-induced cancer in another animal.

The consistent finding in animal studies that resistance is dominant and susceptibility is recessive, is in agreement with expectations based on the classical model for immune response gene control⁶². Another unambiguously established conclusion is that all MHC associations, with the notable exception of the grc effect in rats, seem to result from immune responsiveness to an oncogenic virus or tumour antigens. The immune responsiveness may be the result of molecular mimicry in the BLV infection in cattle and Rous sarcoma in chickens. Although not noted in all studies, a gender effect suggesting a stronger effect of the MHC in susceptibility to virus-induced neoplasms in males appears to be consistent.

Human studies

By the beginning of the 1990s, studies on the role of the HLA system in human leukaemia susceptibility reached a stage where a large amount of data was generated and many questions were raised but few firm answers were available. The field of HLA-disease associations, in general, has been characterised by a combination of a large number of reports, unknown mechanisms and unconsidered implications. Unfortunately, the literature is now saturated with unconfirmed association reports most of which were spurious associations as would be expected from the analysis of such a polymorphic system. As early as 1972, Walter Bodmer had pointed out that "the search for associations of red blood cell groups with diseases has been a most frustrating one producing inconsistencies and statistical pitfalls. The problems of these studies should be

a warning to investigators looking for associations between the HLA-A polymorphism and diseases" ⁶³. This warning did not have a great effect. Today, similar warnings are issued for pharmacogenetic studies quoting the current state of HLA and disease association studies ⁶⁴. There is still no standardisation in the methodology of HLA and disease association studies and statistical analysis of such data largely depends on arbitrary choices of methods. The statistical aspects of HLA and disease associations are discussed in the statistics section.

HLA association studies: The first HLA study on human leukaemia found an increased frequency of HLA-A2 in ALL in 1967 ⁶⁵. In the same year, another HLA disease association was reported also with a haemopoietic malignancy; this time between the HLA-B broad specificity 4c and Hodgkin's disease ⁶⁶. No doubt that the initial studies concentrated on leukaemias because of the finding of the first MHC association in mouse leukaemia in 1964 ⁶. A second study in ALL extended the first HLA-A2 association to HLA-A2B12 haplotypal association in 1970 ⁶⁷. Although not confirmed in all studies ⁶⁸, the HLA-A2 association is one of the few leukaemia associations noted in more than one study and in fact, a significant association in pooled data ⁶⁹.

A number of other studies have investigated HLA associations in childhood and adult leukaemias some of which included the analysis of HLA-DR/DQ antigens or genes ⁷⁰⁻⁸⁹. Most of these studies examined large number of patients and controls and employed the best available typing techniques of the time. It was, however, hardly the case that two studies ever showed the same result. The largest HLA association study in leukaemia was carried out on the International Bone Marrow Transplant Registry data ^{77;78}. These studies analysed a total of 1,834 patients with ALL, AML, and CML treated between 1969 and 1985. These studies showed that HLA-Cw3 and -Cw4 are both susceptibility markers for all of the three major leukaemias. It appears that the very large sample helped to get this result which yielded RRs no greater than 2.49. It is not easy to explain that two different alleles of the same locus are associated with increased susceptibility to the same disease. To complicate the issue further, a following single-centre study showed an association also with HLA-Cw7 in ALL ⁷⁹. Such a situation may arise if these HLA-C alleles share a broad specificity or a common epitope such as those recognised by NK cells. In the molecular era, an RFLP study suggested that HLA-Cw4 association with CML might be due to its linkage disequilibrium with BF*F which is the main association ⁸⁵. A recent PCR study in childhood ALL failed to show any association with HLA-C alleles, and more specifically with the two sequence variants (NK1 and NK2) which are ligands for NK cell receptors ⁸⁹. It appears that the reported HLA-C associations may be the consequence of a large sample effect and expression differences between leukaemic and healthy cells. Confirmed and consistent HLA associations found in leukaemias are shown in Table 1.

While HLA class I studies showed associations with HLA-A2(B12) and several HLA-C alleles, MHC class III loci was the subject of only one study in childhood ALL ⁹⁰. A serological North American study of Factor B phenotypes revealed a significant association with BF*F (RR = 3.62; p<0.0001). This finding has not been confirmed in another study in childhood ALL. The only other MHC class III study in leukaemia was

published by Dorak *et al.*⁸⁵. This RFLP study in CML showed a BF*Fb association which was restricted to the patients below the median age (RR=3.1; p<0.01). This allele also had a significantly higher frequency in the male patients below median age compared to the older patients (p=0.014). This study was the first to suggest that MHC associations in leukaemia may differ in genders as in some animal studies showed^{14;51-53}.

The first HLA-DR studies in childhood ALL showed an increase for HLA-DR7⁷⁰⁻⁷². In one study, adult ALL showed a weak association with HLA-DR4⁸⁰ and in another CLL was associated with HLA-DR53⁸¹. The strongest association in leukaemia reported to date has been presented by Seremetis *et al.*⁸². They used a monoclonal antibody specific for the HVR3 epitope of HLA-DR53 and found a RR of 7.88 ($P<0.000005$) for AML. Also, in the only molecular studies examined the HLA-DRB loci, directly or indirectly, in ALL, CML and CLL, the homozygous genotype for HLA-DRB4*01 (-DR53) had an increased frequency in patients^{85-87;91}. In childhood ALL, this effect was observed only in male patients^{86;91}. It thus appears that in the HLA-DR region, the HLA-DRB4 locus is the susceptibility marker for all major leukaemias (AML, ALL, CML, CLL) examined so far. Despite being a consistent association in the recent studies, this association has not been noted before as strongly as it is^{70-72;80;81} or has been missed completely^{73-76;78;79;83;84;88;92}. The explanation may be that none of the previous studies examined homozygosity, supertypes and gender effect simultaneously.

A possibility has been raised that the then putative HFE gene could be a susceptibility gene for childhood leukaemias⁹³. Dorak *et al.* noted the similarities between the HLA-A3-related susceptibility genotypes for hereditary haemochromatosis (HH) and childhood ALL (as well as CML). Considering some epidemiological links between HH and cancer, the genetic mapping of an HLA class I-like gene (TCA) associated with leukaemias to the area where HFE was thought to map, the possibility of a physical interaction between the TCA protein and the transferrin receptor (TfR), they proposed that HFE would be a HLA class I-like gene, map to the area telomeric to HLA-A, and be relevant in susceptibility to childhood ALL. Since the HFE gene is now found about 5 cM away from the HLA-A gene⁹⁴, it has been possible to test this hypothesis. Indeed, it has been shown in two different populations that the C282Y mutation of the HFE gene is a susceptibility marker for childhood ALL but only in males⁹⁵.

Other studies regarding the HLA system: In addition to HLA associations in leukaemias, patients with leukaemia have an increase in the frequency of HLA-identical siblings^{85;92;96-98}, in overall homozygosity for HLA antigens^{98;99}, and in HLA identity with their mothers where parents share one DR antigen⁹⁸. Parental HLA sharing is also increased in leukaemic families which is most significant for the HLA-DR antigens^{92;98-102}.

Chan *et al* found that 35% of unaffected siblings have the same HLA genotypes as the leukaemic patient although 25% was the expected frequency⁹². Dorak *et al* reported the same in CML but only in the patients with a young-onset⁸⁵. Patients below the median age (number of siblings = 97) had a 37.7% HLA-identical sibling frequency whereas it was 24.7% for older patients (number of siblings = 81). In another study, HLA genotypes

were established in patients with AML and in their first-degree relatives ⁹⁸. Besides the excess of identical antigens (especially HLA-DR) between parents, there was a genetic distortion favouring HLA homozygosity in the affected offspring. The distribution of HLA-DR antigens shared by the parents markedly favoured phenotypic HLA-DR identity of patients with their mothers as compared to fathers. This can be possible only if the paternally shared (and disease-associated) haplotype has a high transmission ratio relative to its homologue haplotype. This suggestion was used by Dorak & Burnett as one of the elements in their hypothesis that a *t-like* HLA haplotype may be relevant in susceptibility to leukaemia ¹⁰³. The mouse *t* haplotypes contain recessive deleterious genes (or embryonic lethals) most of which are within the MHC ^{104;105}, and high transmission rates due to several transmission ratio distorter (tcd) genes the strongest of which (tcd-2) is again within the MHC or very close to it ¹⁰⁶⁻¹⁰⁸.

Von Fliedner *et al* reported the behaviour of HLA-DR antigens in the families of 55 children with ALL ⁹⁹. Their findings were: (i) increased identity at HLA-B and -DR loci between parents; (ii) twice as many as expected homozygotes among the patients where parents shared HLA antigens, and (iii) significantly more heterozygotes for the shared antigens among healthy siblings. Like the above-mentioned ones, these results point out high transmission ratios for HLA-associated HLA haplotypes too. The same researchers also reported an HLA-DR7 association in childhood ALL with an increased homozygosity rate for it ⁷¹. Neither in their studies nor in any other study, transmission ratios of particular antigens were studied in leukaemia.

The first study examined the transmission ratios of HLA haplotypes was by Albert *et al* ¹⁰⁹. That study included 535 families and concluded that although unexpected, the HLA-A2B12 haplotype had a high transmission ratio (60.0%). A more recent but much smaller study provided some evidence that this initial finding may hold in a different population ^{110;111}. Komlos *et al* reported that the HLA-A2Bw4 haplotype had a higher transmission ratio when inherited from fathers. In another small study, the haplotype HLA-B44DR4 showed a high transmission ratio (65.8%) in healthy families ⁸⁵. It appears that the commonest HLA-DR53 group haplotype or its parts may indeed show segregation distortion in a *t-like* manner. This issue is far from being concluded but it is certainly feasible to pursue it further.

The overall interpretation of these observations is that HLA associated leukaemia susceptibility is a recessive trait and most probably HLA-DR-related. The recessive nature of MHC and leukaemia / cancer associations has also been shown in animal studies in mice, chickens and cattle as outlined above. Another interpretation is that (paternal) transmission ratio for the leukaemia-associated HLA haplotypes is increased which results in the violation of Mendelian expectations in HLA-identical sibling frequency. This would also cause the increased maternal identity where parents share an HLA-DR antigen ^{85;103}. There is some suggestion that the haplotypes of the main leukaemia susceptibility allele, HLA-DR53, may show segregation distortion which would result in increased homozygosity in patients.

Finally, an important but not well-recognised study investigated the correlation between HLA class I antigen frequencies and cancer incidences (measured as mortality rates) in 26 different populations¹¹². The most striking correlation was found between HLA-B8 antigen frequencies and intestinal cancer mortality rates in most of the populations examined. It is important to note that Scotland has the highest correlation while England and Wales also show a very high correlation. This study suggests that in general the HLA system may be relevant in cancer susceptibility and this varies among different populations. Unfortunately, this study has not been repeated for HLA class II antigens.

HLA expression in malignancies: Although genetically determined polymorphism in the MHC genes is relevant in determination of susceptibility to leukaemia and other cancers, there is another role played by the MHC molecules in cancer. In cancers, MHC expression may be altered. This may be an increase, aberrant expression, down-regulation or abrogation. The functional significance of these changes varies¹¹³. Cancer cells that present tumour antigens to the immune effector cells may elicit antitumour responses. In the absence of a co-stimulatory signal, engagement of the TCR by MHC class II and peptide complex may lead to selective anergy of tumour-specific clones. Alternatively, expression of class II antigens may promote tumour growth as it may recruit T cells that deliver stimulatory cytokines. Such changes usually occur after the transformation and may be influential in the determination of the fate of cancer cells. The correlation of expression of HLA-DR and -DQ molecules with favourable prognosis in breast cancer¹¹⁴⁻¹¹⁶ and cervical cancer¹¹⁷ has been reported.

Changes in HLA class I expression patterns have also been noted in a variety of tumours^{118;119}. HLA class I alterations may occur at a particular step between the development of an *in situ* lesion and an invasive carcinoma¹¹⁸. The loss of class I expression results in the loss of immunosurveillance leading to invasion and/or metastasis in colon cancer¹²⁰, cervical cancer^{117;121;122}, breast cancer^{114;123} and malignant melanoma¹²⁴. It is a consistent finding that while HLA class II expression may generally be a favourable sign, loss of HLA class I expression is a poor prognostic sign. An interesting feature of down-regulation of HLA expression in tumours is its association with oncogene activation^{125;126}.

Table 1. HLA system in Leukemia

1. Allelic associations
HLA-A2; HLA-Cw3, -Cw4; HLA-DR53
2. Increased parental HLA-DR sharing; increased maternal HLA-DR identity in such families
3. Increased homozygosity for HLA-DR antigens
4. Increased HLA-identical sibling frequency
5. Aberrant expression or loss of expression
6. Increased miscarriage frequency in the mothers of leukaemia patients (relationship to the HLA system has not been investigated)

(see also [HLA-DR53 fact file](#))

References

1. Aird I, Bentall HH, Roberts JAF. A relationship between cancer of stomach and the ABO blood groups. **British Medical Journal** 1953; **1**: 799-801.
2. Gross L. Susceptibility to suckling-infant, and resistance of adult, mice of the C3H and of the C57 lines to inoculation with AK leukemia. **Cancer** 1950; 1073-1087.
3. Gross L. Viral (egg-borne) etiology of mouse leukemia. **Cancer** 1956; **9**: 778-791.
4. Gorer PA. Some recent work on tumour immunity. **Advances in Cancer Research** 1956; **4** : 149-186.
5. Gross L. Viral etiology of mouse leukemia. **Advances in Cancer Research** 1961; **6**: 149-180.
6. Lilly F, Boyse EA, Old LJ. Genetic basis of susceptibility to viral leukaemogenesis. **Lancet** 1964; **ii**: 1207-1209.
7. Lilly F. The inheritance of susceptibility to the Gross leukemia virus in mice. **Genetics** 1966; **53**: 529-539.
8. Boyse EA, Old LJ, Stockert E. The relation of linkage group IX to leukemogenesis in the mouse. In: Emmelot P, Bentvelzen P, eds. RNA Viruses and Host Genome in Oncogenesis, Amsterdam: North Holland Publishers Co., 1972: 171-185.
9. Faraldo MJ, Dux A, Muhlbock O, Hart G. Histocompatibility genes (the H-2 complex) and susceptibility to spontaneous lung tumors in mice. **Immunogenetics** 1979; **9**: 383-404.
10. Chesebro B. Influence of the major histocompatibility complex (H-2) on oncomavirus-induced neoplasia in mice. In: Kaiser HE, ed. Neoplasms - comparative pathology of growth in animals, plants, and man, Baltimore: Williams and Wilkins, 1981: 475-482.
11. Oomen LC, Van der Valk MA, Hart AA, Demant P, Emmelot P. Influence of mouse major histocompatibility complex (H-2) on N-ethyl-N-nitrosourea-induced tumor formation in various organs. **Cancer Research** 1988; **48**: 6634-6641.
12. Lilly F, Pincus T. Genetic control of murine viral leukemogenesis. **Advances in Cancer Research** 1973; **17**: 231-277.
13. Meruelo D, McDevitt HO. Recent studies on the role of the immune response in resistance to virus-induced leukemias and lymphomas [Review]. **Seminars in Hematology** 1978; **15**: 399-419.
14. Nowinski RC, Brown M, Doyle T, Prentice RL. Genetic and viral factors influencing the development of spontaneous leukemia in AKR mice. **Virology** 1979; **96**: 186-204.
15. Vlug A, Schoenmakers HJ, Melief CJ. Genes of the H-2 complex regulate the antibody response to murine leukemia virus. **Journal of Immunology** 1981; **126**: 2355-2360.
16. Vasmel WL, Zijlstra M, Radaskiewicz T, Leupers CJ, de Goede RE, Melief CJ. Major histocompatibility complex class II-regulated immunity to murine leukemia virus protects against early T- but not late B- cell lymphomas. **Journal of Virology** 1988; **62**: 3156-3166.
17. Lonai P, Haran Ghera N. Resistance genes to murine leukemia in the I immune response gene region of the H-2 complex. **Journal of Experimental Medicine** 1977; **146**: 1164-1168.
18. Miyazawa M, Nishio J, Chesebro B. Genetic control of T cell responsiveness to the Friend murine leukemia virus envelope antigen. Identification of class II loci of the H-2 as immune response genes. **Journal of Experimental Medicine** 1988; **168**: 1587-1605.
19. Snell GD. The H-2 locus of the mouse: observations and speculations concerning its comparative genetics and its polymorphism. **Folia Biologica (Praha)** 1968; **14**: 335-358.
20. Zijlstra M, Vasmel WL, Radaskiewicz T, Matthews E, Melief CJ. The H-2 complex regulates both the susceptibility to mouse viral lymphomagenesis and the phenotype of the virus-induced lymphomas. [Review]. **Journal of Immunogenetics** 1986; **13**: 69-76.
21. Demant P, Oomen LC, Oudshoorn Snoek M. Genetics of tumor susceptibility in the mouse: MHC and non-MHC genes. **Advances in Cancer Research** 1989; **53**: 117-179.

22. Lonai P, Katz E, Haran Ghera N. Role of the major histocompatibility complex in resistance to viral leukemia; its effect on the preleukemic stage of leukemogenesis. **Springer Seminars in Immunopathology** 1982; **4**: 373-396.
23. Zijlstra M, Melief CJ. Virology, genetics and immunology of murine lymphomagenesis. [Review]. **Biochimica et Biophysica Acta** 1986; **865**: 197-231.
24. Xu A, van Eijk MJ, Park C, Lewin HA. Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. **Journal of Immunology** 1993; **151**: 6977-6985.
25. Heinzelmann EW, Zsigray RM, Collins WM. Cross-reactivity between RSV-induced tumor antigen and B5 MHC alloantigen in the chicken. **Immunogenetics** 1981; **13**: 29-37.
26. Han R, Breitburd F, Marche PN, Orth G. Linkage of regression and malignant conversion of rabbit viral papillomas to MHC class II genes. **Nature** 1992; **356**: 66-68.
27. Albright AL, Gill TJI, Geyer SJ. Immunogenetic control of brain tumor growth in rats. **Cancer Research** 1977; **37**: 2512-2521.
28. Rao KN, Shinozuka H, Kunz HW, Gill TJI. Enhanced susceptibility to a chemical carcinogen in rats carrying MHC-linked genes influencing development (GRC). **International Journal of Cancer** 1984; **34**: 113-120.
29. Melhem MF, Rao KN, Kunz HW, Kazanecki M, Gill TJI. Genetic control of susceptibility to diethylnitrosamine carcinogenesis in inbred ACP (grc+) and R16 (grc) rats. **Cancer Research** 1989; **49**: 6813-6821.
30. Melhem MF, Kunz HW, Gill TJI. Genetic control of susceptibility to diethylnitrosamine and dimethylbenzanthracene carcinogenesis in rats. **American Journal of Pathology** 1991; **139**: 45-51.
31. Melhem MF, Kunz HW, Gill TJI. A major histocompatibility complex-linked locus in the rat critically influences resistance to diethylnitrosamine carcinogenesis. **Proceedings of the National Academy of Sciences U S A** 1993; **90**: 1967-1971.
32. Lu D, Kunz HW, Melhem MF, Gill TJI. Cell lines from grc congenic strains of rats having different susceptibilities to chemical carcinogens. **Cancer Research** 1993; **53**: 4089-4095.
33. Gill TJI, Kunz HW. Gene complex controlling growth and fertility linked to the major histocompatibility complex in the rat. **American Journal of Pathology** 1979; **96**: 185-206.
34. Gill TJI. The borderland of embryogenesis and carcinogenesis. Major histocompatibility complex-linked genes affecting development and their possible relationship to the development of cancer. **Biochimica et Biophysica Acta** 1984; **738**: 93-102.
35. Gill TJI. Role of the major histocompatibility complex region in reproduction, cancer, and autoimmunity. **American Journal of Reproductive Immunology** 1996; **35**: 211-215.
36. Le Bouteiller P. HLA class I chromosomal region, genes, and products: facts and questions. **Critical Reviews in Immunology** 1994; **14**: 89-129.
37. Gill TJI, Natori T, Salgar SK, Kunz HW. Current status of the major histocompatibility complex in the rat. **Transplantation Proceedings** 1995; **27**: 1495-1500.
38. Mirsky ML, Olmstead C, Da Y, Lewin HA. Reduced bovine leukaemia virus proviral load in genetically resistant cattle. **Animal Genetics** 1998; **29**: 245-252.
39. Stear MJ, Dimmock CK, Newman MJ, Nicholas FW. BoLA antigens are associated with increased frequency of persistent lymphocytosis in bovine leukaemia virus infected cattle and with increased incidence of antibodies to bovine leukaemia virus. **Animal Genetics** 1988; **19**: 151-158.
40. van Eijk MJ, Stewart-Haynes JA, Beaver JE, Fernando RL, Lewin HA. Development of persistent lymphocytosis in cattle is closely associated with DRB2. **Immunogenetics** 1992; **37**: 64-68.
41. Zanotti M, Poli G, Ponti W, et al. Association of BoLA class II haplotypes with subclinical progression of bovine leukaemia virus infection in Holstein-Friesian cattle. **Animal Genetics** 1996; **27**: 337-341.

42. Kaufman J, Milne S, Gobel TWF, et al. The chicken B locus is a minimal essential major histocompatibility complex. **Nature** 1999; **401**: 923-925.
43. Briles WE, Briles RW, Taffs RE. Resistance to a malignant disease in chickens is mapped to a subregion of major histocompatibility (B) complex. **Science** 1983; **219**: 977-979.
44. Hepkema BG, Blankert JJ, Albers GA, et al. Mapping of susceptibility to Marek's disease within the major histocompatibility (B) complex by refined typing of White Leghorn chickens. **Animal Genetics** 1993; **24**: 283-287.
45. Schat KA, Taylor RL, Jr., Briles WE. Resistance to Marek's disease in chickens with recombinant haplotypes to the major histocompatibility (B) complex. **Poultry Science** 1994; **73**: 502-508.
46. Sato K, Abplanalp H, Napolitano D, Reid J. Effects of heterozygosity of major histocompatibility complex haplotypes on performance of Leghorn hens sharing a common inbred background. **Poultry Science** 1992; **71**: 18-26.
47. Collins WM, Briles WE, Zsigray RM, et al. The B locus (MHC) in the chicken: association with the fate of RSV-induced tumors. **Immunogenetics** 1977; **5**: 333-343.
48. Plach J, Jurajda V, Benda V. Resistance to Marek's disease is controlled by a gene within the B-F region of the chicken major histocompatibility complex in Rous sarcoma regressor or progressor inbred lines of chickens. **Folia Biologica (Praha)** 1984; **30**: 251-258.
49. Plach J, Benda V. Location of the gene responsible for Rous sarcoma regression in the B-F region of the B complex (MHC) of the chicken. **Folia Biologica (Praha)** 1981; **27**: 363-368.
50. Bacon LD, Crittenden LB, Witter RL, Fadly A, Motta J. B5 and B15 associated with progressive Marek's disease, Rous sarcoma, and avian leukosis virus-induced tumors in inbred 154 chickens. **Poultry Science** 1983; **62**: 573-578.
51. Martin A, Dunnington EA, Briles WE, Briles RW, Siegel PB. Marek's disease and major histocompatibility complex haplotypes in chickens selected for high or low antibody response. **Animal Genetics** 1989; **20**: 407-414.
52. Pinard MH, Janss LL, Maatman R, Noordhuizen JP, van der Zijpp AJ. Effect of divergent selection for immune responsiveness and of major histocompatibility complex on resistance to Marek's disease in chickens. **Poultry Science** 1993; **72**: 391-402.
53. Collins WM, Dunlop WR, Zsigray RM, Briles RW, Fite RW. Metastasis of Rous sarcoma tumors in chickens is influenced by the major histocompatibility (B) complex and sex. **Poultry Science** 1986; **65**: 1642-1648.
54. Bacon LD. Influence of the major histocompatibility complex on disease resistance and productivity. **Poultry Science** 1987; **66**: 802-811.
55. Collins WM, Brown DW, Ward PH, Dunlop WR, Briles WE. MHC and non-MHC genetic influences on Rous sarcoma metastasis in chickens. **Immunogenetics** 1985; **22**: 315-321.
56. Collins WM, Zervas NP, Urban WE, Jr., Briles WE, Aeed PA. Response of B complex haplotypes B22, B24, and B26 to Rous sarcomas. **Poultry Science** 1985; **64**: 2017-2019.
57. Brown DW, Collins WM, Ward PH, Briles WE. Complementation of major histocompatibility haplotypes in regression of Rous sarcoma virus-induced tumors in noninbred chickens. **Poultry Science** 1982; **61**: 409-413.
58. Schierman LW, Collins WM. Influence of the major histocompatibility complex on tumor regression and immunity in chickens. **Poultry Science** 1987; **66**: 812-818.
59. Plachy J, Pink JR, Hala K. Biology of the chicken MHC (B complex). **Critical Reviews in Immunology** 1992; **12**: 47-79.
60. Heinzelmann EW, Zsigray RM, Collins WM. Increased growth of RSV-induced tumors in chickens partially tolerant to MHC alloantigens. **Immunogenetics** 1981; **12**: 275-284.
61. Yoo BH, Sheldon BL. Association of the major histocompatibility complex with avian leukosis virus infection in chickens. **British Poultry Science** 1992; **33**: 613-620.
62. Benacerraf B, McDevitt HO. Histocompatibility-linked immune response genes. **Science** 1972; **175**: 273-279.

63. Bodmer WF. Evolutionary significance of the HLA system. **Nature** 1972; **237**: 139-145.
64. Cuzick J. Molecular epidemiology: carcinogens, DNA adducts, and cancer - still a long way to go. **Journal of the National Cancer Institute** 1995; **87**: 861-862.
65. Dausset J. The major histocompatibility complex in man. **Science** 1981; **213**: 1469-1474.
66. Amiel JL. Study of the leukocyte phenotypes in Hodgkin's disease. In: Curtioni ES, Mattiuz PL, Tosi RM, eds. *Histocompatibility Testing 1967*, Copenhagen: Munksgaard, 1967: 79-81.
67. Walford RL, Finkelstein S, Neerhout R, Konrad P, Shanbrom E. Acute childhood leukaemia in relation to the HL-A human transplantation genes. **Nature** 1970; **5231**: 461-462.
68. Albert ED, Nisperos B, Thomas ED. HLA antigens and haplotypes in acute leukemia. **Leukemia Research** 1977; **1**: 261-269.
69. Tiwari JL, Terasaki PI. *HLA and Disease Associations*. New York: Springer-Verlag, 1985;
70. de Moerloose P, Chardonens X, Vassalli P, Jeannet M. [HL-A D antigens from B-lymphocytes and susceptibility to certain diseases]. **Schweizerische Medizinische Wochenschrift - Journal Suisse De Medecine** 1977; **107**: 1461-1461.
71. Von Fliedner VE, Sultan-Khan Z, Jeannet M. HLA-DRw antigens associated with acute leukemia. **Tissue Antigens** 1980; **16**: 399-404.
72. Casper JT, Duquesnoy RJ, Borella L. Transient appearance of HLA-DRw-positive leukocytes in peripheral blood after cessation of antileukemia therapy. **Transplantation Proceedings** 1980; **12**: 130-133.
73. de Jongh BM, van der Dose-van den Berg A, Schreuder GM. Random HLA-DR distribution in children with acute lymphocytic leukaemia in long-term continuous remission. **British Journal of Haematology** 1982; **52**: 161-163.
74. Caruso C, Cammarata G, Sireci G, Modica MA. HLA-Cw4 association with acute lymphoblastic leukaemia in Sicilian patients. **Vox Sang** 1988; **54**: 57-58.
75. Orgad S, Cohen IJ, Neumann Y, et al. HLA-A11 is associated with poor prognosis in childhood acute lymphoblastic leukemia (ALL). **Leukemia** 1988; **2**: 79S-87S.
76. Michel K, Hubbel C, Dock NL, Davey FR. Correlation of HLA-DRw3 with childhood acute lymphocytic leukemia [letter]. **Archives of Pathology & Laboratory Medicine** 1981; **105**: 560-560.
77. D'Amaro J, Bach FH, van Rood JJ, Rimm AA, Bortin MM. HLA C associations with acute leukaemia. **Lancet** 1984; **2**: 1176-1178.
78. Bortin MM, D'Amaro J, Bach FH, Rimm AA, van Rood JJ. HLA associations with leukemia. **Blood** 1987; **70**: 227-232.
79. Muller CA, Hasmann R, Grosse Wilde H, et al. Significant association of acute lymphoblastic leukemia with HLA- Cw7. **Genetic Epidemiology** 1988; **5**: 453-461.
80. Navarrete C, Alonso A, Awad J, et al. HLA class I and class II antigen associations in acute leukaemias. **Journal of Immunogenetics** 1986; **13**: 77-84.
81. Dyer PA, Ridway JC, Flanagan NG. HLA-A,B and DR antigens in chronic lymphocytic leukaemia. **Disease Markers** 1986; **4**: 231-237.
82. Seremetis S, Cuttner J, Winchester R. Definition of a possible genetic basis for susceptibility to acute myelogenous leukemia associated with the presence of a polymorphic Ia epitope. **Journal of Clinical Investigation** 1985; **76**: 1391-1397.
83. Caruso C, Lo Campo P, Botindari C, Modica MA. HLA antigens in Sicilian patients affected by chronic myelogenous leukaemia. **Journal of Immunogenetics** 1987; **14**: 295-299.
84. Linet MS, Bias WB, Dorgan JF, McCaffrey LD, Humphrey RL. HLA antigens in chronic lymphocytic leukemia. **Tissue Antigens** 1988; **31**: 71-78.
85. Dorak MT, Chalmers EA, Gaffney D, et al. Human major histocompatibility complex contains several leukemia susceptibility genes. **Leukemia & Lymphoma** 1994; **12**: 211-222.
86. Dorak MT, Owen G, Galbraith I, et al. Nature of HLA-associated predisposition to childhood acute lymphoblastic leukemia. **Leukemia** 1995; **9**: 875-878.

87. Dorak MT, Machulla HK, Hentschel M, Mills KI, Langner J, Burnett AK. Influence of the major histocompatibility complex on age at onset of chronic lymphoid leukaemia. **International Journal of Cancer** 1996; **65**: 134-139.
88. Dearden SP, Taylor GM, Gokhale DA, et al. Molecular analysis of HLA-DQB1 alleles in childhood common acute lymphoblastic leukaemia. **British Journal of Cancer** 1996; **73**: 603-609.
89. Ghodsi K, Taylor GM, Gokhale DA, et al. Lack of association between childhood common acute lymphoblastic leukaemia and an HLA-C locus dimorphism influencing the specificity of natural killer cells. **British Journal of Haematology** 1998; **102**: 1279-1283.
90. Budowle B, Acton RT, Barger BO, et al. Properdin factor B and acute lymphocytic leukemia (ALL). **Cancer** 1982; **50**: 2369-2371.
91. Dorak MT, Lawson T, Machulla HKG, Darke C, Mills KI, Burnett AK. Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. **Blood** 1999; **94**: 694-700.
92. Chan KW, Pollack MS, Braun D, Jr., O'Reilly RJ, Dupont B. Distribution of HLA genotypes in families of patients with acute leukemia. Implications for transplantation. **Transplantation** 1982; **33**: 613-615.
93. Dorak MT, Burnett AK, Worwood M. Thymus-leukaemia antigens: the haemochromatosis gene product? **Immunology & Cellular Biology** 1994; **72**: 435-439.
94. Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. **Nature Genetics** 1996; **13**: 399-408.
95. Dorak MT, Sproul AM, Gibson BE, Burnett AK, Worwood M. The C282Y mutation of *HFE* is another male-specific risk factor for childhood ALL [letter]. **Blood** 1999 [in press].
96. De Moor P, Louwagie A. Distribution of HLA genotypes in sibs of patients with acute leukaemia. **Scandinavian Journal of Haematology** 1985; **34**: 68-70.
97. De Moor P. Distribution of HLA genotypes in sibs of patients with acute lymphoblastic leukemia. **European Journal of Haematology** 1989; **42**: 317-318.
98. Carpentier NA, Jeannet M. Increased HLA-DR compatibility between patients with acute myeloid leukemia and their parents: implication for bone marrow transplantation. **Transplantation Proceedings** 1987; **19**: 2644-2645.
99. Von Fliedner VE, Merica H, Jeannet M, et al. Evidence for HLA-