Basic Concepts of Human Genetics

- The genetic information of an individual is contained in 23 pairs of chromosomes.

  Every human cell contains the 23 pair of chromosomes.

- One pair is called sex chromosomes

  Male: XY

  Female: XX

- Other 22 pairs of homologous chromosomes are called autosomes.

- The autosome chromosome pairs are called homologous pair. Two chromosomes in the same pair are called homologous chromosomes.

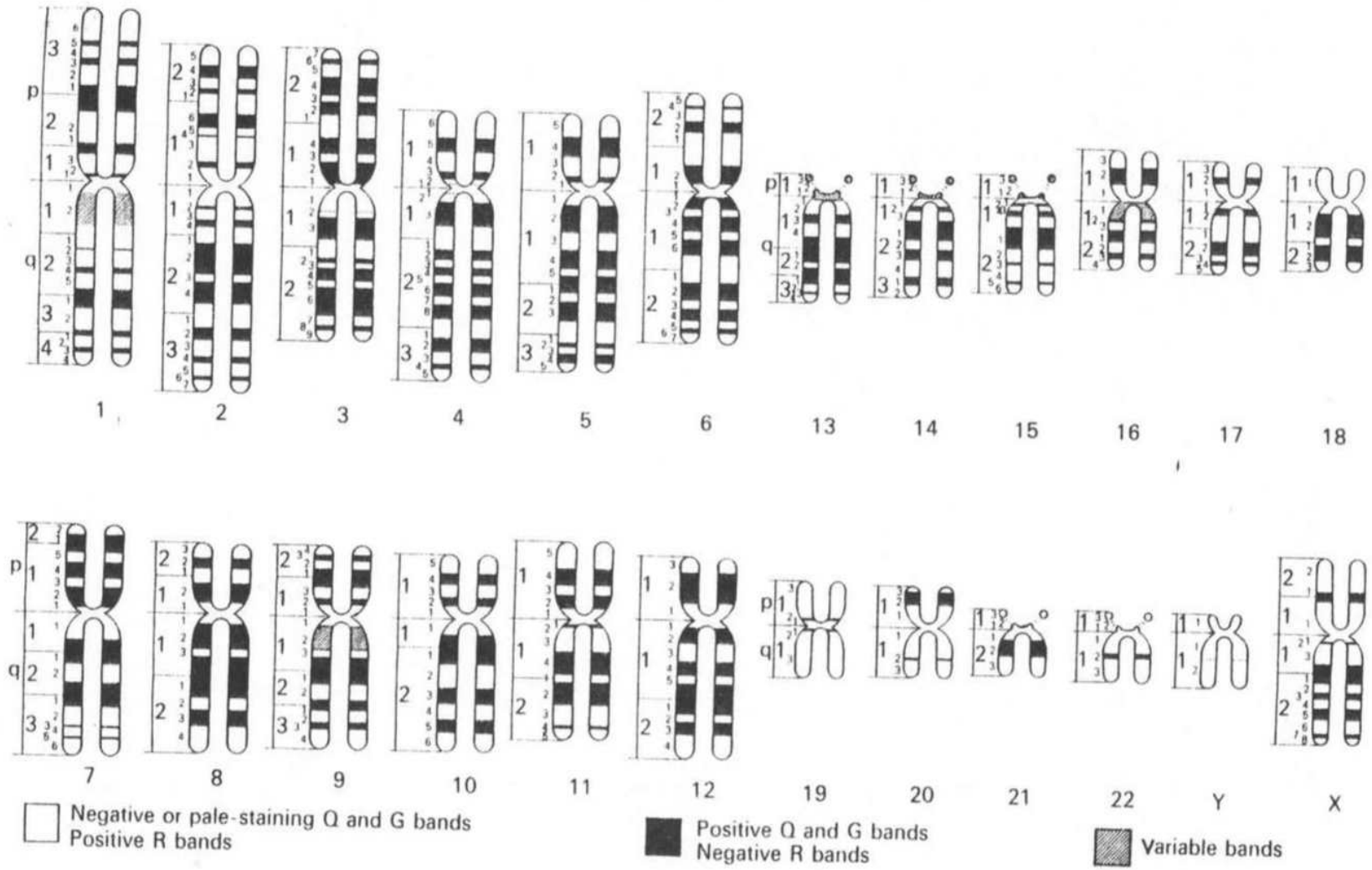
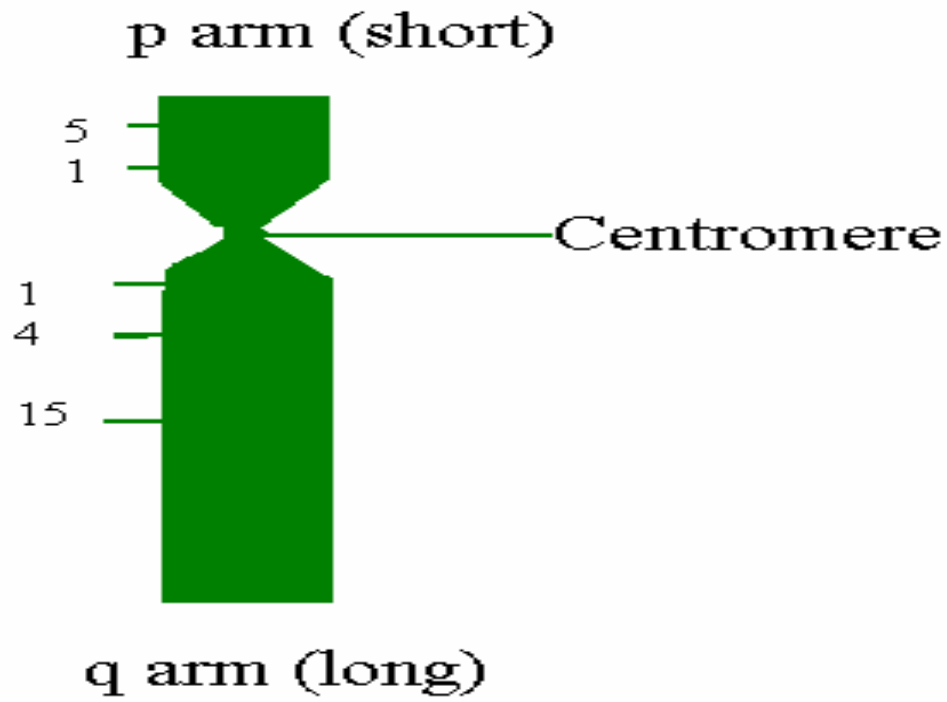
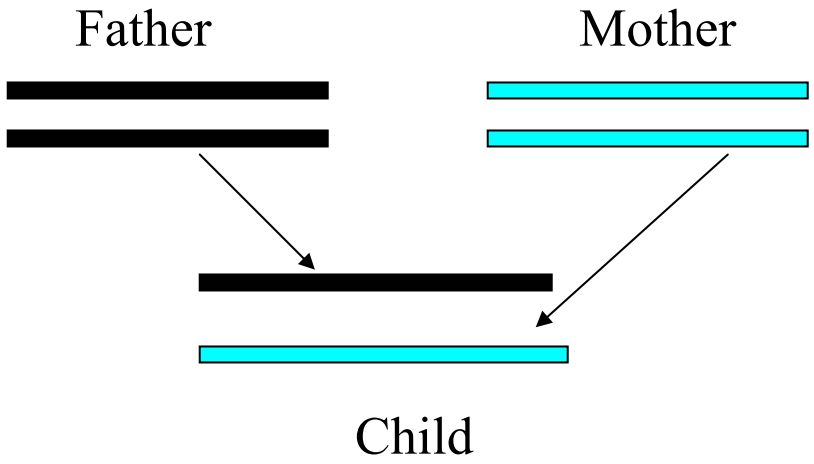


Figure 1.1. Human karyotype, showing a standard numbering system based on chemical staining of a chromosome preparation. Different staining methods yield different levels of detail that are used in refining the system (not shown). (From Thompson and Thompson, 1986.)

• One member of each chromosome pair is from mother; the other is from father. Father or mother transmits each of the two chromosomes with equal probability.

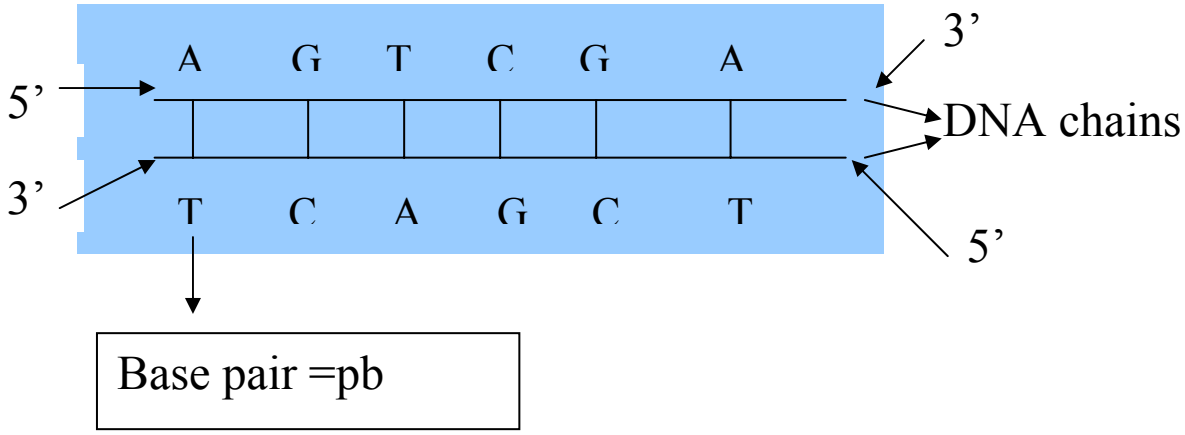




• The location of large scale at chromosome usually use the symbols like 10q5.6 (means at the long arm of chromosome 10, band 5.6) or 5p8.7 etc.

- There are two DNA chains in one chromosome

— DNA has four bases A, G, T and C. A combined with T and G combined with C



— bp is also used as length unit of chromosome or DNA sequence

— DNA sequence has direction. There are two sides (ends) called 5’ side and 3’ side.

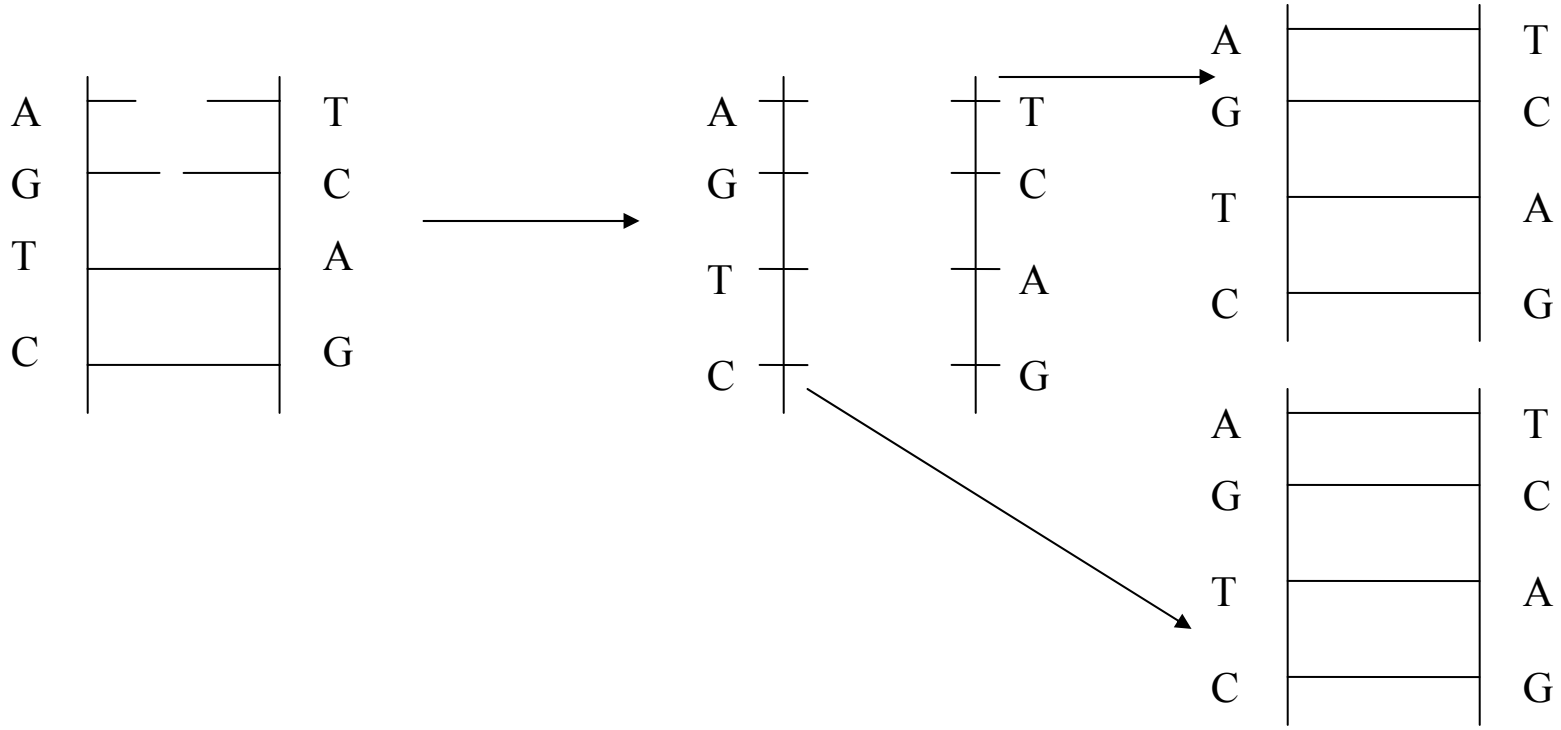
— The homologous chromosomes (chromosome 1, for example) have exactly same length for every individual.

Normal cell division—Mitosis

Old cells died and new cells grow. The new cells grow through normal cell division.

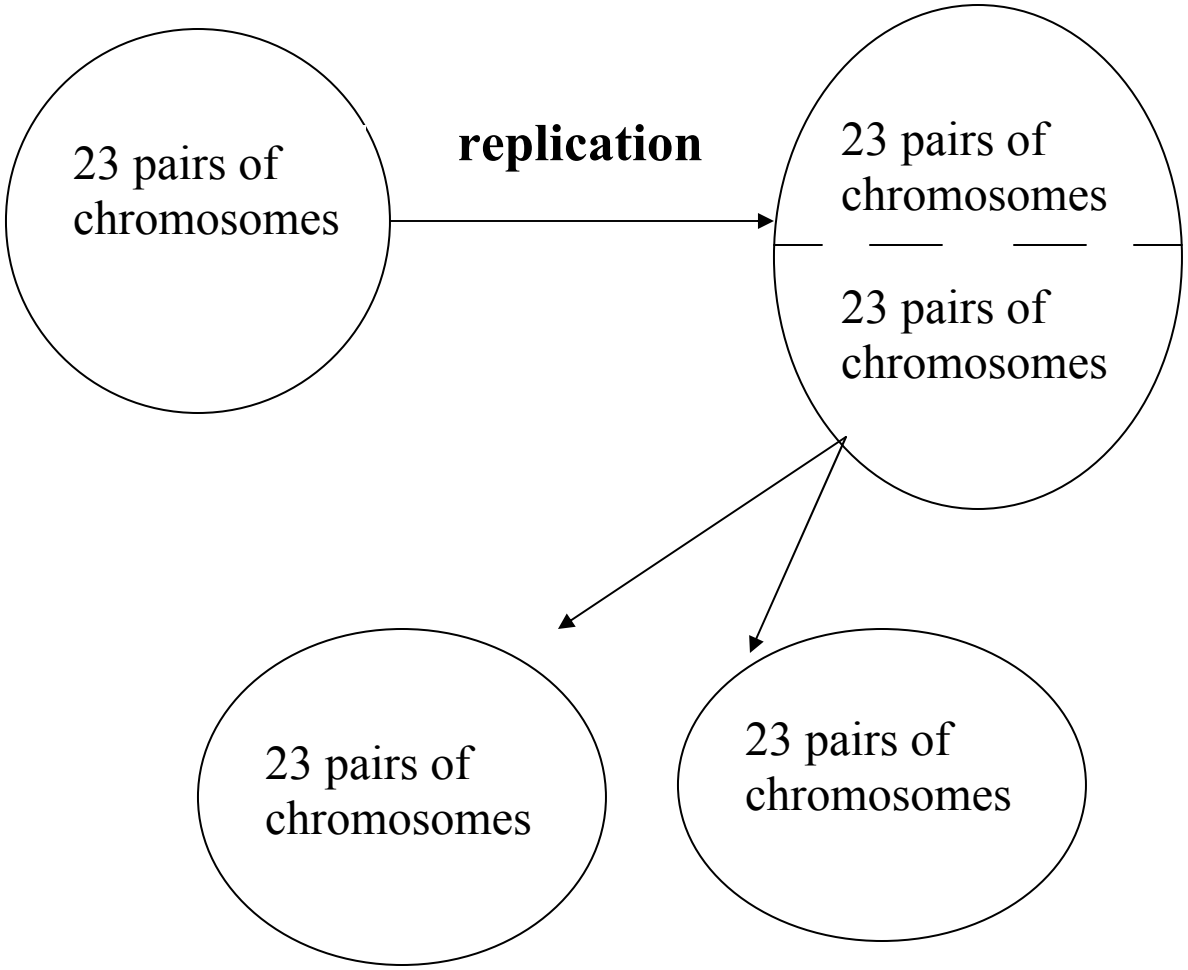
● The normal cell division—Mitosis process:

(1). The double DNA strands in each of the chromosomes split into two single-strands.



(2). DNA replication. After step (2), each chromosome produces another identical one.

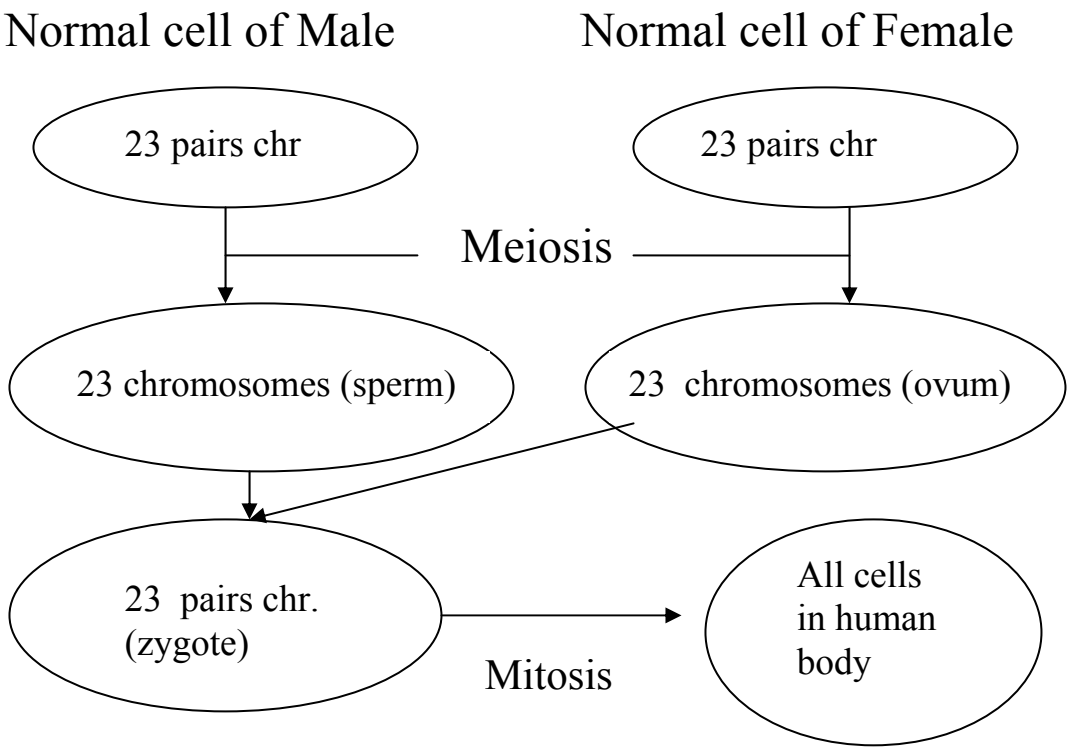
- All the 23 pairs of the chromosomes undergo this process of replication, producing two identical sets of 23 chromosome pairs. The two sets of chromosomes are separated and distributed into two daughter cells.



split to two daughter cells

- **The inheritance of chromosomes—Meiosis**
  1. The 23 pairs of chromosomes in the cell are duplicated every time a cell division occurs.
  2. The only exceptions to this rule are gametes (ovum and sperm), which are produced by sex organ.
  3. Gametes are produced by a special cell division called ***Meiosis***.
  4. **Meiosis** gives rise to daughter cells (ovum or sperm) which contain only a haploid (single chromosome, not pair) set of 22 autosomes and a sex chromosome.

— The procedure of inheritance--Meiosis:



Genetic Terminology

- Some genetic concepts are potentially confusing, such as gene. The reason is that some concepts were introduced prior to the discovery of DNA.

— **Gene**

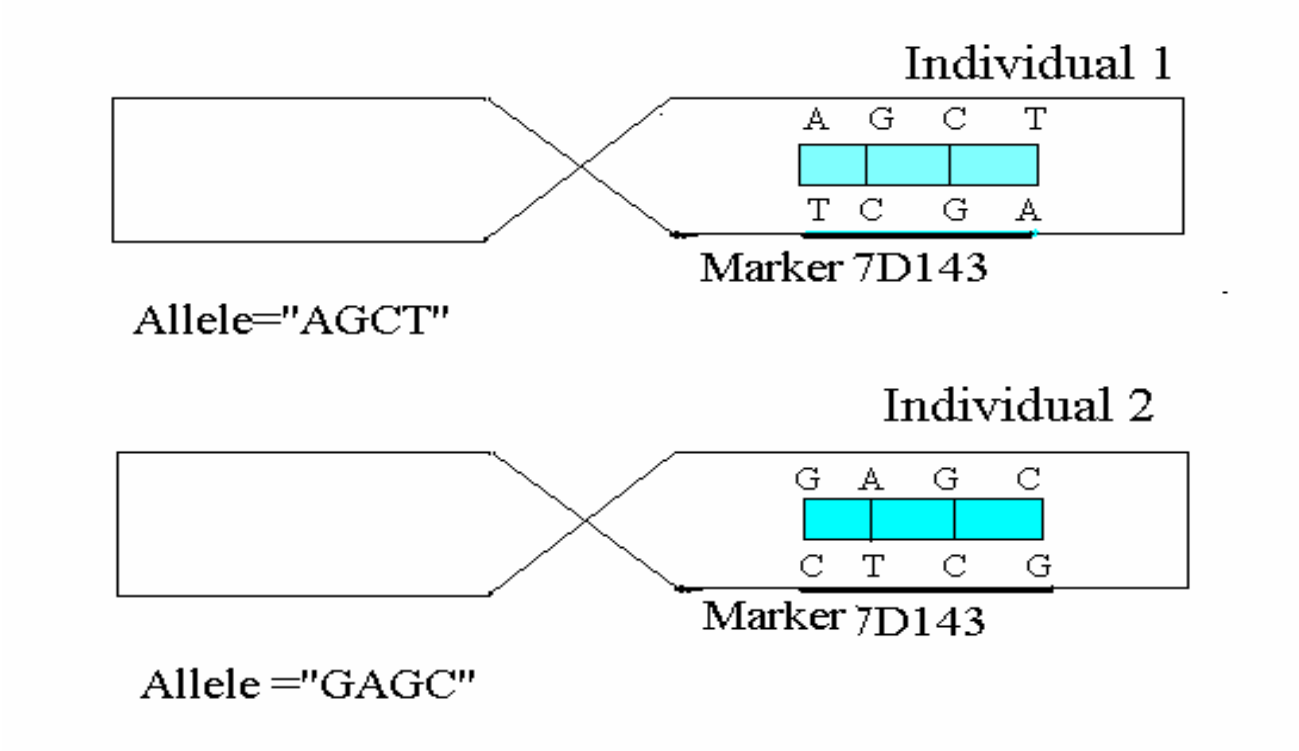
♦ A segment of DNA within a chromosome which has a specific genetic function.

♦ Length from several *bps* to several *kbs*.

♦ Gene is not a smallest unit of genetic material

♦ Before the discovery of DNA, people believe that gene is the smallest genetic unit.

— **Marker (locus):** A specific position in chromosome. It may be 1 *bp* or several hundred *bp*s in length.



— **Alleles:** DNA sequences within a marker or locus (**gene** and **allele** sometimes have same meaning)

• Now , there are mainly two kinds of makers used today.

(1) SNPs (Single nucleotide polymorphism): 1 bp in length; there are two alleles in human population. The two alleles may be any two from {A,G,C,T}.

- Most of the SNPs were found from DNA sequences of a few individuals (Human Genome project). Here is an example of ten sequences from 5 individuals. The positions of second, 8^{th}, 12^{th} columns are SNPs.

GTCGAATTGGAATTGG
GTCGAATTGGAATTGG
GTCGAATTGGAATTGG
GCCGAATAGGAATTGG
GTCGAATTGGAATTGG
GTCGAATAGGAATTCG
GCCGAATTTGATTTGG
GTCGAATTGGAATTGG
GTCGAATTGGAATTGG
GTCGAATTGGAATTGG

(2) Macrosatellite markers: Length from several bps to several hundred bps; many possible alleles; usually denoted by 1, 2, 3, ….

Human Genome

- The totality of DNA characteristic of all the 23 pairs of chromosomes.

— The human genome has about $3x10^9$ *bps* in length.

— 97% of the human genome is non-coding regions called introns. 3% is responsible for controlling the human genetic behavior. The coding region is called extron.

— There are totally about 40,000 genes, over 5000 have been identified. There are much more left

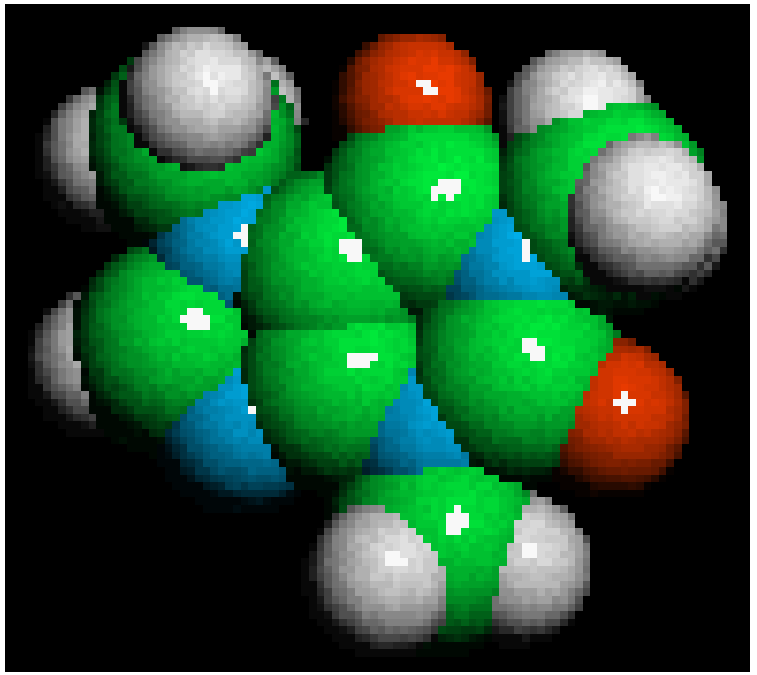
— Human Genome Project is to identified the DNA sequence (every bp) of human genome (only a few individuals)

— For human being, most of the place in human genome are the same. Only a very small part is different among different individuals.

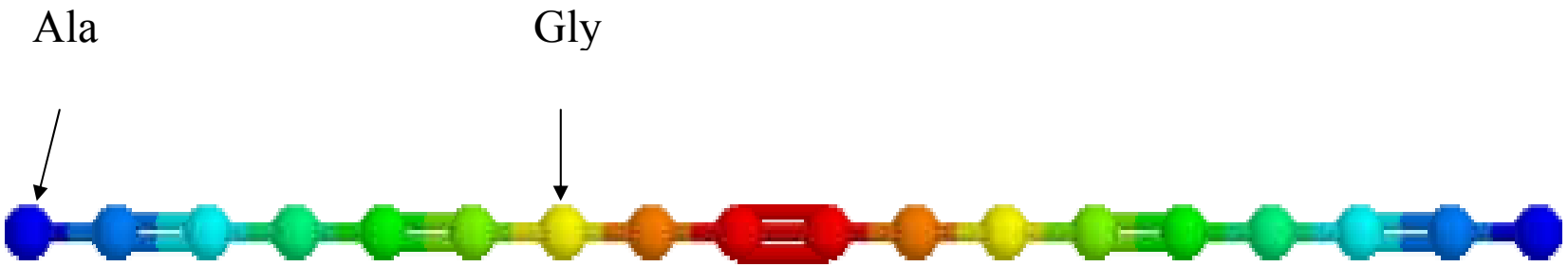
●Gene to protein

Genes or DNA sequences them self are not control the phenotypes. Genes or DNA sequences control the phenotypes through protein.

Protein: like the DNA molecule that is a chain of base pair, each protein molecule is a linear chain of subunits called *amino acids*.



A protein Molecule



Amino acids has 20 different forms which usually denoted by (The firs three letters of different amino acids except in the case of Asn (asparagines); Gln (glutamine); IIe (isoleucine); and Trp (tryptophane)

Ala	Arg	Asn	Asp	Cys
Gln	Glu	Gly	His	Leu
IIe	Lys	Met	Phe	Pro
Ser	Stop	Thr	Trp	Val

The physical and chemical properties of a protein molecule are largely determined by its sequence of the amino acids and its shape structure.

- The DNA sequence specifies *amino acids* sequence (protein), and therefore the structure and function of protein.

- The decoding of the information in the DNA into proteins involve two steps called transcription and translation.

***Transcription*:**

- A single strand of DNA synthesis a single-stranded RNA

- RNA: similar to DNA but have letter U instead of T. So, RNA sequence contains four letters A, U, G and C.

- Before translation, RNA transcripts are processed by the deletion of certain non-coding sequence. The processed RNA chain is called messenger RNA (mRNA).

***Translation*:**

- This step translate mRNA to protein.

- There is a direct relationship between the base sequence of mRNA and the amino acid sequence of its protein product. This relation called genetic code.

Coding rule

- Each "codeword" (or codon) is a triplet of nucleotides (each word is consist of three letters in length).

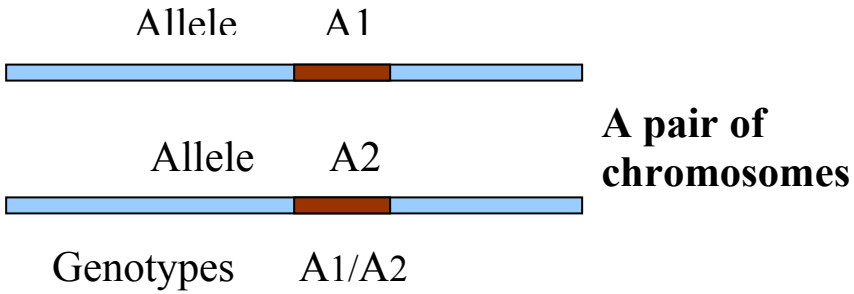
- So, there are 4^3 =64 possible "code words". Since there are only 20 different amino acids, the genetic coding is degenerate, that is, some mRNA code words translate to the sample amino acid. The relation is given below:

Table 1. Genetic coding

First position 5' end	Second position				**Third Position 3' end**
	U	**C**	**A**	**G**	
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Trp Trp Stop Stop	**Cys** Cys Stop Trp	**U** **C** **A** **G**
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	**Arg** Arg Arg Arg	**U** **C** **A** **G**
A	IIe IIe IIe Met	Thr Thr Thr Thr	Asn Asn Lys Lys	**Ser** Ser Arg Arg	**U** **C** **A** **G**
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	**Gly** Gly Gly Gly	**U** **C** **A** **G**

Genotype, phenotype and haplotype

- **Genotype**: At a specific locus there is an allele in each of the two homologous chromosomes. The two alleles together are called genotype.



- **Phenotype:** Observable, such as height, color of eye, etc.

Example: Blood type (ABO locus, three allele A,B and O)

Phenotype	Genotype
A	A/A, A/O
B	B/B, B/O
AB	A/B
O	O/O

Here, A and B both mask the presence of the O allele.

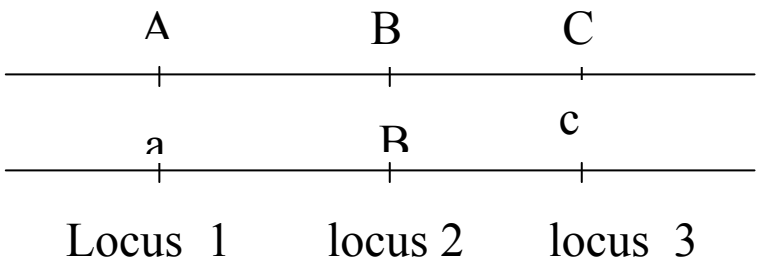
— A and B are said to be **dominant** to O;

— O is **recessive** to A and B.

— A and B are **codominant.**

Now, almost of all the markers with codominant alleles. In this case, we can say that Genotype is observable.

Haplotype: Sequence of alleles along a chromosome



Genotype A/a B/B C/c

The two haplotypes are ABC and aBc.

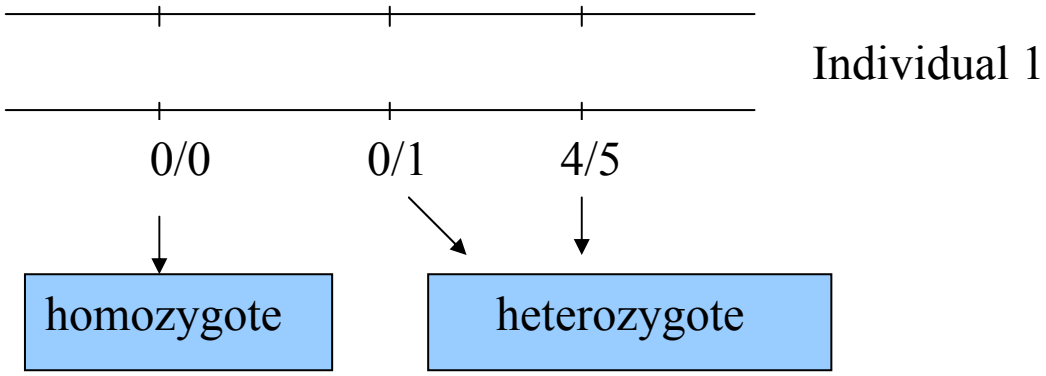
- In practice (for codominant alleles), we can only observe multilocus genotype {A/a B/B C/c}. So the possible haplotype pairs are { ABC, aBc} and {ABc, aBC}

- **Typical data set as follow:**

Ind.	M1	M2	M3
1	0/0	0/1	4/5
2	1/0	0/2	8/7
3	1/1	2/1	9/6
4	0/1	3/2	8/8
5	0/0	0/3	6/5
6	1/0	2/2	7/7
	2 alleles	4 alleles	6 alleles

For individual 6, two haplotypes are {1,2,7; 0,2,7}.

For individual 1, we do not know the two haplotypes. Tow possibilities are {0,0,4; 0,1,5} or {0,1,4; 0,0,5}



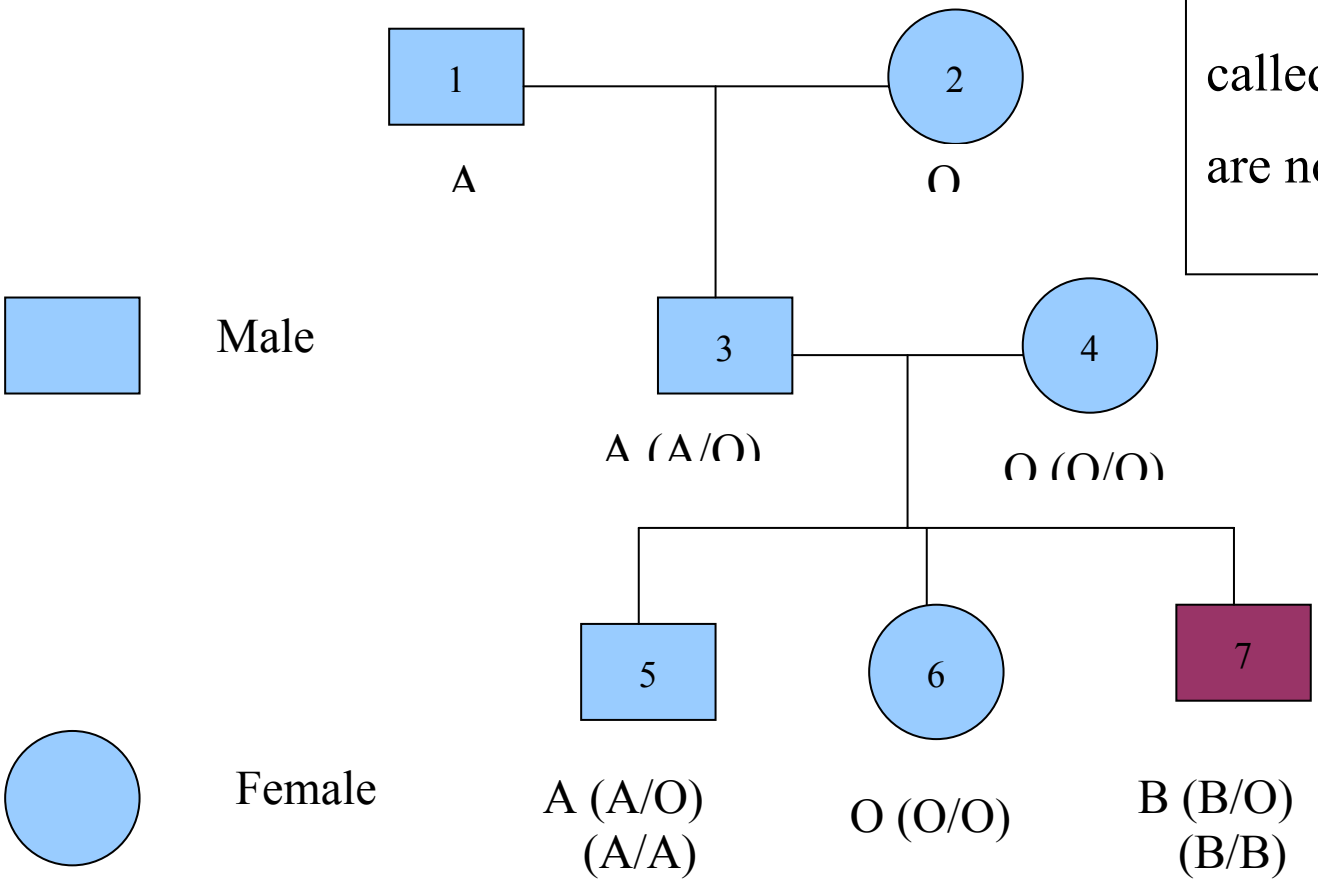
Example: We sample n individuals from a population. Each individual has genotypes at m bi-allelic markers (a marker with two possible alleles). For $n=m=3$, the genotypes are as following

1. A/a B/b C/c
2. A/A B/B C/c
3. A/A B/b C/c

Questions:

(1) Is there any formula for the number of possible haplotype pairs.
(2) How to estimate the haplotype frequencies in the population.

Pedigree (ABO locus)

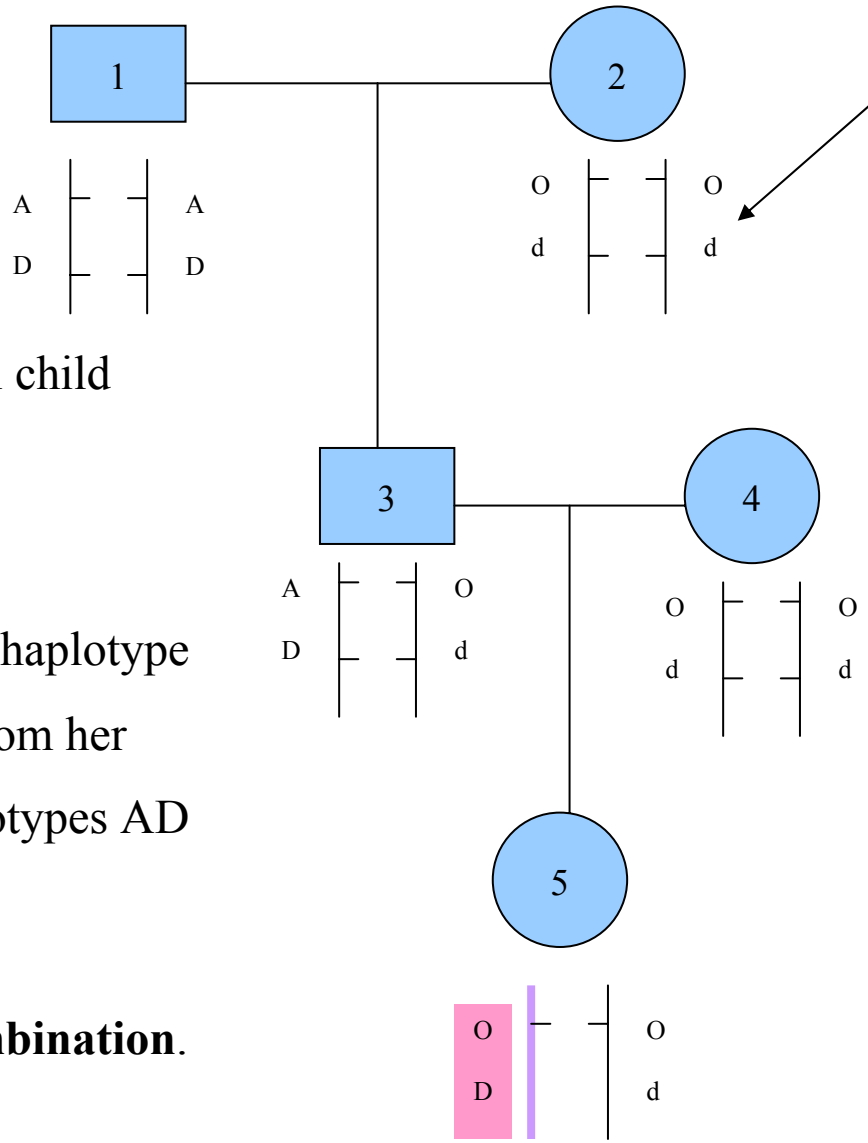


♦ Individuals 1, 2 and 4 are called found whose parents are not in the pedigree.

♦ Individual 7 is not a biological child of the parents 3 and 4 (Mendelian in consistency)

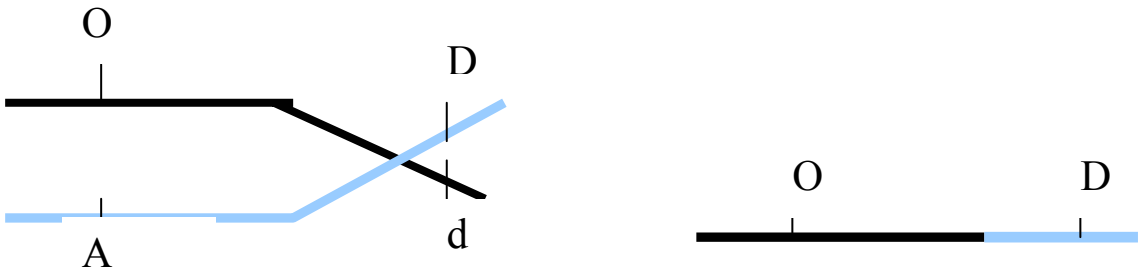
Pedigree (ABO locus and another marker**)**

Consider more than one marker



- Is 5 the biological child of 3 and 4? Yes !
- Why 5's paternal haplotype OD is different from her father's two haplotypes AD and Od.

This is called **recombination**.



- The probability of recombination between two markers is called **recombination rate** between these two markers.
- The recombination rate increases with the physical distance becoming larger.
- The genetic distance between two markers means the **recombination rate**

Relationship between Physical distance and Genetic distance

- **Physical distance*: unit*** *bp, kb*

- **Genetic distance:** Using recombination rate (cM—centi-Morgan, Morgan)

  1 cM = 1% of recombination rate.

- **Relationship** (When physical distance is small , that is, < 1000 *kb*)

  1 cM ≈ 1000 kb

Linkage

Let θ denote the recombination rate between two markers *M* and *m*. If θ<1/2, marker *M* and *m* are said to have **linkage,** otherwise (θ<1/2) the two markers are said to be in **linkage equilibrium. (recombination rate either <1/2 or =1/2)**

Example:

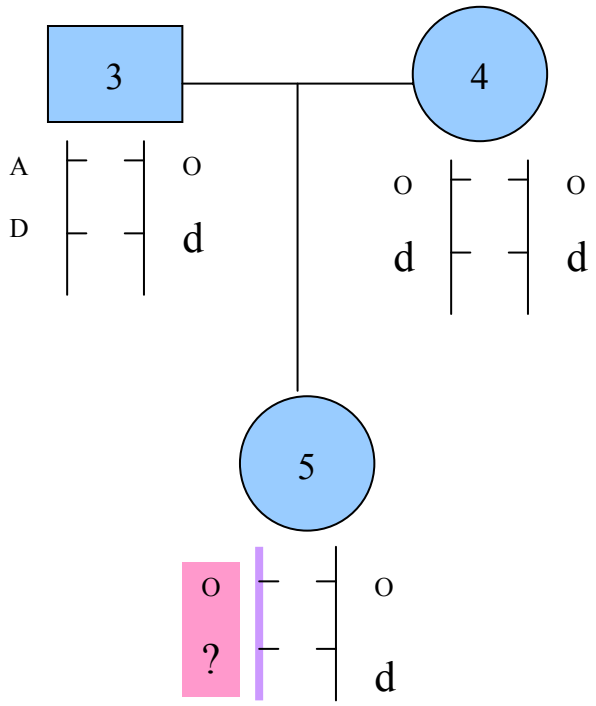
If θ=0, then ?=d,

If θ<1/2, then the

?=d with probability >1/2.

If θ=1/2, ?=d and ?=D with

Equal probability.



Reference

[1] Kenneth Lange (1997) Mathematical methods for genetic Analysis. Springer, New York.

(Chapter 1)

[2] Pak Sham (1998) Statistics in human Genetics. Arnold, New York.

(Chapter 1)

[3] Any text book of "Human Genetics" or "molecular Genetic".

Homework

1. Suppose that segments of DNA sequences (one chain of the DNA) of Chromosome 1 for 10 individuals are as follow:

ATCGAATTGGAATTGG
GTCGAATTGGAATTGG
ATCGAATTGGAATTGG
ACCGAATTGGAATTGG
GTCGAATTGGAATTGG
ATCGAATTGGAATTGG
ACCGAATTTGATTTGG
ATCGAATTGGAATTGG
ATCGAATTGGAATTGG
ATCGAATTGGAATTGG

Which positions are SNPs?

2. What are the major differences between mitosis and meiosis?

3 Given a definition of Gene, Marker (locus), allele, recombination rate, genotype, haplotype, and phenotype. If possible, give a example.