Introduction to some aspects of molecular genetics

Julius van der Werf
(partly based on notes from Margaret Katz)
University of New England, Armidale, Australia

Genetic and Physical maps of the genome

A layout of the order of genes (loci) as well as the distance between them is called a genetic map. The distances in a genetic map are determined according to the recombination fraction between two loci. The unit of measure is Morgans (or Centi-Morgans-cM), representing the recombination frequency between the two locations. One cM is one recombination event per 100 meiosis on average. Thus if two regions of the genome are 10cMs apart, you would expect a recombination event between these two regions in 10 out of 100 meiosis. The genetic map distance between two genes therefore determines the frequency in which those genes are expected to recombine.

A genetic map can be contrasted to a physical map of a chromosome, where the distance between two genes are measured in base pairs (or kilo-base pairs : kb) . In humans, 1 cM on a genetic map corresponds to about 1-2 Mb of DNA (1 to 2 million base pairs). Note that there is considerable variation in thsi figure between different areas of the genome. To construct a physical map, clones containing very large fragments (about 1Mb) of DNA inserted in Yeast Artificial Chromosomes (YAC's) or other suitable vectors. A set of clones is known as a gene library. A complete library contains the whole genome in tens of thousands of individual clones.

If you could compare the sequence of bases in the chromosome of two individuals of the same species, most of the base pairs would be identical. Fore example, it is estimated that differences in DNA sequence occur only 1 in 1000 bp in humans. However, no two individuals (except identical twins) have the same exact sequence of bases in each of the millions of base pairs that make up the genetic material of a mammal. At certain sites along the chromosome the sequence would vary between individuals. These sites, where differences in DNA sequence occur, are known as *molecular markers*.

Molecular markers

When differences in DNA occur within genes, the differences have the potential to affect the function of the gene and hence the phenotype of the individual. Genetic markers which have been used a lot in the past include blood groups and polymorphic enzymes. We have relatively few such markers, but this has been overcome with the advent of new types of markers.

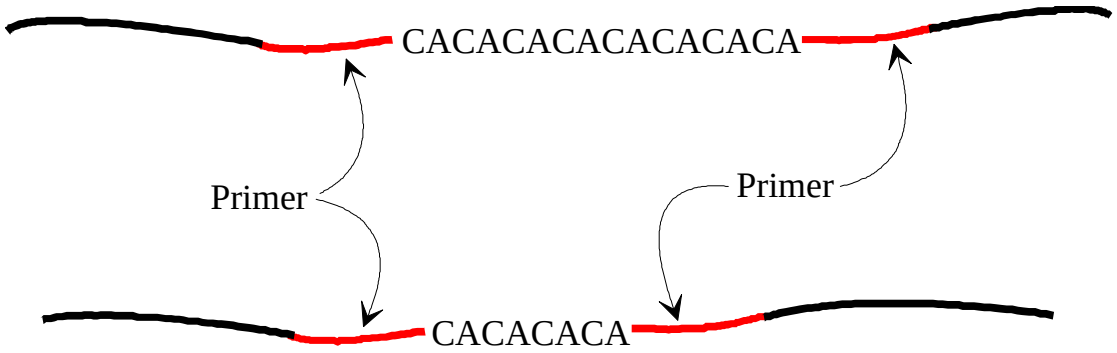
However, most molecular markers are not associated with a visible phenotype. The main types of molecular markers are VNTRs, RFLPs and RAPDs, AFLPs and SNPs.

VNTR's

Scattered at various locations in the human genome are regions that are highly variable. These regions contain a type of DNA sequence called Variable Number Tandem Repeat. Tandem repeats are multiple copies of a sequence of base pairs arranged in head to tail fashion. For example, a frequently found tandem repeat is CA, and one strand containing this type of repeat reads CACACA….. , notated as $(CA)_n$ The other strand would read GTGTGT… In this example, the number of repeating basepairs is two, but it can be more. When the repeating unit is less than four, the VNTR is called a **microsatellite** and when the repeating unit is longer it is a **minisatellite**.

Microsatelites

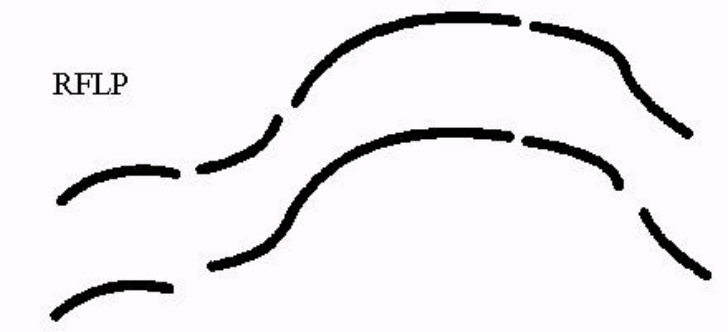
Mictrosatelites are DNA regions with variable numbers of short tandem repeats flanked by a unique sequence. Microsatelites make good genetic markers because they each have many different 'alleles' - ie. there can be many different lengths of the repeat region. An allele is defined by the number of repeats there are at the same location. With many alleles, most individuals are heterozygous, giving power to note association between marker allele and performance in progeny inheriting a favourable linked QTL allele.



Through the PCR reaction (see below), which uses the unique sequences either side of the repeat sequences as primer binding sites, microsatellite DNA can be specifically amplified. The alleles an individual carriers at a particular microsatellite loci can then be determined by accessing the size of the amplifed fragment through agarose gel electrophoresis.

Restriction Fragment Length Polymorhisms (RFLP's).

Restriction enzymes enzymes cut DNA wherever they find the appropriate nucleotide sequence (eg. Eco R1 cuts at the 'recognition sequence' GAATTC). If there is a mutation at this sequence, no cut is made and the resulting DNA fragment is longer. Also mutation to give a new recognition sequence gives a pair of shorter fragments. Genetic differences (polymorphisms) of this type are known as Restriction Fragment Length Polymorphisms.



In this diagram we can see there are different length DNA fragments from each of these 2 chromosomes, according to cutting or lack of cutting at four sites. Using a battery of restriction enzymes (there are more than 100), thousands of RFLP markers can be generated.

The following markers are based on the PCR technique, which will be discussed in more detail in the next section.

RAPD's

Random Amplified Polymorphic DNA (RAPD) markers are DNA fragments generated in PCR reactions that use a single short primer (in normal PCR a primer-pair is used). The primer must be complementary to sequences that are on opposite strands within a small number of base pairs (say 2000). The DNA strand between these two sites is amplified in a PCR. Polymorphism is determined by individuals who have mutations at those sites, and therefore will not show a product on the gel.
The advantage of RAPD's is that we do not need to know the DNA sequence of the species studied. A primer has a certain chance of *randomly* generate a PCR product. Hence, RAPDs are cheap markers to develop. The disadvantage is that RAPDs either give or do not give a product and therefore, we can not distinguish between homo- and heterozygotes.

AFLPs

Amplified Fragment Length Polymorphism (AFLP) is based on PCR amplification of selected restriction fragments. Like RAPDs, AFLPs require no prior knowledge of DNA sequences (unlike microsatellites). The advantage of AFLPs over RAPDs is that they are more reliable and reproducable (depend less on DNA quality and lab conditions). Also, the number of polymorhpic loci (molucular markers) that can be detected is 10-100 times greater with AFLPs than with microsatellites or RAPDs

SNPs

Single Nucleotide Polymorphisms are based on single base pair polymorphisms. A SNP is a position at which two alternate bases occur at appreciable frequency. In humans they may number greater than one in a thousand base pairs. SNPs can be detected by a number of methods, however a relatively new technology, using DNA chips, can be used for large scale screening of numerous samples in a minimal amount of time.

Other Terms

A couple of more terms worth mentioning are STSs and ESTs. STS stands for sequence tag site. A STS is a *unique* sequence with a known chromosomal location. They are often used to put together mapping information from different laboratories. An STS is usually 200-400 base pairs long, and amplifiable by PCR. Microsatellites are an example of an STS.

An mammalian sequences have much repetitive DNA conformation that an STS represents a unique sequence can be very time consuming. One approach to increase the chance of isolating a unique STS is to use expressed sequence tags (ESTs). Generation of ESTs is shown below. First cDNA, a DNA copy of mRNA, is generated. Then 200-400 base pairs of the cDNA ends are sequenced. These are the ESTs. As mRNA is largely free of repeat sequences, the ESTs are likely unique.

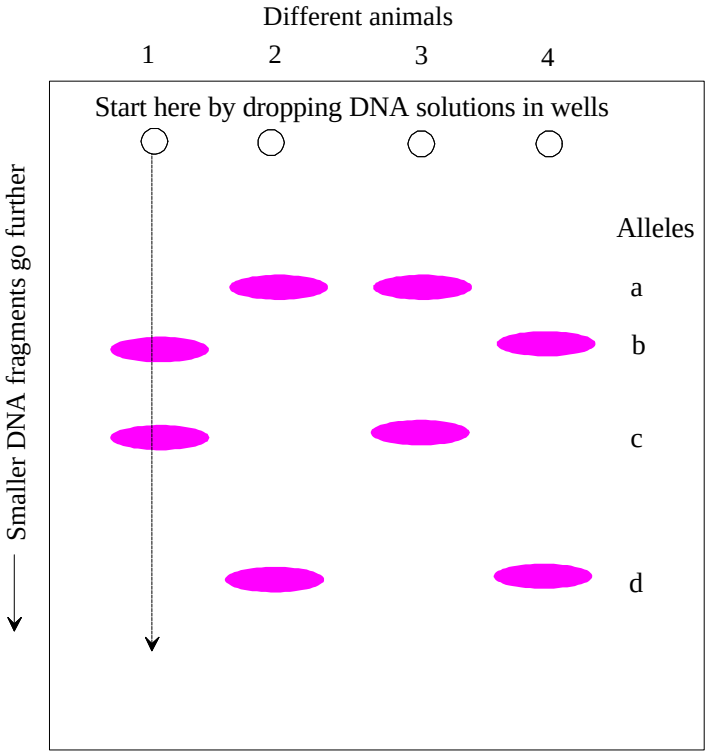
Detecting molecular markers

Molecular techniques (such as PCR or restriction enzyme digestion, followed by gel electrophoresis) can be used to identifying different alleles resulting from DNA polymorphisms.
Different alleles from a VNTR will have different size and similarly, RFLP’s have different sizes (as defined by their name!).

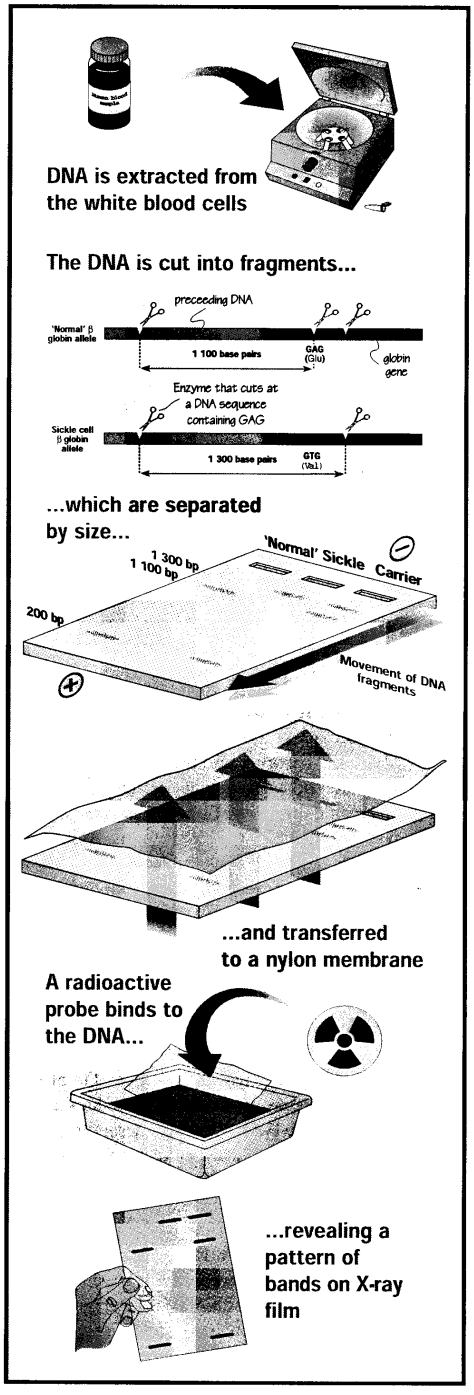
Gel electrophoresis.

Gel electrophoresis seperates DNA according to size. A gel is essentially a slab of gelatinous material. DNA is applied to 'wells' at the top of the gel (which is submerged in a tank containing some buffer), and an electrical current applied. DNA is negatively charged and is drawn towards the positive electrode. Smaller fragments will move down the gel faster, as it is easier for them to move throught the gel matrix.

The figure represents the electrophoretic gel in which microsatelite DNA fragments of different size (different alleles) have been run. Smaller fragments of DNA have migrated further. In this example there are four alleles, and of course each individual can carry only two! The genotypes of the four animals are deduced to be *bc*, *ad*, *ac* and *bb*, respectively.



Southern Blot



A southern blot involves the transfer of DNA from a gel (where it has been seperated according to size) to a special type of membrane. The DNA on the member (which is in a denatured or single stranded state) is exposed to a probe. A probe is a short sequence of DNA that is complimentary to, and thus binds to, a DNA sequence of interest. Probe bound to the membrane is then visualized: this can be achieved by labelling the probe with radiation and exposing the membrane to X-ray film.
A Southern Blot will usually show the alleles of VNTR's on all chromosomes, giving a complex pattern known as a DNA fingerprint.

Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) amplifies a specific region of DNA as defined by two primer sequences. It can thus be used to examine one particular region of the genome. Because many copies of one specific section of the genetic material are generated , it is possible to use this technique with very, very small amounts of DNA as starting material (e.g. a single hair root or a small blood stain)

PCR is a three stage process. Firstly the DNA is denatured (made single stranded), secondly the primers bind or anneal to their complentary sequence, and thirdly the primers are extended by the addition of nucletodies complentary to that on the template sequence (this requires the action of an enzyme called DNA polymerase). This three stage process is then repeated 20-40 times. The first few cycles of PCR as shown below. The end result is amplification of the sequence between and including the primer sequence
A schematic representation of the PCR reaction is given on the next page

Summarizing comments about different molecular markers

PCR based techniques (microsatellites, RAPDs, AFLPs) need only very small amounts of DNA and genotyping is easily automated.

RAPDs and AFLPs require no prior knowledge of DNA sequence of the species under study and are therefore easier and cheaper to develop than microsatellites

AFLPs can be detected at many more polymorphic sites than RAPDs or microsatellites.

SNPs are genetic markers that in the near future may allow cheap large scale genotyping as in DNA chips

Starting template – double stranded DNA

Round 1

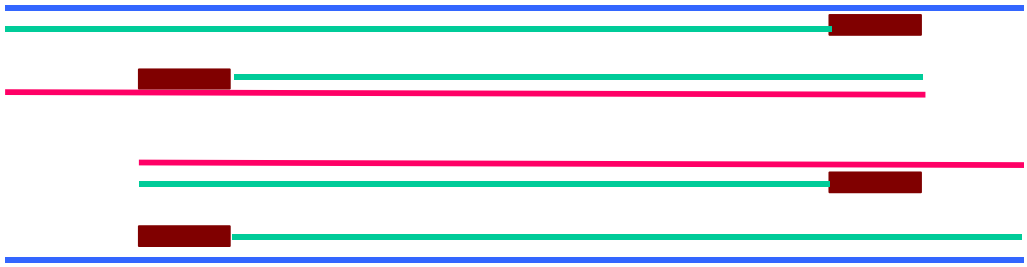
- Denaturation

- Annealing of primers

- Extension

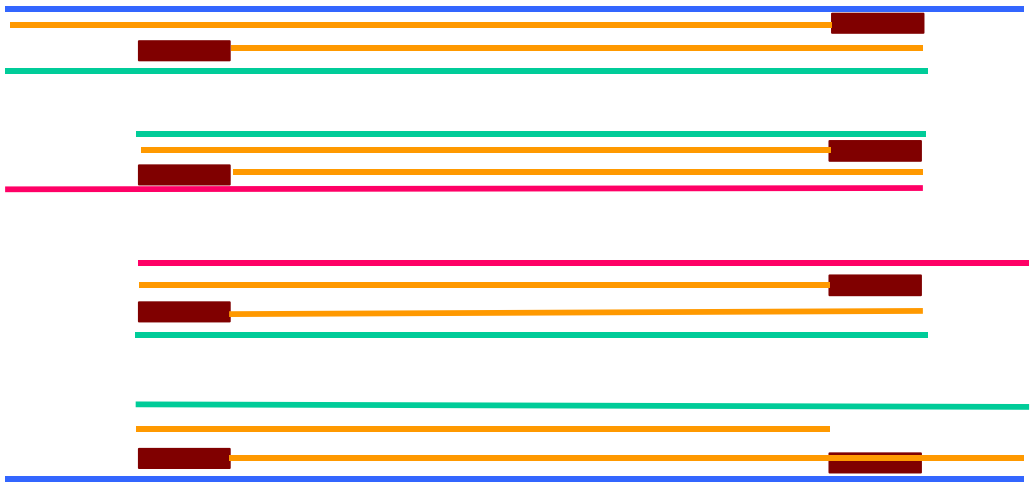
Round 2

- Shown after denaturation / annealing / extension



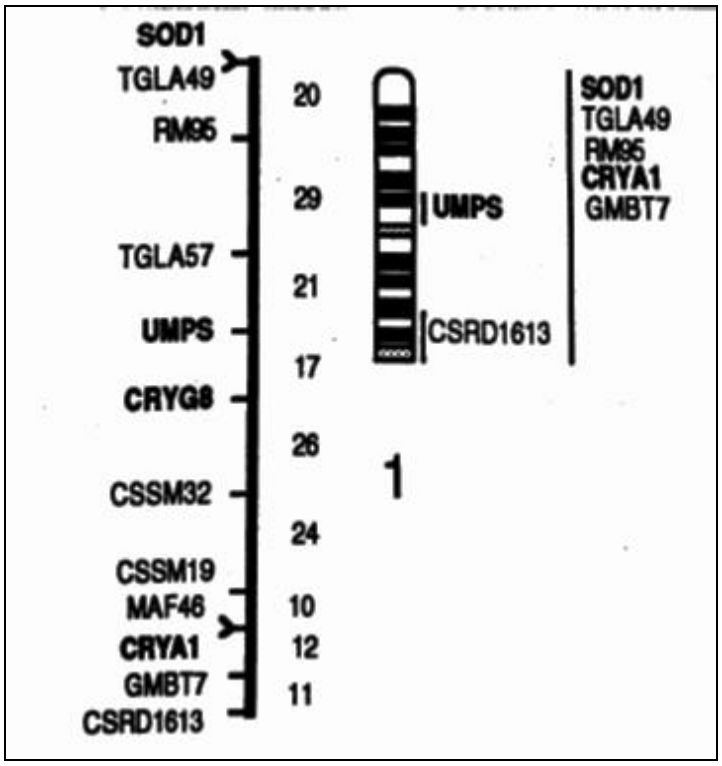
Round 3

- Shown after denaturation / annealing / extension



Linkage maps.

All molecular markers that are found need to be mapped. based on linkage analysis (see next chapter) their position on the genetic map needs to be determined. The next Figure illustrates a linkage map of chromosome 1 in cattle (Barendse et al., 1994).



A linkage map of chromosome 1 of cattle

On the left is the linkage map derived from studies on co-inheritance of alleles. Most loci are microsatelites or RFLPs. Distances between loci are given in units of centimorgans (100 units is equivalent to a 50% recombination fraction).

In the center is a representation of the chromosome with its visible banding regions. Some loci have been localised through, for example, observing fluorescing DNA from the gene of interest 'sticking' to particular regions.

On the right is a list of loci associated with chromosome 1 through work with somatic cell hybrid lines. For example, alleles at these loci might all be present in a cell line containing only chromosome 1 from cattle, the rest from, say, hamsters.

References

Barendse et al. 1994. Nature Genetics 6:227-235.
Katz, M. 1998. Lecture notes GENE422. Univ. of New England.

NOTE: Useful information may be found on the WWW, e.g. www.reading.ac.uk/NCBE