Module 3

Genetic mapping

Thierry Huguet, György B. Kiss, Attila Kereszt,

Dong-Jin Kim and Doug Cook

Local organiser : Pascal Ratet

Genetic mapping

Sections 1, 2 and 3 are adapted from :

EMBO Practical Course on Genetic and Molecular Analysis of *Arabidopsis*
Module 2 : MAPPING MUTATIONS USING MOLECULAR MARKERS
Jérôme Giraudat, Nathalie Beaudoin, Carine Serizet

1. INTRODUCTION

Mapping a marker or a mutation to a well-defined chromosomal region is an essential step in the genetic analysis of a plant, and is also (unless the mutant is tagged) a prerequisite for molecular cloning of the corresponding gene. Determining the map position of a marker or a gene (as identified by its mutant phenotype) consists in testing linkage with a number of previously mapped markers. Once linkage with a specific marker is detected, a refined mapping can be achieved by analysing linkage relations to more markers in that region.

Historically, mapping in plants primarily utilised morphological markers such as mutants with an easily scorable phenotype and a defined map position. Typically, the mutant of interest is crossed to another mutant used as phenotypic marker, the resulting F1 double heterozygote is allowed to self, and the segregation of the two phenotypes is analysed in the F2 population.

The mutation used as marker should of course not interfere with the phenotype of the mutant to be mapped. The genetic distance is the number of meiotic recombination events that occur between the two loci in 100 chromosomes. The genetic distance is expressed in centiMorgans (cM), and can range from 0 cM (absolute linkage) to 50 cM (non-linked loci). It remains however difficult to score many different phenotypes in a single population. Hence, detailed mapping using morphological markers is tedious because it requires numerous crosses.

In contrast, a single cross can be used to analyse linkage with an essentially unlimited number of molecular markers. DNA markers were incorporated into mapping strategies once it was recognised that distantly related individuals differ in DNA sequence throughout their genome [Botstein, 1980]. Molecular markers exploit the natural differences between distinct lines. For instance, in *Arabidopsis*, it has been estimated that the widely used Landsberg erecta and Columbia ecotypes differ by approximately 0.5 to 1 % at the DNA sequence level [Chang,1988 ; Hauser, 1998]. These local differences or polymorphisms of the DNA sequence are due to point mutations, insertions or deletions that randomly occurred in one ecotype and not in the other. These DNA polymorphisms can be conveniently visualised by several methods.

To map a novel mutation that was generated in line A, this mutant is crossed with a wild-type plant of a polymorphic line B, and the F1 progeny is allowed to self. The resulting F2 population can then be used to analyse the linkage between the mutation of interest and any DNA marker that distinguishes lines A and B. As compared to morphological markers, an additional advantage of molecular markers is that in most cases homozygous and heterozygous individuals can be readily distinguished (see below).

During this course, the bases of genetic mapping techniques will be presented. In addition, data will be generated using *M. truncatula* Recombinant Imbred Lines (RILs) in combination with the SSR (microsatellites) method. Data will be analysed using the MAPMAKER program or the colormapping procedure.

2. DIFFERENT TYPES OF DNA MARKERS

The types of markers that will be used during the practical course are indicated with *. However, other types of DNA markers that can be used for genetic mapping are also described. By definition we call "DNA marker" a DNA fragment which exhibit polymorphism between two lines.

2.1. Restriction Fragment Length Polymorphisms (RFLPs)

The first type of DNA markers that were used for genetic mapping were RFLPs. The DNA sequence differences between polymorphic lines may create differences in the length of restriction fragments derived from genomic DNA. For instance a given restriction site may be present in one line and not in the other. As illustrated in Figure 1, this polymorphism can be revealed by genomic DNA blot hybridisation (Southern) using as probe a DNA fragment corresponding to that region. The polymorphic bands can then be used as genetic markers to distinguish the two lines. Multiple RFLP markers can be identified and assembled into genetic maps.

An advantage of the RFLP mapping approach is that RFLP markers are co-dominant. Distinct patterns are indeed obtained for plants that are homozygous or heterozygous for the parental alleles (Figure 1). Hence, all the chromosomes of a given F2 population can be scored. In contrast, a main disadvantage is that RFLP mapping necessitates relatively large amounts of DNA because distinct RFLP markers may require digestion of genomic DNA samples with different diagnostic restriction enzymes.

Figure 1. Principle of RFLP markers. This figure illustrates an RFLP marker which utilises a site for the restriction enzyme (E) which is present in line A and not in line B.

2.2. Cleaved Amplified Polymorphic Sequences (CAPS)

The principle of CAPS markers is very similar to that of RFLP markers. The main difference is that PCR is used instead of DNA blot hybridisation to detect a restriction site polymorphism. As illustrated in Figure 2, a genomic DNA region is amplified by PCR using specific primers and those amplified fragments are then digested with a diagnostic restriction enzyme to reveal the polymorphism. Hence, whereas RFLP probes can be anonymous clones, CAPS markers require sequence information to design the specific PCR primers.

Like RFLPs, CAPS markers are co-dominant (Figure 2). CAPS markers are based on PCR for detection, and thus require only small quantities of genomic DNA. Typically, a single leaf will provide enough DNA for analysis with multiple CAPS markers.
Finally, CAPS markers can be easily assayed using standard agarose gel electrophoresis.

Ecotype A   **Ecotype B**

**PCR amplification
Digestion with restriction enzyme E
Agarose gel electrophoresis**

A/A   A/B   B/B
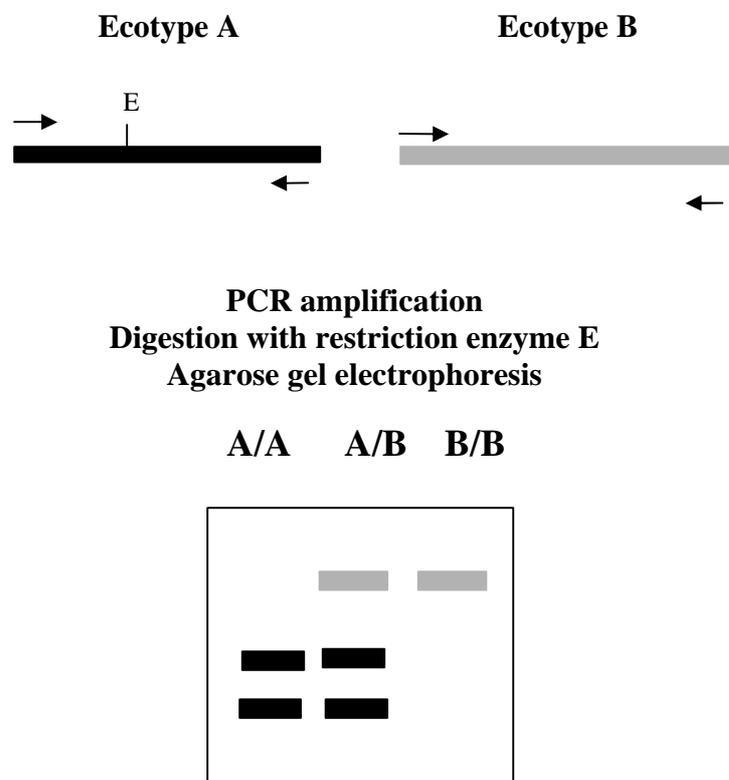
Figure 2. Principle of CAPS markers. This figure illustrates a CAPS marker which utilises a restriction enzyme (E) that cleaves the amplified fragment at one site in line A and not at all in line B.

2.3. Random Amplified Polymorphic DNA (RAPD)

RAPD markers are another type of PCR-based markers that have been used for genetic mapping [Williams, 1993]. This approach is based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence. The oligonucleotide (around 10-bp long) is used for PCR at low annealing temperatures. When the oligonucleotide hybridises to both DNA strands at sites within an appropriate distance from each other, the DNA region

delimited by these two sites will be amplified. Small nucleotide changes (polymorphism) at one of the two sites may prevent hybridisation of the oligonucleotide and hence also prevent DNA amplification [Williams, 1990]. Typically a RAPD primer will amplify a given fragment from line A and not from line B. It will thus be impossible to distinguish an homozygous individual AA from an heterozygous individual AB. In other words, RAPDs are dominant markers and are thus less efficient than co-dominant markers in extracting information from a given F2 population. Another limitation of RAPD markers is that because of the low annealing temperatures used, the amplification of a given polymorphic band seems to be highly sensitive to PCR conditions and hence less consistently reproducible in different laboratories.

2.4. Amplified Restriction Fragment Length Polymorphism (AFLP)

AFLP TM is a patented technology developed by KeyGene, Wageningen, The Netherlands [Vos et al., 1995]. In this procedure, the genomic DNA is digested by two different restriction enzymes, a rare cutter and a frequent cutter. Double-stranded adapters are then ligated to the ends of the restriction fragments. The fragments are then amplified by PCR using primers that correspond to the adapter and restriction site sequences. These primers have additional nucleotides at the 3' ends extending into the restriction fragments, in order to limit the number of fragments that will be amplified. The AFLP products are detected by labelling one of the two primers, and the labelled DNA fragments are separated by electrophoresis in denaturing polyacrylamide gels (similar to sequencing gels). Typically, 50 to 100 amplification products are detected in a single lane. Polymorphic bands can be identified by comparing the amplification products derived from two lines. Like RAPDs, AFLPs are typically dominant markers.
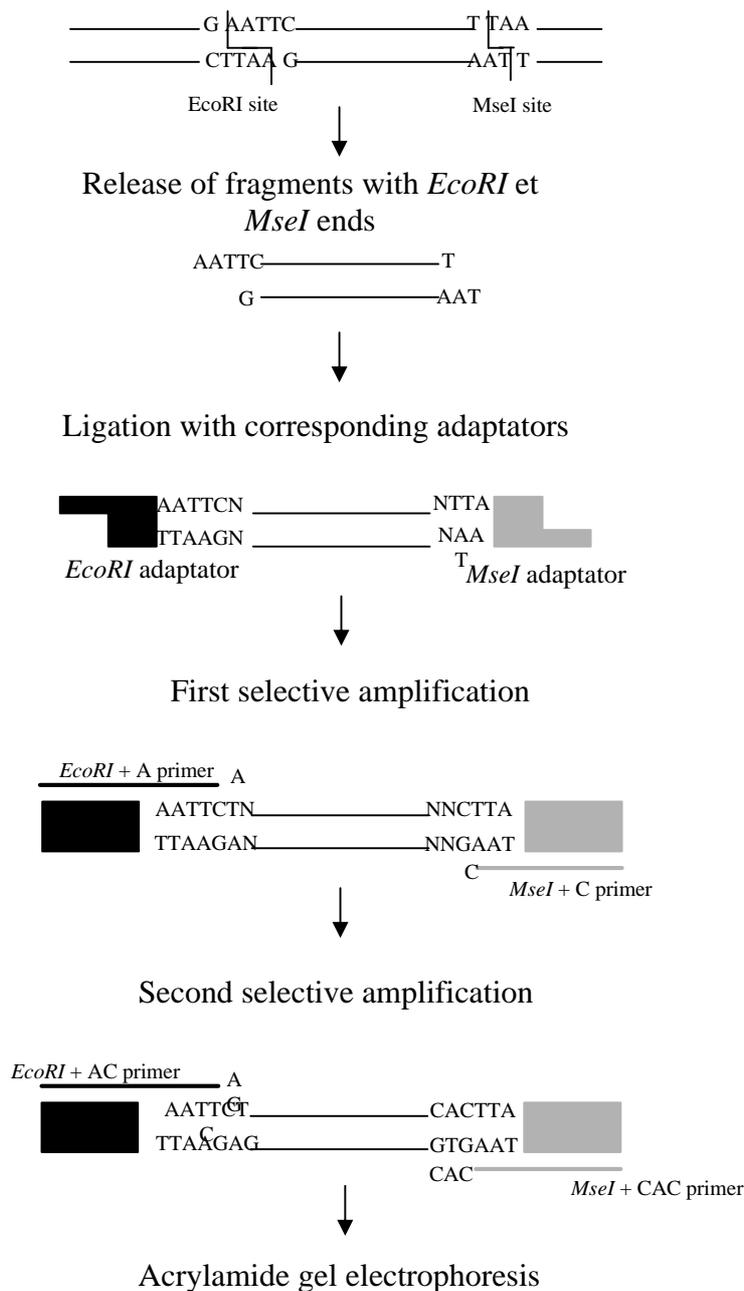
Digestion of genomic DNA by *EcoRI* et *MseI* enzymes

G AATTC   T TAA
CTTAA G   AAT T

EcoRI site   MseI site

Release of fragments with *EcoRI* et *MseI* ends

AATTC—————T
G—————AAT

Ligation with corresponding adaptators

AATTCN   NTTA
TTAAGN   NAA
T
EcoRI adaptator *MseI* adaptator

First selective amplification

EcoRI + A primer A
AATTCTN   NNCTTA
TTAAGAN   NNGAAT
C
MseI + C primer

Second selective amplification

EcoRI + AC primer A
AATTCT   CACTTA
TTAAGAG   GTGAAT
CAC
MseI + CAC primer

Acrylamide gel electrophoresis

Figure 3. Principle of AFLP markers from D. de Vienne (1998, INRA éditions)

2.5. Simple Sequence Repeats (SSR) : Microsatellites

Like other eukaryotic genomes, the plant genome contains tandem repeats of one- two- or three-nucleotide motifs. These microsatellite repeat sequences are usually polymorphic in different lines because of variations in the number of repeat units. These polymorphisms are called SSR, and can be conveniently used as co-dominant genetic markers. As illustrated in

Figure 3, specific primers are used to PCR amplify a small genomic region (150 to 250 bp) that contains a polymorphic microsatellite sequence. The size of the amplified fragment will vary depending on the number of repeats present in a given line. These polymorphic fragments can be separated and visualised by electrophoresis in agarose or polyacrylamide gels.

As compared to CAPS markers, SSR offer the additional advantage that they do not involve the use of restriction endonucleases and thus avoid the problems associated with partial digestions.
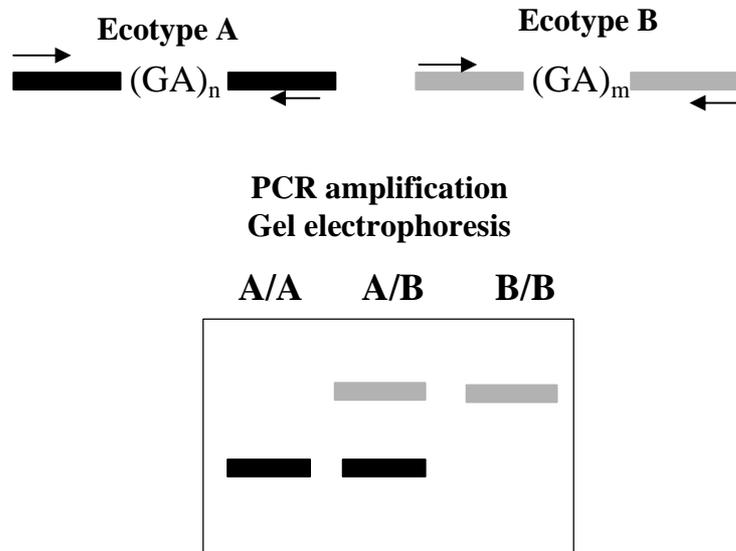
Figure 4. Principle of SSR markers. This figure illustrates an SSR marker which utilises the fact that the number of (GA) repeat units is higher in line B than in line A.

2.6. Single Nucleotide Polymorphisms (SNPs)

The most common class of DNA polymorphisms present both in natural lines and after induced mutagenesis is single nucleotide polymorphisms (SNPs). As described above, the RFLP and CAPS methods can detect only the SNPs which alter a recognition site for a restriction enzyme. The RAPD and AFLP methodologies can in principle detect any type of SNPs, however these two techniques are not very convenient to target a selected genomic region. In contrast, plant genome sequencing is generating a wealth of sequence information which provides a starting point for the development of PCR-based markers. In other words, once the sequence of a region is known, primers can be synthesised to amplify alleles of interest which can then be analysed to find allele-specific polymorphisms in that region.

Although they have not been extensively used in plants thus far, a large number of techniques have been developed to scan a defined region of DNA for SNPs (reviewed by [Cotton, 1997]). In particular, single-strand conformation polymorphism (SSCP) is based on the fact that a strand of single-stranded DNA folds differently from another if it differs by a single base, which leads to different mobilities of these two strands in non-denaturing gel electrophoresis. Heteroduplex analysis is based on the different mobilities of homo- and hetero-

duplexes in non-denaturing gel electrophoresis, or in slightly denaturing high-performance liquid chromatography. It should be noted that methods to detect SNPs can also be extremely useful in the last step of a positional cloning of a mutant locus, namely to locate the mutation within the DNA region that has been delimited by mapping.

Finally, an efficient method has been described that allows to create a PCR-based marker for any known point mutation. This technique is called derived cleaved amplified polymorphic sequence (dCAPS), and has been recently applied to *Arabidopsis* [Michaels, 1998 ; Neff, 1998]. The dCAPS method is primarily used when the point mutation of interest does not alter an existing restriction site. In this case, the dCAPS technique consists in designing a primer with one or two mismatches which, together with the mutation, will create a unique restriction site in one only of the two alleles. A second primer (usually without mismatch) is used to PCR amplify the region, and the amplification products are digested with the appropriate restriction enzyme, exactly as for CAPS markers. dCAPS are useful for genetic mapping, and to follow known mutations in segregating populations.

3. GENETIC MAPPING

When selecting DNA markers on genetic maps, or later in interpreting linkage data to these markers, it is important to bear in mind how these reference genetic maps were constructed. A genetic map only displays the relative genetic distances between the particular set of markers that were analysed in a given mapping population. Hence, a given marker will typically have different map positions in genetic maps that were constructed independently, or in successive versions of the same map as new markers are incorporated. In order to construct a better reference map, in particular with respect to the relative order of markers, it is essential that all markers are mapped in the same population. Even with the development of PCR-based markers which require a smaller amount of genomic DNA than RFLP markers, only a limited number of markers and phenotypic traits can be mapped in a given F2 population. This provided the impetus for the generation of populations of recombinant inbred (RI) lines for mapping.

To construct an RI population, individual F2 plants are selfed, and for each F3 family a single F3 plant is selected at random and allowed to self. This process, called single-seed descent, is repeated to the F7 (Figure 4). At each generation, the average level of heterozygocity is reduced by 50 %. Hence, F7 lines are over 98 % homozygous. These RI lines thus constitute permanent mapping populations because they are near-homozygous and can therefore be multiplied indefinitely, enabling multiple laboratories to use the same mapping population.
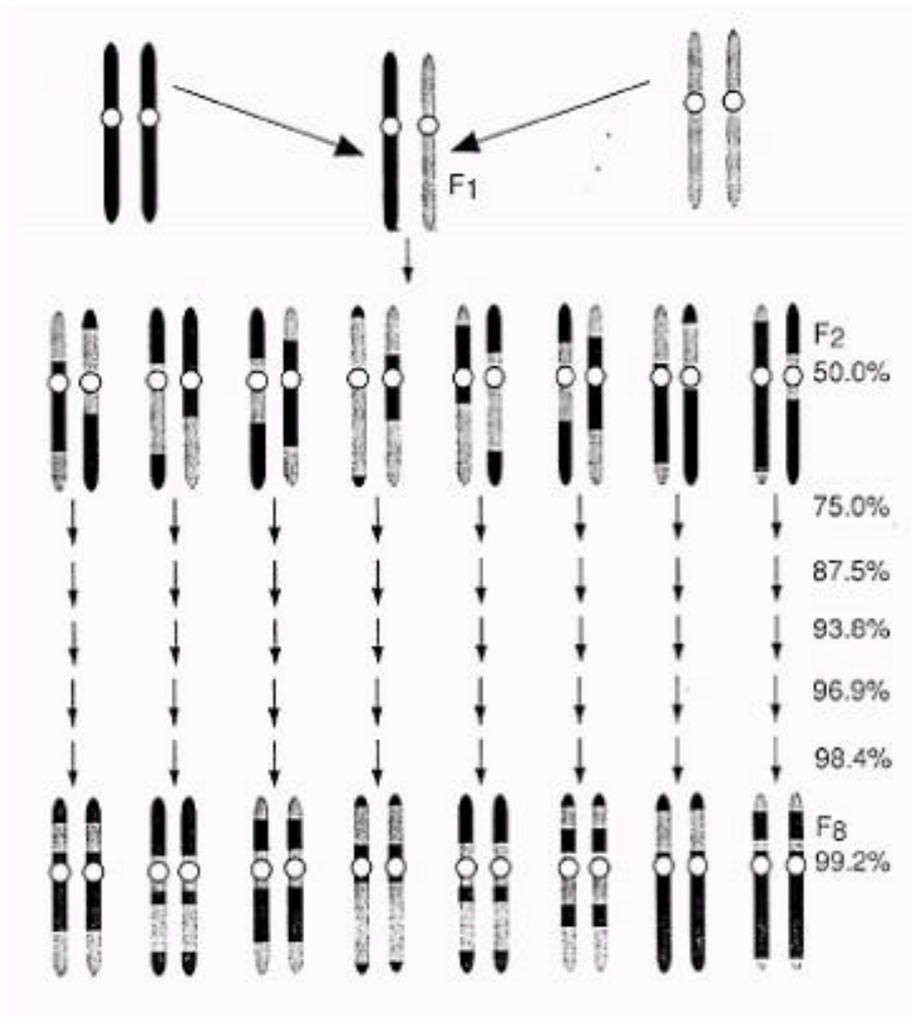
Figure 5. Development of RI lines by single-seed descent from individual F2 plants. At each generation, the expected level of homozygocity is indicated on the right (from Reiter et al., 1992b).

4. EXPERIMENTAL PROCEDURES TO IDENTIFY SSR MARKERS

During this course, we have chosen to use SSR markers because these DNA markers present a number of advantages. (i) Being PCR-based, they require a small amount of DNA which in principle allows the study of a large number of markers on a small sample of DNA. (ii) SRR markers are co-dominant which allows the characterisation of heterozygote as well as homozygote individuals. (iii) They do not require the use of DNA digestion by restriction enzymes and thus avoid the problems associated with partial digestions.

A number of microsatellite markers (SSR) will be mapped using RILs populations of *Medicago truncatula* which have been generated from crosses between the Jemalong line (female parent) and a genotype (DZA315.16) from an Algerian population (T. Huguet, unpublished results).

4.1. *Medicago truncatula* DNA extraction protocol

Adapted from : Stewart CN and Via LE (1993) " A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications " Biotechniques 14(5) 748-751.

Day 1
- Put 3 trifoliate leaves in a 2 ml Eppendorf tube containing ~8 sterilised glass beads.
- Leave tubes open. Dry overnight at 65°C.

Day 2
1. Reduce the dried leaves in powder by vortexing the tubes (30 to 50 sec).
2. Add 1 ml of extraction buffer.
3. Incubate in a water bath at 65°C for 20 min. While incubating, mix a few times by inverting the tubes.
4. Add 600 µl of chloroform and shake vigorously for 15 min at room temperature (RT).
5. Spin in a microfuge for 10 min at full speed (RT).
6. Transfer aqueous upper phase in another 2 ml Eppendorf tube (avoid taking debris from the pellet).
7. Add 600 µl of isopropanol (RT). Mix gently and centrifuge immediately in a microfuge for 20 sec at full speed (RT). Do not work with more than 6 or 8 tubes at a time.
8. Pour out the supernatant and rinse the pellet with 500 µl of 70 % ethanol (RT).
9. Spin in a microfuge for 10 mn at full speed (RT). Discard the supernatant. Recentrifuge briefly and eliminate with a micropipet any trace of supernatant. Dry the pellet under hood for 5 minutes (Do not let the pellet get too dry, otherwise it will be difficult to redissolve DNA).
10. Resuspend the pellet with 50 µl of sterile distilled water (gentle shaking, no vortex) and leave at 4°C overnight before diluting at 1/10 with water. Use for PCR reactions 1 µl of this dilution.

Extraction Buffer :
(Stock at room temperature)

For 100 ml :
2 g of hexadecyltrimethyl ammonium bromide (CTAB)
10 ml of Tris-HCl 1M pH8
28 ml of NaCl 5M
4 ml of EDTA 0.5M pH8
distilled water

Autoclave
0.5 ml of betamercapto-ethanol (to be added after autoclave)

4.2. PCR protocol

PCR can be made either in individual tubes (preferentially in thin wall tubes) or in a 96 well plate. In both cases, use a final volume of 20 µl.

The $MgCl_2$ concentration must be tested for each marker analysis (usually 1.5 to 2 mM). The annealing temperature depends on the Tm of the primers (usually 45 to 65°C). The time at the different temperature depends on the length of the fragment to amplify and on the power of the thermocycler. If the thermocycler has a heated lid, it is not necessary to add an oil layer.

4.2.1. Mix composition

Usually, a common mix for all the samples is prepared. It is distributed 19 µl into each tube or well, then it is added 1 µl of DNA diluted solution and a drop of mineral oil if necessary.

	1 sample (µl)	**Final concentration**
Distilled water	11.4	
Buffer 10X	2	1X
$MgCl_2$ (50 mM)	0.64	1.5 mM
dNTPs (1.25 mM each)	3.2	0.2 mM (each)
Primers (50 ng/µl each)	0.8 (each)	2 ng (each)
Taq polymerase (5 U/µl)	0.16	0.8 units
DNA (1/10 dilution of stock solution)	1	
	Total = 20 µl	

4.2.2. PCR buffer composition
- 20mM Tris-HCl (pH 8.4)
- 50 mM KCl

4.2.3. PCR program
- 4 mn at 94°C (denaturation)   1x
- 30 sec at 94°C (denaturation) ⌝
- 30 sec at 55°C (annealing)   40x
- 30 sec at 72°C (elongation)   ⌟
- 6 mn at 72°C (final elongation)   1x

4.3. Agarose gel electrophoresis

The volume of the gels and the number of loaded samples depend on the equipement. The agarose concentration depends on the length of the amplified fragments.
Besides the samples to analyse, add a DNA ladder to be used as a molecular weight standard DNA as well as a control without DNA, in order to check if self-amplification or accidental contamination occured.

4.3.1. Gel preparation

3.5 % agarose gel for amplified products of about 150bp

a.   10.5 g agarose

b.   300 ml of Tris Borate EDTA buffer (TBE) 0.5X

c.   Melt

d.   12 µl of 1 % Ethidium bromide (to be added just before pouring the gel and wear gloves !)

TBE buffer

(Stock at room temperature)

To be used to prepare gel and migration buffer

Tris-borate 44 mM

EDTA 1.25 mM

pH 8.0

4.3.2. Migration

Either rapid : 2 to 6 hours at high voltage (3.5V/cm) or slow : overnight at low voltage (1 to 1.5V/cm).

4.3.3. Revelation

The gel is photographed in UV light and the DNA analysis is performed on the picture.
Score the genotype of each F2 individual by the polymorphic bands. Use as standards control samples corresponding to DNA from each of the two parental lines, and DNA from the heterozygote (or an equimolar mixture of DNA from both parents).

5. GENETIC MAPPING METHODS

5.1. Mapmaker software

In large-scale mapping projects, analysis of the segregation data and generation of the genetic map cannot be achieved without a computer-implemented procedure [Koornneef, 1998].

The MAPMAKER [Lander, 1987] software will be used during this course. This software can be freely downloaded (http://www-genome.wi.mit.edu/ftp/distribution/software/mapmaker3/).

An online tutorial is also available (http://linkage.rockefeller.edu/soft/mapmaker/).

Hence, for a given DNA marker, each individual can be scored as homozygous for the female Jemalong parent (A) or homozygous for the male DZA315.16 allele (B), or heterozygous (H). Linkage is detected when the recombination frequency is significantly lower than 50 %. Once linkage is detected, it is necessary to convert the recombination frequency to map distance [Koornneef, 1992]. This conversion is needed to account for two facts : i) chromosomes in which two recombination events occurred between the marker and the locus of interest are counted as having no recombination event; and ii) recombination events can influence the probabilities of a second recombination event occurring in the vicinity, a phenomenon called interference. A reasonable estimate of map distance is given by the Kosambi function : $D = 25 \ x \ ln \ [(100 + 2r) / (100 - 2r)]$, where r is the recombination frequency expressed as a percentage, and D is the map distance in centiMorgans (cM) [Koornneef, 1992]. Map distances over adjacent intervals are additive whereas recombination frequencies are not.

5.2. Colormapping : a non-mathematical procedure for genetic mapping

Materials : Apple or IBM compatible computers, Excel program.

Methods :

In order to construct a colormap the determination of the genotypes of the markers for the individuals in a segregation population is needed. The genotypes which were used to deduce the colormap from numerical characters are taken from the raw data file generated during the EMBO course.

Conversion of the numerical genotypes to color symbols : The numerical genotypes are as follows : 1 and 3 maternal and paternal homozygous, respectively ; 2, heterozygous ; 4, paternal dominant genotypes ; 5, maternal dominant genotypes ; 0, missing data. These genotypes are converted to colors by *EXCEL* program using short Macro programs utilizing the *"Format\Cells...\Patterns\Cell shading\color"* functions. These Macro programs convert the white shading of the cells containing the numerical characters to colors as follows : Character 0, 1, 2, 3, 4, and 5, are converted to gray, yellow, green, purple, metal-blue, and light green, receptively.

Procedure of colormapping using *EXCEL* program : Colormapping is carried out by the computer program *EXCEL* which is available for both IBM compatible and Macintosh/Apple computers. Data are interchangeable between the computer systems through ASCII text file format. Genotypes are converted to colors (see above) which results in a color pattern (see Results). Care has to be taken to introduce the genotype scores of a new marker in the same order as the order of the individual plants in the colormap to be used. Mapping manipulation starts by opening a window with two horizontal parts, one is steady, this contains the color pattern of the new marker in one row (usually at the bottom of the screen), the other is rolled and contains as many rows of the markers from the colormap as possible or convenient. The markers in the later window are scrolled from the first marker until the last one by using the mouse on the scrolling arrowhead on the right side of the window. The human eyes recognize easily when the color pattern of the new marker and that of the markers in the colormap are similar. When best fit is achieved, scrolling is stopped and the new marker is inserted into the colormap by the *"Cut"* and the *"Insert cut cells"* functions of the *"Edit"* menu. This operation takes less then one minute after having the combined color genotype data of the new marker introduced into the spreadsheet. Program *EXCEL* for IBM compatible computers allows to handle 256 columns and 16384 rows that is there is room for more than 4,000,000 genotype data. The best way to handle the data is to have a master file ("frame colormap") containing only limited number of core markers (preferentially codominant ones), as well as individual files for each LG with all markers mapped. Conveniently "rough" mapping is done first in the "frame colormap". Once a new marker has been mapped then it is transferred to the file containing the appropriate LG where "fine" mapping can be performed. It is of choice whether the newly mapped marker becomes the member of the core markers.

Generation of color pattern from raw data, and the colormap : Genetic mapping has been based till now on mathematical analysis of the incidence rate of the genotypes for the loci segregated in the individuals of a mapping population. The individual genotype for a locus can be homozygous for the maternal allele, homozygous for the paternal allele or heterozygous. This kind of genotype allocation can only be used when the evaluation is codominant. In the case of dominant inheritance, however, heterozygous configuration is indistinguishable from the dominant homozygous. Traditionally in diploid organisms the following symbols were introduced taking $\mathbf{a}_1$ and $\mathbf{a}_2$ as the maternal and paternal alleles, respectively : $\mathbf{a}_1\mathbf{a}_1$, maternal homozygous ; $\mathbf{a}_2\mathbf{a}_2$, paternal homozygous ; $\mathbf{a}_1\mathbf{a}_2$, heterozygous ; $\mathbf{a}_1\textbf{-}$, maternal dominant ($\mathbf{a}_1$ is dominant over $\mathbf{a}_2$, therefore either $\mathbf{a}_1\mathbf{a}_1$, or $\mathbf{a}_1\mathbf{a}_2$) ; $\mathbf{a}_2\textbf{-}$, paternal dominant ($\mathbf{a}_2$ is dominant over $\mathbf{a}_1$, therefore either $\mathbf{a}_2\mathbf{a}_2$, or $\mathbf{a}_1\mathbf{a}_2$). These scoring symbols obey the traditional designation highlighting the genotype of both chromosomes of the homologue pair. The application of computer programs for the calculation of genetic linkages for several markers, as well as the ambiguity which may arise from the genetic configuration of two markers with heterozygous genotypes ($\mathbf{a}_1\mathbf{a}_2/\mathbf{b}_1\mathbf{b}_2$ is indistinguishable from $\mathbf{a}_2\mathbf{a}_1/\mathbf{b}_1\mathbf{b}_2$) single character symbols representing combined genotypes or scores have been introduced (see for example MAPMAKER/EXP 3.0). These symbols (numeric or alphabetic) can be defined for example as follows : 1 = maternal homozygous ($\mathbf{a}_1\mathbf{a}_1$) ; 3 = paternal homozygous ($\mathbf{a}_2\mathbf{a}_2$) ; 2 = heterozygous ($\mathbf{a}_1\mathbf{a}_2$ or $\mathbf{a}_2\mathbf{a}_1$) ; 4 = paternal dominant ($\mathbf{a}_2\textbf{-}$) ; 5 = maternal dominant ($\mathbf{a}_1\textbf{-}$).

For the estimation of the recombination frequencies between marker pairs mathematical calculations are generally applied (most frequently the maximum likelihood method is used ; ALLARD 1956). For these calculations (which are computerized nowadays), the genotypes of the loci of the individuals in a segregation population have to be fed into the computer one after another. Genotypes are scored as numerical or alphabetical characters (see above) and stored favorably in a spreadsheet file (in our case program *EXCEL* is used). The created file contains the genotypes of the markers for the individuals in a mapping population

Properties of the colormaps : The colormap presents the marker order and the combined genotypes of the appropriate individuals for the given genetic markers. This primary colormap, however, does not reflect the proper genetic distances between the markers. Markers and their pertaining genotypes can be arranged according to the genetic distances between the marker pairs. Genetic distances have to be calculated from the recombination frequencies using the Haldane or Kosambi map functions. Missing genotypes of the chromosomal regions located between the mapped markers are represented by gray. One can see that the ordered genetic markers determine fixed positions on the chromosomes to serve as reference points (i.e. core markers) for subsequent mapping.

A highly saturated genetic map possesses thousands of markers, yet these reference points are far from each other in terms of physical distance. Core markers are flanked by hundreds of genes with undetermined genotypes, that is fixed points represent only a very tiny portion of the whole genome. To handle the LGs as continuos genotype segments instead of interrupted ones, the genotypes of the flanking regions can be predicted as described. The prediction can be carried out easily on the colormap to get a continuos color pattern and thereby a continuos genotype strip. The principle of the predictions are as follows :

(1) if the combined genotype of the two flanking markers was the same the following assumption was made : the genotypes of the intermediate genes (chromosomal segment) were inferred from that of the flanking markers because no recombination event was supposed (the principle of maximum parsimony). It means that if the flanking markers display maternal, paternal homozygous, or heterozygous genotypes than the genotypes of the intermediate genes are most likely maternal or paternal homozygous, or heterozygous, respectively. However, it is important to keep in mind that even, but not odd number of recombination could have occurred between the markers, consequently an "island of a different genotypic segment" (abbreviated as island afterwards) could interrupt the region. These hidden islands can be revealed by further fine genetic mapping in the region. The shorter the distance between the reference markers, the more probable that no even number of recombination took place, that is no island was present.

(2) if the combined genotype of the two flanking markers are different it is supposed that only one recombination took place (the principle of maximum parsimony) somewhere between the markers. Recombination is supposed to take place halfway between the flanking markers, therefore sequences located in one or the other side of the assumed recombination site "inherit" the genotype of the closer marker. Three or more odd number of recombination would result in island(s). The shorter the genetic distance between the markers, the more probable that no three or more odd number of recombination occurred. Determination of the genotype of additional markers in the region makes the localization of the recombination spots(s) more accurate.

(3) if the last genotype at the end of the LG was missing no recombination was supposed, consequently the same genotype was predicted that the closest proximal flanking marker had. Similarly, the genotype of the chromosomal region located distal from the last marker to the telomer was inferred from the genotype of the last marker. It is not excluded that further mapping reveals recombination. In these cases corrections have to be made. The closer the end marker to the telomer, the less probable the recombination event is.

(4) Dominant markers located between two codominant markers provide more reliable prediction of genotypes and recombination events. On the other hand, if the genotype is maternal dominant (combined genotype is 5, that is pale green), no additional information is obtained since this genotype is heterozygous or maternal homozygous, therefore the predicted recombination should be placed in halfway between marker.

It is advised that determined and predicted genotypes should be distinguished. For that purpose numerical characters (except 0, 1, 2, 3, 4, 5,) can be used in the data sheet where scores are stored. However, when computer calculation is to be carried out, these characters have to be converted either to zero (missing data), or to number 1, 2, 3, 4, 5. Zero replacement gives the so called non-predicted raw data, while predicted data will yield predicted values for the calculation of linkage.

Colormapping as a novel procedure for genetic mapping of new markers : Colormaps displayed on a monitor by the help of the *EXCEL* program can be used efficiently for genetic mapping. The principal role of colormapping is to find the best match between the color pattern of the new marker to be mapped and that of the colormap. Program *EXCEL* makes it possible to divide the opened window on the monitor in two horizontal parts at the same time. This feature of the program is used to find the correct map position of the new marker. The first window contains as many marker rows of a LG as possible including all or several individuals (columns) in the mapping population. This can be achieved by adjusting the column height and row width to the desired value. Below this, another narrow window is opened containing the color genotypes of the individuals for the new marker to be mapped (the order of the individuals must be the same in both window). Mapping is done by visual comparison of the color pattern of the markers in the upper and lower windows. To find the position of the new marker, the rows of the upper window are rolled upwards to show the color pattern of the upcoming markers until the best match of the pattern is found. If the end of the LG is reached without unambiguous matching, the next LG is started to be screened, and so on. During this procedure visual evaluation is needed to catch the region of the map to which the new marker shows the highest degree of color pattern match (thereby the less recombination event). When this region is found the new marker is inserted between those two markers where the new marker fits the best. By inserting a new marker into this region of the colormap the minimal number of color transition (recombination) is generated. This mapping procedure takes usually less time than running any of the computer programs which calculate recombination frequencies and map distances using mathematical formulas. On the other hand, colormapping produces the same order of loci as mathematical approaches without determining the genetic distance.

Troubleshooting with colormap : One of the most advantageous properties of the colormaps is their powerful troubleshooting capacity. Troubleshooting is usually done on the raw data introduced as numerical genotypes into a matrix. On one hand, the genotypes should be checked again on the autoradiograms or on the photos, on the other hand, the genotypes have to be verified by proofreading the characters one by one. These procedures are time consuming, laborious, and still doubtful. Whereas, a newly introduced marker into the colormap according to the position of its best fit immediately give a spectacular color view of the new marker and its neighborhood. One can obviously perceive whether the color of the new marker matches unambiguously their flanking regions (markers) for each individual or not. The following situations may occur :

(1) the genotypes of the flanking markers are the same, and the new marker matches this pattern ; no discrepancy that is no recombination occurred.

(2) the genotypes of the flanking markers are the same but the genotype of the new marker differs from both ; in this case either recombination events occurred on both sides of the new marker (see for example the genotype of the U286 for individual 70 and 103 on LG 7 in Figure 5.), therefore islands were generated, or errors were made.

(3) the genotypes of the flanking markers are different and the genotype of the new marker coincide with one of the flanking genotypes. In this case, the recombination event has occurred between the new marker and the marker with the diverse genotype.

(4) the genotypes of the flanking markers are different and the genotype of the new marker does not coincide with none of the flanking genotypes. In this case, recombination events have occurred on both sides of the new marker or miss-genotyping occurred.

The presence of an island can be confirmed by mapping new marker(s) with the same genotype next to the marker by which the island was generated in the appropriate individual, or thrown out if repeated experiments disproved the genotype. In the case of an island one should keep in mind that it may be the result of miss-genotyping caused by personal mistakes or technical errors. The color display of the genotypes (that is the colormap) is extremely powerful to highlight ambiguities, consequently facilitates troubleshooting. In the case of discrepancy the genotypes of only those marker/individual combinations have to be rechecked or re-determined which produced conflicting results. This kind of troubleshooting helped us tremendously during our mapping work to pick up discrepancies and to correct genotypes, consequently experimental mistakes could be eliminated or reduced to a large extent.

Discussion :

As compared to the concept of graphical genotype described by YOUNG and TANKSLEY (1989), colormaps have a major advantage: the alleles appeared in a locus of the homologue chromosome pairs are integrated into a combined genotype and this is displayed as **colors**. Maternal and paternal homozygous, heterozygous, or maternal and paternal dominant configurations are displayed by different colors. The resulting color picture, that is the genotypes of the ordered markers of the individuals in a segregation population is the colormap. Colored genotypic pattern of any genomic segment can be analyzed more easily as compared to either graphical or numerical patterns. Recently the representation of the

genotypes by color was described by Boutin et al. (1995) which was used for marker-based pedigree analysis.

Colormaps are extremely useful in plant genetics and breeding when whole genome analysis is required. The selection of individuals with a desired genotype among the progeny is simple and straightforward even if more than one loci is looked at, e.g. in the case of multiple loci selection. Colormaps display the genetic composition of the individuals in a segregation population giving a concise and comprehensive color image where the entire genome is highlighted. Taking the color genotype of an individual plant, transition from one color to another mirrors recombination event. Similarly to the graphical genotype, cis or trans heterozygous configuration of the alleles can not be revealed by colormaps, consequently cis or trans heterozygous configuration of the alleles has to be determined, if necessary, by progeny analysis.

Colormaps can be used for genetic mapping. This new approach is called colormapping since the location of the new markers is found by the best match of the color pattern of the new marker compared to the already existing colormap. Strikingly, colormapping is at least as efficient, fast and powerful as mathematical calculations to find the position of the new markers. This is achieved by the fact that the perception of color patterns by the human eyes is extremely sensitive (Chaparro *et al.,* 1993). Colormapping is carried out by the help of a computer spreadsheet program called *EXCEL* which is available for both Macintosh and IBM computers, in addition *EXCEL* is compatible with other spreadsheet programs e.g. *LOTUS 1-2-3*. The new marker to be mapped is shown as a color patterned row in a window separate of the other one displaying the already established colormap, which is followed by finding the best match between the patterns of the new marker and the appropriate region of the colormap as described in Results. After inserting the new marker into its adequate position it is becoming a new component of the map. Since colormapping does not determine the genetic distances, a colormap displays the order of the markers disproportionately. Genetic distances have to be calculated by other methods and the appropriate distances can be incorporated afterwards into the colormap as shown in Figure 3B. It has to be emphasize, that the construction of colormaps can be started from scratch. Similar color pattern of markers can be separated one by one, and by continuing this sorting out LGs can be established as was performed with scrambled markers for the eight LG of alfalfa (data not shown).

Colormap can be used to find linkage relationship between markers and partial linkage groups if mathematical approaches give ambiguous results as was demonstrated for LG 6 and 7 in this study, and in the paper by Kaló *et al.* (2000). If the segregation ratio of given markers are extremely distorted as in the case of some chromosomal regions in the outcrossing diploid alfalfa mapping population (Kiss *et al.* 1993), calculating the recombination frequencies may result in false values (heterozygous marker pairs are not neutral but enhance linkage). Displaying the appropriate color patterns visual sensation can perceive the more important homozygous links or color transitions, by which linkage or non-linkage can be strengthened or rejected.

An extremely important advantage of colormaps to our experience is the immediate realization of surprising results or experimental mistakes. The appearance of a different color that is an "island" in a genetically uniform region may indicate double recombination events or mistakes in the determination of the genotype or typing. Repeating the determination of the

genotype for the appropriate locus can confirm or reject the genotype of the island. This is of great help when genotyping is carried out by RAPD or other PCR based techniques with non-specific primers, since PCR amplifications may result in false patterns for different reasons.

RFLP patterns can also produce incorrect genotypes if sampling mistakes, slot shift or mixing up individual plants occurred. In these cases colormaps highlights strikingly the irregular pattern. Improper genotyping can also occur if wrong allele allocation was made. This can happen when the parents of the F2 population were heterozygous for the appropriate locus and the origin of the alleles can not be determined unambiguously. In this case, colormap highlights discrepancies by showing opposite color pattern, and the appropriate genotype can be corrected.

Troubleshooting is extremely important in genetic mapping. Some experimental mistakes can not be detected by computer when recombination frequencies are calculated using mathematical algorithms. Most of the above errors would be resulted in an outcome called "unlinked" and some mistakes would give longer genetic distances. Analyzing the color pattern of colormaps genotype errors can be corrected after which more reliable values can be calculated.

Colormap(ping) is not restricted only to plant genomes, it is applicable for any Genome Projects like human, *Drosophila, Caenorhabditis*, yeast and others, which could benefit from it by displaying the combined color genotypes making possible genome analysis, comparison, troubleshooting and mapping.

6. REFERENCES

- Allard, R.W. (1956) Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia **24 :** 235-278.
- Botstein, D., White, R.L., Skolnick, M., and Davis, R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32** : 314-331.
- Boutin, S.R., Young, N.D., Lorensen I.L., and Shoemaker R.C. (1995) Marker-based pedigrees and graphical genotypes generated by suprgene software. Crop Sci. **35 :** 1703-1707.
- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., and Meyerowitz, E.M. (1988) Restriction fragment length polymorphism linkage map for Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA **85** : 6856-6860.
- Cotton, R.G. (1997) Slowly but surely towards better scanning for mutations [published erratum appears in Trends Genet 1997 May ;13(5) : 208]. Trends Genet. **13** : 43-6.
- Hauser, M.T., Adhami, F., Dorner, M., Fuchs, E., and Glossl, J. (1998) Generation of co-dominant PCR-based markers by duplex analysis on high resolution gels. Plant J. **16** : 117-25.
- Kalo, P., Endre, G., Zimanyi, l., Csanadi G., and Kiss, G.B. (2000) Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). Theor. Appl. Genet. **100** : 641-657.
- Kiss, G.B., Csanadi G. Kalman K., Kalo P., and Ökresz L. (1993) Construction of a basic genetic map for alfalfa using RFLP, RAPD, isozyme and morphological markers. Mol. Gen. Genet. **238 :** 129-137.

- Koornneef, M., Alonso-Blanco, C., and Stam, P. (1998) Genetic analysis. In Arabidopsis Protocols, J.M. Martinez-Zapater and J. Salinas, eds (Totowa, New Jersey: Humana Press), pp. 105-17.

- Koornneef, M., and Stam, P. (1992) Genetic analysis. In Methods in Arabidopsis research, C. Koncz, N.-H. Chua and J. Schell, eds (Singapore: World Scientific Publishing Co), pp. 83-99.

- Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newburg, L. (1987) MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1** : 174-81.

- Michaels, S.D., and Amasino, R.M. (1998) A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. Plant J. **14** : 381-5.

- Neff, M.M., Neff, J.D., Chory, J., and Pepper, A.E. (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms : experimental applications in *Arabidopsis thaliana* genetics. Plant J. **14** : 387-92.

- Reiter, R.S., Williams, J.G., Feldmann, K.A., Rafalski, J.A., Tingey, S.V., and Scolnik, P.A.(1992a) Global and local genome mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs. Proc. Natl. Acad. Sci. USA **89** : 1477-81.

- Reiter, R.S., Young, R.M., and Scolnik, P.A. (1992b) Genetic linkage of the Arabidopsis genome : Methods for mapping with recombinant inbreds and random Amplified Polymorphic DNAs (RAPDs). In Methods in Arabidopsis research, C. Koncz, N.-H. Chua and J. Schell, eds (Singapore : World Scientific Publishing Co), pp. 170-190.

- Vos, P. (1998) AFLP TM fingerprinting of Arabidopsis. In Arabidopsis Protocols, J.M. Martinez-Zapater and J. Salinas, eds (Totowa, New Jersey : Humana Press), pp. 147-155.

- Vos, P., Hogers, R., Bleeker, M., reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995) AFLP : a new technique for DNA fingerprinting. Nucl. Acids Res. **23** : 4407-4414.

- Williams, J.G., Reiter, R.S., Young, R.M., and Scolnik, P.A. (1993) Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. Nucl. Acids Res. **21** : 2697-702.

- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. **18** : 6531-6535.

- Young, N.D. and Tanksley, S.D. (1989) Restriction fragment length polymorphism maps and the concept of graphical genotypes. Theor. Appl. Genet. **77 :** 95-101.

7. TIME SCALE FOR MODULE 3

<u>Day 1 (22d November)</u> :
- Preparation of plant material for DNA extraction : (30 min)

<u>Day 2 (23rd November)</u> :
- DNA isolation (4 hours)

<u>Day 3 (26th November)</u> :
 - DNA quantitation, preparation of primers and definition of amplification conditions (2 hours)

<u>Day 4 (27th November)</u> :
- PCR amplification and electrophoresis (4 hours : not continuously)

<u>Day 5 (28th November)</u> :
- Reading the electrophoresis patterns, preparing data (1 hour)

<u>Day 6 (29th November)</u> :
- Genetic mapping (2 hours)

<u>Day 7 (30th November)</u> :
- Genetic analysis of data and conclusions (2 hours)

8. BAC FINGERPRINTING AND ALLIED METHODS

8.1. Introduction

The ability to clone and manipulate large segments of DNA has become a key factor in the success of many plant genomics projects. The availability of large insert cloning systems has had particular utility for map-based cloning projects, whole genome physical map assembly, and complete genome sequencing efforts in organisms such as *Arabidopsis*, humans and *Caenorhabditis elegans*. Among the large insert cloning vehicles in use today, the bacterial artificial chromosome system, or BAC, has been shown to have the greatest utility, both in terms of insert size and stability, ease of manipulation, and the low frequency of chimeric cloning products (Shizuya, 1992).

The BAC system is based on the *Escherichia coli* F factor. Strict control of F factor replication in the bacterial host limits copy number of BAC DNA to between 1 and 2 copies per cell, which is presumed to reduce the potential for inter-plasmid recombination. BAC libraries of plant DNA often contain inserts exceeding 200Kbp in size, although mean size is typically in the range of 100 to 150Kbp. In *Medicago truncatula*, with average BAC insert sizes near 120 kbp and gene densities in the range of 6.5 kb/gene (based on limited surveys to date), individual BAC clones may contain an average of 20 predicted genes. As a consequence, genome walking typically proceeds by tens of genes at a time, and moderate sized BAC contigs can contain hundreds of genes. From a genetic perspective, short chromosome walks in genome regions of average recombination frequency (i.e., 300 kbp/cM) can easily encompass a few centimorgans of genetic distance. Thus, the advent of BAC libraries, combined with advances in DNA sequencing technology, now enables individual laboratories to easily undertake detailed characterization of candidate genome regions.

One of the key features in the utility of BAC libraries is the fact that they are curated in ordered microtiter plates, typically with 384 samples per plate. Robots such as the Genetix Q-bot facilitate library replication, and they also provide the means to produce high-density nylon filters for the identification of BAC clones based on DNA-DNA hybridization. A typical high-density filter contains approximately 18,000 individual BAC clones printed in duplicate, with the ability to represent greater than 4X coverage of the *Medicago truncatula* genome on a single filter. The ordered microtiter plate format also facilitates preparation of DNA multiplexes that allow rapid screening of several-fold genome coverage by PCR-based methods. In combination with BAC end sequencing to characterize flanking regions of BAC clones and contigs, high-density filters and PCR-multiplexes provide particularly valuable tools for the identification of physically adjacent BAC clones during chromosome walking projects.

As mentioned above, BAC libraries have also proven invaluable for the construction of whole genome physical maps in species ranging from Arabidopsis (Marra et al., 1999 ; Mozo et al., 1999) to humans (McPherson et al., 2001). In contrast to targeted chromosome walking projects, efforts to map entire genomes depend on computational tools, such as Contig FPC (Soderland et al., 2000), to establish the overlap and relative order of BAC clones. A key protocol in many whole genome physical mapping projects is restriction enzyme digestion of BAC clones (Mara et al., 1997) and the separation and sizing of the resulting DNA fragments on high-percentage agarose gels. Although alternative fingerprinting methods are available, the

simplicity of the agarose gel method has established it as the method-of-choice in most laboratories. When the restriction patterns, or "fingerprints", of two BAC clones are substantially similar, the clones are predicted to form a contig of overlapping clones. For complex genomes, such as Arabidopsis and humans, the assembly of whole genome physical maps has required the inclusion of additional data types to assess BAC clone overlap and to resolve ambiguities in fingerprint data. One of the most common data types used in conjunction with DNA fingerprints is the demonstration of homologous relationships between BAC clones. Such cross homology between BAC clones is often based on hybridization (e.g., using high-density filters and ^{32}P-dCTP labeled probes) or PCR criteria (e.g., DNA multiplex). The regions of homology between BAC clones are also referred to as "sequence tagged sites" or STSs. When STSs also represent genetic markers, they have the increased value of anchoring emerging BAC contigs to the genetic map. In addition to providing another criterion for ordering contigs, linking the genetic and physical map can serve to locate BAC contigs within the genome well in advance of complete assembly. Finally, the availability of a complete physical map of *Medicago truncatula* will ultimately preclude the need for chromosome walking. In this case, the order of BAC clones will be prior knowledge, and one simply needs to correlate genetic markers with BAC clones to delimit physical intervals.

Whole genome physical maps are also extremely valuable for genome sequencing projects. In the case of a BAC-by-BAC strategy, knowledge of the order and extent of overlap between BAC clones is essential to identify a minimum tiling path for efficient sequencing. Alternatively, in the case of a random shotgun approach, the combination of a physical map with complete BAC end sequence information can provide an important scaffold for placement and assembly of emerging sequence contigs.

During this section of the course, you will become familiar with the following techniques :

（4） DNA fingerprinting and contig assembly of BAC clones.

（5） Hybridization to high-density filters to identify *Medicago truncatula* BAC clones.

（6） The use of a BAC DNA multiplex to identify and distinguish paralogous members of a small gene family.

8.2. Overview of Activity

We will fingerprint and assemble BAC contigs based on the analysis of a set of 96 BAC clones. Each group will work with the identical set of BAC clones, and thus we will have an opportunity to compare the uniformity of data between gels. In any whole genome mapping effort, gel-to-gel variation must be kept to a minimum

We will begin by growing BAC clones and extracting DNA in a 96 well format. The purified DNA will serve as substrate for restriction enzyme digestion with HindIII. The resulting DNA fragments will be resolved on a 1.0 % agarose gel and stained with a fluorescent dye to view the DNA and record the data. Fingerprints will be compared between clones to predict overlap and assign contigs.

In parallel to the analysis of DNA fingerprints, we will also perform DNA hybridization to high-density nylon membranes to establish overlap between BAC clones. The results of DNA hybridization analysis will be compared to those from a DNA multiplex PCR analysis of the same BAC clones. Both the DNA hybridization and PCR results will be used in conjunction with DNA fingerprint data to enhance contig assembly. We will also consider the relative merits of hybridization versus PCR methods for the identification of BAC clones with homology to known genes.

List of primary protocols
1. BAC DNA Isolation (96 well)
2. Restriction Enzyme Digestion of BAC DNA
3. High Resolution Agarose Gel Electrophoresis
4. Filter hybridization (protocol to be provided)
5. DNA Multiplex for PCR (protocol to be provided)

List of secondary protocols
1. Individual BAC DNA Isolation
2. BAC End Sequencing

8.3. BAC DNA isolation

8.3.1. Protocol 1. 96-well BAC DNA Extraction Protocol for Fingerprinting

This protocol is suitable for analysis of large numbers of BAC clones, in particular where computational methods will be used for contig assembly.
1. Fill each well of a Qiagen 96-well round bottom culture block with 1.5ml LB/12.5 µg/ml chloramphenicol.
2. Thaw a single 384-well BAC library plate and use the 96-pin replicator tool to inoculate four 96-well blocks, as prepared in step 1.
3. Incubate the bacterial cultures for18 hours with shaking at 350 rpm in a HiGrow shaking incubator, or equivalent.
4. Spin the 96-well block for 15 minutes at 3000 rpm/4°C to pellet the bacteria.
5. Decant the supernatant into a waste container and use a stack of paper towels to blot the surface of each culture block until dry.
6. Using a multi-channel pipette, add 100 µl of solution 1 (see below) to each of the 96 samples and vortex for 1 minute to resuspend the bacterial pellet.
7. Place the samples on ice and add 200 µl of solution 2. Swirl the sample block firmly twenty times. Avoid vigorous shaking, so as not to shear the high molecular weight BAC DNA. Wait 5 to 10 minutes to allow efficient cell lysis.
8. Add 150 µl of solution 3 and swirl as above, twenty times.
9. Place the sample block in a -20°C freezer for 15 minutes.
10. Centrifuge the sample block for 20 minutes at 3000 rpm/ 4°C to pellet debris.
11. Prepare a 96-well filter apparatus by stacking a Whatman filter plate (800 µl, 25 µM MBPP + PP 0.45 um) on top of a Whatman 750 µl polypropylene receiving plate

("uniplate"). [It may be necessary to cut off the lower rim of the receiving plate to allow the stacked plates to fit into the centrifuge basket.]

12. Transfer 440 µl of supernatant from step 10 to the filter apparatus prepared in step 11. Be careful to not transfer precipitate.

13. Centrifuge the filter apparatus for 30 minutes at 3000 rpm/4°C to recover the DNA-containing filtrate.

14. Check for clogged filters – if filters are clogged re-spin the plate for additional 20 min.

15. Add 310 µl isopropanol to the flow-through that is now contained in the receiving plate. Apply an adhesive plate seal, and invert plate 5 times. Avoid getting isopropanol on the surface of the plate, as this will break down the adhesive and cause cross contamination of samples. (We use a firm foam pad cut to the size of the receiving plate, in conjunction with another receiving plate, to press tightly against the seal while inverting.)

16. Chill the isopropanol mixture overnight at -20°C.

17. Spin the receiving plate for 30 minutes at 3000 rpm/4°C to collect the DNA pellet.

18. Carefully decant the supernatant and blot the plate onto a clean paper towel.

19. Add 500 µl of 70 % EtOH and shake for 5 minutes on an orbital shaker.

20. Spin for 15 minutes at 3000 rpm/4°C.

21. Carefully decant the supernatant and blot the plate onto a clean paper towel.

22. Add 200 µl of 95 % EtOH and spin for 15 minutes at 3000 rpm/ 4°C.

23. Carefully decant the supernatant and blot the plate onto a clean paper towel.

24. Let plates dry on their side for at least 1 hr in a laminar flow hood.

25. Add 20 µl TE to each of the 96 wells and store overnight at 4°C to re-dissolve the DNA pellet.

26. Vortex for 1 minute to ensure complete resuspension of DNA.

27. For storage longer than 24 hours, place samples at -20°C.

Solution 1		Final Concentration
MilliQ water	955 ml	
Glucose	9.0 g	50 mM
EDTA	20 ml 0.5 M stock	10 mM
Tris-HCL (pH 8.0)	25 ml 1.0 M stock	25 mM

Autoclave and store at 4°C

Solution 2		Final Concentration
Sterile MilliQ water	8/10 volume	
NaOH	1/10 volume 2N stock	2N
SDS	1/10 volume 10 % w/v stock	1 % w/v

Solution 3
Ammonium Acetate   578 g
Dissolve in 500 ml of purified water and complete to 1l

8.3.2. Alternative Protocol 2 : Isolation of individual BAC clones using Qiagen Plasmid Mini Kit columns

This protocol is suitable for analysis of individual BAC clones and BAC end sequencing.

Important notes before starting :
• Add the provided RNase A solution to Buffer P1 before use. Use one vial of RNase A (spin down briefly before use) per bottle of Buffer P1, to give a final concentration of 100 µg/ml.
• Check Buffer P2 for SDS precipitation due to low storage temperatures.
If necessary, dissolve the SDS by warming to 37°C.
• Pre-chill Buffer P3 to 4°C.

1. Pick a single colony from a freshly streaked selective plate and inoculate a starter culture of 2-5 ml LB medium containing the appropriate selective antibiotic.
2. Inoculate 0.2-0.5 ml of starter culture into 100 ml of selective LB medium in a 250 or 500 ml flask. Incubate for 14-16 h at 37°C with vigorous shaking (250-300 rpm). Using a tube or flask with a volume of at least 2.5 times the volume of the culture, improves cell growth.
3. Divide the culture into two 50-ml disposable polypropylene tubes, and harvest the cells by centrifugation at 3500 rpm for 30 min in a Beckman-Coulter Allegra 6KR swing bucket centrifuge. Remove all traces of supernatant by inverting the open centrifuge tube until all medium has been drained.
    •If you wish to stop the protocol and continue later, freeze the cell pellets at –20°C.
4. Resuspend the bacterial pellet in 5 ml/tube Buffer P1. Pool the duplicate pellets of a clone into a single tube to get ~ 10ml resuspension per clone. Save the empty centrifuge tube for step 8. Ensure that the RNase A has been added to Buffer P1. The bacteria should be resuspended completely by vortexing or pipetting up and down until no cell clumps remain.
5. Add 10 ml Buffer P2, mix gently but thoroughly by inverting 4–6 times, and incubate at room temperature for 5 min.
    Do not vortex as this will result in shearing of genomic DNA. The lysate should appear viscous. Do not allow the lysis reaction to proceed for more than 5 min.
    After use, the bottle containing Buffer P2 should be closed immediately to avoid acidification of Buffer P2 from CO_2 in the air.
6. Add 10 ml of chilled Buffer P3, mix immediately but gently by inverting 4–6 times, and incubate on ice for 15-20 min.
    Using chilled Buffer P3 and incubating on ice enhances precipitation. After addition of Buffer P3, a fluffy white material forms and the lysate becomes less viscous. The precipitated material contains genomic DNA, proteins, cell debris, and SDS. The lysate should be mixed thoroughly to avoid localized potassium dodecyl sulfate precipitation.

7. Centrifuge at 3500 rpm for 30-60 min in a Beckman-Coulter Allegra 6KR swing bucket centrifuge 4°C. Promptly transfer supernatant containing plasmid DNA into duplicate 50 ml polypropylene tube saved from step 4.

8. Re-centrifuge at 3500 rpm for 30-60 min in a Beckman-Coulter Allegra 6KR swing bucket centrifuge 4°C.

    This second centrifugation step should be carried out to avoid applying suspended or particulate material to the QIAGEN-tip. Suspended material (causing the sample to appear turbid) can clog the QIAGEN-tip and reduce or eliminate gravity flow, thereby significantly slowing progress through each of the subsequent steps that utilize gravity flow.

9. Equilibrate a QIAGEN-tip 100 by applying 4 ml Buffer QBT, and allow the column to empty by gravity flow.

    Flow of buffer will begin automatically by reduction in surface tension due to the presence of detergent in the equilibration buffer. Allow the QIAGEN-tip to drain completely. QIAGEN-tips can be left unattended, since the flow of buffer will stop when the meniscus reaches the upper frit in the column.

10 Apply the supernatant from step 8 to the QIAGEN-tip and allow it to enter the resin by gravity flow. Filtering the supernatant by placing a cone of miracloth into the Q-100 column will exclude large particulates, and is recommended. The supernatant should be loaded onto the QIAGEN-tip promptly. If it is left too long and becomes cloudy due to further precipitation of protein, it must be re-centrifuged or filtered through miracloth before loading to prevent clogging of the QIAGEN-tip.

11. Wash the QIAGEN-tip with 2 x 10 ml Buffer QC.

    Allow Buffer QC to move through the QIAGEN-tip by gravity flow. The first wash is sufficient to remove all contaminants in the majority of plasmid DNA preparations. The second wash is particularly necessary when large culture volumes or bacterial strains producing large amounts of carbohydrates are used.

12. Elute DNA with 3-5 aliquots of 1.0-1.5 ml (to a total of 5.0-5.5 ml elution volume) of Buffer QF pre-warmed to 65°C. Using smaller aliquots keeps the elution buffer from excessive cooling, and thereby improves DNA recovery. Collect the eluate in a 10 ml glass tube, or into 15 ml disposable polypropylene tube.

13. Aliquot the 5 ml of eluate into six 1.5 ml eppendorf tubes @ 0.90 ml/tube. Precipitate DNA by adding 0.7 volume (0.65 ml) of room-temperature isopropanol to the eluted DNA. Mix and store tubes overnight at 4°C, or centrifuge immediately at 13-15,000 rpm in benchtop centifuge for 30 min at 4°C. Carefully decant the supernatant.

    All solutions should be at room temperature in order to minimize salt precipitation, although centrifugation is carried out at 4°C to prevent overheating of the sample.

    Isopropanol pellets are also more loosely attached to the side of the tube, and care should be taken when removing the supernatant.

14. Wash DNA pellets with 1 ml of room-temperature 70 % ethanol, and centrifuge at max speed in bench top centrifuge (>15,000 rpm for 10 min. Carefully decant the supernatant without disturbing the pellet. Quick spin the microfuge tubes, and remove supernatant with a pipette, taking care not to disturb the pellet.

The 70 % ethanol removes precipitated salt and replaces isopropanol with the more volatile ethanol, making the DNA easier to redissolve.

15. Air-dry the pellet for 15–30 min, and redissolve the DNA in 15-25 µl of 10 mM Tris·Cl, pH 8.5. Allow the pellet to dissolve in resuspension buffer for 1-15 min followed by pipetting the DNA up and down, or brief vortexing. Excessive pipetting or vortexing may cause shearing. Overdrying the pellet will make the DNA difficult to redissolve. DNA dissolves best under slightly alkaline conditions ; it does not easily dissolve in acidic buffers.

16. Pool eluates of each clones prep into one of the eppendorf tubes (will be 90-150 µl).

Determination of yield

Use 1-4 µl of eluate to determine the yield UV spectrophotometry. Based on spec. quantification, run 1-5 µl (~50-100 ng), on a 0.8% agarose gel to check for DNA quality.

Typical yields from a 100 ml culture are >10 µg (and up to 30 µg) of BAC DNA.

8.4. Restriction enzyme fingerprinting

8.4.1. Restriction digestion of BAC DNA

1. Gently shake DNA in microtiter plate and quick-spin in centrifuge before aliquoting into digestion plates. If sample is frozen prior to use, be certain that each well is thawed completely before use.

2. Prepare a master mix of all components, except DNA (DNA must be put into well individually.)

Components	per well	2 plates	8 plates	16 plates	20 plates
Buffer C	*1.0 µl*	*220 µl*	*880 µl*	*1760 µl*	*2000 ml*
100 x BSA	*0.1 µl*	*22 µl*	*88 µl*	*176 µl*	*200 ml*
sterile ddH₂0	*3.4 µl*	*803 µl*	*3212 µl*	*6424 µl*	*8000 ml*
Hind *III (HC=80U/µl)*	*0.5 µl*	*55 µl*	*220 µl*	*440 µl*	*550 ml*
Loading Dye	*1.0 ml*	*220 ml*	*880 ml*	*1760 ml*	*2000 µl*
total volume (w/o DNA)	*6.0 µl*	*1320 µl*	*5280 µl*	*10560 µl*	12750 µl

3. Aliquot 6µl of mastermix per well of 96 well Thermowell PCR plate. The same tips can be used for all wells when adding master mix.

4. Add 5 µl of each DNA sample. Change tips between each sample. Quick-spin in centrifuge. Cover with parafilm and rubber plate sealers.

5. Place sealed plates into a hermetically sealed Tupperware container whose bottom has been lined with paper towels moistened with distilled water. Incubate 37°C overnight (at least 10 hrs).

8.4.2. High resolution agarose gel electrophoresis

1. Gels are 250 ml of 1 X NEB and 1 % agarose. To an Erlenmeyer flask add 250 ml of 1X TAE and 2.5 g of agarose. Weigh the flask plus ingredients and record total weight. Microwave to dissolve agarose completely, mixing intermittently to prevent boiling over. Reweigh the flask and adjust to original weight with distilled water. Mix well but avoid bubbles. Place flask in 56°C water bath to cool.
    *1 X NEB is 0.1M Tris, 12.5mM Sodium Acetate, 1mM EDTA/pH 8.0. Prepare as a 10X stock and dilute to 1X before use.

2. While agarose is cooling, level the gel tray and insert the well combs into the tray. Carefully but quickly pour the cooled agarose into the prepared gel tray, avoiding bubbles and remove particulates and bubbles before the gel solidifies. Allow gel to harden for at least 45 minutes before loading samples. Gels may be stored overnight at 4°C after wrapping in plastic wrap.

3. Use approximately 4.5 l of 1X TAE for each gel box and store at 4°C until used. Place the buffer and gels in the gel boxes and turn on pump to start circulation.

4. Load 0.5 µl of Marker Mix (see recipe below) in every fifth lane starting with lane 1. Then mix each sample by pipetting twice and load 2 µl in the appropriate well with a multi channel pipette.

5. Run gels at 100 volts for 16 hours. Maintain the buffer at 10°C by means of a chiller system, or place the gel apparatus in a cold room at 4°C during electrophoresis.

6. After gels are run, place each gel in 120 ml of 1X TAE containing 12 µl of SybrGold (10,000X) stain. Stain for 30 minutes in a pan covered with aluminum foil with gentle movement. Be sure to wear gloves and remember that SybrGold is optically active (light sensitive).

7. Gently place gel into imaging device and scan gel according to the manufacturer's instructions. Be sure to use several different scan times. Save as TIFF files.

8.5. BAC end sequencing

For sequencing, use 700-1000 ng (500 ng also works well) of BAC DNA as template in a 20 µl final vol sequencing reaction using 8 µl of sequencing cocktail.

Sequencing Reaction (20 µl volume) :
Perkin Elmer BigDye   8.0 µl
SP6 or T7 primer (5 pmol/µl)   1.0 µl
BAC DNA + H2O   to make up to 20 µl

Sequencing :
1 cycle   95°C   3 min

35 - 40 cycles   95° C   15-30 sec
                      53-55°C   10-15 sec
                      60°C   4 min
4° C hold

Following the sequencing reaction, pass samples through sephadex columns to remove unincorporated primer, dNTPs and ddNTPs.

Dry the eluate in a speedvac set to low-medium heat setting, until completely dry.

Resuspend reactions in 1-3 μl of sequencing gel running buffer.
• Unlike sequencing of plasmids with smaller (1-few kbp) inserts, the amount of sequenced products from BAC templates is typically low, resulting in lower signal intensities. Consequently, it is best to minimize the resuspension volume to get the most amount of the reaction on the sequencing gel.

8.6. References

- Mara, M.A., Kucaba, T.A., Dietrich, N.L., Green, E.D., Brownstein, B., Wilson, R.K., McDonald, K.M., Hillier, L.W., McPherson, J.D., and Waterston, R.H. (1997) High throughput fingerprint analysis of large-insert clones. Genome Research **1** : 1072-1084.
- Mara, M., et al. (1999) zA map for sequence analysis of the Arabidopsis thaliana genome. Nature Genetics **22** : 265-270.
- McPherson, J.D., et al., (2001) A physical map of the human genome. Nature **409** : 934-941.
- Mozo, T., Fischer, S., Meier-Ewert, S., Lehrach, H., and Altman, T. (1998) Use of the IGF BAC library for physical mapping of the Arabidopsis thaliani genome. Plant Journal, **16** : 377-384.
- Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S, Altmann T. (1999) A complete BAC-based physical map of the Arabidopsis thaliana genome. Nat Genet **22** : 271-5.
- Shizuya, H., Birren, B., Kim, U-J, Mancino, V., Slepak, T., Tachiri, Y., and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. PNAS USA **89** : 8794-8797.
- Soderlund, C., Humphray, S., Dunham, A., and French, L. (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Research **10** : 1772-1787.