AFLP: Principle and Application

1. Molecular Markers: General Introduction 85

1.1. Definitions 85

1.2. What is a good molecular marker for a population geneticists? 87

2. AFLP 88

2.1. Summary of the method 88

2.2. Basic steps of AFLP fingerprinting 90

2.2.1. DNA extraction 90

2.2.2. Restriction 92

2.2.3. Ligation of oligonucleotide adapters 92

2.2.4. Pre-amplification 92

2.2.5. Amplification 93

2.2.6. Electrophoresis 94

3. How to develop AFLP markers 95

3.1. Primer choice 95

3.2. Repeatability 99

3.3. Protocols 101

3.3.1. DNA extraction 101

3.3.2. Restriction/Ligation 103

3.3.3. Preamplification 103

3.3.4. Amplification 105

3.3.5. Electrophoresis 105

3.4. Critical phases of the process & troubleshooting 105

3.4.1. DNA degradation 106

3.4.2. Simultaneous isolation of RNA 107

3.4.3. Isolation of polysaccharides 108

3.4.4. Isolation of polyphenols 108

4. Genotyping 108

5. Interpretation of results 109

Literature cited 111

AFLP: Principle and Application

1. Molecular Markers: General Introduction

In recent years, molecular markers and especially DNA-based markers, have been extensively used in many areas such as gene mapping and tagging (kliebenstein et al., 2001; Karp and Edwards, 1997), characterisation of sex (Flachowsky et al., 2001; Martinez et al., 1999), analysis of genetic diversity (Erschadi et al., 2000; Palacios et al., 1999; Lerceteau and Szmidt, 1999; Godt and Hamrick, 1999) or genetic relatedness (Mace et al., 1999; Roa et al., 1997; Brookfield, 1992). In population genetics, protein-based markers (allozymes) were the first markers developed and widely used (see Hamrick and Godt, 1990 for review). DNA-based methodologies are now the method of choice to differentiate closely related organisms (Widen et al., 1994; Ouborg et al., 1999; Avise, 1994). Moreover, the use of DNA-based markers allows efficient comparisons because genetic differences are detectable at all stages of development of the organism unlike allozymes which may show age dependent changes.

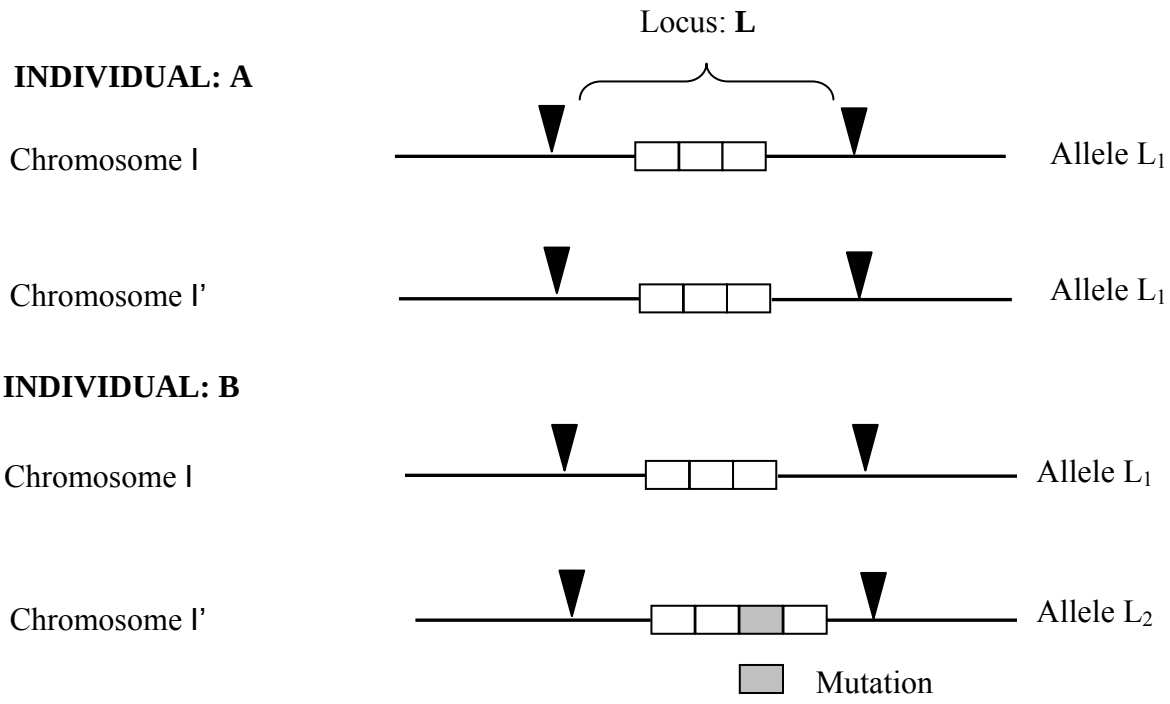
1.1. Definitions

According to Stansfield (1986), the term MARKER is usually used for "LOCUS MARKER". Each gene has a particular place along the chromosome called LOCUS (Box 1). Due to mutations, genes can be modified in several forms mutually exclusives called ALLELES (or allelic forms). All allelic forms of a gene occur at the same locus on homologous chromosomes. When allelic forms of one locus are identical, the genotype is called HOMOZYGOTE (at this locus), whereas different allelic forms constitute a HETEROZYGOTE. In diploid organisms, the GENOTYPE is constituted by the two allelic forms of the homologous chromosomes. All these definitions are illustrated in the Box 1 with a codominant marker as example.

Thus, MOLECULAR MARKERS are all loci markers related to DNA (markers can also be biochemical, or morphological).

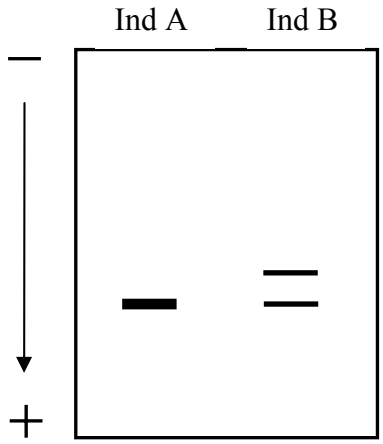
Box 1: Molecular Markers: an example of a codominant marker

LOCUS: DNA fragment comprised between the two arrows (▼).



In this example, the locus **L** has two different **ALLELES**: L_1 and L_2.
GENOTYPE of individual A is "L_1 L_1", and genotype of Individual B "L_1 L_2"

In a case of a codominant marker, the pattern observed after a gel electrophoresis will be:



Individual A is **HOMOZYGOTE** for the locus L

Individual B is **HETEROZYGOTE** for the locus L

1.2. What is a good molecular marker for a population geneticists?

A good molecular marker should be/have:

1/ Mendelian inheritance: transmit from one generation to another.
2/ Polymorphic: present several alleles at the locus investigated (multiallelic).
3/ Codominant: allow the discrimination between homo and heterozygotes.
4/ Neutral: all alleles have the same fitness
5/ Not epistatic: one can determine the genotype of a phenotype irrespective of the genotype of the other loci.
6/ Independent of environment: no phenotypic plasticity.
7/ Frequent occurrence in the genome
8/ Even distribution throughout the genome
9/ Highly reproducible

The most frequently used markers in population genetics are allozymes (biochimical), RAPD (Random Amplified Polymorphic DNA; Williams et al., 1990), RFLP (Restriction Fragment Length Polymorphism, Botstein et al., 1980), AFLP (Amplified Fragment Length Polymorphism, Zabeau and Vos, 1993), minisatellite fingerprints, microsatellites and SSR (Single Sequence Repeats, Tautz and Renz, 1989).

The choice of a specific molecular marker depends of its suitability to answer a particular ecological question. For this purpose, the main difference among molecular markers is their degree of dominancy. Co-dominant markers enable for an easy estimation of allele frequencies. Therefore, they are suitable to estimate gene flow between populations or will be preferred, for example, to study dispersal. On the other hand, dominant markers can estimate genotypes but not the allele frequencies. Dominant markers are preferably used as fingerprints (Mueller and Wolfenbarger, 1999; Hongtrakul et al., 1997; Weising et al., 1995) and can be helpful in the identification of clones.

To study *Cirsium arvense* genetic diversity, we choose to develop a recent DNA marker called "AFLP" (Amplification Fragment Length Polymorphism). They full fill all the previous characteristics of "good molecular markers", except for the codominance. AFLP markers are dominant markers. Nevertheless, because of the high amount of polymorphism they can detect (Mueller and Wolfenbarger, 1999), AFLP markers were a priori the most

efficient markers for our study; e.g. to identify individual genotypes at the landscape scale, in a species described to be highly clonal (Moore, 1975; Bostock and Benton, 1979; Donald, 1994) and where genotypes were thus expected to be closely related. Moreover, from a technical point of view, no prior DNA sequence information is needed, many markers can be analysed in a short time, and only small amount of DNA is needed.

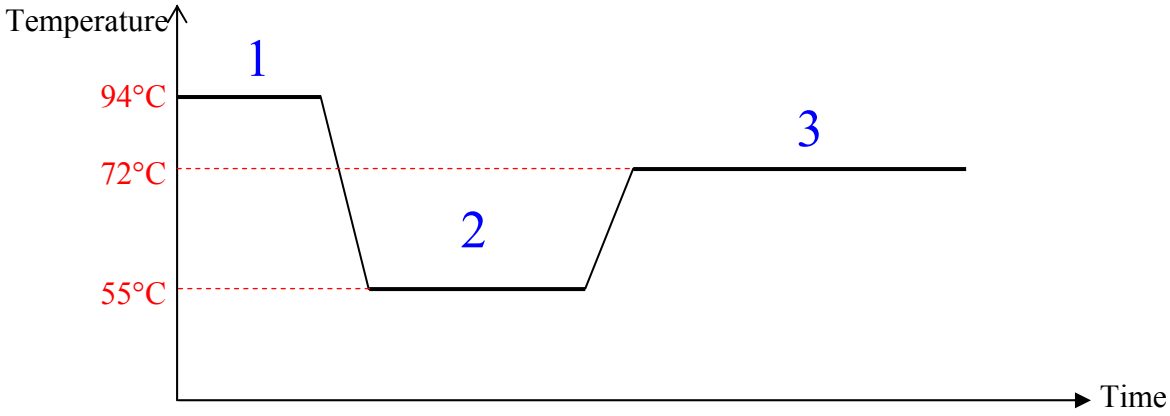
2. AFLP

2.1. Summary of the method

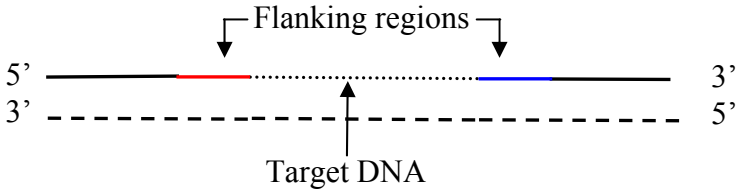
Amplified Length Fragments Polymorphism is a recent DNA fingerprinting technique developed by Zabeau and Vos (1993); but see also Vos et al. (1995) and Vos and Kuiper, 1997). This method is based on PCR amplification of selected restriction fragments of a total digested genomic DNA. Once labelled, amplified products are separated by electrophoresis. DNA fragments obtained range from 60 to 500 base pairs.

To be visualised, DNA polymorphism, which is usually made of small DNA fragments of few base pairs (up to 500), must be amplified. This amplification is commonly done by Polymerase Chain Reaction (Mullis et al., 1986; Mullis and Faloona, 1987). The PCR method can amplify specific DNA fragments through a precise priming of the polymerisation reaction occurring at each end of the target DNA. This precise priming is done by short oligonucleotidic sequences (Primers) able to anneal to the template DNA in the target zone. Primers are 18-24 base pairs long, synthesised in laboratory and correspond to a complementary DNA sequence designed in the flanking regions of the heavy strand of the target DNA (Box 2). The Polymerase Chain Reaction starts first with a high temperature phase (denaturation) that produces single-stranded DNA. Then, once temperature has reached the TM (Box 2), primers will bind to the template DNA. The

Box 2: Illustration of the Polymerase Chain Reaction

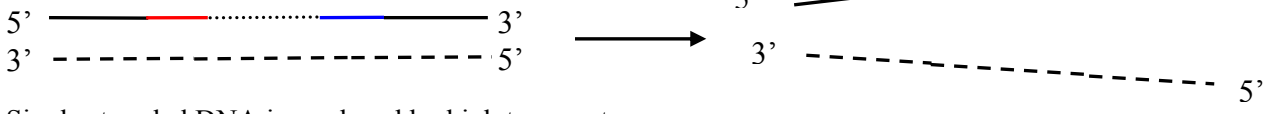


Primers are designed one DNA strand



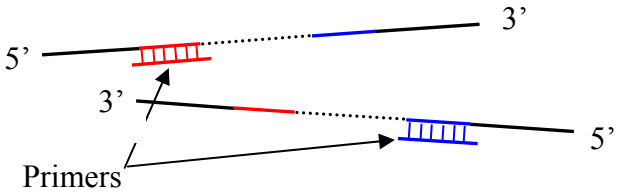
Dashed line (3'-5') represents the complementary DNA strand

1: Denaturation of the DNA Template



Single stranded DNA is produced by high temperature

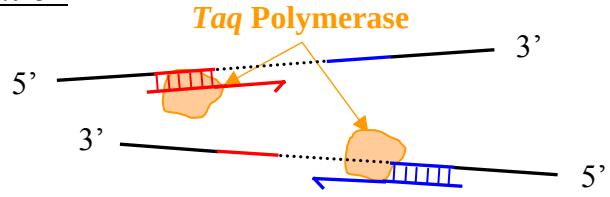
2: Annealing of the primers on the DNA template



The annealing temperature is defined by the oligonucleotides composition of the primers.

TM = 2(N# of AT) + 4(N# of GC)

3: Elongation



The *Taq* polymerase optimally works at 72°C.

During the elongation, each DNA strand is polymerised in the direction 5'->3'

Taq polymerase recognises each double-stranded DNA as a start of synthesis and will continue the polymerisation reaction in the direction 5' → 3' as soon as the temperature has reached 72°C (optimal elongation temperature).

Therefore, in order to design specific primers, the sequences of the flanking regions of the target DNA must be known. This supposes detailed knowledge about the genome or further elaborated investigations to get it. This step usually requires high laboratory equipments and are most of the time, time consuming.

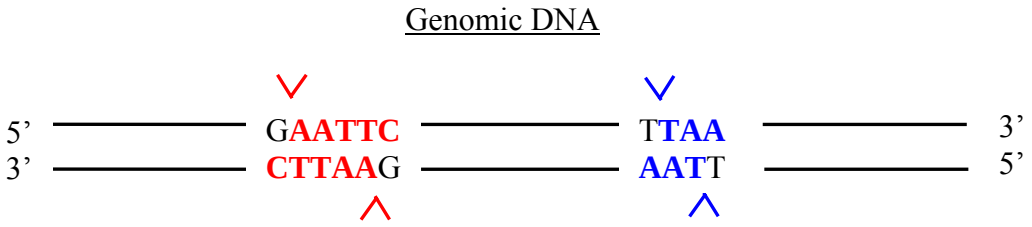
The originality of the AFLP method was to design and synthesise arbitrary primers first, and then to ligate them to target DNA fragments (Box 3). The AFLP arbitrary primers are called "adapters" and consist of a known sequence of 20 nucleotides. The target DNA sequences are DNA fragments generated by restriction enzymes. Fragments are produced from total genomic DNA by the combined action of two restriction enzymes. Then, adapters are ligated at each end of a restriction fragment by a protein ligase (New England Biolab®). Finally, adapters are used in a PCR as priming sites to amplify the restriction fragments. AFLP markers reveal a "restriction site" polymorphism and must be treated as dominant markers, since homozygotes and heterozygotes cannot be established unless breeding/pedigree studies are carried out to determine inheritance patterns of each fragment. However, the large number of fragments gives an estimate of variation across the entire genome, which thus gives a good general picture of the level of genetic variation of the studied organism.

2.2. Basic steps of AFLP fingerprinting

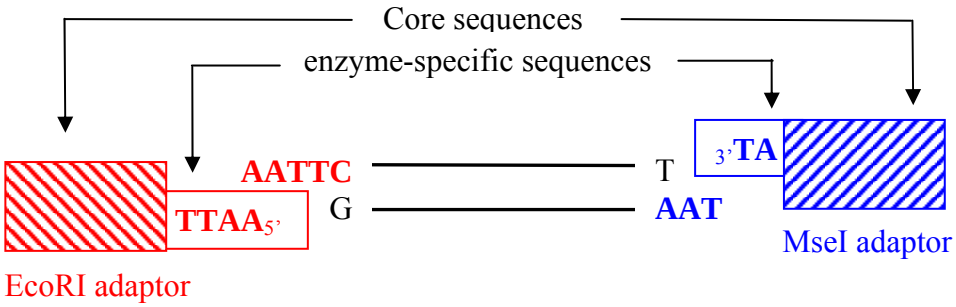
2.2.1. DNA extraction

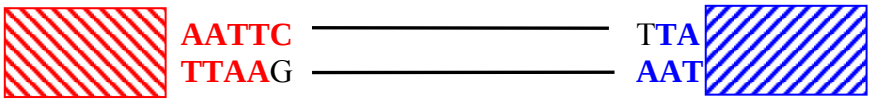
Clean and high molecular weight DNA is a prerequisite for AFLP. In our study, we extracted DNA according to Doyle & Doyle method (Doyle and Doyle, 1988). This method is based on the CTAB procedure. For more details, refer to the protocol (3.3) and troubleshooting (3.4) parts.

Box3: AFLP: Principle

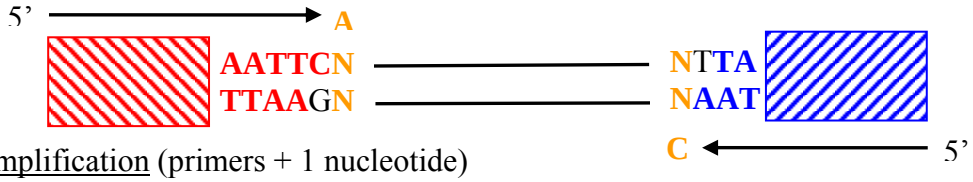


Action of the Restriction Enzymes **EcoRI** (>) & **MseI** (>)

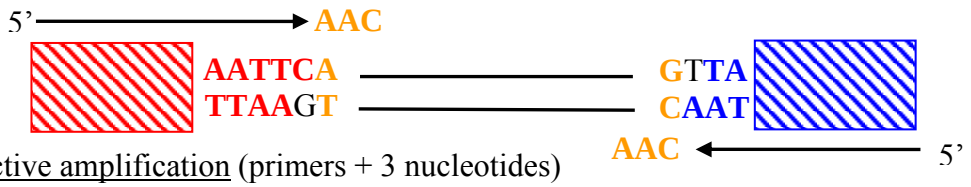




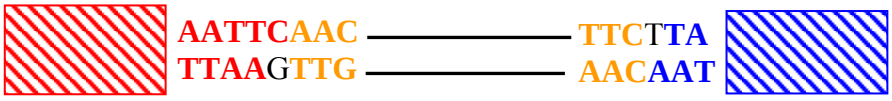
Ligation of the adaptors (known DNA sequence = futur primers)



Preamplification (primers + 1 nucleotide)



Selective amplification (primers + 3 nucleotides)



Amplified fragments will be separated by electrophoresis

2.2.2. Restriction

Restriction fragments of the genomic DNA are produced by using two different restriction enzymes: a frequent cutter (the four-base restriction enzyme *Mse*I) and a rare cutter (the six-base restriction enzyme *Eco*RI) (Box 3). The frequent cutter serves to generate small fragments, which amplify well and which have the optimal size range for separation on a sequence gel, whereas the rare cutter limits the number of fragments to be amplified.

2.2.3. Ligation of oligonucleotide adapters

Double-stranded adapters consist of a core sequence and an enzyme-specific sequence (Box 3). Therefore, adapters are specific for either the *Eco*RI site or the *Mse*I site. Usually restriction and ligation take place in a single reaction. Ligation of the adapter to the restricted DNA alters the restriction site in order to prevent a second restriction from taking place after ligation has occurred. The core sequence of the adapters consists of a known DNA sequence of 20 nucleotides, which will be used later as primer in the PCR.

2.2.4. Pre-amplification

This step is a normal PCR where the adapters are used as primers. This first PCR, called preamplification, allows a first selection of fragments by only amplifying the DNA restriction fragments that have ligated an adapter to both extremities. Additionally to the adapter sequences, the primers used for the pre-selective amplification have a supplementary base. This extra base enables another first selection by amplifying ¼ of the fragments that have ligated an adapter to both extremities. These first three steps (DNA extraction, restriction/ligation and preamplification) can be run and visualised on a 1.6 % agarose gel (Figure 1).

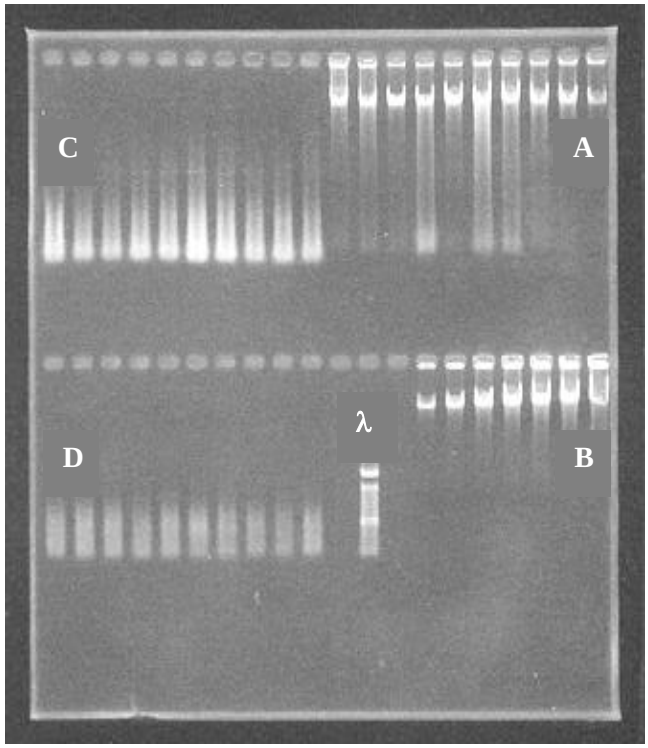


Figure 1: The three first steps of the AFLP process (DNA extraction, restriction/ligation and preamplification) run on a 1,6% agarose gel. The 10 first lanes (**A**) represent genomic DNA of 10 *Cirsium arvense* samples. According to the Lambda DNA concentration standards (**B** from left to right: 0.125; 0.25; 0.5; 0.75; 1; 1.5; 2 µg) *C. arvense* DNA concentration can be estimated to be in around 5ng per lane (5µl DNA load). (**C**) restriction/ligation of the same 10 samples. Genomic DNA was restricted with Mse I and EcoR I enzymes. The restriction produces a large quantity of small size DNA fragments. During the preamplification (**D**), only a part of the restricted fragments are amplified. The preamplification leads to a homogeneous DNA smear ranging from 100 to 800 bp. (λ) represents a DNA size marker.

2.2.5. Amplification

The aim of this step is to restrict the level of polymorphism and to label the DNA. For this second amplification, we added three more nucleotides at the 3' end of the primer sequence used for the preamplification (= adapters sequence + 3 nucleotides; Box 3). These two additional nucleotides make the amplification more selective and will decrease the number of restriction fragments amplified (polymorphism). Moreover, one of the primers (usually the EcoRI primer) is labelled with a fluorescent dye, and will allow the visualisation of DNA during the migration.

2.2.6. Electrophoresis

The PCR products are denaturated and run on acrylamide gel (DNA sequencer). In our study, samples were run on an ABI Prism 310 (Figure 2).

Figure 2: ABI Prism sequencer

A thin capillary containing a polymer replaces the usual acrylamide gel. The electrophoresis conditions we used for fragments analysis can resolve DNA fragments differing just by one base pair. Samples are loaded in a track, and run one after the other through the capillary.

All fragments are separated with regard to length, smaller fragments running first. Once passing the laser, a dye attached to the primer is excited and emits a fluorescent signal that is then collected by a computer. The results of fluorescence are visualised on the computer as peaks, called Electropherogramms (Figure 3). Each peak corresponds to a band on a normal acrylamide gel. Amplified fragments range from 30 to 400 base pairs.

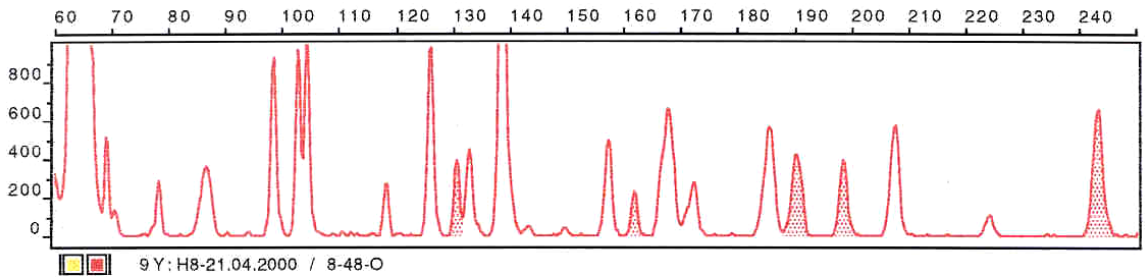


Figure 3: AFLP electropherogramm. On the figure, each peak corresponds to a DNA restriction fragment of a precise length. From the left to the right, the 5 shaded peaks correspond to DNA fragments having respectively 131; 161; 190; 198 and 242 base pairs.

3. How to develop AFLP markers

3.1. Primer choice

Choice of number and sequence of primers that will be used for the selective amplification is an important step of the AFLP process because they will later determine the level of polymorphism accessible in the studied species.

To test for good primer combination we used the *Selective Amplification Modules for Small Plant Genomes* from Applied Biosystems. The *Selective Amplification Modules for Small Plant Genomes* from Applied Biosystems provided eight EcoRI-labelled primers and eight MseI primers. Thus, 64 selective primer combinations were available. An initial screening using all these 64 primer combinations was performed on a random sample of 10 individuals across all sampled populations.

For our study, we choose primers according to their level of inter- and intrapopulation polymorphism (figures 4 and 5). A good primer must discriminate individuals coming from different populations, but should still be polymorphic enough at the population level to precisely identify clones.

From the screening of the 64 primers, two primer combinations EcoRI-ACC / MseI-CTG and EcoRI-ACG / MseI-CTT appeared efficient at the inter-population level, and sufficiently polymorphic to discriminate clones within populations.

Figures 4 and 5 represent respectively inter and intrapopulation polymorphism obtained with the primer combination EcoRI-ACC / MseI-CTG.

In Figure 4, about interpopulation polymorphism, the electropherograms of the same colour represent two individuals coming from the same population. From up to down: individuals 8-48-O, 8-44-O, 47-5-O, 47-6-O, N2-1-O, N2-2-O were respectively 2 individuals of populations 8, 47 and N2. In each case, both individuals of the same population present an identical electropherogramm. Therefore, according to the AFLP patterns they have identical genotypes and thus belong to the same clone. In figure 4, shaded peaks indicate interpopulation polymorphism. Shaded peaks are just present in the population they belong, and offer a good discrimination between populations.

Figure 5 illustrates an example of intrapopulational polymorphism. In figure 5, the electropherograms of the same colour represent two individuals coming from the same population and having different genotypes. From up to down: 8-48-O, 8-45-O, 47-6-O, 47-10-O, N2-2-O, N2-3-O were respectively 2 individuals of populations 8, 47 and N2. Shaded peaks indicate intrapopulation polymorphism, e.g. peaks that are specific to one individual within a population. These results show that the primer combination EcoRI-ACC / MseI-CTG was also efficient to discriminate between individuals of a same population.

Remarks: We found similar results with the primer combination EcoRI-ACG / MseI-CTT (results not shown).

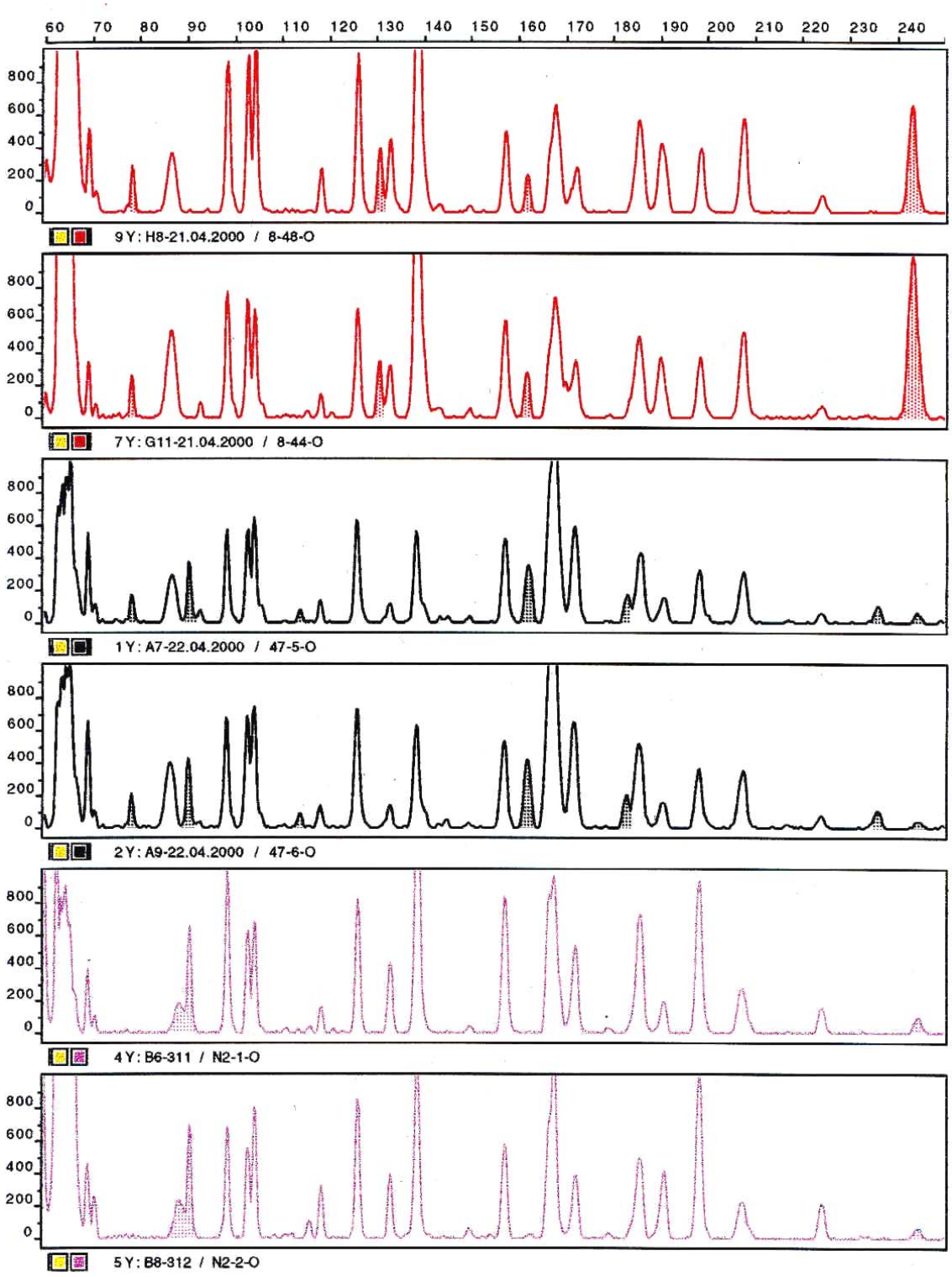


Figure 4: Example of polymorphism interpopulational obtained with the primer combination EcoRI-ACC / MseI-CTG. Electropherograms of the same color represent two individuals coming from the same population and having the same genotypes (clones). Shaded peaks indicate interpopulation polymorphism.

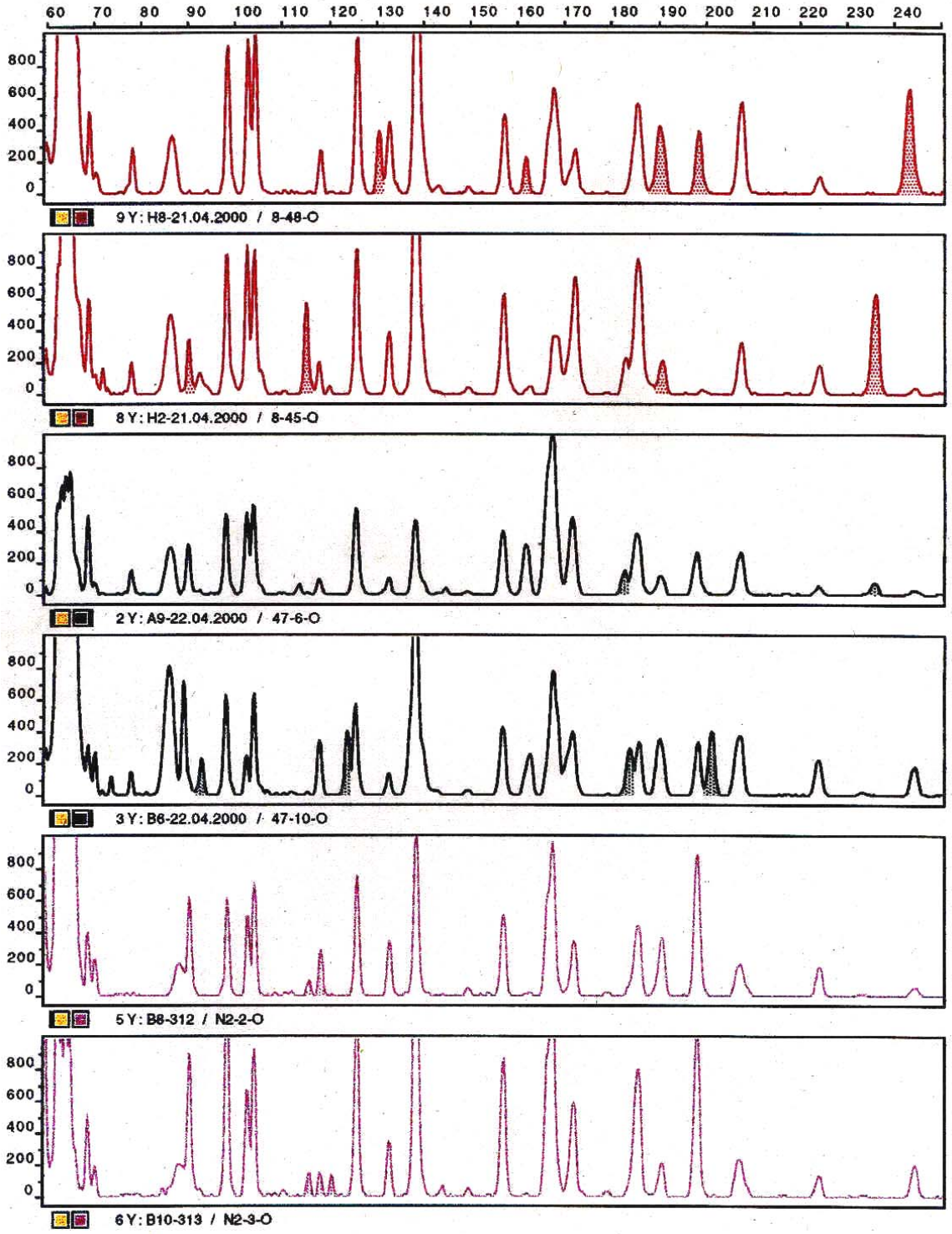


Figure 5: Example of polymorphism intrapopulational obtained with the primers EcoRI-ACC / MseI-CTG. Electropherograms of the same colour represent two individuals coming from the same population and having different genotypes. Shaded peaks indicate intrapopulation polymorphism.

3.2. Repeatability

Compared with RAPD, the AFLP technique is highly reproducible (Mueller and Wolfenbarger, 1999). In order to check for the reproducibility of our AFLP method, we repeated the whole process with different DNA extractions. Results are presented in figure 6.

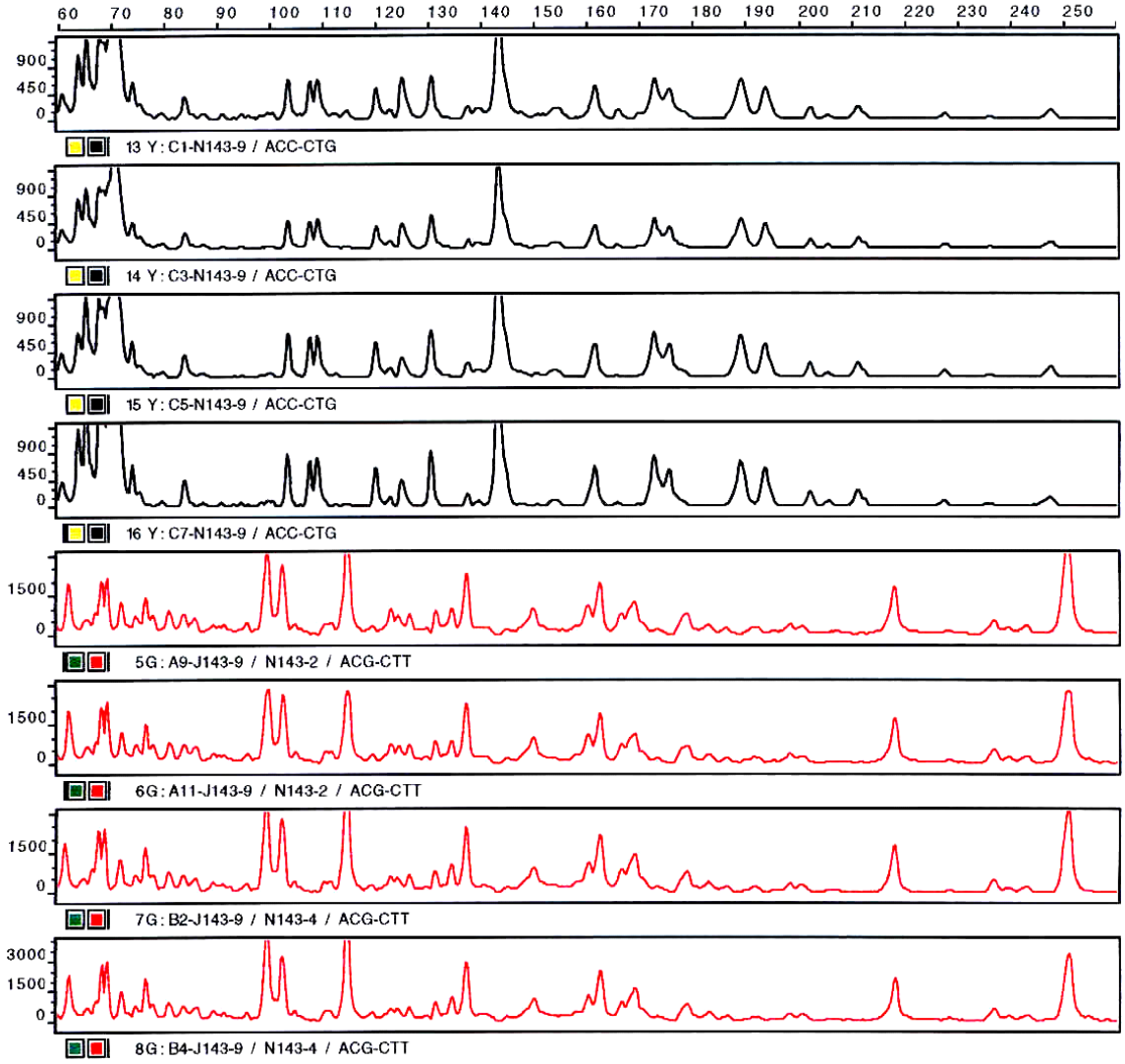


Figure 6: Repeatability test. The figure shows four independent AFLP processes realised from four DNA extractions of the same sample (N1-143-9) for the primers ACC-CTG (in black) and for the primers ACG-CTT (in red).

Box 4: DNA extraction: components and their effects

Components	Name	Nature	Effects
pH			- Inhibition of degradative enzymes. (ex: DNases act at pH=7)
Tris		Buffer	- Maintains the pH
EDTA	Ethylenediamine-tetraacetate		- Chelation of divalent cations (Ca^{2+}, Mg^{2+}...) - Inhibition of metal-dependant enzymes (ex: nucleases, DNases)
Na or K		Salt	- Stabilisation of nucleic
Proteinase K		Enzyme	- Digestion of proteins
SDS	Sodium Dodecyl Sulfate	Anionic detergent	- Solubilisation of cellular membranes - Denaturation of proteins
CTAB	Hexadecyltrimethyl ammoniumbromide	Cationic detergent	- Solubilisation of cellular membranes - Denaturation of proteins - Formation of a complex with the DNA
CIA	Chloroform-isoamyl-alcohol		- Extraction proteins
Isopropanol			- Precipitation of the CTAB-DNA complex
ß-mercaptoethanol		Reducing agents (antioxidants)	- Inhibition of the oxidation processes - Protection of DNA from quinones, disulfites, peroxidases, polyphenoloxydases
Glutathione			
Cysteine			
DDT	Dithiothreitol		
Ascorbic acid			
$Na_2S_2O_5$	Sodium bisulfite		
PVP	Polyvinylpyrro-lidone	Polyphenols absorbents	- Decrease the effect of polyphenols, quinones, tanins - BSA: denaturation of degradative enzymes
BSA	Bovine Serum Albumin		
Diethylpyrocarbonate		Polyamines	- Protection from RNases
Bentonite			
Spermine			
Spermidine			
DIECA	Diethydithiocar-bamic acid	Phenoloxidase inhibitors	- Cu^{2+} chelation
Cyanide			- Protection from heavy metal oxidases
NH			- Protection from H^+
Aurintriarbolylic acid			- Inhibition of nucleases

3.3. Protocols

3.3.1. DNA extraction

We extracted DNA according to the Doyle & Doyle method (Doyle and Doyle, 1988). This method is based on the CTAB procedure (Webb and Knapp 1990). The CTAB (Hexadecyltrimethylammoniumbromid) is a cationic detergent (Box 4) that forms a complex with the DNA. The CTAB-DNA complex is then separated from the cellular debris by chloroform. In this step, we can observe two phases: a superior clear aquatic phase containing the DNA, and a denser inferior one containing the chloroform and all the secondary components (polysaccharides, proteins etc…). After centrifugation, cellular debris can usually be observed at the interface. The purification by chloroform can be repeated several times (2, 3 times).

DNA is soluble in water and can be precipitate with salts (NH_4^+, Na^+, NaOAc, NaCl, NH_4OAC...) and ethanol 100% or isopropanol. To precipate DNA, one can equally use one of the following salts: Na^+(3M) or NH_4^+(7.5M), with one of the following alcohols: ethanol 100% or isopropanol. The precipitation is done in these proportions: 10% and 50% of the final volume for Na^+ and NH_4^+ respectively, and 250% and 70% of the final volume for ethanol and isopropanol respectively.

After the precipitation, the DNA molecule must be washed with ethanol to remove the salts used for the precipitation. DNA is then dissolved and stored in a Tris/EDTA buffer (TE buffer: Tris (10mM) / EDTA (1mM)) pH 8.

The protocol we used to extract DNA was the following. About 150 mg of plant tissue were first frozen in liquid nitrogen and ground in a mixer mill. Then, we added 700 µl of extraction buffer (Box 5) to the ground material and incubated the mixture 45 minutes at 60°C. At this stage, it is important to vortex to get a homogenous solution free of clumps to prevent DNA degradation (3.4.1). At the beginning of the incubation, DNA is still in the cells and can not be broken by vortexing. Afterwards, vortexing should be avoided.

Box 5: DNA extraction: buffers and stock solutions

Extraction buffer

	Final concentration	Stock solution	For 500 ml of buffer
NaCl	1.4M	5 M	140 ml
Tris-HCl pH 8.0	100 mM	1 M	50 ml
EDTA pH 8.0	20 mM	0.5 M	20 ml
CTAB	2 %		10.0 g
$Na_2S_2O_5$	1 %		5.0 g

Autoclave the solution and then add ß-Mercaptoethanol (0.2 %).
Check the pH again.

CIA-solution (Chloroform:Isoamylalcohol 24:1)

	Final concentration	For 250 ml of CIA
Chloroform	96 %	240 ml
Isoamylalcohol	4 %	10 ml

Store the solution at 4°C.

TE-Buffer (TE 10/1)

	Final concentration	Stock solution	For 250 ml of TE
Tris-HCl pH 8.0	10 mM	1 M	1.0 ml
EDTA pH 8.0	1 mM	0.5 M	0.2 ml

After incubation, we added 700 µl of CIA 4°C (Box 5), mixed carefully and let the solution 5 min at room temperature. We centrifuged 10 min at 10.000 rpm, and recuperated the floating phase and repeated this purification step several times (2 times).

DNA was then precipitated with a 3M Na^+ salt (10%) and 0.7 volume of cold isopropanol (-20°C). We recuperated the DNA pellet after centrifugation (5 min 10.000 rpm), washed for 15 min in 500 µl of 75% and 100 % ethanol, dried at room temperature and then redissolved in 100 µl of TE buffer.

For each sample, we checked DNA quality and quantity on a 1.6% agarose gel.

3.3.2. Restriction/Ligation

The Mix prepared for the restriction / ligation is presented in the Box 6. Because of their high concentrations, we first diluted the enzymes in order to have reasonable volumina for pipetting. As enzymes are conserved in glycerol, in order to limit the amount of glycerol and to prevent the reaction from inhibition, the final volume of enzymes should not exceed 5% of the total volume (0.55 µl). Therefore, we divided this volume in three, and got the optimal volume of 0.18 µl for each enzyme. Then, we diluted enzymes within their respective buffers and prepared relevant stock solutions to have just to pipette 0.18 µl to get the right final concentration (Box 6, Restriction / Ligation Mix). Then, we added 11µl of the Mix to 5.5 µl of DNA containing 500 ng of genomic DNA. The reaction mixture was incubated 2 hours at 37°C in a MWG-Biotech Primus 96 thermocycler.

3.3.3. Preamplification

Products of the restriction ligation were diluted 1:2. We used 4 µl of the restriction / ligation as DNA template for the preamplification PCR (Box 6, preamplification). Amplification reaction was performed in a MWG-Biotech Primus 96 thermocycler with the following program: 2’ 72°C, (20’’ 94°C, 30’’ 56°C, 2’ 72°C)X30, 30’ 60°C, 4°C for ever.

Box 6:
Master mix:
Restriction/Ligation, Preamplification and Selective Amplification

Restriction / Ligation

	Stock solution	Final concentration	For 5.5 µl (one reaction)
H_2O			0.21 µl
T4 DNA ligase buffer	X 10	X 1	1.1 µl
NaCl	0.5 M	0.05 M	1.1 µl
BSA	1 mg/ml	0.05 mg/ml	0.55 µl
MseI adaptor	50 pmol/µl	5 pmol/µl	1 µl
EcoRI adaptor	5 pmol/µ	0.5 pmol/µl	1 µl
MseI	5.5 U/µl	1 U	0.18 µl
EcoRI	27.77 U/µl	5 U	0.18 µl
T4 DNA Ligase	372.22 U/µl	67 U	0.18 µl
final Volume			*5.5 µl*

Preamplifiction

	Stock solution	Final concentration	For 20 µl (one reaction)
PCR buffer	X 10	X 1	2 µl
dNTPs	2mM (X 10)	200 µM	2 µl
Primers MIX	2mM (X 10)	200 µM (X 1)	1 µl
Taq polymerase	5U / µl	2U / 50 µl	0.16 µl
H_2O			10.84 µl
ADN			4 µl
Volume final			*20 µl*

Selective Amplification

	Stock solution	Final concentration	For 20 µl (one reaction)
PCR buffer	X 10	X 1	2 µl
dNTPs	2mM (X 10)	200 µM (X 1)	2 µl
MseI primer	2mM (X 10)	200 µM (X 1)	1 µl
EcoRI primer	2mM (X 10)	200 µM (X 1)	1 µl
Taq polymerase	5U / µl	2U / 50 µl	0.16 µl
H_2O			10.84 µl
ADN			3 µl
Volume final			*20 µl*

3.3.4. Amplification

Products of the preamplification were diluted 1:10 and we used 3 µl as DNA template for the selective amplification PCR (Box 6, selective amplification). Amplification reaction was performed in a MWG-Biotech Primus 96 thermocycler with a touchdown PCR.

In a touchdown PCR the annealing temperature decreases by 1°C every cycle to a 'touchdown' annealing temperature which is then used for further cycles. Touchdown PCR is usually used to reduce non-specific amplification by optimising the optimal annealing temperatures. The idea of a touchdown PCR is that any differences in TM (Box 2) between correct and incorrect annealing gives a 2-fold difference in product amount per cycle. Therefore, touchdown PCR enriches for the correct product over any incorrect products. The touchdown PCR program we used was the following: 2' 72°C, (20'' 94°C, 30'' (66°C decreased by 1°C during 10 cycles), 2' 72°C)X10, (20'' 94°C, 30'' 56°C, 2' 72°C)X20, 4°C for ever.

3.3.5. Electrophoresis

We used 1µl of each selective amplifications (two primer combinations) that we mixed to 12 µl deionised formamide and 0.3 µl of Genescan-500 (ROX) size standard. ROX 500 is made of labelled DNA fragments of known size and is run with each sample. After a run, the internal standard will allow the calibration of all AFLP signals produced by primer dies recorded during the electrophoresis. The use of an internal standard makes comparisons between independent samples feasible.

To get a good signal it is important to mix properly the PCR products with the formamide and the size standard (quickly vortex). DNA was then denaturated by heating each tube to 95°C for 5 minutes and quick chill on ice. Samples were finally loaded in the ABI Prism 310. The electrophoresis was performed at 60°C, DNA was injected into the capillary during 20'' and we used the GS STR POP 4 module to detect peaks.

3.4. Critical phases of the process & troubleshooting

The AFLP process comprises five consecutive steps (DNA extraction, restriction/ligation, preamplification, amplification, electrophoresis) and only a success of all these steps will lead to good electropherograms (e.g. good signal intensity and quality of peaks). Compared with

other molecular markers, like microsatellites or RAPD that just consist in a single PCR, AFLP markers are more sophisticated. The implication of enzymes (restriction and ligation enzymes) makes the method particularly sensitive and sometimes problematic. In particularly, presence of secondary compounds in the DNA (like polyphenols or polysaccharides) can damage DNA, inhibit restriction enzymes or the *Taq* polymerase. Therefore, in order to avoid further technical problems, the first step (DNA extraction) appears to be of the most importance. This section is about the most common problems encountered during DNA extraction, and advices to resolve them.

3.4.1. DNA degradation

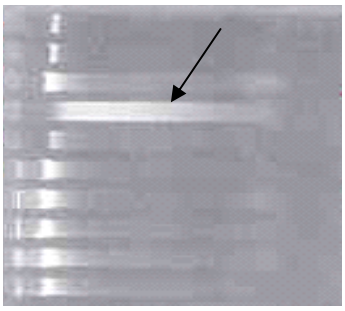


Figure 7: Detail of a 1.6% agarose gel of *C. arvense* genomic DNA. Whereas all lanes show a clear single band attesting for a high molecular weight DNA, the lane number 4 (cf arrow on the picture) presents a degradated DNA. No clear band of DNA is visible; DNA appears in a smear

DNA degradation is characterised by the generation of DNA fragments through the action of endogenous nucleases. Endogenous nucleases are active at room temperature, but inhibited are low and high temperature. Therefore, it is important to pay attention of temperatures all over the DNA extraction. For example, the presence of clumps in the solution containing ground tissues and the extraction buffer during the 60°C incubation phase (cf. 3.3.1.) can lead to a decrease of the temperature within a clump and therefore activate the endogenous nucleases. However, DNA degradation can also be due to a break of the cold chain before the extraction.

3.4.2. Simultaneous isolation of RNA

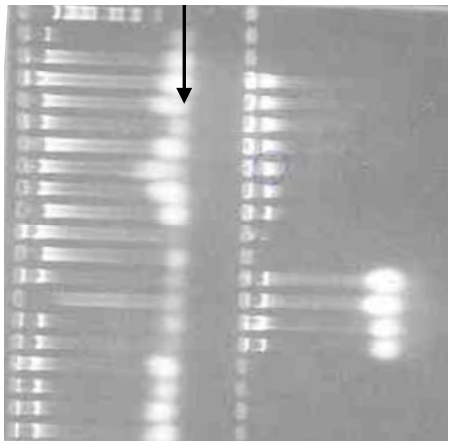


Figure 8: Detail of a 1.6% agarose gel of *C. arvense* genomic DNA showing a simultaneous extraction of RNA. On the picture, genomic DNA is represented by a clear single band closed to the lanes, whereas spots at the end of the lanes (cf arrows on the picture) represent small size RNA.

In some cases, RNA can be extracted simultaneously with DNA and then disturb the AFLP process. To avoid extracting RNA it is important to check the pH of the extraction buffer. A too acidic pH could lead to RNA extraction. RNA can also be removed by an additional step including a RNase A. The RNase A should be prepared as follows:

> Dilute 10 mg/ml of RNase (use a DNase-free RNAses to avoid a digestion of the DNA) in a RNase buffer (RNase buffer: 10 mM Tris HCl; 15 mM NaCl; pH: 7.5). Boil the solution for 15 min (95°C); cool it at room temperature, stored at -20°C. RNase should be used with a final concentration of 100ug/ml.

The RNase step can be done "as preventive" during the incubation of tissues with extraction buffer at 60°C, or just after the protein extraction by CIA and before the precipitation of the CTAB-DNA complex. In both cases, adequate volume of RNase leading to a final concentration of 100ug/ml must be added.

In case RNA has already been extracted (present on the agarose gel), the RNase step can still be done after the DNA extraction. For that, add the adequate volume of RNase to the extracted DNA (TE buffer and DNA), mix and incubate 30min room temperature (DNA can eventually be precipitated again).

3.4.3. Isolation of polysaccharides

Polysaccharides can easily be recognised because of their high viscosity. To remove polysaccharides one can either increase the CTAB concentration, or increase the concentration of NaCl (2M) prior to the first alcohol precipitation (polysaccharides remain soluble in high salt concentration). If DNA is already extracted, one can remove polysaccharides by redissolving them by increasing the NaCl concentration of the TE buffer up to 2M. Then, DNA must be precipitated again with 2 volumes of ethanol.

3.4.4. Isolation of polyphenols

Polyphenols lead to brown colouration of the DNA. They can damage DNA, inhibit restriction enzymes or *Taq* polymerase. Polyphenols can be removed by increasing the concentration of polyphenol absorbents up to 2%. One can also add phenoloxidase inhibitors, or elevate the concentration of antioxidants.

4. Genotyping

Genotyping of *C. arvense* electropherograms was one of the most time consuming phase of the study. First, we used the *Genotyper* software of ABI PRISM *PE biosystems*. In *Genotyper*, it is possible to standardise electropherogram peaks into categories. Each peak is defined in a category according to its width and its intensity (e.g. to its high). Categories are recorded and then, runs are screemed and automatically genotyped. Such an automatic labelling of peaks appears relevant especially when numerous samples must be treated (307 in our study). Although we spent a lot of time to define all possible categories, we did not manage to genotype automatically our data. This inability of using *Genotyper* can be explained by a too high sensitivity of the program coupled with a heterogeneity among runs. Because of the size reconstruction method of fragments (3.3.5) each AFLP peak gets a very precise position: for example: 55.6; 55.9; 56.1 base pairs and 56.2; 56.4; 56.7 base pairs. In the case that the first three peaks belong to the length “55 bp” and the next 3 peaks belong to the length 56 bp, it is difficult to define 2 independent categories which do not overlap. This example, repeated over 307 samples and 93 AFLP markers, made an automatic genotyping impossible.

Therefore, in a second approach, we only used the program *Genotyper* to produce a preliminary presence / absence matrix, which we subsequently checked manually.

5. Interpretation of results

This section will be about how to estimate the "consistency" of the markers.

Each same-sized fragment from different electropherograms, are supposed to be homologous (e.g. have the same origin) but this condition is not fully filled in AFLP. An absence of fragment in one run can either be due to a mutation in the restriction site of the enzymes, a mutation in the primers sequence, or a mutation in the restriction fragment. All these processes leading to the same phenotype: an absence of peak. Nevertheless, the numerous markers produced by AFLP can compensate for eventual homoplasy.

According fits Kimuras infinite allele model (Kimura and Crow, 1964) mutations happen randomly throughout the genome and tend to form a new allele from an ancestral allele, rather than an allele already existing. This random mutation process can lead to an asymmetric distribution of alleles (relatively high occurrences at high and low frequencies) when one looks at the allele class frequency. Checking the allele class frequency of AFLP could thus be a way to estimate the quality (randomness) of the makers (Figure 9).

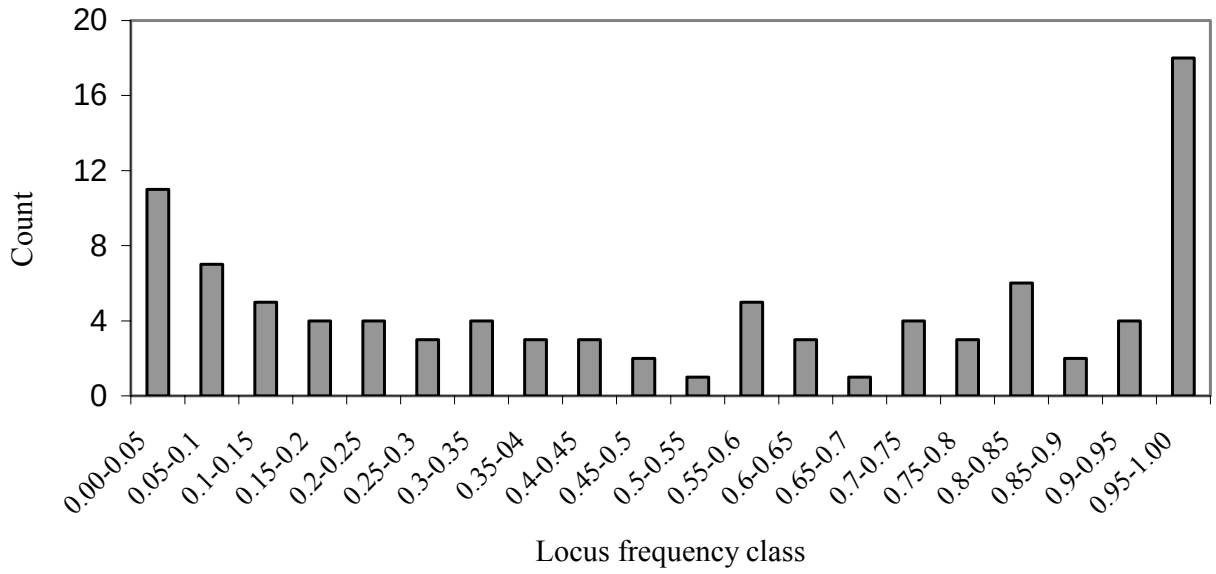


Figure 9: Locus frequency class distribution of 93 AFLP loci over 307 *C. arvense* individuals.

Therefore, quality of AFLP data was assessed by calculating the allele frequency class distribution for the 93 loci (42 for EcoRI-ACC / MseI-CTG and 51 for EcoRI-ACG / MseI-CTT) (Figure 9). The graph shows relatively high occurrences at high and low frequencies (common and rare bands). This asymmetric distribution fits Kimuras infinite allele model (Kimura and Crow, 1964) and leads to a bimodal distribution with common and rare alleles. This pattern of class frequency distribution of AFLP alleles has already been found in *Arabidopsis thaliana* (Miyashita et al., 1999; Sharbel et al., 2000) and attests the quality of our data.

Literature cited

Avise JC 1994 Molecular markers, natural history and evolution. Chapman & Hall, New York.

Bostock SJ, RA Benton 1979 The reproductive strategies of five perennial Compositae. Journal of Ecology 67:91-107.

Botstein D, RL White, M Skolnick, RW Davis 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314-331.

Brookfield JFY 1992 DNA Fingerprint in clonal organisms. Molecular Ecology 1:21-26.

Donald WW 1994 The biology of Canada Thistle (*Cirsium arvense*). Reviews of Weed Science 6:77-101.

Doyle JJ, JL Doyle 1988 Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Erschadi S, G Haberer, M Schoniger, RA TorresRuiz 2000 Estimating genetic diversity of *Arabidopsis thaliana* ecotypes with amplified fragment length polymorphisms (AFLP). Theoretical and Applied Genetics 100:633-640.

Flachowsky H, E Schumann, WE Weber, A Peil 2001 Application of AFLP for the detection of sex-specific markers in hemp. Plant Breeding 120:305-309.

Godt MJ, JL Hamrick 1999 Population genetic analysis of *Elliottia racemosa* (Ericaceae), a rare Georgia shrub. Molecular Ecology 8:75-82.

Hamrick JL, MJ Godt. 1990 Allozyme diversity in plant species. Pages 43-63 *in* Brown AHD, MT Clegg, AL Kahler, and BS Weir eds. Plant population genetics, breeding and genetic resources. Sinauer Associates Inc., Sunderland.

Hongtrakul V, GM Huestis, SJ Knapp 1997 Amplified fragment lengh polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theoretical and Applied Genetics 95:400-407.

Karp A, KJ Edwards. 1997 DNA markers: a global overview. Pages 1-13 *in* Caetano-Anollés G and PM Gresshoff eds. DNA markers: protocols, applications and overviews. Wiler-VCH, New York.

Kimura M, JF Crow 1964 The number of alleles that can be maintained in a finite population. Genetics 49:725-738.

kliebenstein DJ, J Gershenzon, T Mitchell-Olds 2001 Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. Genetics 159:359-370.

Lerceteau E, AE Szmidt 1999 Properties of AFLP markers in inheritance and genetic diversity studies of Pinus sylvestris L. Heredity 82:252-260.

Mace ES, CG Gebhardt, RN Lester 1999 AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). Theoretical and Applied Genetics 99:634-641.

Martinez EA, C Destombe, MC Quillet, M Valero 1999 Identification of random amplified polymorphic DNA (RAPD) markers highly linked to sex determination in the red alga *Gracilaria gracilis*. Molecular Ecology 8:1533-1538.

Miyashita NT, A Kawabe, H Innan 1999 DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. Genetics 152:1723-1731.

Moore RJ 1975 The biology of Canadian weeds. 13. *Cirsium arvense* (L.) Scop. Canadian Journal of Plant Sciences 55:1033-1048.

Mueller UG, LL Wolfenbarger 1999 AFLP genotyping and fingerprinting. TREE 14:389-394.

Mullis KB, FA Faloona 1987 Specific sythesis of DNA *in vitro* via a polymerase-catalized chain reaction. Methods in Enzymology 155:335-339.

Mullis KF, FA Faloona, R Scharf, G Saiki, G Horn, H Erlich 1986 Specific enzymatic amplification of DNA in vitro: THe polymerase chain reaction. Cold Spring Harbor Symp.Quant.Biol. 51:263-

Ouborg NJ, Y Piquot, J Van Groenendael 1999 Population genetics, molecular markers and the study of dispresal in plants. Journal of Ecology 87:551-568.

Palacios C, S Kresovich, F Gonzáles-Candelas 1999 A population genetic study of the endangered plant species *Limonium dufourii* (Plumbaginaceae) based on amplified fragment length polymorphism (AFLP). Molecular Ecology 8:645-657.

Roa AC, MM Maya, MC Duque, J Tohme, AC Allem, MW Bonierbale 1997 AFLPs analysis of relationships among cassava other *Manihot* species. Theoretical and Applied Genetics 95:741-750.

Sharbel TF, B Haubold, T Mitchell-Olds 2000 Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization in Europe. Molecular Ecology 9:2109-2118.

Stansfield WD 1986 Theory and Problems of Genetics. McGraw-Hill Book Company, New York.

Tautz D, M Renz 1989 Simple sequences are ubiquitous components of eukaryotic genomes. Nucleic Acids Res 18:6531-6535.

Vos P, R Hogers, M Bleeker, M Reijans, T van de Lee, M Hornes, A Frijters, J Pot, J Peleman, M Kuiper, M Zabeau 1995 AFLP: a new technique for DNA fingerprinting. Nucleic Acid Research 23:4407-4414.

Vos P, M Kuiper. 1997 AFLP analysis. Pages 115-131 *in* Caetano-Anollés G and PM Gresshoff eds. DNA markers: protocols, applications and overviews. Wiley-VCH, New York.

Weising K, H Nybon, K Wolff, W Meyer 1995 DNA Fingerprinting in Plants and Fungi. Boca Raton.

Widen B, N Cronberg, M Widen 1994 Genotypic diversity, molecular markers and spatial distribution of genets in clonal plants, a literature survey. Folia Geobotanica Phytotaxonomica 29:245-263.

Williams JGK, AR Kubelik, JA Livak, JA Rafalski, SV Tingey 1990 DNA polymorphisms amplified by arbitraty primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Zabeau M, P Vos1993Selective Restriction Fragment Amplification: A general method for DNA fingerprints