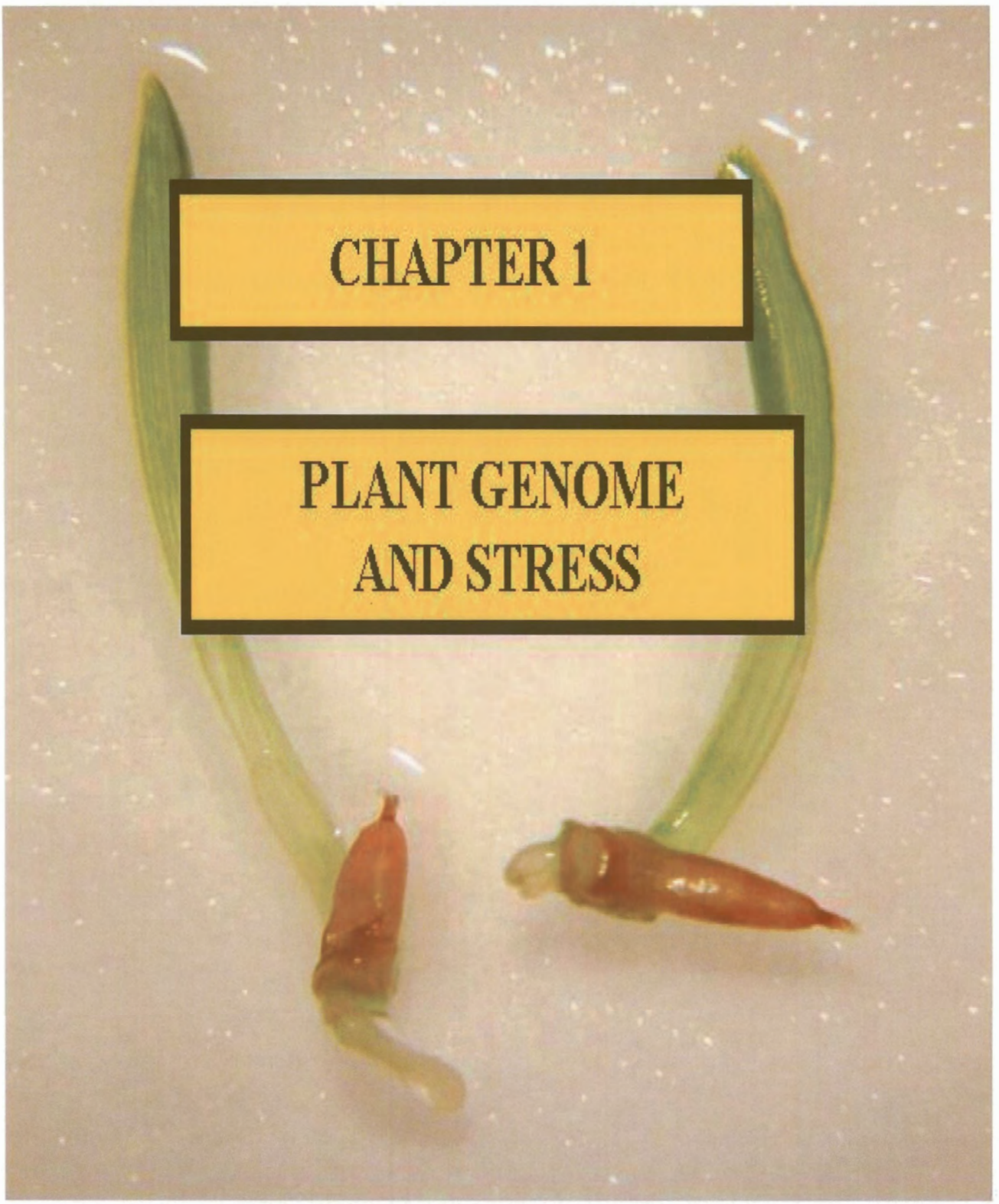


CHAPTER 1

PLANT GENOME AND STRESS

1.1 Abstract

Conditions of growth are seldom optimal and any change in an environmental condition that results in a response of an organism might be considered as stressful with the potential for modifying genome composition, growth and development of the organism. A plant needs to adapt to stress in order to survive. Any change in an environmental condition potentially affects the genome of the plant. However, there are programmed responses, such as variation of the gene expression and also non programmed responses, that might result in chromosomes breakage, DNA mutations and ultimately changed in gene expression. Quantitative modifications of repetitive DNA, DNA methylation excision and insertion of transposable elements, gene amplification or deletion and histone-acetylation are points of control of these challenges on the DNA level. Detection of such genome variation has been investigated with a variety of methods at the morphological, cytological, cytochemical, biochemical and molecular levels.

1.2 The plant genome

The plant cell has three genomes, which are the chloroplast genome (cpDNA), mitochondrion genome (mtDNA) and the nuclear genome (nDNA) (Dean and Schmidt, 1995) (Figure 1.1). These three genomes interact with each other in the plant cell. The cpDNA and mtDNA are very conservative having changed little on the molecular level over billions of years in comparison to the nDNA. The chloroplast genome of all species also carries nearly the same complement of genes arranged in very nearly the same order, contains roughly one hundred different gene functions and the size of the cpDNA is very similar even for diverse species. The cpDNA is composed of a single, circular chromosome of double-stranded DNA (Sugiura, 1992). It comprises typically of four segments: a large region of single-copy genes (LSC), a small region of single-copy genes (SSC), and two copies of an inverted repeat that separate the single copy IR_A and IR_B regions (Sugiura, 1992). The size and arrangement of the plant mitochondrion DNA is highly variable while the genetic content is conserved among plant species. Part of the

variability results from an unusual accumulation of non-coding DNA sequences in the regions between genes (Schuster and Brennicke, 1994).

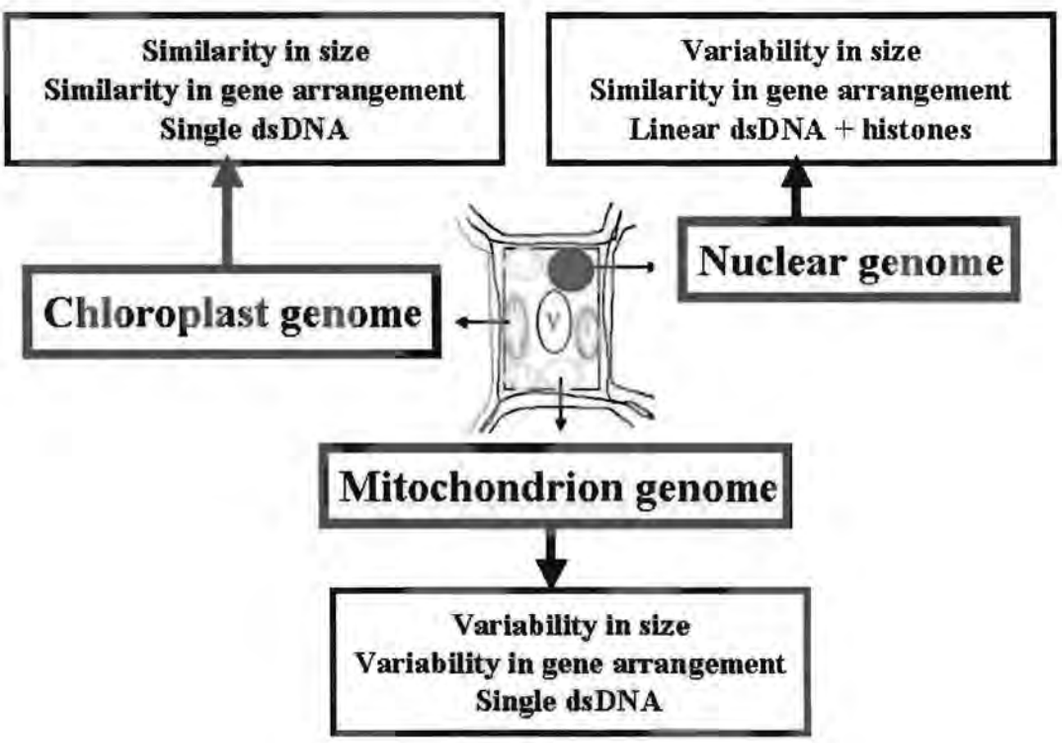


Figure 1.1: The plant cell genomes and their characteristics (V = Vacuole).

Mitochondrion DNA does not code for many genes, as most of the enzymes required for DNA replication, transcription and translation are encoded by the nucleus. The nuclear DNA (nDNA) has evolved dramatically and may vary in size by several orders of magnitude even among closely related species of flowering plants. The nuclear DNA contains both unique, single copy sequences and repetitive DNA. Repetitive DNA contributes to the character and function of specialized structures and plays a role in genome organization (Franklin and Cande, 1999). Repetitive DNA can be subdivided into tandem repeats including sequences associated with centromeres, telomeres and knobs and dispersed repeats. Dispersed repeats include transposable elements and retro-transposons (Franklin and Cande, 1999). In comparison to the chloroplast DNA

(cpDNA) and mitochondrion DNA (mtDNA), the nuclear genome (nDNA) consists of linear double-stranded DNA molecules bound to histones. Every eukaryotic species has further a characteristic amount of nuclear DNA and the DNA amount in the haploid cell of a species is called the C-value. However, no correlation exists between C-values and the phenotypic complexity.

Sequences within the genomes can be further classified according to a number of criteria. The most important of these is functionality and the largest class of functional DNA consists of coding sequences within transcription units, which function on behalf of the organism. The functional class of DNA elements also includes a number of specialised sequences that play roles in chromosome structure and transmission. The best-characterised structural elements are associated with the centromeres and telomeres (Sun *et al.*, 1997; Wright *et al.*, 1996; Pardue *et al.*, 1997). However, most of the genome appears to consist of DNA sequences that have no apparent function. This includes pseudo-genes that derive from specific genes but are not themselves functional with a lack of transcription or translation. For the most part non-functional DNA is present in the context of long lengths of apparently random sequence and repetitive elements. Repetitive DNA is especially abundant around the centromeric regions (Copenhaver and Preuss, 1999). Non-transcribing repeats (NTR)-DNA is an integral part of most plant genomes and its amount is proportional to the genome size (Flavell *et al.*, 1974). For example, regions present between two gene-rich regions are composed of NTR-DNA as well as regions present near the tip of chromosome arms are deficient in genes (Sandhu and Gill, 2002; Figure.1.2). NTR-DNA, which is unevenly distributed in the plant genome, is primarily composed of retro-transposons and pseudo-genes (Bennetzen *et al.*, 1998). The composition of plant NTR-DNA seems to be the result of multiple invasions by retro-transposons that display a high degree of sequence variability (Marillonnet and Wessler, 1998) and in most cases retro-transposons represent elements that have lost the ability to transpose.

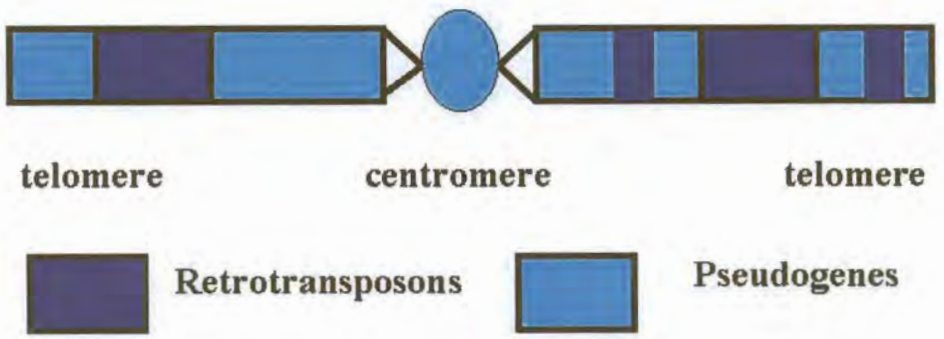


Figure 1.2: Genome organization and gene distribution in a plant genome (Sandhu and Gill, 2002).

Copy number can also distinguish between both functional and non-functional sequences. Sequences in a genome that do not share homology with any other sequences in the same genome are considered as single copy elements. This single copy class contains both functional and non-functional elements. Sequences that do share homology with one or more other genomic regions are considered to be repeated or multi-copy elements. The most abundant multi-copy elements found within the genome of plants are retro-transposons. Highly repetitive DNA tends to accumulate only in regions of low recombination, such as the centromeres and telomeres, where recombination is suppressed. In contrast, repeats occurring in regions of high recombination are much more susceptible to crossing-over and tend to be more variable in copy number relative to their array length. Much of moderately repeated DNA consists of transposable elements.

1.3 Plant genome variation

1.3.1 Induction of genome variation

Plants frequently encounter stresses or external conditions that adversely affect growth, development, or productivity. Stresses can be biotic, imposed by other organisms, or abiotic, arising from an excess or deficit in the physical or chemical environment. Physical factors and their interactions are important in determining performance and distribution of plants. Of foremost relevance are temperature and water availability, additional interferences are light intensity, lack of nutrients and organic/inorganic pollutants. Temperature, water, radiation and nutrient stresses are responsible for as much as 50% reduction in crops yields. Synthesis, accumulation and storage of proteins are mostly affected by temperature stress whereas growth is slowed by almost all stresses. In general, drought stress and high temperatures are accepted to be the most widespread abiotic stresses experienced by crop plants.

The genome has been considered for long as stable to stress despite occasional changes in chromosome structure or inversions. However, recent research suggests that the genome is rather flexible and can undergo changes, which are often referred to as plasticity (Capy, 1998). Such changes might occur naturally over long time periods during evolution. Since plants are unable to move and search for favorable growth conditions, they have to adapt their genome to the changing environment. Walbot and Cullis (1983 and 1985) proposed that once the ordinary physiological responses to an environmental stress are exhausted, the plant genome has to adapt to the new environment by rearranging its DNA in limited genomic regions, which might be related to phenotypic effects.

Genome variation in plants as a response to stress can further be either genetic or epigenetic (Kaeppler *et al.*, 2000; Cassells and Curry, 2001; Abe *et al.*, 2002). Genetic changes include both chromosomal gross rearrangements and changes in the DNA sequence, whereas epigenetic changes are primarily alterations in DNA methylation. All parts of the genome may not be equally susceptible so that variation in the genome is

dependent on a particular stress being experienced. However, some regions in the DNA sequence of the genome may be more susceptible during stress than others and therefore might alter irrespective of the inducing stress (Figure 1.3).

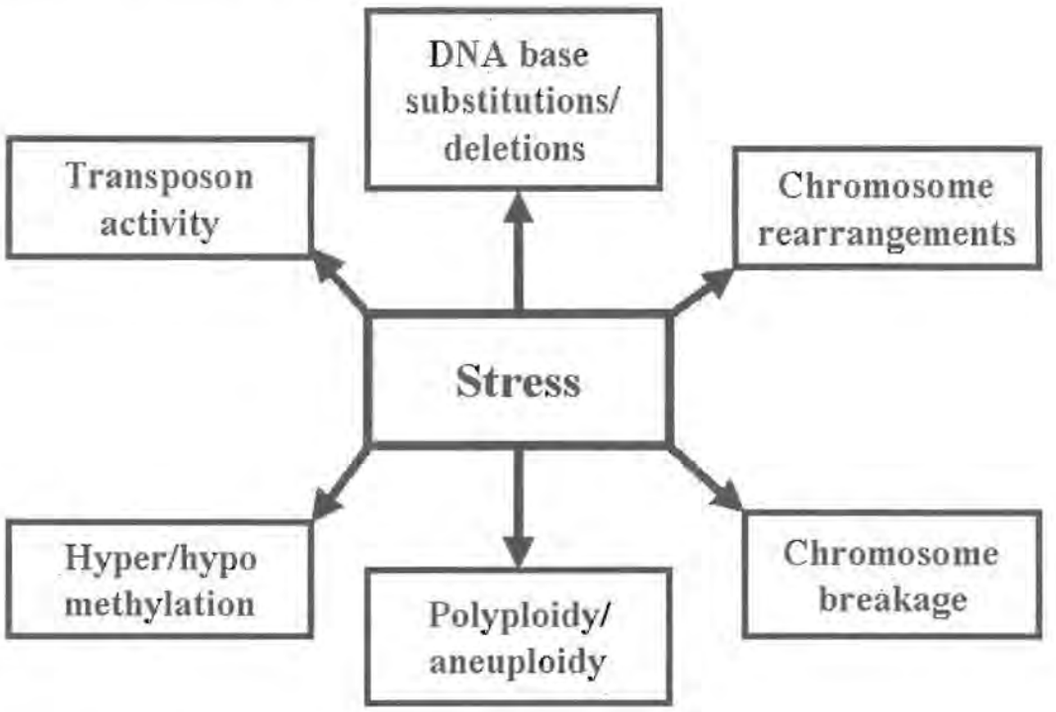


Figure 1.3: Stress- induced genome variation.

1.3.2 Repetitive DNA and DNA sequence variation

Most plant and also animal genomes consist largely of repetitive DNA. Stretches of nucleotide sequence that occur one or only a few times in the genome of a plant can represent as little as 5% of the DNA, while repetitive sequences, typically one to 10 000 nucleotides long, are present in hundred or thousands of copies in the genome (Schmidt and Heslop-Harrison, 1998). Among the repetitive DNA, retro-elements have been found in the genomes of all plant species that have been examined and they seem to be highly abundant in species with large genomes. This suggests that retro-elements, particularly

retro-transposons, account for most of the great variation in plant genome size (SanMiguel *et al.*, 1996) (Figure 1.4).

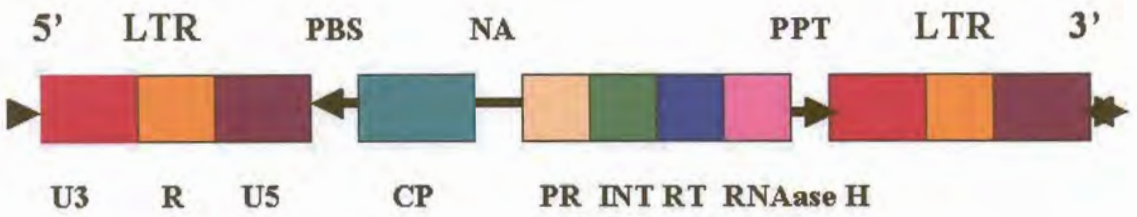


Figure 1.4: General structure of a retro-element (TY1-copia retro-transposon) with long terminal repeats (LTR) in direct orientation at each end. Within the LTRs are U3, R, and U5 regions that contain signals for initiation and termination of transcription. Also showing the unique gene order within two open reading frames encoding for a capsid-like proteins (CP); Protease (PR); Integrase (INT); Reverse transcriptase (RT); and RNAase H. Other sequences featured are PBS (primer binding sites), PPT (polypurine tracts), NA (nucleic acid binding moiety) (after Kumar and Bennetzen, 1999).

Retro-elements transpose without excision and their mobility will always increase their copy number and thereby increase the genome size. The magnitude of transposable elements in genomes of plants, such as maize, can make up more than 50% of the genome (SanMiguel *et al.*, 1996). In maize, the presence of multiple low-copy number families further indicates that a great number of distinct retro-element families exist (Voytas, 1996). However, any genetic damage caused by the integration of retro-elements has to be minimized, because the host's survival is vital for persistence of the element. In yeast for example, in which retro-transposons have been studied extensively, five families of retro-transposons have been found with a strong bias for sites in the genome where they integrate. Retro-elements have been found particularly upstream from tRNA

genes or at the telomeres and regions targeted are typically devoid of open reading frames (Voytas, 1996).

Repetitive DNA sequences are especially sensitive to stress-related DNA changes and account for a large portion of variation in sequence copy numbers. Chemical stress, such as application of an auxin-type plant growth regulator, can amplify AT-rich satellite DNA, whereas exposure to the plant hormone gibberellic acid can increase GC-rich fractions (Nagl and Rucker, 1976). Highly repeated sequences were amplified up to 75-fold in rice suspension cultures (Zheng *et al.*, 1987) and reduction in copy number of a highly repetitive DNA sequence in plant tissue culture of alfalfa (*Medicago sativa*) was also recently reported (Pluhar *et al.*, 2001).

Ribosomal RNA sequences are another highly repetitive sequence family, which can vary (Blundy *et al.*, 1987). Ribosomal RNA (rRNA) is transcribed from DNA as a large RNA precursor that is subsequently processed. Two types of ribosomes are known in higher plants as the 70S and 80S ribosomes (Ting, 1982). The 80S ribosomes are located in the cytoplasm and the 70S are located in the chloroplast and mitochondria. These ribosomes contain smaller subunits and are repeated and arranged in one or more tandem arrays (Nierras *et al.*, 1997). With the exception of some legumes, almost all plant chloroplast genomes, including tobacco, contain two copies of a large inverted repeat, with a size of between 20 and 25 kb. The inverted repeat regions contain the 16S, 23S and 5S rRNA genes as well as some tRNA and ribosomal protein genes, and separate the large single-copy (LSC) and small single-copy regions (SSC) (Lu *et al.*, 1996). In contrast, the rRNA unit in the cytosol consists of the 18S, 5.8S and 25S rRNA coding regions with non-coding spacers with the 5S rRNA genes being present as tandem arrays elsewhere in the genome (Haberer and Fischer, 1996). In the mitochondrion, rRNA is made up by the 18S, 5S and 26S coding units and non-coding spacers (Heldt, 1997). Copy numbers of rRNA genes are highly variable between plants species ranging from a few hundred to thousands of copies per haploid genome, for example *Linum usitatissimum* (flax) contains about a 1000 copies per haploid genome, while *Arabidopsis thaliana* contains about 570 repeats per haploid genome (Cullis, 1979; Pruitt and Meyerowitz, 1986). Some of these

rRNA genes, such as the 5S rRNA gene, are highly conserved in the coding region and are useful tools to study evolutionary relationships in organisms. Although the ribosomal RNA sequences are highly variable, stress-induced DNA changes in these regions have not been investigated in great detail. So far, only a decrease in ribosomal RNA genes in callus culture of flax and changes in the amount of rDNA and peroxidase isozyme band patterns in flax exposed to stress have been reported (Cullis, 1981; Blundy *et al.*, 1987).

1.3.3 Detection of genome variation

Genome variation can be detected by several molecular techniques, which are either non-polymerase-chain reaction (PCR) based, such as Restriction Fragment Length Polymorphisms (RFLP), or based on a PCR reaction. The introduction of the polymerase chain reaction (PCR) has enabled molecular biologists to measure more efficiently with a molecular marker variation on the genome level. By measuring genotype, rather than phenotype, a genetic marker avoids complicating environmental effects and provides ideal tools for assessing genetic variation, identification and defining genetic relationships (O'Hanlon *et al.*, 2000). In PCR, two oligonucleotide primers are hybridized to the opposite DNA strands and allowing amplification of a target DNA sequence. The elongation of the primers is catalyzed by a heat-stable DNA polymerase via a series of temperature cycles, which involve DNA template denaturation, primer annealing, and extension of the annealed primers by Taq polymerase an exponential accumulation of a specific DNA fragment is achieved.

The genomes of closely related plants or varieties might be identical except for differences in a few coding genes or in minor genome re-organizations. Among the PCR-based techniques that are being used in the differentiation of such plants are the analyses of r-DNA intergenic regions (Scribner and Pearce, 2000), simple sequence repeats (SSRs), which are also known as microsatellites, random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphic DNA (AFLP). The two most widely molecular techniques currently used to detect plant variation with a PCR-based technique are Random Amplified Polymorphism DNA (RAPD) analysis, which detects DNA

polymorphisms amplified by arbitrary primers (Williams *et al.*, 1990) and Amplified Fragment Length Polymorphisms (AFLPs) (Vos *et al.*, 1995). In the following a brief description of widely used techniques is given, which includes the technique of Representational Difference Analysis used in this study. Figure 1.5 summarizes the characteristics of the different techniques, which have been used for detection of genome variation.

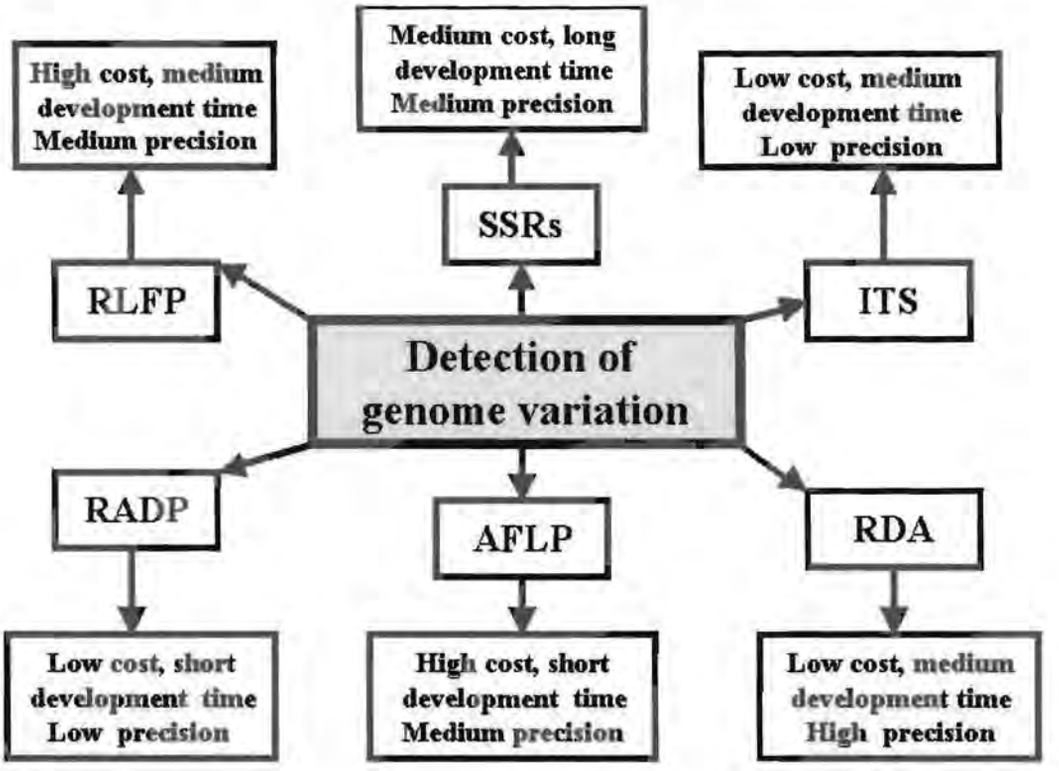


Figure 1.5: A summary of the qualitative characteristics of DNA based techniques to detect genome variation in plants. Cost = financial requirements to prepare a laboratory work and obtain results. Development time = time required to develop genetic assays that depend on the availability of primers. Precision = diversity present within a sample. (adapted from O'Hanlon *et al.*, 2000).

1.3.3.1 *ITS*

The nuclear genes coding for ribosomal RNAs (rRNA) occur hundreds of times as tandem repeats on one or more chromosomes of a haploid set (Figure 1.6). They have been used for a variety of molecular studies including phylogenetics of angiosperms (Vodkin and Katterman, 1971). Each rRNA gene is transcribed into one continuous primary transcript, from which the ribosomal 18S, 5.8S, and 25S RNAs are cut. An "external transcribed spacer" (ETS) and two short "internal transcribed spacers" (ITS1 and ITS2) are discarded. The ribosomal RNAs (rRNA) transcription units are separated by "non-transcribed spacers" (NTSs), and intergenic spacers (IGSs).

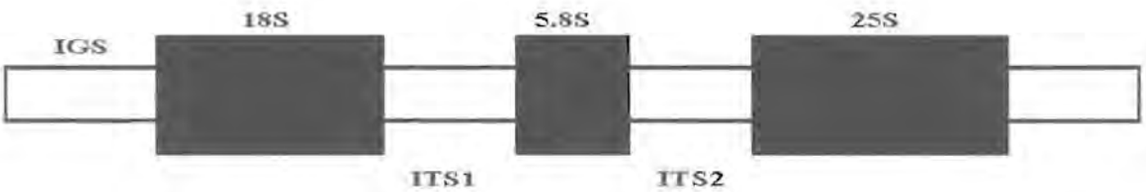


Figure 1.6: Ribosomal RNA genes in the cytosol. Arrangement of the 18S - 5.8S- 25S RNA gene complexes. IGS = intergenic spacer; ITS = internal transcribed spacer (Henry, 1997).

The clusters of 5S rRNA genes are particularly suitable for the analysis of genetic variation using PCR. This is because the genes occur in all of the eukaryote organisms investigated to date in tandemly repeated units comprised of a 120 bp coding region with a non-transcribed spacer of variable length between them (Dvorak *et al.*, 1989). Whilst the genes themselves show a very high degree of conservation (Long and David, 1980), the non-transcribed spacer sequence can vary widely as it is apparently not under the same rigorous selection pressure. PCR to examine 5S-rRNA gene clusters and their spacer have been also used to assess variation and to judge its value in identifying plants from DNA in wheat (*Triticum aestivum*) (Cox *et al.*, 1992; Ko *et al.*, 1994); barley (Kanazin *et al.*, 1993, Ko *et al.*, 1994); rye (Ko *et al.*, 1994), maize, sorghum and oat (Ko

et al., 1994). Each intergenic spacer typically contains simple sequences tandem repeats of 80-325 bp in length in different species. For the ITS region, Fernandez *et al.* (2001) have found that the region was considerably variable within *Doronicum* species corresponding to 265 bp in ITS1 and 231 bp in ITS2. A molecular phylogeny was further constructed using nucleotide sequence of the ITS regions of nuclear ribosomal DNA to elucidate the evolutionary history of the genus *Neolaugeria* (Rubiaceae) (Moynihan and Watson, 2001); *Dillantia* (Compositae: Liabeae; Funk and Robinson, 2001); *Styrax* (Styracaceae; Fritsch, 2001).

1.3.3.2 *RFLP*

RFLP analysis has been available for approximately 15 years. RLFP analysis relies on differences in DNA sequence that affect the position of restriction enzyme recognition site on the DNA at which it cuts double-stranded DNA. Homologous DNA fragments from different individuals cut to different lengths by a restriction endonuclease constitute RFLPs. The fragments are separated by gel electrophoresis (Figure 1.7.A) and blotted onto a filter and then probes are hybridized to the target DNA.

RFLPs give highly reproducible patterns but variations in fragments length between individuals or species can arise either when mutations alter restriction sites, or result in insertions/deletions between them (Burr *et al.*, 1983). As a source of RFLPs, random genomic clones or clones from cDNA library can be used (Sambrook *et al.*, 1989). For the chloroplast genomes of several species from diverse angiosperm families there exists complete sets of probes covering the entire molecule among the dicotyledonous, the lettuce (Jansen and Palmer, 1987) and the monocotyledonous *Oncidium excavatum* (Chase and Palmer, 1989). Variation among the maps from related species is due to the gain or the loss of the restriction site due to a point mutation in this way, RFLPs are a statistically characterized random sample of sequence variation all across the cpDNA (Chase and Palmer, 1989). RFLP analysis of nuclear DNA is never as complete and usually not as precise (Song *et al.*, 1988 a, 1988 b). It involves the detailed comparison of

a gene or the search for RFLPs at a random set locus across the genome (Song *et al.*, 1990).

1.3.3.3 *RAPD*

The RAPD technique developed by Williams *et al.* (1990) utilizes short oligo-nucleotides (usually less than 10 nucleotides) for DNA amplification. Large numbers of fragments, which can be polymorphic, are amplified with a single primer by this method at a relatively low primer annealing temperature (Figure 1.7B).

RAPD analysis is useful for detection of genetic variability among different cultivars and varieties. In general the technique is simple and sensitive and provides a PCR fingerprint for related organisms based on the genome. The RAPD technique offers several advantages. It can produce more polymorphisms than for example the non-PCR based RFLP technique. It is simple to use as well as relatively fast, and does not require radioisotopes. A large number of bands can be produced for a single primer and a range of primers are commercially available. The major disadvantage of this technique is the inconsistency of reproducibility. Furthermore, it only detects dominant markers (Williams *et al.*, 1990).

1.3.3.4 *AFLP*

AFLP is based on selective amplification of digested genomic DNA by a series of extended primers and is used to visualize hundreds of amplified DNA restriction fragments simultaneously. AFLP technology combines the power of RFLPs with the flexibility of PCR-based technology by ligating primer-recognition sequences (adaptors) to restricted DNA (Vos *et al.*, 1995). The first step involves restriction digestion of the genomic DNA with two specific enzymes, one a rare cutter (*Mse*I) and the other a frequent cutter (*Eco*RI). Adaptors are then added to the ends of the fragments to provide a known sequence for PCR amplification. If fragments should be amplified, not all the fragments would be resolvable on a single gel (Karp *et al.*, 1997). Primers are thus

designed to incorporate the known adaptor sequence with one to three additional base pairs. The additional base pairs are referred to as selective nucleotides. Because of the added base pairs, PCR amplification can only occur where the primers are able to anneal to fragments that have the adaptor sequence plus the complimentary base pairs to the selective nucleotides (Karp *et al.*, 1997). This kind of amplification results in 50-100 fragments, which can easily be separated using poly-acrylamide gel electrophoresis (Figure 1.7C). More than three additional nucleotides will result in the non-specific amplification of fragments (Vos *et al.*, 1995). Several polymorphisms are detected in a single assay.

Radiolabeled primers can be used to visualize the amplified products with exposure to a X-ray film, but the cost and danger involved make non-radiolabeled and silver staining techniques preferable (Karp *et al.*, 1997). Some advantages of AFLPs are that only small amounts of DNA are needed. Unlike RAPDs that use multiple, arbitrary primers and lead to unreliable, non-reproducible results, the AFLP technique uses only two primers and gives reproducible results. Many restriction fragment subsets can be amplified by changing the nucleotide extensions on the adaptor sequences and hundreds of markers can be generated reliably. High resolution is obtained because of the stringent PCR conditions. No prior knowledge of the genomic sequence is required. The AFLP technique also works on a variety of genomic DNA samples making it very flexible (Karp *et al.*, 1997).

All the evidence so far indicates that AFLPs are as reproducible as RFLPs. They need more DNA and are technically more demanding than RAPDs. Because of the speed and efficiency of the technique, compared to RFLP and RAPD, it is now being used more widely for comparative purposes.

1.3.3.5 *SSR*

Tandem Nucleotide Repeat Markers are the most informative used for studying diversity. They utilize hyper-variable regions of the genome comprised of tandem repeated simple

sequence (Jeffrey *et al.*, 1985). These repeats vary in number (and, hence, length) and are therefore called variable number tandem repeats (VNTRs), although the terms micro-satellites or simple sequence repeat (SSRs) and mini-satellites are used where the basic repeat unit is around two to eight base pairs in length or longer.

Micro-satellites or simple sequence repeats (SSRs) are tandem repeats of short motifs (2-6 bp) inherited in a single locus, co-dominant, and Mendelian manner. Di- (CA) n, tri- (AAT) n and tetra-nucleotide (GATA) n repeats are the most common DNA sequences (Litt and Luty, 1989; Figure 1.7D). These repeats were first observed in human sequences. Their used is being expanded now in plants. The most frequent motifs in plants seemed to be AT, GT, AG/TC, TAT (rice GAA, GCG). Most micro-satellite loci are located between genes or within introns. They are extremely abundant markers. Although most work to date has involved repeats of di-nucleotides, especially the (GT) n repeat, other simple tandem repeats are present such as (T) n and (TTTA) n (Litti and Luty, 1989; Moore *et al.*, 1991). These repeats are highly polymorphic, even among closely related cultivars, due to mutations causing diversity in the number of repeating units. Micro-satellites may have arisen due to unequal meiotic exchange or slippage during replication. The source of variation for micro-satellites is the number of repeats within a block of tandem repeats. This number can vary greatly so that any given locus may possess a large number of alleles. An example of SSRs is shown in Figure 1.7D.

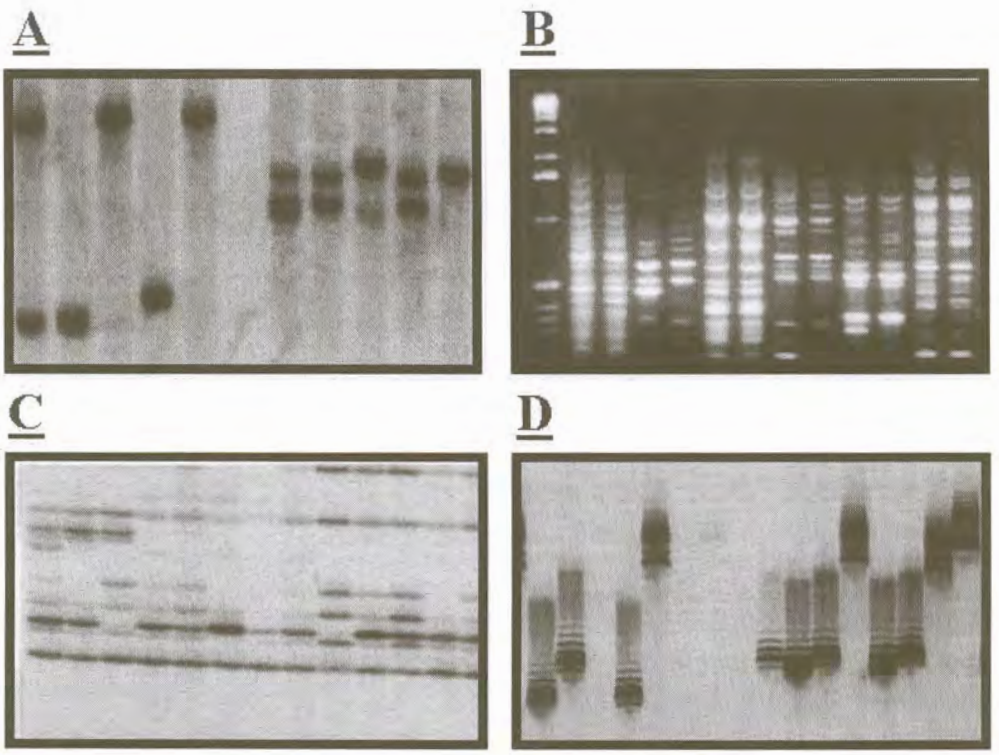


Figure 1.7: Comparative patterns of some DNA based techniques used to detect genome variation in plants. (A) represents an example of a RFLP obtained from five different and unrelated, inbred maize lines digested with *Eco*RI (five tracks on the left) or *Hind*III (five tracks on the right) and probed with a ^{32}P- labeled single copy maize sequence (Karp *et al.*, 1996). (B) shows some RAPD patterns obtained from two near isogenic lines of maize amplified with one of eight operon primers (Karp *et al.*, 1996). (C) shows a typical AFLP profile obtained from different breeding lines of sugar beet amplified with a ^{33}P-labeled (*Pst*I) primer and a non-labeled (*Mse*I) primer (Karp *et al.*, 1996). (D) shows an example of SSRs profile obtained from different inbred lines of sunflower using a single ^{33}P-labeled micro-satellite primer pair (Karp *et al.*, 1996).

1.3.3.6 *RDA*

Representational difference analysis (RDA) has attracted much attention as a technique to isolate and characterize genome differences in plants as well as to identify genome

variation in plants (Cullis and Kunert, 1999; Cullis and Kunert, 2000; Kunert *et al.*, 2002; Vorster *et al.*, 2002; Figure 1.8). The RDA technique is a powerful DNA-based molecular subtractive technique to isolate labile hyper-variable DNA regions of the plant genome, which might have changed due to adaptation to the environment. Essentially, the method consists of a subtraction of all sequences that are held in common between two individuals, which might be morphologically identical, but differ for example significantly in their tolerance to environmental conditions or in the expression of certain morphological or biochemical characteristics. Technically, the RDA technology combines representation, subtractive hybridization, and kinetic enrichment. Representation means a production of the sub-population of DNA fragments derived from a given DNA population, such that the complexity is lower than the sequence complexity of the initial DNA. Representations, which reduce complexity at least ten-fold over the complexity of the genome of a higher organism with genomes as complex as grasses, are generally required for the success of the subsequent steps. The representations derived from the target DNA is designed as tester, while the control material is known as driver. A tester is generated by removal of the R-adaptor used to generate the representations, and the ligation of a new adaptor of unphosphorylated 12 and 24 base oligo-nucleotides. Only the 12-mer provides the appropriate end structure to permit ligation of the 24-mer to the digested DNA, and is not linked to the DNA. The representation of two nearly identical genomes will differ not only when there are absolute differences between the sequence content in the two genomes, but also sometimes when rearrangements or point mutations alter restriction endonuclease fragment lengths. But, because the representation does not have the complexity of the whole genome, not all of the potential differences between two genomes will be found. The amplified representation then serves as a starting material for successive rounds of subtraction and amplification. Subtractive hybridization can be explained as elimination of similar sequences by hybridization between two representations and obtaining of unique sequences present in only one of the representations. The driver and tester are then mixed at different ratios, melted, and allowed to anneal under optimal conditions. Three types of hybrids can be formed. The abundant driver/driver hybrids are formed most frequently, but they lack adaptors and cannot generate a primer-binding site during

the initial fill-in reaction and are therefore incapable of amplification in the subsequent amplification. Driver/tester hybrids are also formed, but in these hybrids the driver strand is unable to generate a primer-binding site, so that these fragments can only amplify in a linear fashion. The tester/tester hybrids that have the primer-binding site on each 5'-ends can amplify exponentially. Part way through the PCR amplification, single stranded DNAs are degraded with mung bean nuclease, thereby eliminating both driver and un-amplified tester DNA (Figure 1.8). Kinetic enrichment is based on the second order kinetics of DNA re-annealing. The rate of formation of double-stranded DNA is higher for DNA species of higher concentration. The first round of RDA is mainly dependent on subtractive enrichment, but subsequent rounds do heavily rely on kinetic enrichment. In RDA, kinetic enrichment and subtractive enrichment are combined in a single step called hybridization /amplification.

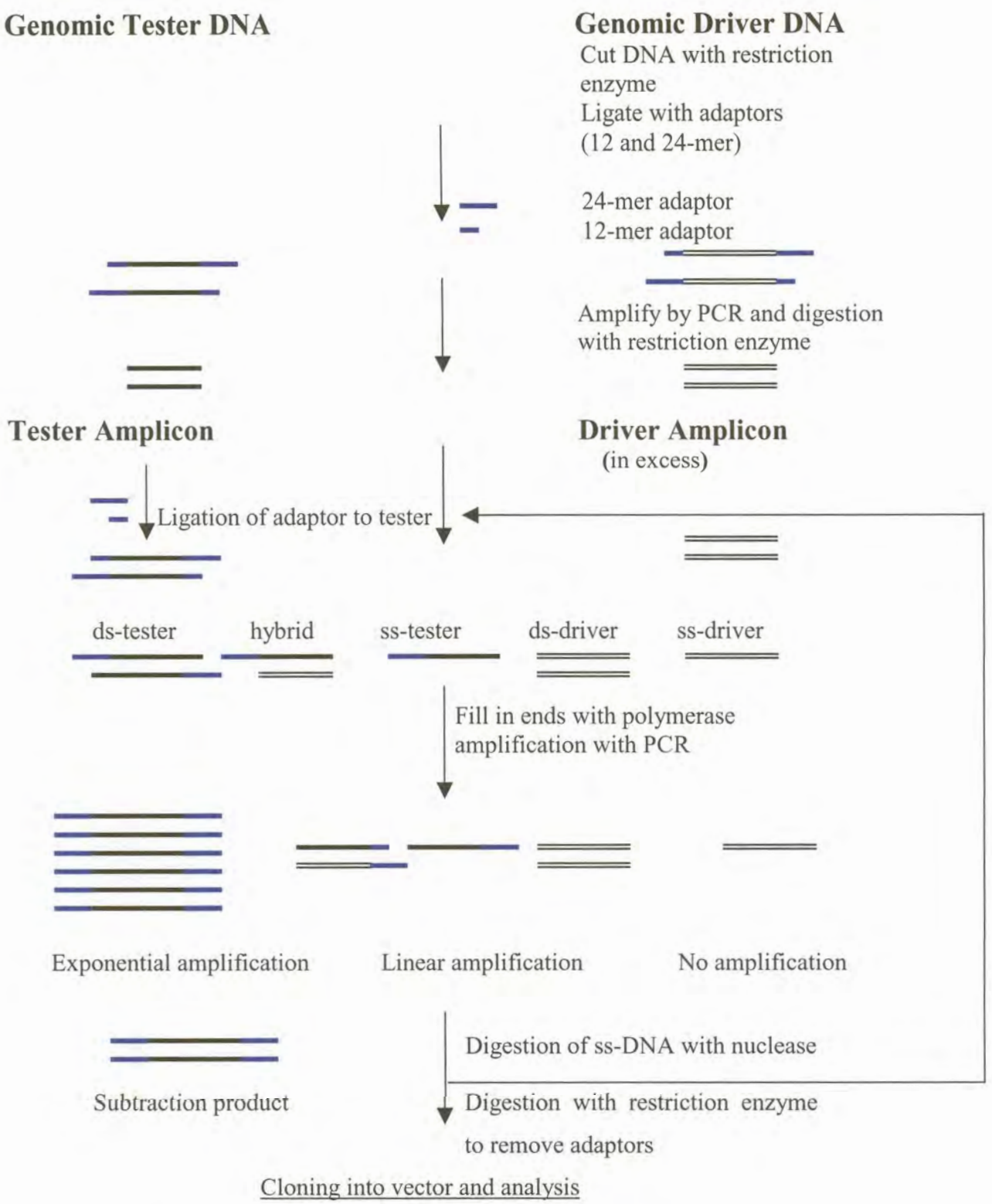


Figure 1.8: Different steps involved in representational difference analysis. (▬) represents adaptor sequence; (▬) represents tester DNA and (═) represents driver DNA.

One of the important features of RDA performed on genomes of different plants is its ability to scan up to 15% of the genome of most plants in each subtraction, in comparison to RFLPs, RAPDs and AFLPs. The use of 300 random primers in RAPD analysis would scan less than 1% of the same genome. The RDA technique has also the potential to give in a relatively short time period direct information at both non-coding and coding regions about genomic losses, rearrangements, amplifications and transposable elements insertion into the genome (Lisitsyn *et al.*, 1993). Another important feature of RDA performed on genomes of different plants is its ability to preferentially isolate families of repetitive sequences that are unique to one of the compared genomes. Such families of repetitive DNA are homoplasy-free characters that can be converted into genetic markers for plant identification in a high throughput PCR-based assay (Nekrutenko *et al.*, 2000).

1.4 Genome variation and grasses

1.4.1 The grass genome

The grasses constitute one of the largest families of the flowering plants. They occur in every habitat available to flowering plants except the seabed, and dominate the vegetation types, which cover about 30% (Shantz, 1954) of the earth's land surface. Grasses are central to human civilization, whether one is considering nutrition, ecology, and aesthetics. Despite its size, the family Poaceae is by flowering plant standards a very coherent one, whose members exhibit characteristic combinations of unusual morphological and anatomical features. In most grasses, genes appear to comprise less than 20% of the genome (Flavell *et al.*, 1977). The huge differences in DNA content of grasses and the differences in the chromosomes number seem to have little or nothing in common with the gene number or order (Devos and Gale, 2000). Much of the difference in genome size is attributable in amounts of repetitive DNA (Flavell *et al.*, 1974). Despite these large differences in DNA content, it has been recognized that the grass genomes maintain a high level of macro-synteny (Gale and Devos, 1998) and a moderately high level of micro-synteny (Tarchini *et al.*, 2000; Keller and Feuillet, 2000). This synteny among the crop grasses suggests that the rice genomic sequence will be more than a tool for understanding the biology of a single species (McCouch, 1998; Gale *et al.*, 2001)

because it can function as a window into the structure and function of genome in the other crop grasses as well (Freeling, 2001; Dubcuvsky *et al.*, 2001). Most genes are expected to function similarly in all grasses and the gene order and synteny are conserved among various Poaceae species (Ahn *et al.*, 1993; Devos *et al.*, 1994; Van Deynze *et al.*, 1995a). This observation of conserved gene content and order in the grasses gave rise to the model that individual grass species could be viewed best as manifestations of a single grass genome and that each of the strengths of studies in different grasses could be used to benefit all individual grass studies (Bennetzen and Freeling, 1993). The numerous rearrangements that do differentiate grass genomes are commonly inversions, translocations, or duplications that involve all or nearly all of a chromosome arm (Moore *et al.*, 1995a). Further, in all Poaceae, the genome is partitioned into gene-rich and gene-poor compartments (Clay and Bernardi, 2001) and the number of genes, gene density, and the extent of recombination vary greatly among the gene-rich regions. The gene-poor regions are composed of retro-transposon-like NTR-DNA and pseudo-genes (Flavell *et al.*, 1974). The NTR-DNA is primarily composed of retro-transposons (Bennetzen *et al.*, 1998; Shirasu *et al.*, 2000; Wicker *et al.*, 2001) and active retro-transposons are a common feature of grass genome (Clayton and Renvoize, 1986; Kellog, 1998; Watson and Dallwitz, 1992).

1.4.2 Characterization of grass genome

Comparative genomics have been performed on economically important grass species such as the staple cereals including rice, wheat, barley, maize, millet, oat and sorghum (Gale and Devos, 1998; Bennetzen, 2000a; Keller and Feuillet, 2000). Restriction fragment length polymorphism (RFLP) was used to produce maps for sorghum (Dufour *et al.*, 1997; Tao *et al.*, 1998). RFLP has also been used to make comparative map from cereals (Binelli *et al.*, 1992; Whitkus *et al.*, 1992). Comparative mapping of cereals using RFLP has also enabled the genomes of barley (*Hordeum vulgare*), wheat (*Triticum aestivum*), and maize (*Zea mays*) to be described in terms of containing rice (*Oryza sativa*) chromosome sections (Moore *et al.*, 1995; Van Deynze *et al.*, 1995a, 1995b).

PCR-based DNA markers composed by tandem repeated short di or tri-nucleotide repeats known as simple sequence repeats (SSRs) markers have also been utilized to map genes in maize (Senior and Heun, 1993; Taramino and Tingey, 1996). The information on the rice-sorghum syntenic regions (Nagamura *et al.,* 1998) gave rise to the development of sequence tagged marked site (STMS) based on simple sequence repeats (SSR) markers for sorghum. 16 loci in the maize genome map were probed by simple sequence repeats (SSRs) (Davis *et al.*, 1999). Further, RAPDs were used to estimate genetic diversity in rice (*Oryza rufipogon*) (Xie *et al.,* 1999) and AFLPs were used for characterizing rice populations (*Oryza sativa*) (Maheswaran *et al.*, 1997). Figure 1.9 summarizes the DNA based techniques used to characterize the grass genome.

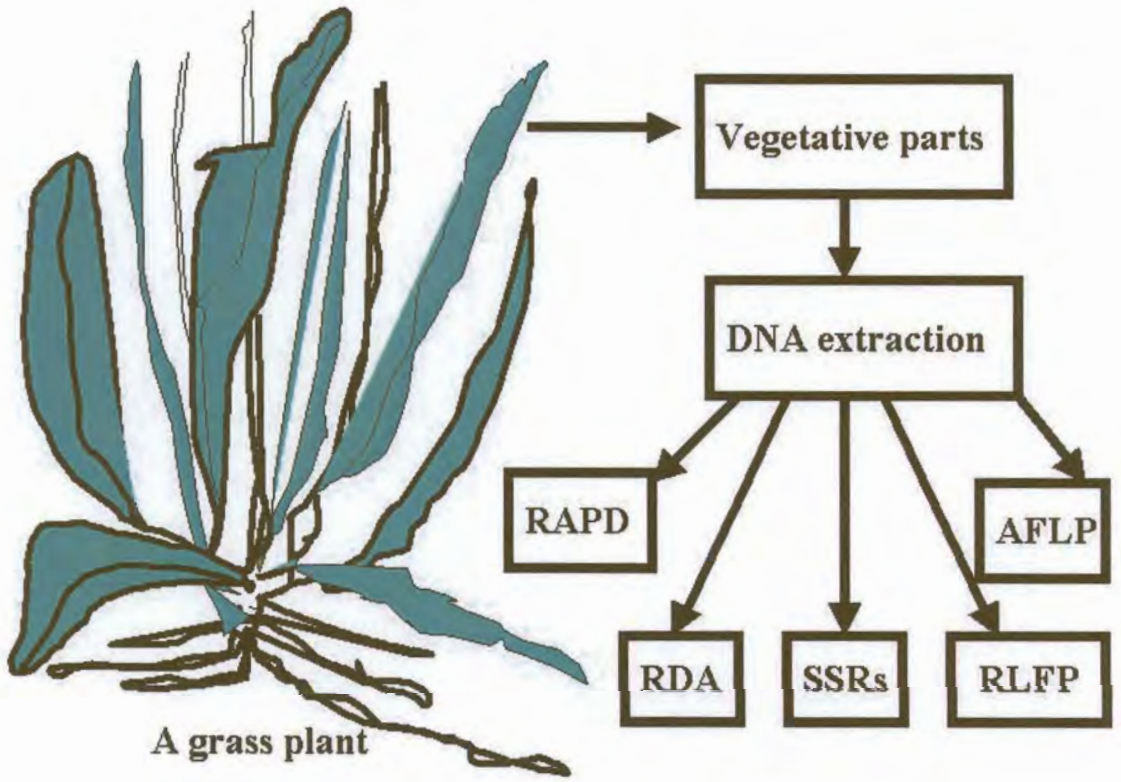


Figure 1.9: A summary of some DNA based techniques used so far to characterize grass genome variation.

1.4.2.1 *Monocymbium ceresiiforme*

In this study, variations in the genome of the grass *Monocymbium ceresiiforme* (Figure 1.10) has been investigated by applying the rather new RDA technique, which preferentially isolates families of repetitive sequences that are unique to one of the compared genomes. *Monocymbium*, with the common name "wild oat grass", has diverged in the tropics along two principal routes. *Monocymbium* can be classified into five species: tropical and southern Africa species; mesophytic species; grassland species; glycophytic species, and savanna species (Clayton, 1972). It is also found in Sudan, Angola, in the West-African rainforest, in Somalia, Ethiopia, and in the southern tropical regions of Africa (Watson and Dallwitz, 1991).

Figure 1.10: Vegetative parts of *Monocymbium ceresiiforme.*

Monocymbium can be diploid or tetraploid (2 n = 20) (Wattson and Dallwitz, 1991). The grass belongs to the grass family Poaceae, which is one of the largest families of flowering plants. The grass has adapted in South Africa (Figure 1.11) to a wide range of ecotypes especially stressful arid habitat. It is widely distributed in open grassland and hillsides in highly leached acid soils in wet places (Watson and Dallwitz, 1992). Flowering from January to June, *Monocymbium ceresiiforme* is an indicator of acid soils. As a late flowering perennial, the grass becomes conspicuous in the South African

Highveld from February until April, when the graceful panicles appear. The whole plant is usually 30-60 cm high, although in wet places it is often taller and more robust (Watson Dallwitz, 1992).

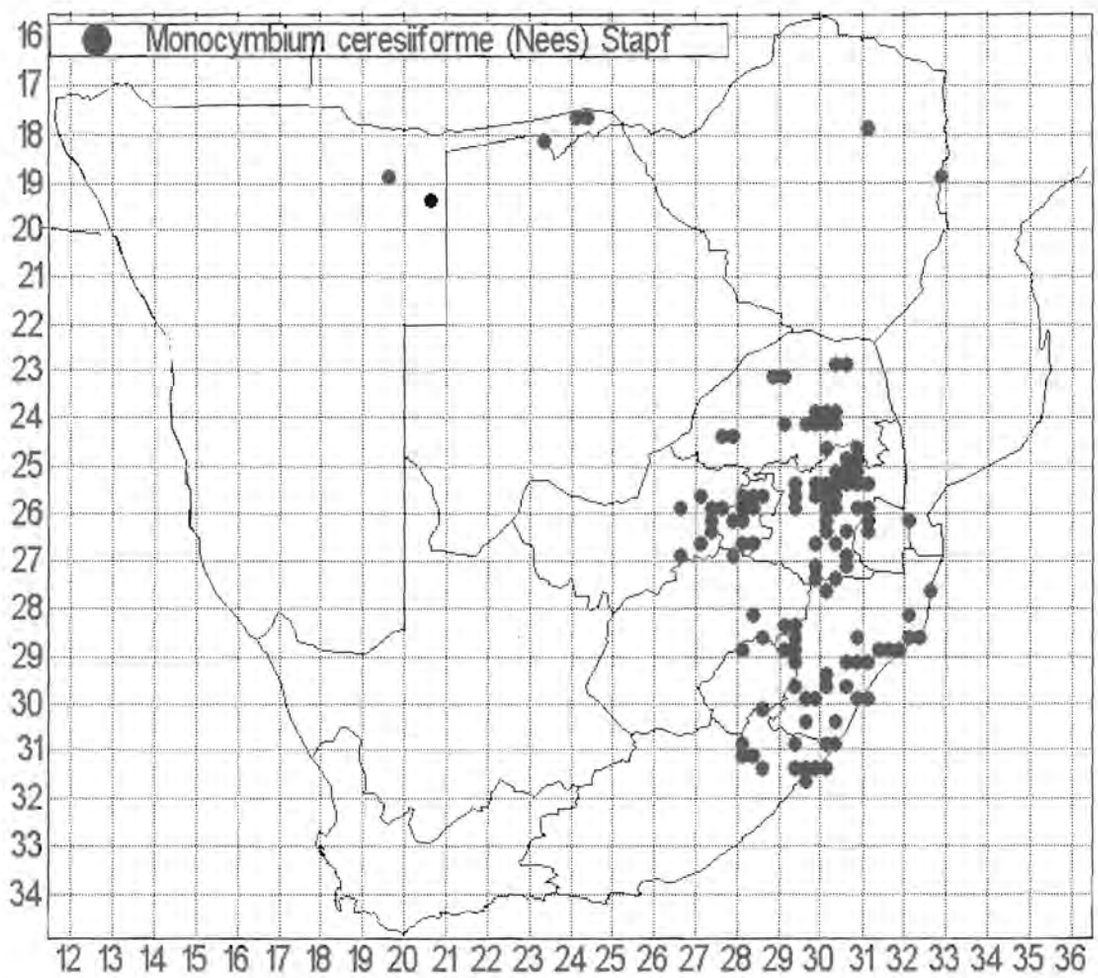
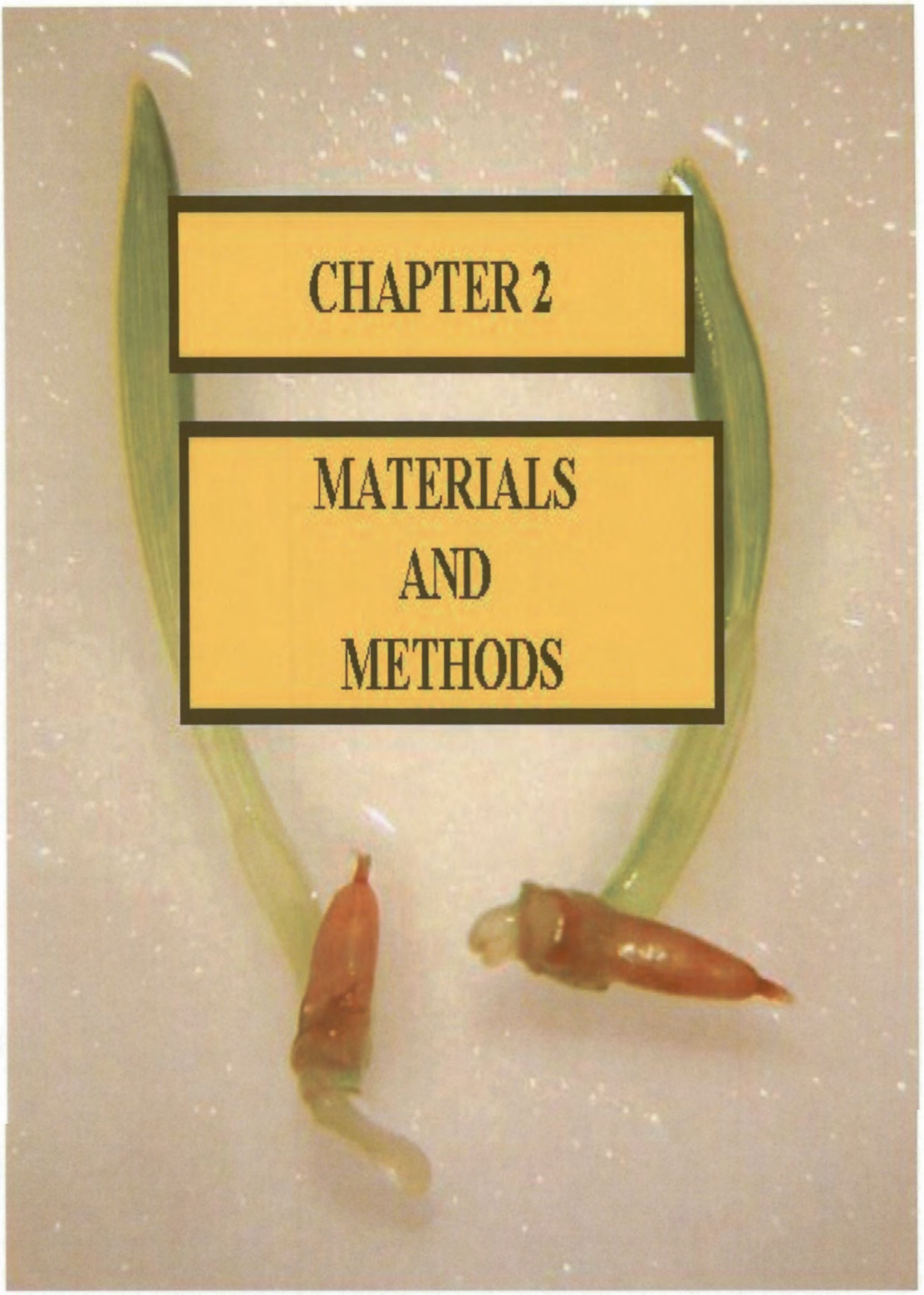


Figure 1.11: *Monocymbium ceresiiforme* distribution map in South Africa (National Botanical Institute).

Despite growing at different altitudes in the savannah and grassland biomes, *Monocymbium ceresiiforme* species have no morphological differences. Therefore, this grass represents an excellent model to identify and characterize possible differences on the genome level. A further interest has been, to evaluate (i) if the RDA technique can detect such differences using for genome subtraction, grasses collected at different

locations in South Africa and (ii) if possibly detected genome differences might indicate any genetic diversity within the grass. For that, the RDA technology was compared with the technique to characterize the ITS region, which has been previously applied to detect genome variation in grasses (Cox *et al.*, 1992; Kanazin *et al.*, 1993; Ko *et al.*, 1994).



CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Plant material

Plant material (*Monocymbium ceresiiforme*) used for analysis in this project was collected randomly in 2001 at three different locations in South Africa. This was at the Golden Gate Highlands National Park, at 2500 meters from the beginning of the Oribi loop, at Drakensberg view. This area is a wet mountain meadow on shallow deep soil. The grass is called 'Drakensberg' in this MSc project. The second grass so called 'Savannah' was collected in the Chimanimani National Park, at an altitude of 2200 meters high, towards the boarder of Mozambique. It is tufted grassland of low ground cover in the poorer white sands, in Maputo land. The third grass so called 'Highveld'was collected in the Rocky Highveld grassland in the dolomite plains of Gauteng at an altitude of about 1500 meters in Pretoria East and near Roodepoort. The collections were done with the help of Profs. K. Kunert, A.Van Wyk (both University of Pretoria) and Prof. L. Mucina (University of Stellenbosch).

2.1.2 Microbial material

A *Bacillus subtilis* sample was provided by Dr Amelita Lombard (Plant Pathology Department, FABI/University of Pretoria). Bacteria were grown overnight at 37°C in 4 ml of LB (Luria-Bertani) medium (LB- broth) comprised of 10 g/l of Bacto-tryptone; 5 g/l Bacto-yeast extract and 10 g/l NaCl, pH 7.4. An undiluted sample (1 ml) was plated onto LB agar medium and incubated without any antibiotic at 37°C for 1 day. Half of the grown bacteria resulting were then collected and re-suspended into 5 ml LB liquid medium and incubated overnight on a shaker at 180 rpm. Grown bacteria were used for bacterial DNA isolation.

2.2 DNA isolation

2.2.1 Grass DNA

2.2.1.1 *Sample preparation*

Genomic grass DNA required to carry out the representational difference analysis (RDA) technique and ITS analysis was extracted from the vegetative part of the grasses according to the method of Gawel and Jarret (1991) which is based on a modification of a combination of methods using CTAB precipitation (Murray and Thompson, 1980; Saghai-Maroof *et al.*, 1984; Webb and Knapp, 1990). The major modifications were increased CTAB concentration, additional mercaptoethanol supplementation, a single extraction with chloroform-isoamyl alcohol, and an additional precipitation of DNA with sodium acetate and ethanol.

For DNA isolation, the grass was cut with a scalpel and placed into a sterile plastic bag and kept in the cold room (4°C) until processing. For DNA isolation about 20 leaves were excised by hand from the culms of field-collected material using sterile gloves to avoid any DNA contamination. For obtaining clean material, sheaths were cut off from the leaf and only the grass blades were used for DNA isolation. For DNA isolation from the flowers, the raceme was first cut off, the awns and reddish spatheoles were then removed from the spikelets. Plant material (1.5 g), which corresponded to about 20 leaves or 60 spikelets, was placed into a sterile Petri dish. Due to the size of the seed, extraction of seeds from the grass was done with a pair of sterile tweezers under the microscope using the highest enlargement for seed removal. All hairy spikelets contained mature seeds, whereas all hairless spikelets were without any seed. For seed isolation, the caryopsis was opened on a sterile Petri dish, the glumes and lemnas were removed and the seed was finally isolated.

2.2.1.2 *Grass DNA isolation*

For DNA isolation, flower or leaf material (1.5 g) was pre-chilled at -80°C and quickly crushed with additional liquid nitrogen in a mortar and pestle and the resulting powder was transferred into a sterile 50 ml plastic centrifuge tube (Sterilin, UK). A pre-heated DNA extraction buffer (7 ml) containing 100 mM Tris-HCl (pH 8); 1.4 mM NaCl; 20 mM $Na_2EDTA.\ 2H_2O$; 1% 2-mercaptoethanol and 3% CTAB) and 2-mercaptoethanol (70 µl) was then added to crushed leaf material and mixed by inverting several times the tube. The mixture was incubated at 65°C for 30 minutes, which was followed by the addition of chloroform:isoamyl alcohol (24:1) mixture (7 ml), and a further incubation for another 5 minutes at room temperature. After centrifugation for 7 minutes at 10000 x g at room temperature, the aqueous phase was filtered through a Miracloth (Amersham Life Science) to remove any remaining cellular debris and an equal volume of ice-cold isopropanol was added to precipitate the DNA. The DNA was collected by centrifugation for 1 minute at 4°C and the pellet was washed with 70% ethanol, dried and re-suspended in 2 ml low TE buffer (10 mM Tris; 0.1 mM $Na_2EDTA.\ 2H_2O$). Any RNA contamination was removed by addition of 2.5 µl of a 10 µg/ml stock solution of RNase and incubation of isolated genomic DNA at 37°C for 30 minutes. DNA was recovered by the addition of 1/10 volume of 3 M sodium acetate (pH 4.8) and 2 volumes of 96% ethanol to the DNA containing solution as outlined by Sambrook *et al.* (1989). The genomic DNA was finally dissolved in 200 µl of low TE buffer.

2.2.1.3 *Grass DNA quantification*

To test for the quality and the quantity of isolated genomic DNA, 20 µl of the DNA solution was run on 1% agarose gel in TAE buffer (0.04 M Tris-acetate; 1 mM $Na_2EDTA.\ 2H_2O$, pH 8) as outlined by Sambrook *et al.* (1989). After separation of DNA on the gel, the gel was stained with ethidium bromide for 15 minutes and the quality of isolated DNA was determined on a UV-trans-illuminator (TFX 20M Vilber Lourmat,

France), photographed with a Grab-IT system (Vacutec, USA) and the DNA concentration of samples was determined by visual comparison with three different known λ DNA amounts (125 ng, 250 ng, 500 ng DNA).

2.2.2 Bacterial DNA

2.2.2.1 Bacterial DNA isolation

Bacterial DNA purification was conducted using a lysis method that ensured recovery from both vegetative cells and spores as described by Spreveslage *et al.* (1996). Cells were pelleted by centrifugation at 13000 x g for 4 minutes and re-suspended in 200 µl of a cold re-suspension solution containing 50 mM glucose; 10 mM Na_2EDTA $2H_2O$; 25 mM Tris (pH 8), supplemented with RNAase A (1 mg/ml). Alkaline-SDS buffer (200 µl) containing 0.2 N NaOH and 1% SDS was added to the mixture. The mixture was mixed gently by inverting the capped tube several times and left for 5 minutes on ice. After addition of cold 3 M NaAc (sodium acetate) (pH 4.8), the mixture was vortexed briefly, kept on ice for 5 minutes and centrifuged for 10 minutes. The supernatant was carefully removed and transferred to a fresh Eppendorf tube. The bacterial DNA was purified with an equal volume of phenol/chloroform and precipitated with 2 volumes of 96% ethanol. The precipitate was incubated on ice for 5 minutes and collected by centrifugation at 13000 x g for 10 minutes. The precipitated DNA was washed with 70% ethanol and collected by centrifugation at 13000 x g for 2 minutes. The pellet was dried under vacuum and re-suspended in 20 µl of dsH_2O.

2.2.2.2 *Bacterial DNA quantification*

Bacterial DNA was quantified by running 10 µl of the isolated DNA extract on 1% agarose gel containing 2 µl of ethidium bromide (10 mg/ml) for staining into a 1xTAE buffer containing (0.04 M Tris-acetate; 1 mM $Na_2EDTA.$ $2H_2O$, pH 8) as outlined by Sambrook *et al.* (1989). The gel was run at ±100V in a (7x10 cm) tray and then visualized on a UV trans-illuminator (TFX 20M Vilber Lourmat, France) to ensure the integrity of the DNA.

2.2.3 Bacterial endophyte

2.2.3.1 Isolation and identification of bacterial endophytes

For the bacterial endophyte isolation, grass seeds were placed in a 1.5 ml Eppendorf tube containing sterile distilled water and crushed with a spatula. The resulting suspension was streaked onto LB medium and incubated at 30°C for 24 hours.

For the bacterial endophyte identification, bacteria isolated from the seed extract were purified and Gram stains performed. The oxidative/fermentative test was also undertaken according to the protocol outlined by De Boer and Kelman (2001).

2.3 DNA amplification and cloning

2.3.1 DNA amplification

Standard DNA amplifications by the polymerase chain reaction (PCR) were carried out in a 100 µl reaction mixture containing 500 mM KCl; 25 mM $MgCl_2$, 100 mM Tris-HCl, (pH 8.3); 25 mM dNTPs and 5 units Takara Taq DNA polymerase (Takara, Japan) in a gene AMP PCR 9600 system (Perkin Elmer, Palo Alto, USA). Primers for PCR were designed using the online tools of molecular Biology Shortcuts (MBS), program "Oligos and Primers" (www.mbshortcuts.com/biotools/index.htm). Primers used in this study were purchased from MWG-Biotech AG Germany and Inqaba-Biotech in South Africa. The standard PCR program consisted of 94°C (5 minutes) to denature the DNA. This was followed by 42 cycles of amplification consisting of denaturing DNA at 94°C (1 minute), primer annealing at 50°C or higher depending on the primer pair (1 minute), and extension of the DNA chain at 72°C (2 minutes). Extension at the last cycle was at 72°C for 5 minutes.

2.3.2 DNA cloning

All PCR-derived DNA fragments from grasses were cloned into the *EcoR*V restriction site of the cloning vector p*MOSBlue* according to the protocol of the supplier (Amersham Life Science, UK). Alternatively, all PCR-derived DNA fragments from bacteria were cloned into the multiple cloning site of the cloning vector *pGEM-T* Easy, which allows cloning of PCR products with a deoxy-adenosine overhang, according to the protocol of the supplier (Promega, USA). For ligation, a ligation buffer (1 µl) containing 66 mM Tris-HCl, (pH 7.6); 6.6 mM $MgCl_2$; 10 mM DDT; 66 mM ATP; and T4 DNA ligase (4 units) was used. Ligations were incubated at 22°C overnight and *MOSBlue E. coli* competent cells (20 µl) (Amersham Life Science, UK) or *JM109 E. coli* competent cells (Promega, USA) were transformed with ligated plasmid DNA by heat shock treatment of cells for 40 seconds at 42°C in a standard procedure as outlined by Sambrook *et al.* (1989). Transformed cells were plated onto LB agar plates containing 10 g/l Bacto-Tryptone; 10 g/l NaCl; and 5 g/l Bacto-Yeast extract. Plates were supplemented with 20 µl of a 100 µg/ml ampicillin, 35 µl of a 50 mg/ml solution of X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and 20 µl of 100 mM IPTG (isopropyl-β-D-thio-galacto-pyranoside) to allow selection of blue/white bacterial colonies. White colonies containing the cloned DNA fragments were randomly picked and plasmid DNA purified from these colonies according to the method outlined by Sambrook *et al.* (1989). Cloned DNA fragments were analyzed after restriction enzyme digestion of plasmid DNA with *BamH*I and *Hind*III (p*MOSBlue*) and *EcoR*I (*pGEM-T* Easy) to release the cloned DNA insert and were finally detected by gel electrophoresis on a 1% agarose gel in TAE buffer.

2.4 DNA analysis

2.4.1 DNA sequencing

DNA sequencing analysis was carried out with the dideoxy chain termination method developed by Sanger *et al.* (1977). Recombinant plasmids were sequenced by primer walking using fluorescent dye terminators and AmpliTaq in a cycle sequencing protocol

according to the recommendations of the manufacturer on an AB1377 automatic DNA sequencer (PE Applied Biosystems). Correctness of DNA sequences was confirmed by Inqaba-Biotech in South Africa. Sequence comparisons and database searches were done with the basic alignment search tool for fast database searching (BLAST and FastA). BLAST emphasizes regions of similarity and FastA emphasizes for similarities between a query sequence and a group of sequences of the same type. The BLAST program was used to compare a nucleotide query sequence against a nucleotide sequence database. The database contains all non-redundant GenBank + EMBL+ DDBJ + PDB sequences (but no EST, STS, GSS, or HTGS sequences).

2.4.2 Southern blot analysis

For Southern blot analysis, the general outline by Sambrook *et al.* (1989) was followed. Total isolated DNA was digested with a respective restriction enzyme and digested DNA was run on a 1% agarose gel in TAE buffer and then blotted overnight onto a nitrocellulose membrane (Stratagene, USA) using capillary forces. For probe labeling, the gene image random prime labeling kit was used (Amersham life science, UK). Labeled probes (subtraction product S40 and DP510) were hybridized to blotted DNA at 60°C overnight, which was followed by different stringency washes and incubation in a liquid blocking solution as recommended by the supplier. Membranes were then incubated for 5 minutes with a 5000-fold diluted anti-fluorescein-AP conjugate to obtain a fluorescence signal. After washing, fluorescence signals on the membrane were finally detected using a gene images CDP-Star detection kit (Amersham life Science, UK), which was followed by exposure to X-ray film for either 15 minutes or 2 hours.

2.4.3 DNA slot blotting

Plasmid DNA (18 ng) of subtraction product was prepared with several dilution series supplemented by 100 ng of tRNA (10 µg/ µl) for 10, 20, 30, 40, 50, 100, and 1000 copy numbers of the subtraction product into the genome. In the dilution procedure, 10 copies of plasmid DNA corresponded to 0.000018 ng or 18.10^{-6} ng; 20 copies corresponded to

0.000036 ng or 36.10^{-6} ng of plasmid DNA; 100 copies corresponded to 0.00018 ng or $18\ 10^{-5}$ ng of plasmid DNA and 1000 copies corresponded to 0.0018 ng or 18.10^{-4} ng of plasmid DNA. The genomic DNA (100 ng) from the three grasses were denatured in a 0.1 N NaOH and 20 µl of 20xSSC for 10 minutes, and immediately chilled on ice. A nitrocellulose membrane (Stratagene, USA) was lightly marked with a pencil to identify each dilution before spotting. The membrane was placed on six layers of Whatman 3 MM paper to allow sucking up the liquid by using capillary forces. Each dilution was dispensed onto the membrane. The membrane was washed twice with 250 µl of 6xSSC and the DNA was then fixed onto the membrane by UV cross-linking. For probe labeling, the gene image random prime labeling kit was used (Amersham life science, UK). Labeled probe (subtraction product DP510) was hybridized to blotted DNA at 60°C overnight, which was followed by different stringency washes and incubation in a liquid blocking solution as recommended by the supplier. Membranes were then incubated for 5 minutes with a 5000-fold diluted anti-fluorescein-AP conjugate to obtain a fluorescence signal. After washing, fluorescence signals on the membrane were finally detected using a gene images CDP-Star detection kit (Amersham life Science, UK), which was followed by exposure to X-ray film for either 15 minutes or 2 hours. Visual comparison of signals obtained from plasmid DNA and genomic DNA determined number of copies.

2.4.4 Colony hybridization

For colony hybridization, LB plates with *E. coli* colonies carrying cloned DNA inserts were pre-chilled for 1 hour at 4°C to prevent the agar sticking to the nitrocellulose membrane. A nitrocellulose membrane (Stratagene, USA) was then placed for 2 minutes onto the agar plates to transfer bacteria onto the membrane. A needle was used to prick the membrane for orientation. For DNA isolation and denaturation from bacteria, the membrane was denatured on a prepared filter paper by soaking into a denaturation solution containing 1.5 M NaCl and 0.5 M NaOH for 15 minutes. This was followed by a neutralization step for 15 minutes where the membrane was placed onto a filter paper soaked in a solution containing 1.5 M NaCl and 0.5 M Tris-HCl, (pH 8). The membrane was then briefly blotted onto Whatman 3 MM paper and placed for 10 minutes onto a

prepared filter paper soaked with 2xSSC buffer containing 0.3 mM NaCl, 30 mM sodium citrate, (pH 7.0). The transferred DNA was finally cross-linked for 4 minutes using an UV-light trans-illuminator (TFX 20M Vilber Lourmat, France). The membrane was used for probe hybridization in a roller bottle by Southern blot analysis. Labeled probe (subtraction product DP510) was hybridized to blotted DNA at 60°C overnight, which was followed by different stringency washes and incubation in a liquid blocking solution as recommended by the supplier. Membranes were then incubated for 5 minutes with a 5000-fold diluted anti-fluorescein-AP conjugate to obtain a fluorescence signal. After washing, fluorescence signals on the membrane were finally detected using a gene images CDP-Star detection kit (Amersham life Science, UK), which was followed by exposure to X-ray film for either 15 minutes or 2 hours.

2.5 DNA subtraction

2.5.1 Amplicon production

For genomic subtraction the RDA technique was applied using the outline reported by Lisitsyn *et al.* (1993) and Vorster *et al.* (2002). In the first step, genomic DNA (2 µg) derived from two types of grasses, where one type served as tester DNA and the other type as driver DNA, was digested in a 100 µl restriction enzyme buffer at 37°C for 90 minutes with 80 Units of the restriction enzyme *Hind*III (Roche, Switzerland). After digestion, digested DNA was analyzed for effective digestion on an ethidium bromide containing 1% agarose gel in TAE buffer. A pair of single-stranded oligonucleotide adaptors of different lengths was used to alter the ends of digested DNA fragments to enable DNA amplification. The longest adaptor was used as the primer for DNA amplification after ligation. For adaptor ligation, digested tester and driver DNA (between 0.5 and 1 µg) were mixed in a total volume of 30 µl with 7.5 µl of a 12-mer and a 24-mer adaptor with a concentration of 58 pmol/µl (RHind 12 and RHind 24, Table A.8). Adaptor DNA was diluted from a 81.9 nmol/ml and 58.2 nmol/ml of adaptors RHind 12 and RHind 24 stock solution respectivelly and the adaptor ligation reaction was carried out in a ligase buffer consisting of 66 mM Tris-HCl (pH 7.6); 6.6 mM $MgCl_2$; 10

mM DDT and 66 μM ATP. To anneal the adaptors, the ligation mixture containing DNA fragments and adaptors were incubated in Eppendorf reaction tubes at 55°C for 5 minutes in a heating block. After heating, the block was immediately placed into a cold room for approximately 2 hours until the temperature dropped in the ligation mixture between 15°C and 10°C. The reaction tubes were then incubated on ice for 3 minutes. After incubation, 4 μl (1 unit/μl) of T4 DNA ligase (Amersham Life Science, UK) was added to the mixture and the ligation mixture was then incubated overnight at 16°C to ligate the adaptors.

For preparation of tester and driver amplicons by PCR, ligated DNA was diluted up to 500 μl with 470 μl of dsH_2O. For DNA amplification, a PCR tube containing a PCR amplification mixture (100 μl), which contained 40 ng of ligated DNA; 372 pmol of the 24-mer adaptor (RHind 24 Table A.8); 10 mM dNTPs (4 μl); 25 mM $MgCl_2$ (6 μl) and PCR buffer consisting of 50 mM KCl; 10 mM Tris-HCl (pH 8.3); 1.5 mM $MgCl_2$ and (0.001% w/v) gelatine, was placed into a pre-warmed (72°C) thermocycler for 10 minutes (Gene amp PCR System, Perkin Elmer, USA). To fill in the oligonucleotide cohesive ends, 6 units of Taq DNA polymerase (2.5 units/μl) (Amersham, Life Science, UK) were added to the PCR amplification mixture after 5 minutes of pre-warming. DNA amplification by PCR was followed using 32 cycles of (11 seconds at 94°C; 2.07 minutes at 72°C) with the last cycle for DNA extension for 10 minutes at 72°C. Approximate total amount of DNA of amplified tester and driver amplicons was determined on a 1.5% agarose gel in TAE buffer with sheared herring sperm DNA as a standard to determine the total amount of amplified DNA produced. Amplified DNA was phenol/chloroform purified and after ethanol precipitation (Sambrook *et al.*, 1989) amplicon DNA was dissolved in TE buffer to obtain a DNA concentration of about 0.5 μg/μl.

2.5.2 First round subtraction and amplification

To cleave adaptors from amplified DNA, driver DNA and tester DNA (40 μg) were digested for 1 hour at 37°C with the 20 units *Hind*III/μg DNA). Yeast tRNA (10 μg) was added to digested DNA, which was then phenol/chloroform purified, ethanol precipitated and finally dissolved in 70 μl of dsH_2O. The tester amplicon DNA (1 μg) from which

adaptors were cleaved was then ligated to a second adaptor pair (JHind 12 and JHind 24; Table A.8) following the procedure outlined under " Ligation of DNA adaptors". Ligated tester amplicon DNA was then amplified following the procedure outlined under "Amplification of DNA by PCR and adaptor removal" but with the addition of 10 extra DNA amplification cycles (42 cycles).

Ligated tester DNA was diluted to 50 ng/µl in a total volume of 70 µl with TE buffer (10 mM Tris-HCl, (pH 8); 0.1 mM $Na_2EDTA.\ 2H_2O$). For hybridization, diluted tester DNA (50 µl) was mixed with driver amplicon (30 µl) with a ratio of driver to tester of 50 to 1. Then 10 M ammonium acetate (12 µl) solution and 96% ethanol (144 µl) were added to the two DNAs and mixed by pipetting up and down using a Finn pipette. The mixture was chilled at -70°C for 10 minutes followed by an incubation period of 2 minutes at 37°C. DNA was precipitated by centrifugation for 10 minutes at 13000 x g and the DNA containing pellet was washed twice with 70% ethanol and dried. The DNA pellet was resuspended in 4 µl EE buffer containing 30 mM EPPS (N- (2-hydroxyethyl piperazine) - N- (3-propene sulfonic acid) (pH 8) and 3 mM Na_2 EDTA. $2H_2O$. The DNA was overlaid with 20 µl of sterile mineral oil and the sample was incubated at 98°C for 5 minutes to denature the DNA. A 5 M sodium chloride solution (1 µl) was directly injected into the DNA drop and the mixture was incubated at 67°C overnight.

The mineral oil was removed and tRNA (10 µg) was added to hybridized DNA and the sample was diluted by adding 200 µl TE buffer to the mixture. To fill the adaptor ends; diluted hybridized DNA (40 µl) was added to 360 µl standard PCR reaction mixtures as outlined under 'DNA amplification'. The solution was divided into 4 separate PCR tubes and 1 µl of Taq DNA polymerase was added in each tube. The solution was incubated at 72°C for 5 minutes after which 4 µl of a 24-mer primer (JHind 24; Table A.8) was added to the solution. Ten cycles of PCR (11 seconds at 94°C and 2.07 minutes at 72°C) were performed using an extension at 72°C for 10 minutes after the last cycle. To evaluate the effectiveness of the hybridization step, 20 µl of the hybridization mixture was amplified for an additional 32 cycles of amplification and any amplification products were visualized on an ethidium bromide containing 1.5% agarose gel in TAE buffer. If the

amplification products were visible, 20 µl of the hybridization were digested with 20 units of mung bean nuclease at 30°C for 30 minutes to remove single-stranded DNA. The reaction was stopped by the addition of TE buffer (80 µl). The digested product was amplified in a standard PCR reaction mixture containing the 24-mer primer (4 µl) (JHind 24; Table A.8). Amplified DNA subtraction products were purified with phenol/chloroform and precipitated with ethanol and finally dissolved in 200 µl of dsH_2O.

2.5.3 Second and third round subtraction and amplification

For the second round DNA subtraction and kinetic enrichment by PCR, the first round subtraction products DNA (5 µg) were digested with 100 units of *Hind*III in a total volume of 100 µl. The digested DNA was phenol/chloroform purified after addition of tRNA (10 µg), ethanol precipitated and re-suspended in dsH_2O to obtain a DNA concentration of 20 µg/ml. DNA (100 ng) was ligated to a third set of adaptors (NHind 12 and NHind 24; Table A.8) in a total volume of 30 µl as described above for first round subtraction and amplification. To ligated DNA, 50 µl of dsH_2O containing tRNA (20 µg/ml) was added so that the mixture (80 µl) contained about 100 ng of DNA. DNA hybridization and kinetic enrichment by a PCR reaction was carried out with 66 ng of ligated DNA (40 µl) and an appropriate amount (100 ng) of driver amplicon DNA (20 µl) as described above.

For the third round DNA subtraction and kinetic enrichment by PCR, the second subtraction products DNA (5 µg) were digested with an appropriate restriction enzyme *Hind*III. The DNA was phenol/chloroform-purified, ethanol-precipitated but was ligated to a second set of adaptors (JHind 12 and JHind 24; Table A.8) as described above. DNA hybridization and kinetic enrichment was carried out with 7 ng of ligated DNA (70 µl) and 10 µg of driver amplicon DNA (20 µl) and the procedure repeated as above.

2.5.4 Fourth and fifth round subtraction and amplification

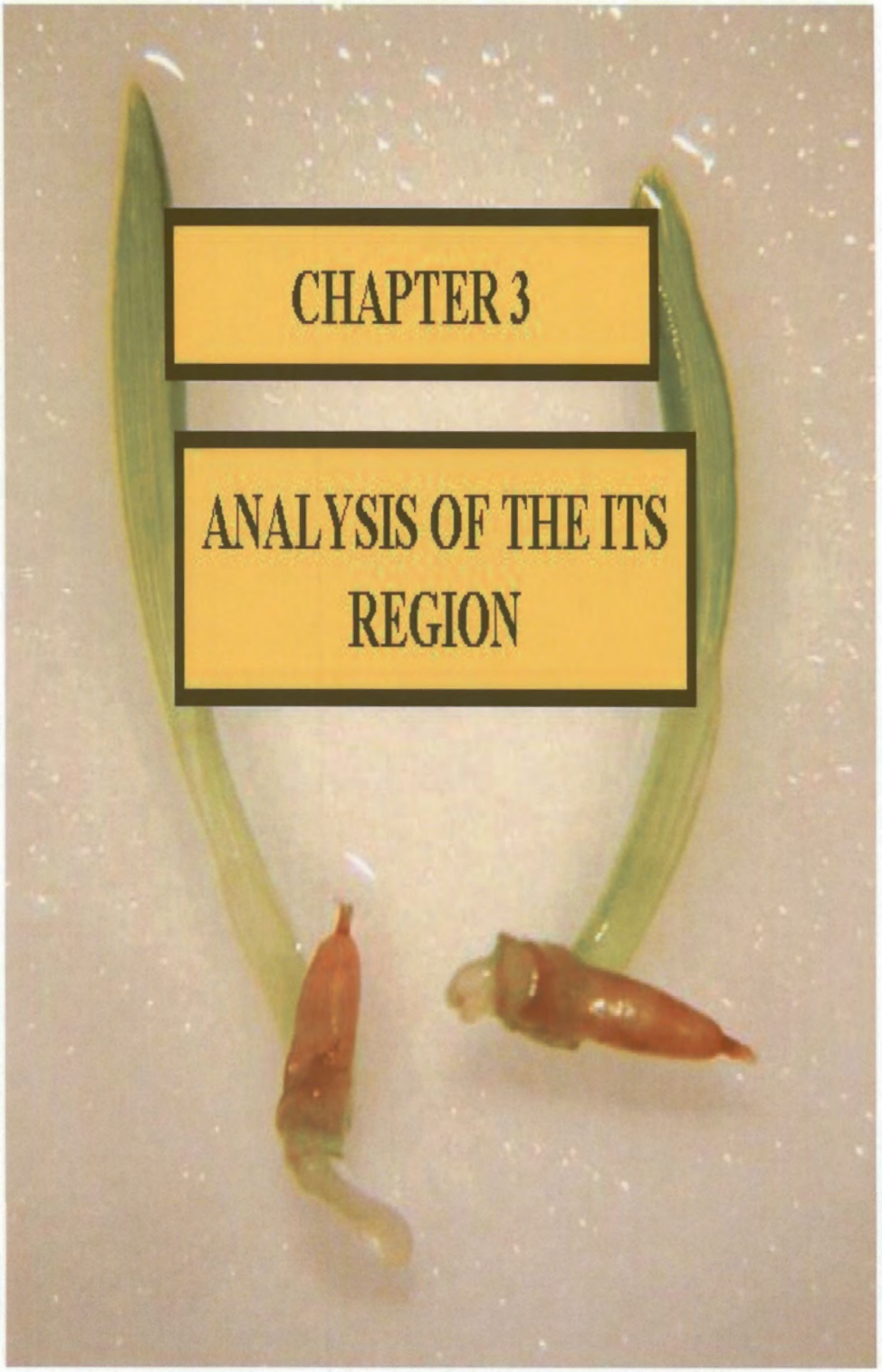
For the fourth round subtraction adaptors of the third round subtraction were changed to a new adaptor set (NHind 12 and NHind 24; Table A.8). The concentration of the ligated

DNA was adjusted to 125 pg/µl using three consecutive dilutions with TE buffer containing tRNA (20 µg/ml). Before DNA hybridization and kinetic enrichment of DNA by PCR, third round subtraction products were digested with *Mse*I to remove partially repetitive DNA. Hybridization and PCR amplification was carried out with 50 pg ligated DNA in 400 µl of low TE buffer and 10 µg of driver amplicon in 20 µl of low TE buffer as described above.

For the fifth round of hybridization and kinetic enrichment of DNA by PCR, the procedure was repeated by using the adaptor pair (JHind 12 and JHind 24; Table A.8) in a mixture of 5 pg of ligated DNA in 40 µl of low TE buffer and 10 µg of driver amplicon in 20 µl of low TE buffer.

2.5.5 Cloning of DNA subtraction products

Final RDA subtraction products were treated with appropriate restriction enzyme to remove ligated adaptors, separated on a 1.5% ethidium bromide containing agarose gel in TAE buffer and visualized on a UV transilluminator. DNA fragments were eluted from the agarose gel and purified using a Sephaglas Band Prep Kit following the protocol given by the supplier (Pharmacia Biotech, USA). Purified DNA fragments were cloned into the *Eco*RV restriction site of the cloning vectors *pMOSBlue* and *pGEM-T EASy*.



CHAPTER 3

ANALYSIS OF THE ITS REGION

3.1 Abstract

Genomic DNA was isolated from the flower parts of the inland grass species *Monocymbium ceresiiforme*, collected at three different locations: Drakensberg grassland (DG), Savannah grassland (SG) and Highveld grassland (HG). Isolated genomic DNA from the flower part was used for amplification and analysis of the internal transcribed spacer (ITS) region and non-coding spacer (NTS) regions. ITS primers amplified a fragment of approximately 500 bp from the genomic DNA of the three grasses, whereas no amplification product was found with NTS primers. A homology search of the amplified ITS sequences with BLAST showed a 99% homology to the internal transcribed spacer 1 (ITS1) and internal transcribed spacer 2 (ITS2) of the 18S rRNA gene, 5.8S rRNA gene, and 28S rRNA gene of the of the fungus species *Cladosporium oxysporium* (accession number AJ3000332.1). These amplified products showed no significant homology to any plant species.

3.2 Objective

The first objective of this part of the study was to isolate high quality genomic DNA from the grasses in order to carry out amplification of the ITS and NTS regions, as well as carrying out the RDA procedure.

3.3 Results

3.3.1 Grass material collection

Figure 3.1 shows the different parts of the inland grass *Monocymbium ceresiiforme*, which were used in the experiments for isolation of genomic DNA. The grasses were randomly selected in the field at different locations in South Africa, namely the Savannah grassland (SG), Drakensberg grassland (DG) and Highveld grassland (HG). Collection

was done in February to April when the flowers of the grasses are visible for identification of the grasses (Figure 3.1A).

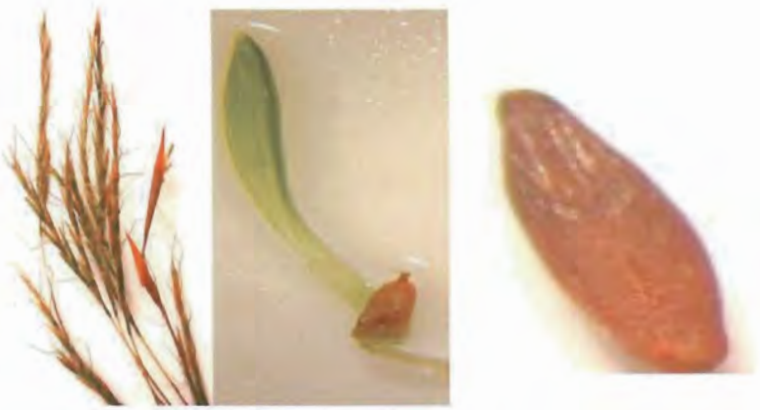


Figure 3.1: Different parts of the inland grass *Monocymbium ceresiiforme*. The vegetative parts shown are: the flower (A), the emerging leaf of the grass (B), and the grass seed (C).

3.3.2 Grass genomic DNA isolation

Genomic DNA was isolated from the leaves, stems, and flowers of SG, DG, and HG *Monocymbium ceresiiforme* grass material. Only non-degraded DNA obtained from the flowers of the grass (Figure 3.2B) was used for amplification of the ITS and NTS regions, and further used for the RDA technique. The different amounts of non-degraded genomic DNAs isolated from the grasses collected at different locations were: 10 µg of genomic DNA per gram fresh flower material from SG grass, 20 µg of genomic DNA per gram of fresh flower material from the DG grass, and 24 µg of genomic DNA per gram of fresh flower material from the HG grass. Non-degraded DNA showed a single high molecular weight DNA band (Figure 3.2B). Degraded DNA derived from either the leaves or stems showed only a smear on an agarose gel and no distinct high molecular DNA band (Figure 3.2A). The degradation of genomic DNA was observed on the gel as a

big smear of DNA without any distinct bands and the non- degradation was observed as a single, high molecular weight DNA band.

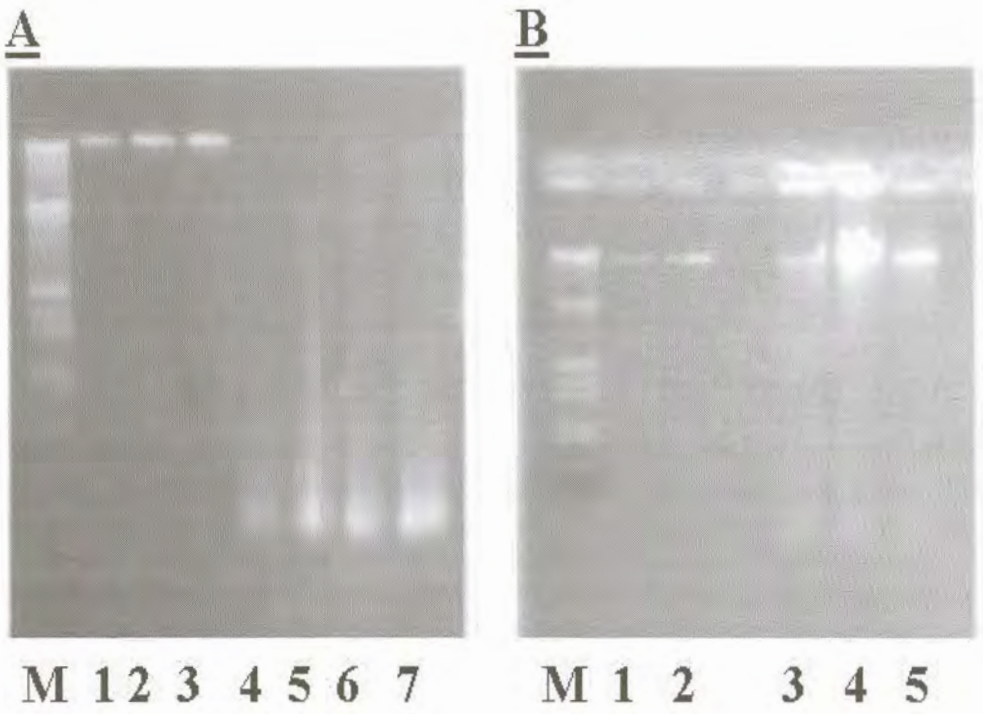


Figure 3.2: Isolated genomic DNA from grasses collected at three different locations, visualized on a 1% agarose gel stained with ethidium bromide. In (A) and (B) lane M represents molecular weight marker III (Roche, Switzerland), while lanes 1 and 2 represent λ DNA of 125 ng and 250 ng respectively. In (A) lane 3 represents 500 ng of λ DNA. (A) Degraded genomic DNA isolated from the leaves and stems indicated by a smear (lanes 4-7). Lanes 4-5 represents DNA isolated from SG grass, lane 6 represents DNA isolated from DG grass, and lane 7 represents DNA isolated from HG grass. (B) Non-degraded genomic DNA isolated from the flowers shown as a high molecular weight band (Lanes 3-5). Lane 3 represents DNA isolated from SG grass, lane 4 represents DNA isolated from DG grass, and lane 5 represents DNA isolated from HG grass.

3.3.3 PCR amplification

The ITS region was successfully amplified out of the total genomic DNA isolated from the flowers of the grasses collected at all three locations (Figure 3.3) using ITS1 and ITS4 primers (Table A.7). An amplification product of about 532 bp was visualized after 42 cycles of amplification. The NTS region, however, could not be amplified from total genomic DNA in any of the grasses.

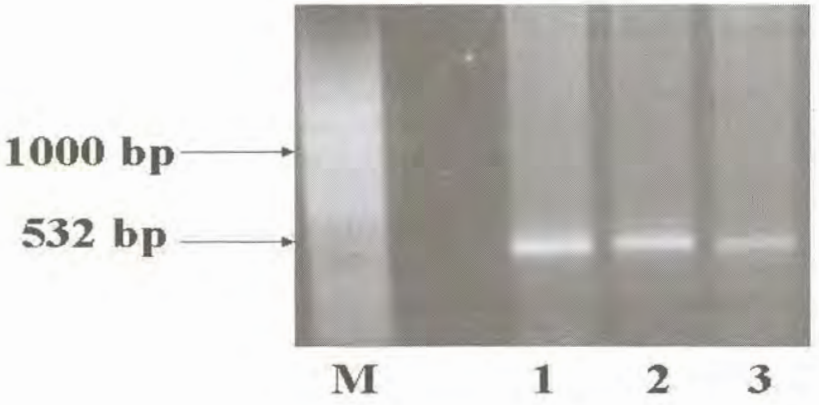


Figure 3.3: The ITS region amplified from genomic DNA of the grasses collected at three different areas visualized on an agarose gel stained with ethidium bromide. Lane M represents molecular weight marker III (Roche, Switzerland). Lane 1 represents the ITS amplified product from the SG grass genomic DNA. Lane 2 represents the ITS amplified product from the DG grass genomic DNA. Lane 3 represents the ITS amplified product from the HG grass genomic DNA. The size of the amplification product (532 bp) is indicated with an arrow.

3.3.4 Sequencing analysis

Since all the amplified DNA products from the three isolates were similar in size, it was decided to sequence only the product obtained from amplification of the ITS region of the SG grass DNA. A sequence of 532 bp was obtained after sequencing of the amplified ITS

region (Figure 3.4). The sequence was analyzed for homology with known sequences using the BLAST online database (Altschul *et al.*, 1990). The homology search of the amplified product showed a 99% homology of the sequence to the 18S rRNA gene, 5.8S rRNA gene, and 28S rRNA gene of the internal transcribed spacer 1 (ITS1) and internal transcribed spacer 2 (ITS2) of the fungus species *Cladosporium oxysporium* (accession number AJ3000332.1). However, the product amplified from the genomic DNA of SG grass showed no significant homology to any plant species.

```
        10         20         30         40         50         60
TCCGTAGGTG AACCTGCGGG GGGATCATTA CAAGTGACCC CGGTCTAACC ACCGGGATGT
        70         80         90        100        110        120
TCATAACCCT TTGTTGTCCG ACTCTGTTGC CTCCGGGGCG ACCCTGCCTT CGGGCGGGGG
       130        140        150        160        170        180
CTCCGGGTGG ACACTTCAAA CTCTTGCGTA ACTTTGCAGT CTGAGTAAAC TTAATTAATA
       190        200        210        220        230        240
AATTAAAACT TTTAACAACG GATCTCTTGG TTCTGGCATC GATGAAGAAC GCAGCGAAAT
       250        260        270        280        290        300
GCGATAAGTA ATGTGAATTG CAGAATTCAG TGAATCATCG AATCTTTGAA GCACATTGCG
       310        320        330        340        350        360
CCCCCTGGTA TTCCGGGGGG CATGCCTGTT CGAGCGTCAT TTCACCACTC AAGCCTCGCT
       370        380        390        400        410        420
TGGTATTGGG CAACGCGGTC CGCCGCGTGC CTCAAATCGA CCGGCTGGGT CTTCTGTCCC
       430        440        450        460        470        480
CTAAGCGTTG TGGAAACTAT TCGCTAAAGG GTGTTCGGGA GGCTACGCCG TAAAACAACC
       490        500        510        520        530        540
CCATTTCTAA GGTTGACCTC GGATCAGGTA GGGATNCCCG CTGAACTTAA GCATATCAAT
       550
AAGCGGAGTG ATGCTGCGTTCTTCATCGATGC
```

Figure 3.4: Sequence data of ITS region amplified from genomic DNA of SG grass. ITS1 and ITS4 primer used for amplification (Table A.7) are underlined.