Isolation and characterization of genome differences in the indigenous grass *Monocymbium ceresiiforme*

By

MARIE CATHERINE ONANENA

Dissertation submitted in partial fulfillment of the requirement for the degree

MAGISTER SCIENTIA

Forestry and Agricultural Biotechnology Institute (FABI)

Department of Botany

In the

Faculty of Natural and Agricultural Sciences

University of Pretoria

SUPERVISOR: PROF KJ KUNERT

CO-SUPERVISOR: DR C VAN DER VYVER

March 2004

TABLE OF CONTENTS

ABSTRACT

In this study genome differences between three types of the grass *Monocymbium ceresiiforme* collected at different locations in South Africa were investigated. For identification and characterization of genome differences, PCR amplification of the ITS region with specific DNA primers designed to amplify the ITS region and Representational Difference Analysis (RDA), which is a rather new subtractive DNA technology for plants, were used. Although PCR products could be amplified with the ITS technique, these products were identified by bio-informatics tools to be of fungal origin possibly due to infestation of seed material with typical grass pathogens. By executing RDA on genomic DNA isolated from grass flowers and application of several rounds of DNA subtraction and kinetic enrichment by PCR reaction, several subtraction products derived from genomic DNA of individual types of grasses were identified and characterized. This included a subtraction product with homology to a highly repetitive maize retro-transposon and a second sequence, called DP510, with homology in part of the sequence to *Bacillus* DNA. Sequence analysis using bio-informatics tools further revealed that DP510 also had homology to genomic *Arabidopsis thaliana* DNA. However, by applying PCR amplification using DNA primers designed to amplify the individual subtraction products, none of the subtraction products was unique to one of the individual grass genomes but was able to identify several variants of DP510. Although experiments were carried out to demonstrate that DP510 has not derived from bacterial contamination of grass DNA, hybridization of labeled DP510 to isolated genomic DNA resulted only in a very weak signal. But, no experiments were carried out for hybridization of bacillus DNA with DP510 by Southern blotting technique. Consequently, there is still a lack of clear indication that DP510 is part of the grass genome.

RESEARCH OBJECTIVES

Molecular tools are increasingly applied for the characterization of plant genomes. Aim of this study was to characterize the genome of the inland grass species *Monocymbium ceresiiforme* occurring in South Africa at different locations with diverse environmental conditions. In particular, the RDA technique has been applied in this study with the objectives (i) to evaluate the potential of the RDA technique for genome characterization and (ii) to identify and characterize possible variations on the genomic level from the grass collected at different locations in South Africa without having any morphological differences, despite growing under different environmental conditions.

DISSERTATION COMPOSITION

Chapter 1 presents a short overview about our current knowledge of the composition of the plant genome including grasses, how stresses can cause variation in the plant genome and where such variation can occur in the genome. This chapter also focuses on the techniques that are widely applied to identify genome variation. Finally, a short introduction is given in this chapter regarding the characteristics of the grass species used in this study. In **Chapter 2** the experimental procedures that have been applied in this study are outlined. **Chapter 3** outlines the results of the different genomic DNA isolation techniques applied to obtain genomic DNA of sufficient quality to carry out a PCR amplification of the ITS region of the different types of grasses investigated. This technique, previously applied by scientists to characterize genome variation in grasses was used as a general molecular technique to identify any advantages of the RDA technique for characterization of genome variations. **Chapter 4** emphasizes on the application of the RDA technique for identification and cloning of putative DNA sequence differences after several rounds of subtractive hybridization and PCR amplification from the different types of grasses used in the study. In **Chapter 5** the results obtained by analyzing the different subtraction products using DNA sequencing and bio-informatics tools for sequence alignments and identification of DNA sequence homologies with known sequence data available in DNA sequence databases are outlined. **Chapter 6** outlines results obtained from PCR amplification of Bacillus DNA sequences and staining techniques for detection of bacterial endophytes in seed extracts to determine possible cross-contamination of grass genomic DNA with bacterial DNA. **Chapter 7** finally outlines the achievements made and also problems experienced in this study regarding the extension of the RDA technique to a further plant species, the identification of repetitive DNA in the grass and the detection of DNA sequences in the plant genome with homology to maize and Bacillus DNA. This chapter also outlines perspectives for future research activities. Finally relevant references listed in this study are listed in **References** and in the **Annexure** details about the composition of buffers, solutions, and other chemicals used in this study are provided.

ACKNOWLEDGEMENTS

First and foremost I would like to thank God, Almighty be all glory, without whom none of this would have been possible. You have guided me thus far; being a companion with your unfailing love, grace and guidance through all lives joys and sorrows. I look forward to traveling this path that you have set for me in your constant presence.

I would like to thank my supervisor, Prof. Karl Kunert for years of patience, enthusiasm and support in teaching. Also for the opportunities he gave to me to share his endless amount of knowledge.

A special thanks goes to M. Du Plessis, whose strong intervention allowed my acceptance at the University of Pretoria.

Thank you to my late mother who passed away the 20th of May 2003. Your death has disturbed deeply this work, but did not discourage me. I would like to dedicate this work to you.

Thank you to my father, my sisters and brother for years of fundamental support and guidance.

This work would not have been achieved without the help of lab colleagues: J Vorster, A. Prins, and A. Kiggundu. Thanks for all your editorial assistance and help with bioinformatics tools to analyze data.

My sincere gratitude goes to Dr C. Van der Vyver for her patience during my lab work, sharing her knowledge and spending her time teaching and guiding me.

I would like to thank the University of Pretoria that allowed me to conduct and complete my research.

My sincere gratitude and appreciation go to Prof. Ladislav (Laco) Mucina, Dept. of Botany, University of Stellenbosch for collecting grasses at Drakensberg.

A big thanks goes also to Prof Braam van Wyk for helping me with the grass distribution and Dr Teresa A. Coutinho for analyzing the bacteria.

Lastly but not the least, I want to thank my husband Boniface, for his constant love, encouragement and confidence in me.

Finally, to my children: Muriel, Raïssa, Glwadys, Donald, Audrey, and Bleuette. I would like through this achievement to be the model of endurance and determination for all of you.

ABBREVIATIONS AND SYMBOLS

°C	Degrees Celcius
%	Percentage
µg	Microgram
µl	Microlitre
A	Adenine
AFLP	Amplified fragment length polymorphism
Amp	Ampicillin
bp	Base pair
BHR	Broad host-range
C	Cytosine
CH_3COOHH_4	Ammonium acetate
cm	Centimeters
CMP	Cytosine monophosphate
cpDNA	Chloroplast DNA
CP	Capside-like protein
CTAB	Cetyltrimethylammonium bromide
DG	Drakensberg grassland
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
dNTP	Deoxynucleoside triphosphate
DP	Difference product(s)
DTT	Dithiothreitol
E. coli	*Escherchia coli*
EDTA	Ethylenediamine tetra acetic acid
EPPS	N- (2-hydroxyethyl) piperazine-N- (3-propane sulfonic acid)
ETS	External transcribed spacer
g	Grams

G	Guanine
HGP	Horizontal gene pool
H	Hours
H_2O	Water
HG	Highveld grassland
HGT	Horizontal gene transfer
IGS	Intergenic spacer
IPTG	Isopropyl-[beta]-D-thiogalactopyranoside
IR_A and IR_B	Inverted repeats
INT	Integrase
ITS	Internal transcribed spacer
KDO	3-deoxy-D-manno-octulosonate
Kpa	Kilopascal
l	Liter
LB	Luria Bertani
LTR	Long terminal repeat
LSC	Large region of single copy gene(s)
M	Molar
mer	Oligomer
mg	Milligrams
MGEs	Mobile genetic element
ml	Millilitres
mm	Millimeters
mM	Millimolar
mtDNA	Mitochondrial DNA
mRNA	Messenger ribonucleic acid
NaCl	Sodium chloride
ng	Nanogram
NaAc	Sodium acetate
NaOH	Sodium hydroxide

NTR	Non-transcribed repeat(s)
NTS	Non-transcribed spacer
nDNA	Nuclear DNA
PCR	Polymerase chain reaction
RAPD	Random amplified polymorphism DNA
RDA	Representational difference analysis
rDNA	Ribosomal DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAase-H	Ribonuclease H
rRNA	Ribosomal ribonucleic acid
rpm	Revolutions per minute
RT	Reverse transcriptase
SDS	Sodium dodecyl sulphate
sdH_2O	Sterile distilled water
sec	Second(s)
SG	Savannah grassland
SP	Subtraction product(s)
ssDNA	Single stranded DNA
SSC	Small region of single copy gene(s)
spp	Species (plural)
SSR	Single sequence repeats
STMS	Sequence tagged marked site
T	Thymine
TAE	Tris-acetate EDTA
T_m	Melting temperature
Tris	2-amino-2- (hydromethyl) propane-1, 3 diol
tRNA	Transfer RNA
UV	Ultraviolet
VNTR	Variable number tandem repeat(s)

X-gal	5-bromo-4-chloro-indol- [beta]-D galactoside

INDEX OF FIGURES

Lane M (A and B) represents a 100 bp ladder (Roche, Switzerland) lane 1 represents (A) SG genomic DNA and (B) amplicon, lane 2 represents (A) HG genomic DNA and (B) amplicon, lane 3 represents (A) DG genomic DNA and (B) amplicon, and lane 4 (A and B) represents 2 µg herring sperm DNA for DNA quantification.

Figure 4.2 **50**

First subtraction products (SP1) of *Monocymbium ceresiiforme* genomic DNA representations visualized on an agarose gel stained with ethidium bromide. Lane M represents a 100 bp DNA ladder (Roche, Switzerland). Lanes 1 and 2 represent the subtraction product resulting from the subtraction of SG and DG grass representations, where SG grass representation was used as tester and DG grass representation as driver (lane 1), and DG grass representation was used as tester and SG grass representation as driver (lane 2). Lanes 3 and 4 represent the resulting products from the subtraction of DG and HG grass representations, where the HG grass representation was used as tester and DG grass representation as driver (lane 3), and DG grass representation was used as tester and HG grass representation as driver (lane 4).

Figure 4.3 **52**

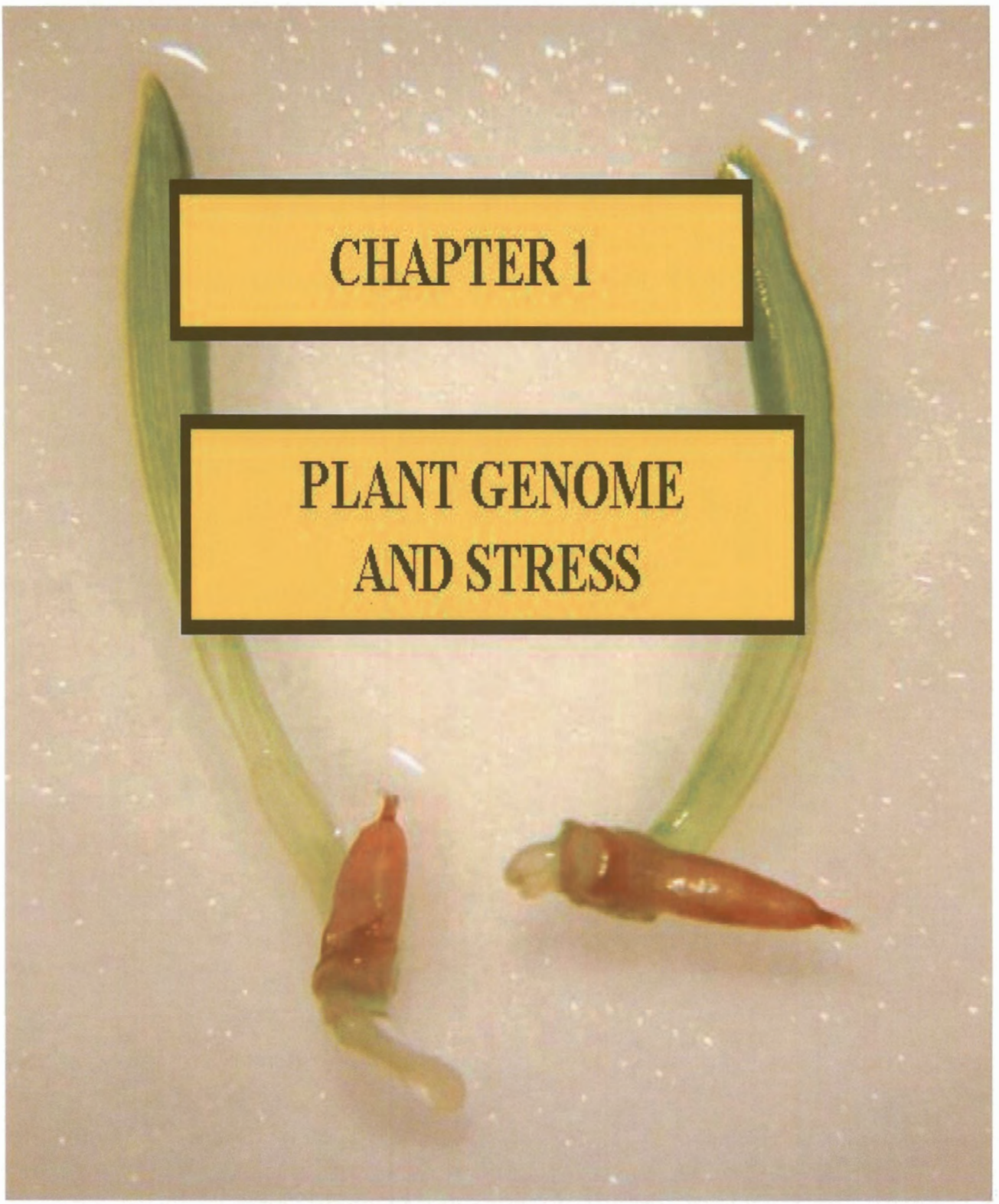
Second subtraction products (SP2) of *Monocymbium ceresiiforme* genomic DNA representations at a 150:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the resulting subtraction products from the subtraction of the DG grass representation and the SG grass representation, where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and DG grass representation was used as tester and the SG grass representation used as driver (lane 2). Lanes 3 and 4 represent the reciprocal subtraction of DG grass representation and HG grass representation, where the HG grass representation was used as tester and the DG grass representation was used as driver (lane 3), and the DG grass representation was used as tester and the HG grass representation was used as driver (lane 4). Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

representations at a 2000000:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the subtraction of the DG grass representation and the SG grass representation where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and the DG grass representation was used as tester and the SG grass representation as driver (lane 2). Lanes 3 and 4 represent the subtraction of the DG grass representation and the HG grass representation, where the HG grass representation was used as tester and the DG grass representation as driver (lane 3), and the DG grass representation was used as tester and the HG grass representation as driver (lane 4). Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

INDEX OF TABLES

"Retrotransposon" primer used to amplify a retro-transposon–like region from the grass species *Monocymbium ceresiiforme*. "DP510" primer set used to amplify a fragment with homology to the *Bacillus halodurans* region and "*Bacillus subtilis*" the primer set used to amplify the *Bacillus subtilis* 16s rRNA region.



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).

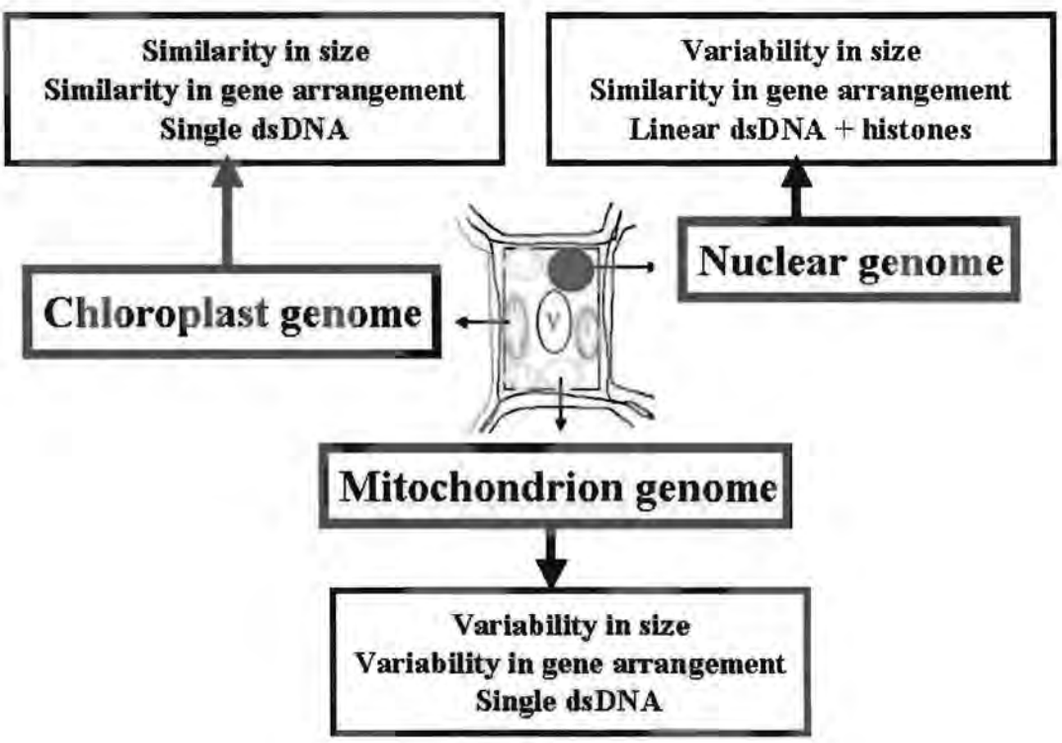


Figure 1.1: The plant cell genomes and their characteristics (V = Vacuole).

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.

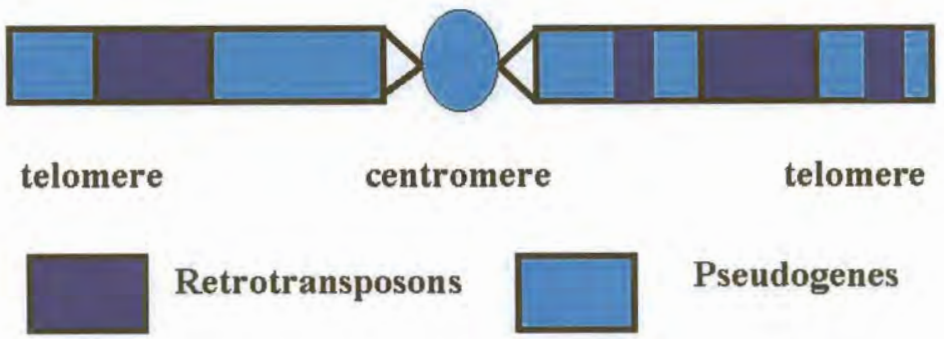


Figure 1.2: Genome organization and gene distribution in a plant genome (Sandhu and Gill, 2002).

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).

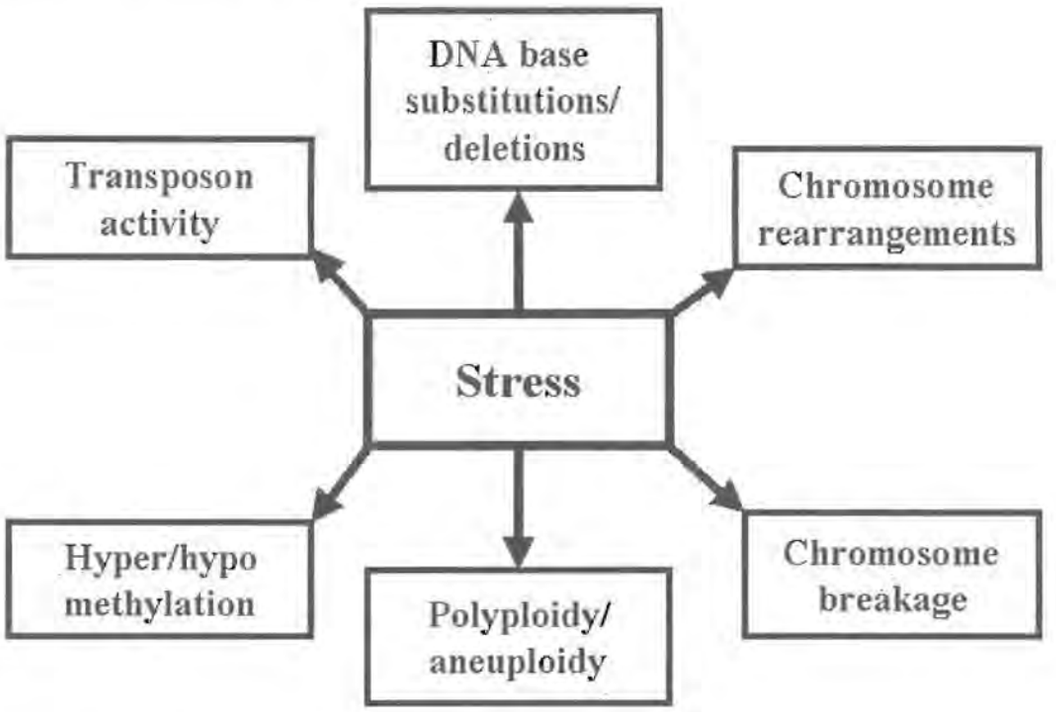


Figure 1.3: Stress- induced genome variation.

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).

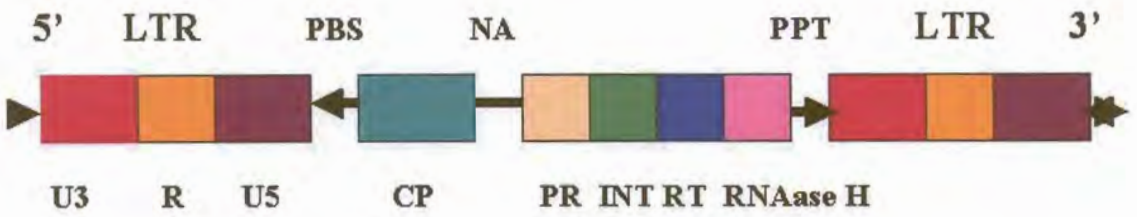


Figure 1.4: General structure of a retro-element (TY1-copia retro-transposon) with long terminal repeats (LTR) in direct orientation at each end. Within the LTRs are U3, R, and U5 regions that contain signals for initiation and termination of transcription. Also showing the unique gene order within two 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.

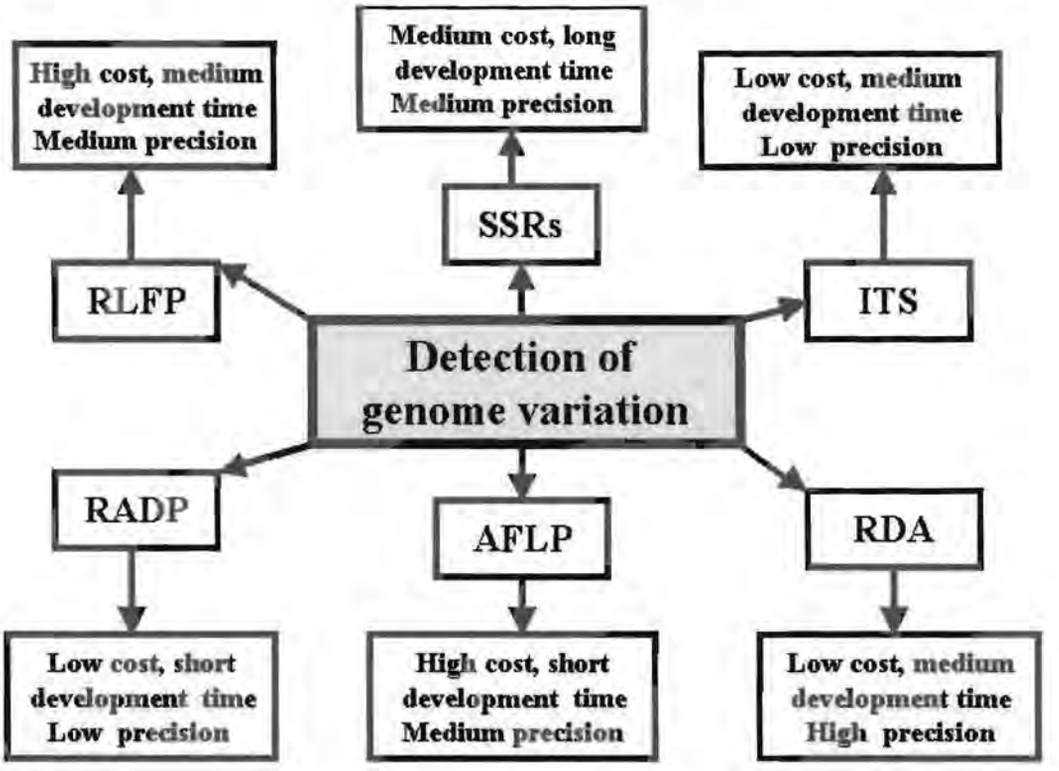


Figure 1.5: A summary of the qualitative characteristics of DNA based techniques to detect genome variation in plants. Cost = financial requirements to prepare a laboratory work and obtain results. Development time = time required to develop genetic assays that depend on the availability of primers. 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).

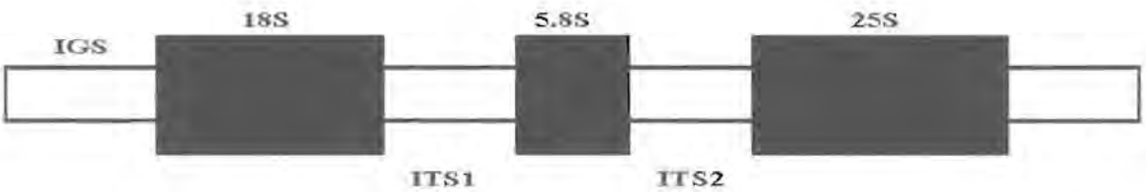


Figure 1.6: Ribosomal RNA genes in the cytosol. Arrangement of the 18S - 5.8S- 25S RNA gene complexes. IGS = intergenic spacer; ITS = internal transcribed spacer (Henry, 1997).

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.

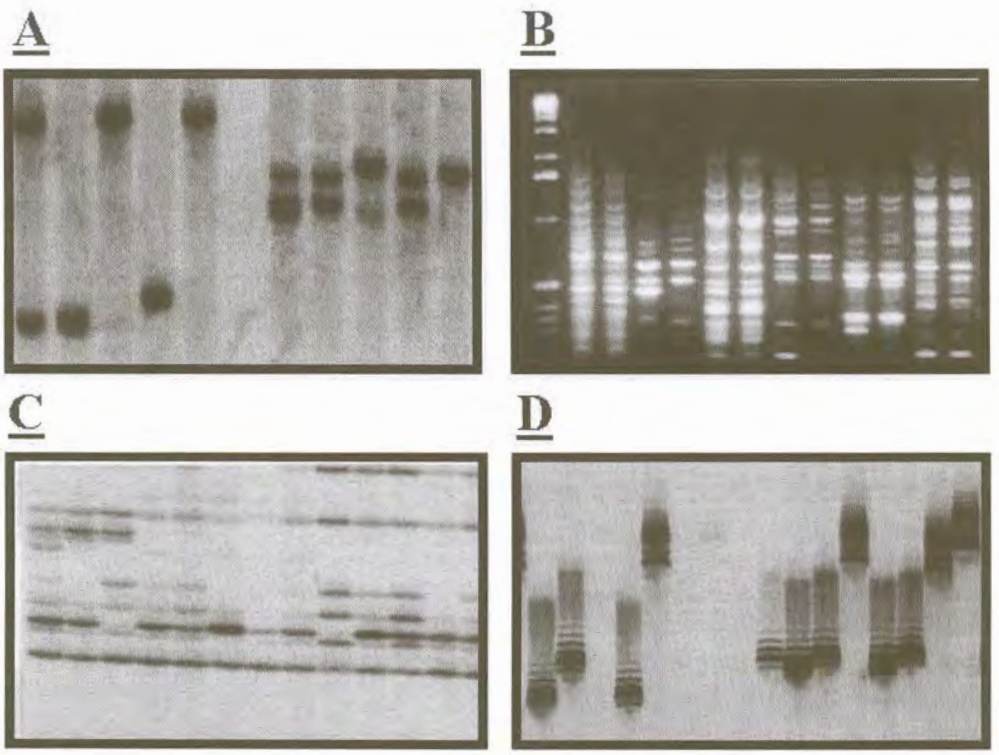


Figure 1.7: Comparative patterns of some DNA based techniques used to detect genome variation in plants. (A) represents an example of a RFLP obtained from five different and unrelated, inbred maize lines digested with *Eco*RI (five tracks on the left) or *Hind*III (five tracks on the right) and probed 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.

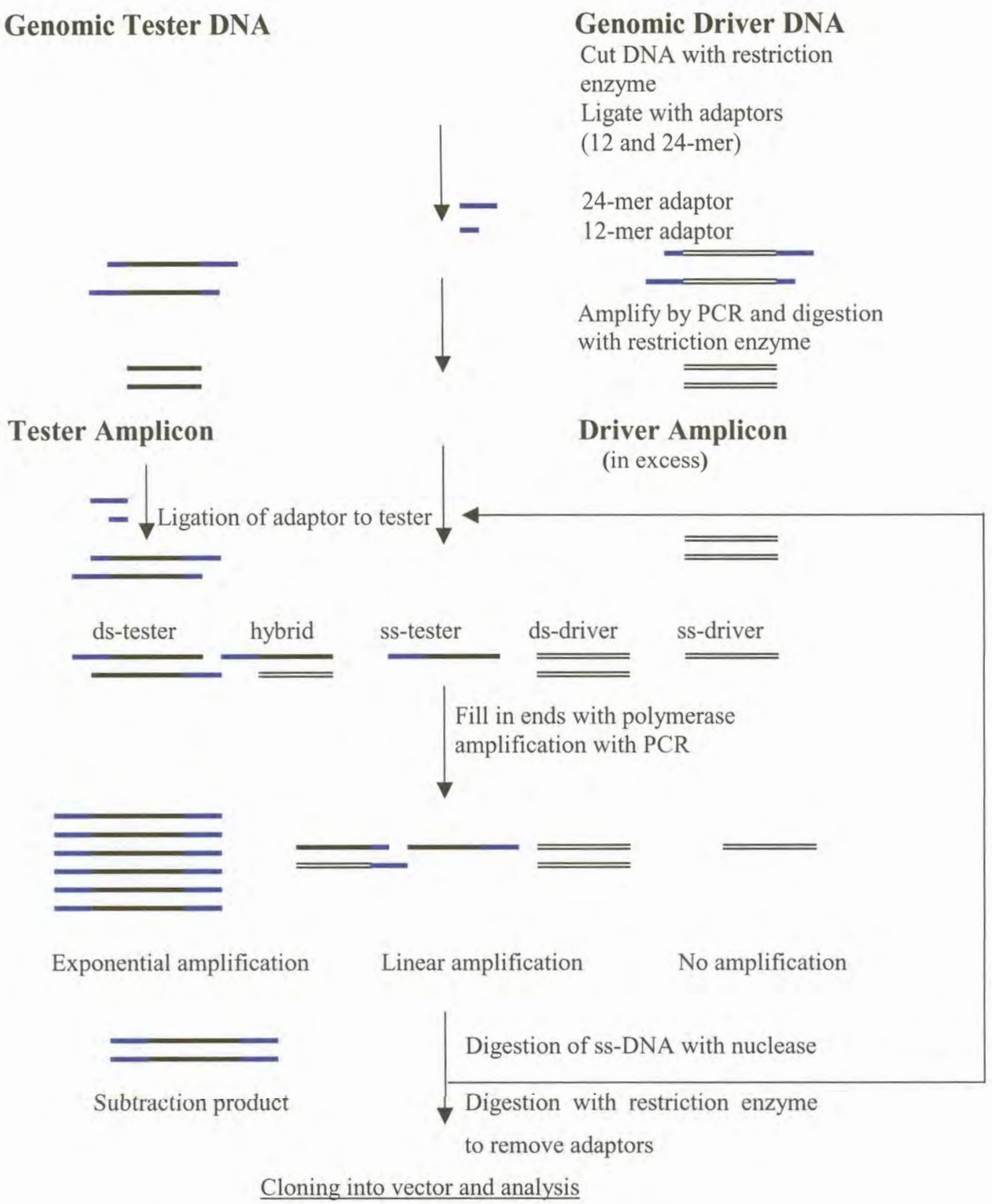


Figure 1.8: Different steps involved in representational difference analysis. (▬) represents adaptor sequence; (▬) represents tester DNA and (═) represents driver DNA.

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.

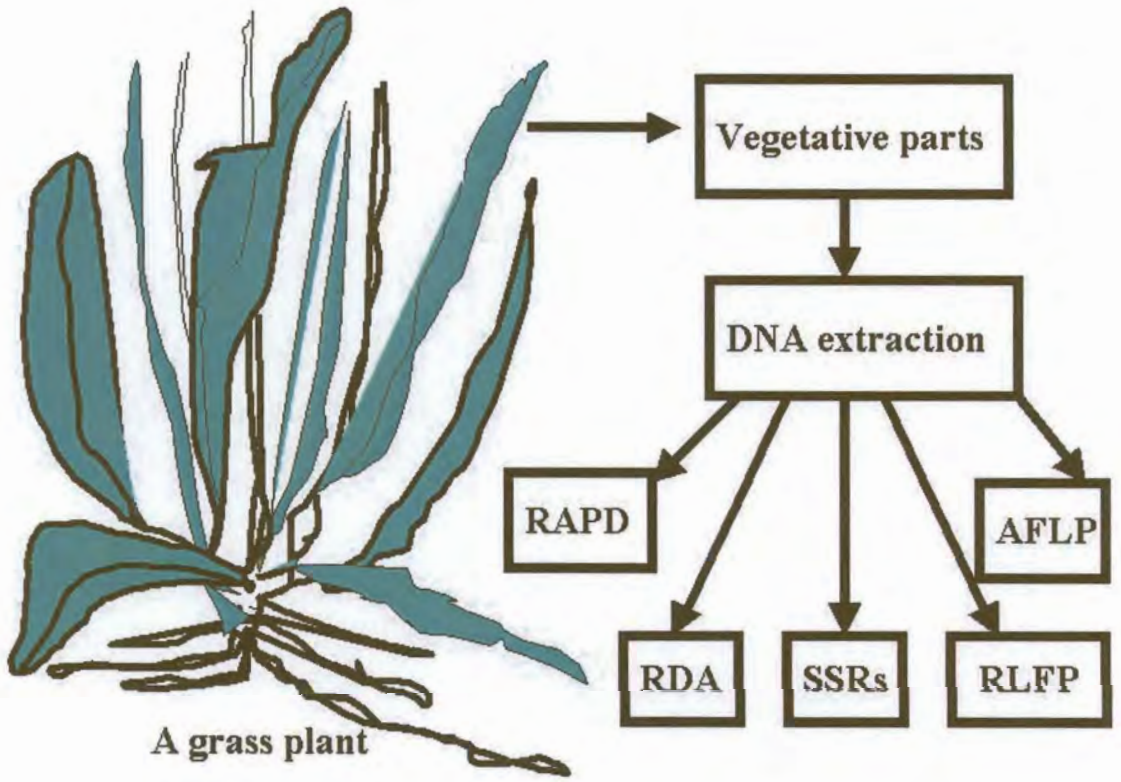


Figure 1.9: A summary of some DNA based techniques used so far to characterize grass genome variation.

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).

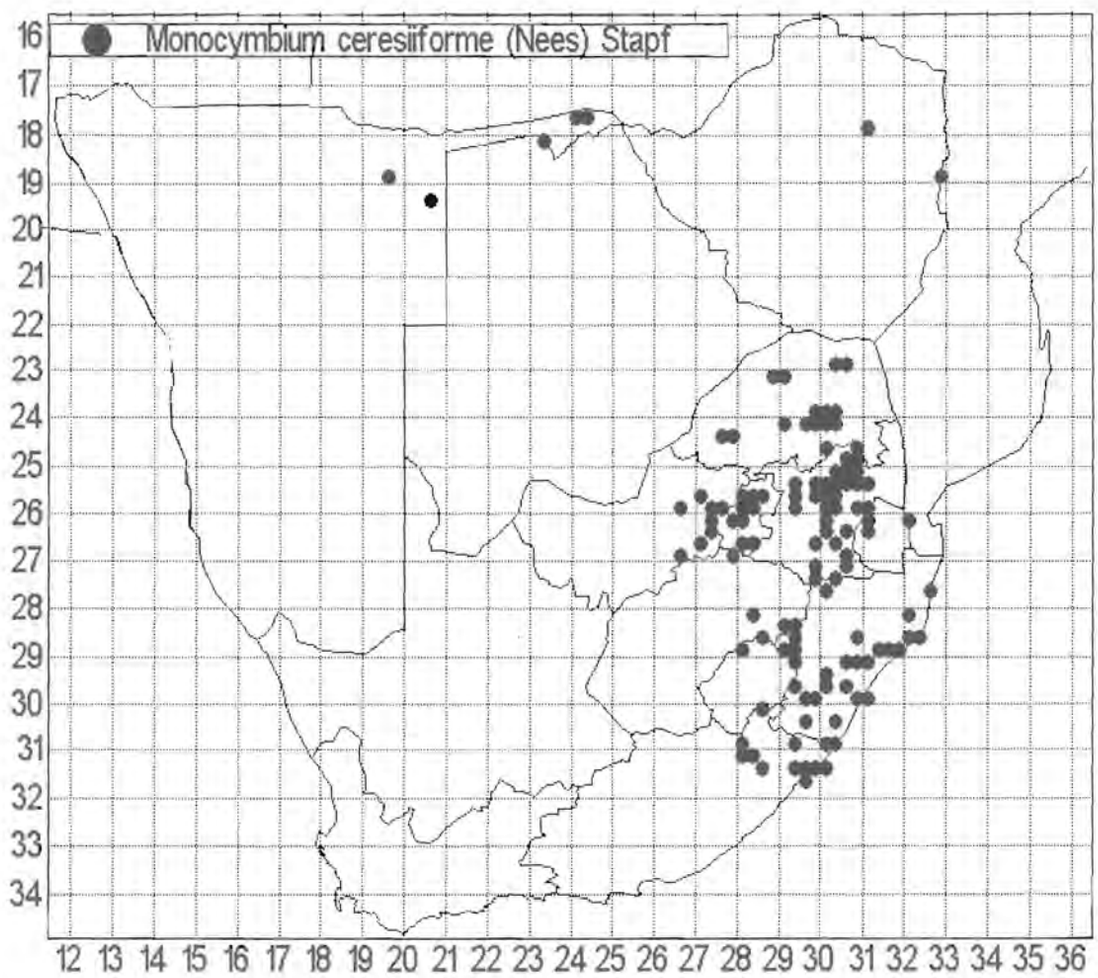
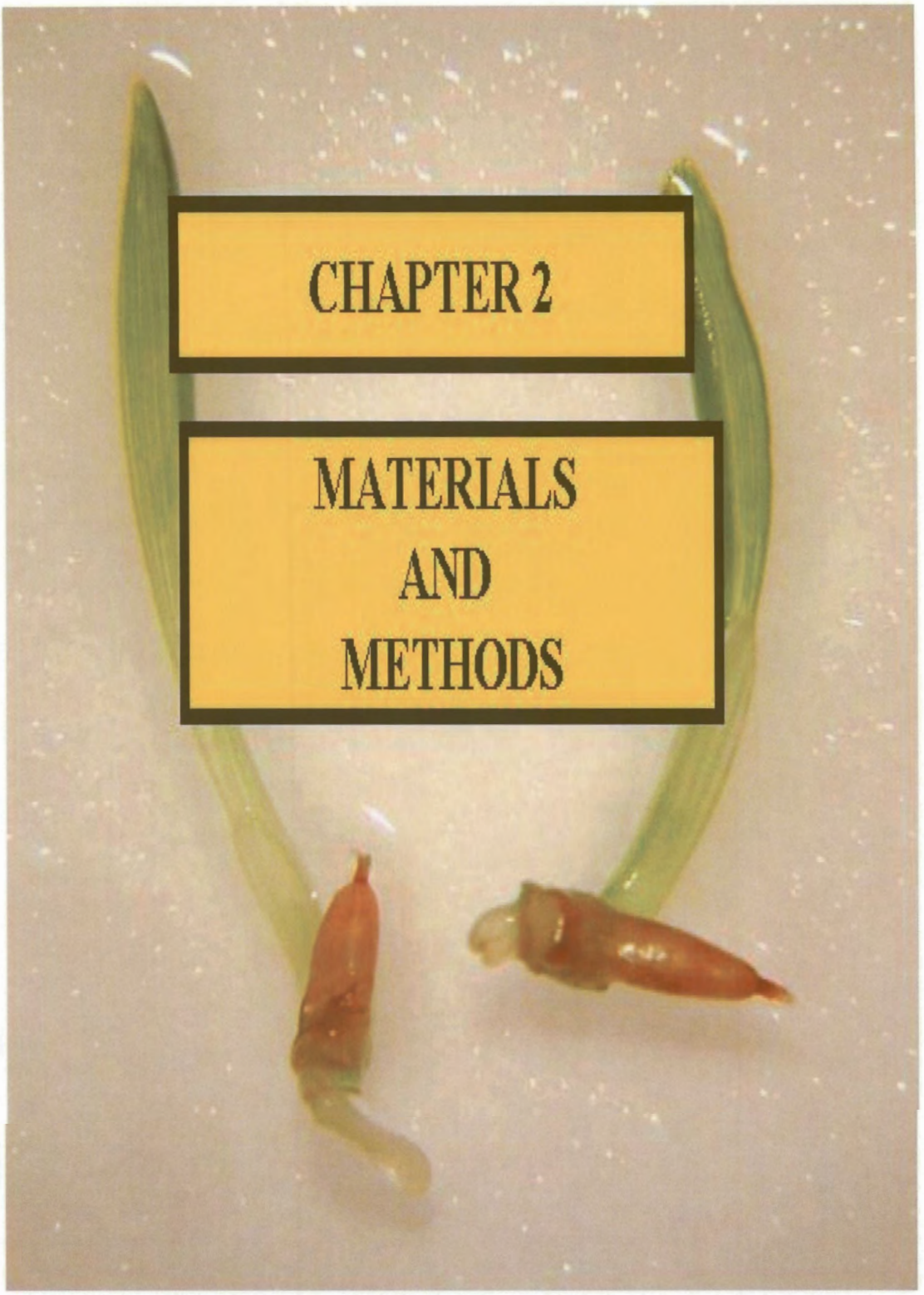


Figure 1.11: *Monocymbium ceresiiforme* distribution map in South Africa (National Botanical Institute).

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

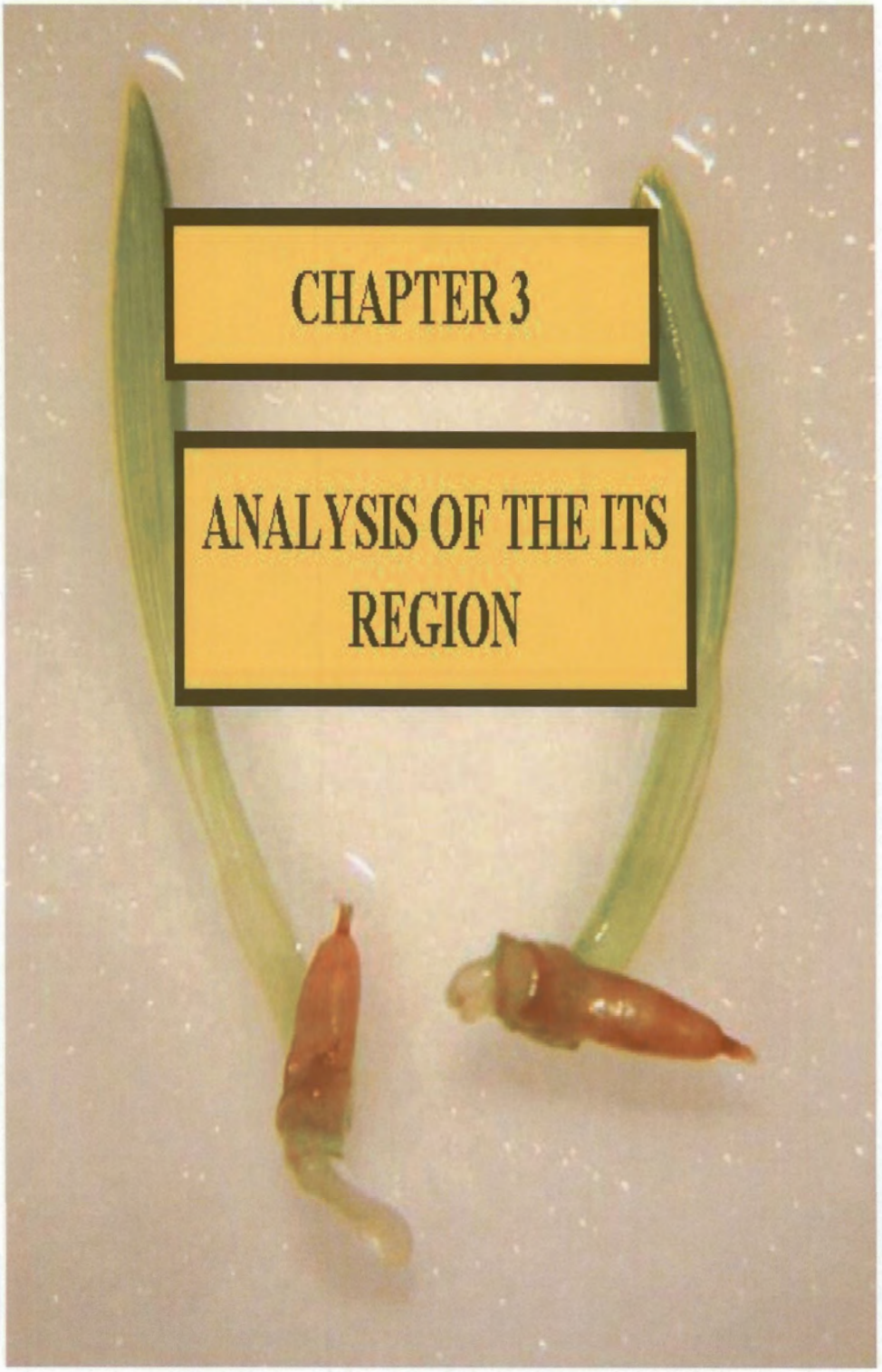
For the fourth round subtraction adaptors of the third round subtraction were changed to a new adaptor set (NHind 12 and NHind 24; Table A.8). The concentration of the ligated

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).

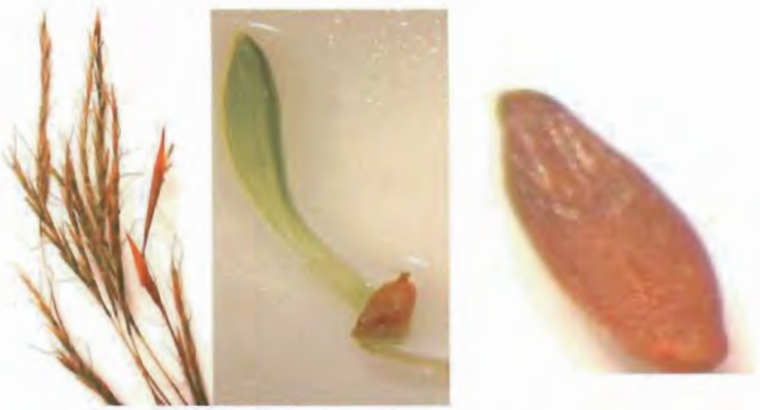


Figure 3.1: Different parts of the inland grass *Monocymbium ceresiiforme*. The vegetative parts shown are: the flower (A), the emerging leaf of the grass (B), and the grass seed (C).

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.

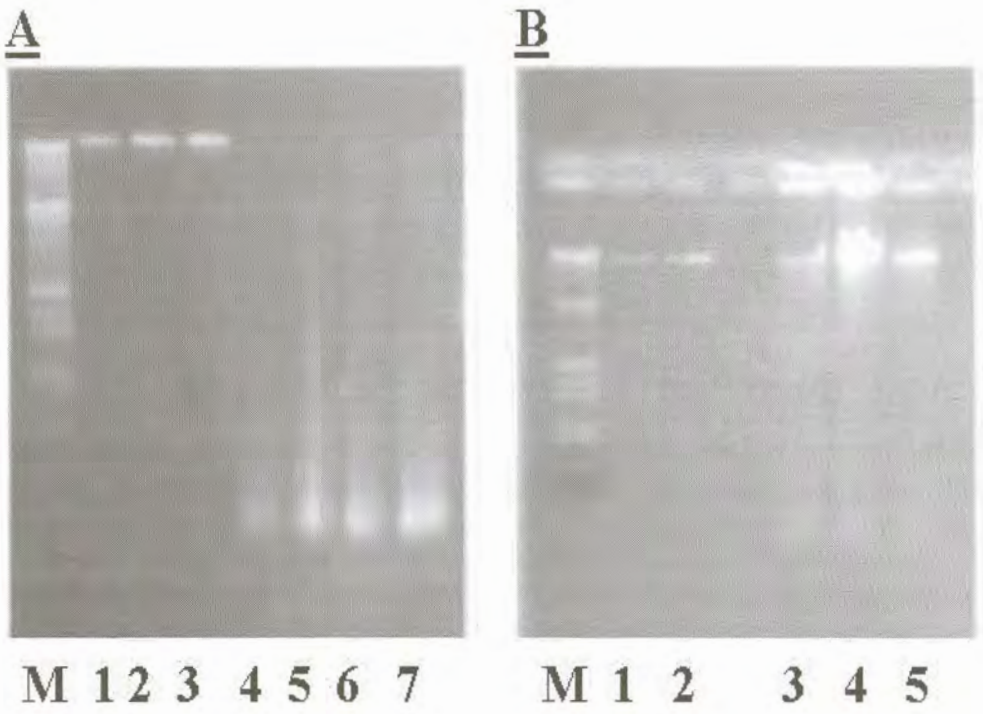


Figure 3.2: Isolated genomic DNA from grasses collected at three different locations, visualized on a 1% agarose gel stained with ethidium bromide. In (A) and (B) lane M represents molecular weight marker III (Roche, Switzerland), while lanes 1 and 2 represent λ DNA of 125 ng and 250 ng respectively. 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.

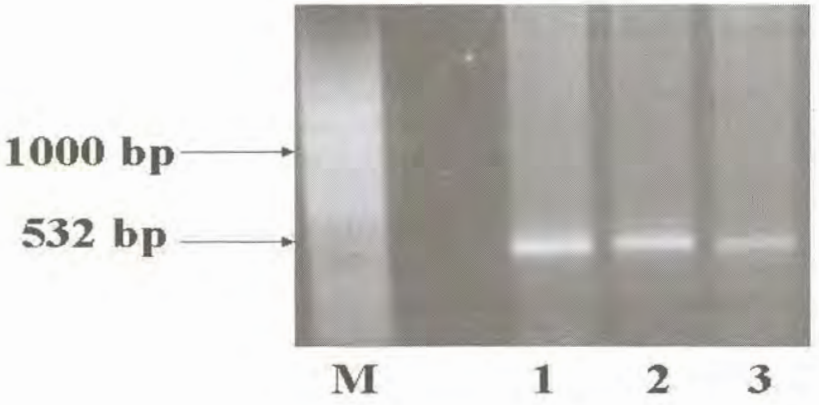


Figure 3.3: The ITS region amplified from genomic DNA of the grasses collected at three different areas visualized on an agarose gel stained with ethidium bromide. Lane M represents molecular weight marker III (Roche, Switzerland). Lane 1 represents the ITS amplified product from the SG grass genomic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

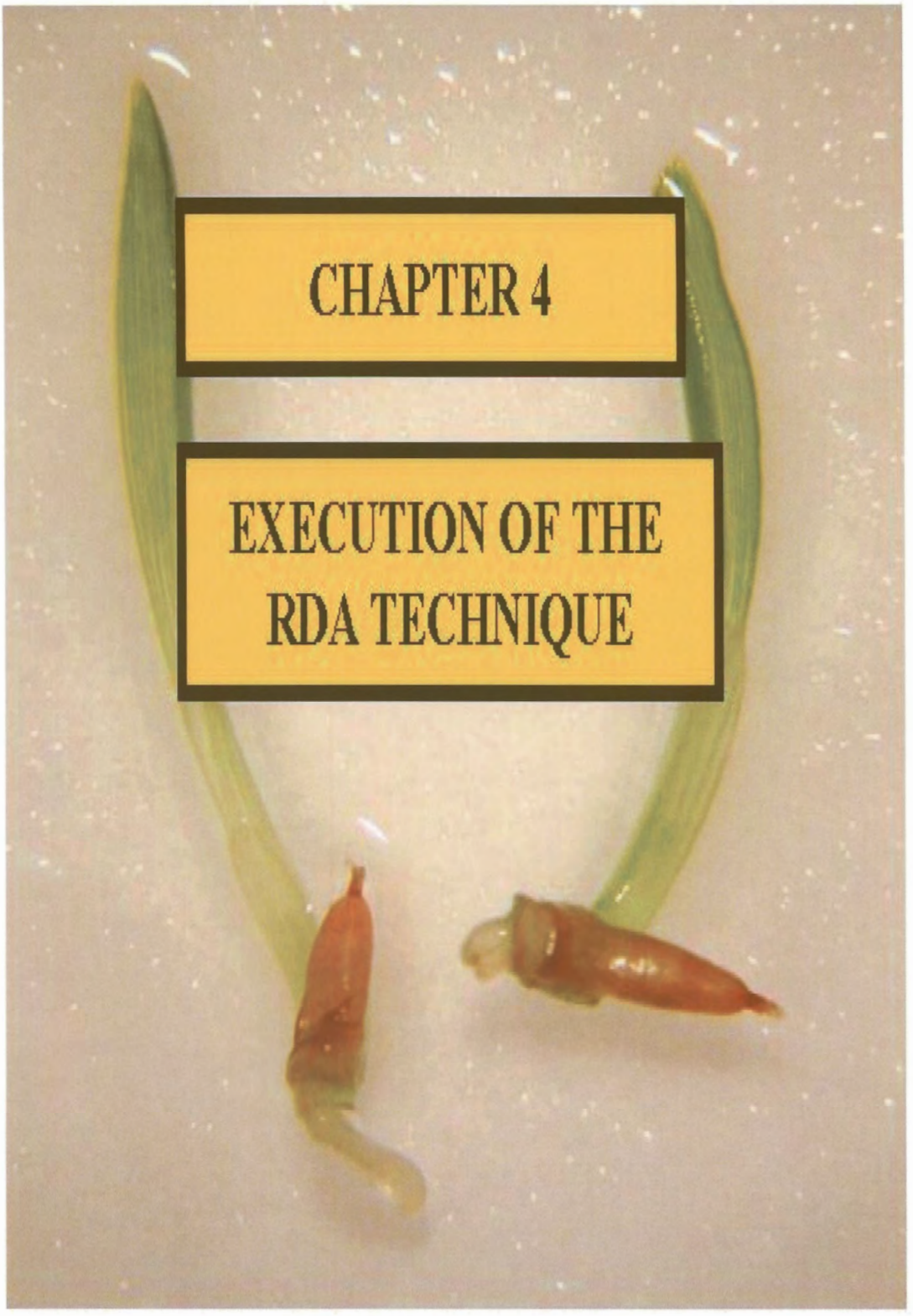
3.3.4 Sequencing analysis

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.


CHAPTER 4
EXECUTION OF THE RDA TECHNIQUE

4.1 Abstract

The RDA technique was applied to the inland grass species, *Monocymbium ceresiiforme*, collected at three different locations in South Africa (Savannah grassland/SG, Drakensberg grassland/DG and Highveld grassland/HG). Although growing under different climatic conditions, the grasses have no obvious, detectable morphological difference. The RDA was performed using reciprocally the three isolates as tester or driver. Five rounds of subtractive hybridization were used at different ratios of driver to tester on *Hind*III-digested genomic DNA isolated from the grass. Different subtraction products were obtained after execution of different rounds of subtraction and amplification. Different subtraction products were cloned into the vectors either *pMOSBlue or pGEM-T Easy* (depending of the round of subtraction) allowing sequence analysis. Uniqueness of products in the different types of grasses was determined by PCR analysis.

4.2 Objectives

The first objective of this part of the study was to produce a sub-population (representation) of DNA fragments derived from genomic DNA isolated from *Monocymbium* collected at three different locations. The second objective was to eliminate all similar DNA sequences by hybridization/amplification between two representations and then to isolate possible unique sequence present only in one of the representations.

4.3 Results

For the execution of the RDA technique amplicons were generated, which are the representations of the genomic DNA isolated from *Monocymbium* plants collected at different locations. For the generation of representations, the genomic DNAs were firstly digested with restriction enzyme *Hind*III (Figure 4.1A). Adaptor pair RHind 12 and RHind 24 (Table A 8) was then ligated to the ends of the digested DNA, with subsequent

amplification by PCR (Figure 4.1B). The amplified fragments represent the amplifiable portion of digests, also called representations. The amplified representations then serve as a starting material for successive rounds of subtraction and amplification.

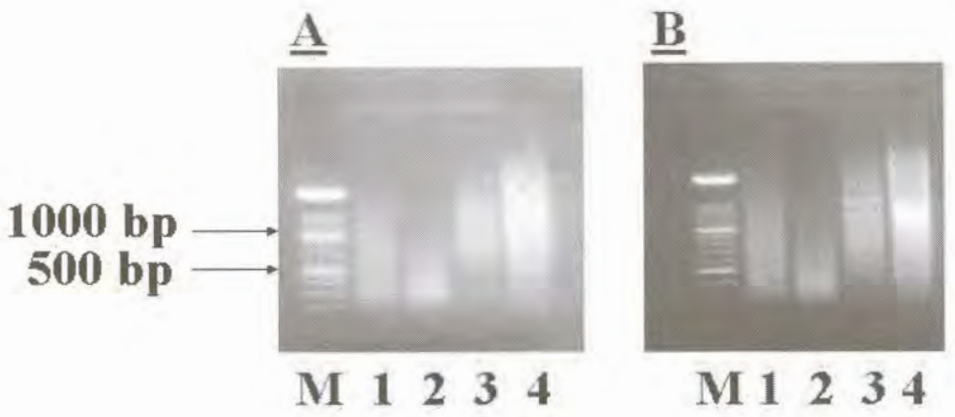


Figure 4.1: Digestion of genomic DNA isolated from *Monocymbium* collected from three different areas with restriction enzyme *Hin*dIII (A) and amplicons obtained from genomic DNA after adaptor ligation and amplification with adaptor pair RHind 12 and 24 (B), visualized on an agarose gel after staining with ethidium bromide. Lane M (A and B) represents a 100 bp ladder (Roche, Switzerland). Lane 1 represents (A) SG genomic DNA and (B) amplicon. Lane 2 represents (A) HG genomic DNA and (B) amplicon. Lane 3 represents (A) DG genomic DNA and (B) amplicon, and lane 4 (A and B) represents 2 µg herring sperm DNA for DNA quantification.

4.4 Production of subtraction products

4.4.1 First subtraction products

Two representations were hybridized reciprocally as tester or driver on the one hand for *Monocymbium ceresiiforme* grass species collected from Savannah and Drakensberg grasslands and on the other hand for grasses collected from Drakensberg and Highveld grasslands. In the first round, subtractive hybridization and amplification reactions were carried out with 10 µg of driver representation and 200 ng of tester representation for a 50:1 driver to tester ratio during which common material is selectively depleted and an amplified subtraction product is produced. The control material, known as driver, is used to 'drive out' fragments common to both samples. After 42 cycles, the resulting first subtraction products (SP1) is generated, which is substantially depleted for fragments that are common in the starting representations, and modestly enriched for differences. The resulting SP1 fragments were visualized on an agarose gel stained with ethidium bromide (Figure 4.2).

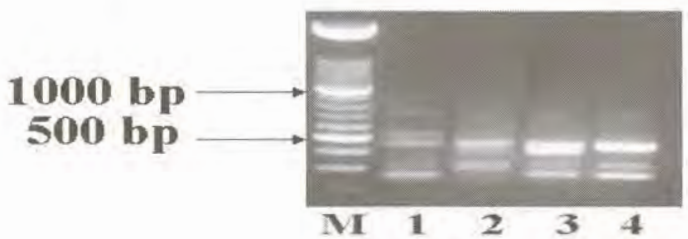


Figure 4.2: First subtraction products (SP1) of *Monocymbium ceresiiforme* genomic DNA representations visualized on an agarose gel stained with ethidium bromide. Lane M represents a 100 bp DNA ladder (Roche, Switzerland). Lanes 1 and 2 represent the subtraction products resulting from the subtraction of SG and DG grass representations, where SG grass representation was used as tester and DG grass representation as driver

(lane 1), and DG grass representation was used as tester and SG grass representation as driver (lane 2). Lanes 3 and 4 represent the resulting products from the subtraction of DG and HG grass representations, where the HG grass representation was used as tester and DG grass representation as driver (lane 3), and DG grass representation was used as tester and HG grass representation as driver (lane 4).

4.4.2 Second subtraction products

In the first round of subtractive hybridization and amplification a number of bands representing differences between the two representations were already observed on an agarose gel (Figure 4.2). But, because random annealing events might have occurred, many amplified tester/tester hybrids do not represent true differences at this stage. For this reason, a second round of subtraction was performed in a similar manner to the first using 10 µg of driver representation and 66 ng of tester representation for a 150:1 ratio driver to tester, generating a second set of subtraction products (SP2) (Figure 4.3). The second round of subtraction ensures that all DNA that is common between the two grasses is ultimately eliminated. The products from the first round of subtractive hybridization/amplification were first enriched in target sequences and then amplified in the second round, giving rise to two DNA fragments of approximately 250 bp and 450 bp (Figure 4.3).

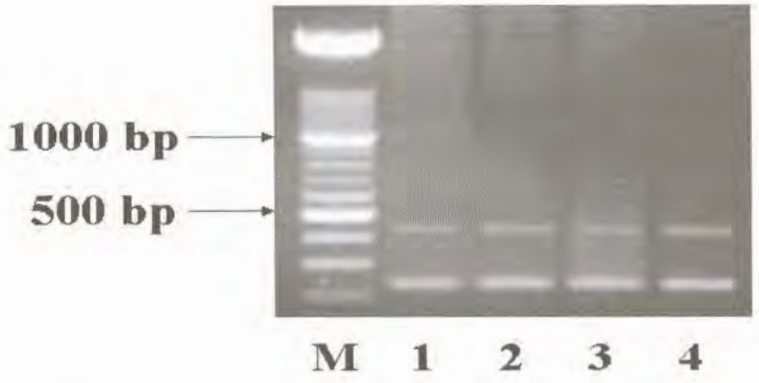


Figure 4.3: Second subtraction products (SP2) of *Monocymbium ceresiiforme* genomic DNA representations at a 150:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the resulting products from the subtraction of DG grass representation and SG grass representation, where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and DG grass representation was used as tester and the SG grass representation used as driver (lane 2). Lanes 3 and 4 represent the reciprocal subtraction of DG grass representation and HG grass representation, where HG grass representation was used as tester and DG grass representation was used as driver (lane 3), and the DG grass representation was used as tester and the HG grass representation was used as driver (lane 4). Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

4.4.3 Third subtraction products

Owing to the low complexity of the material compared with the genomic DNA, a third round of hybridization was required. During the third round of subtraction 10 µg of driver representation and 7 ng of tester representation were used for a 1500:1 driver to tester ratio which allowed the removal of the majority of unwanted material. After the third round of subtraction, a number of subtraction products were observed on an agarose gel (Figure 4.4). All three the representations (SG, DG, and HG) were successfully subtracted and several subtraction products could be amplified. Two clear bands per

subtraction were seen with similar sizes of approximately 250 bp for the smallest band and approximately 450 bp for the biggest band. No major differences in size between the products obtained in the second and third subtraction could be observed on an agarose gel.

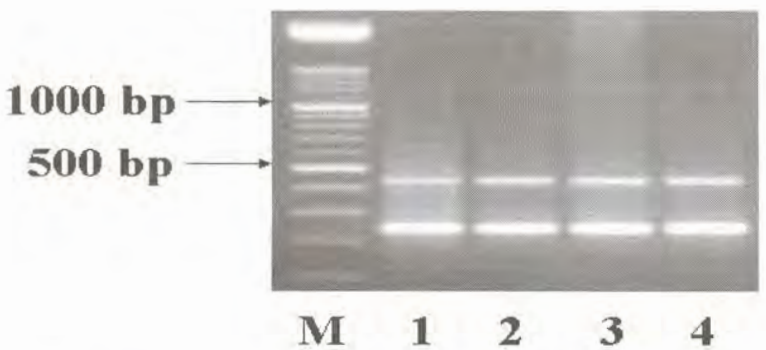


Figure 4.4: Third subtraction products (SP3) of *Monocymbium ceresiiforme* genomic DNA representations at a 1500:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the resulting products from the subtraction of DG grass representation and SG grass representation where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and DG grass representation was used as tester and the SG grass representation used as driver (lane 2). Lanes 3 and 4 represent the subtraction of the DG grass representation and the HG grass representation, where the HG grass representation was used as tester and the DG grass representation was used as driver (lane 3), and the DG grass representation was used as tester and the HG grass representation was used as driver (lane 4). Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

4.4.4 Fourth subtraction products

The subtraction products obtained after the third subtraction were shown a similar and highly repetitive DNA. To eliminate those repetitive DNA and ensure the isolation of more unique subtraction products, a fourth round of subtractive hybridization was applied. Before carrying out the fourth round of hybridization, the third subtraction

products were digested with the frequent cutter *Mse*I, which cuts preferably in repetitive DNA. The fourth round of subtraction was carried out with 10 µg of driver representation and 50 pg of tester for a 200000:1 driver to tester ratio. A number of subtraction products were obtained after this round of subtraction (Figure 4.5). The digestion of subtraction product with *Mse*I also made a clear difference in the profile of subtraction products after the fourth round of subtraction, indicating that the third subtraction products contained a significant portion of repetitive DNA.

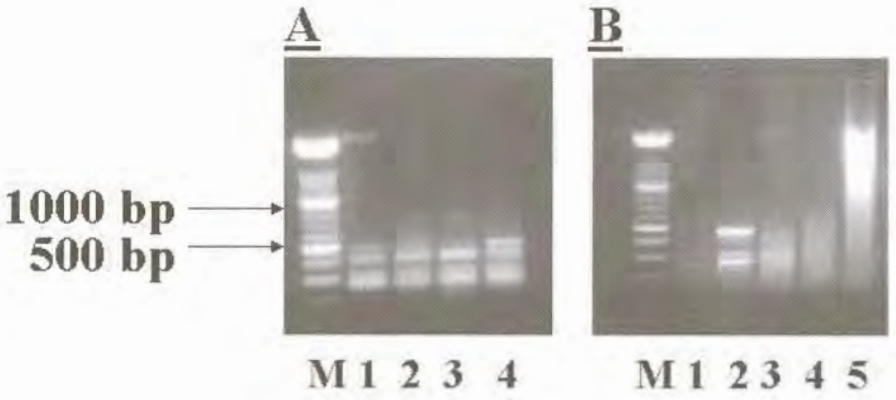


Figure 4.5 Fourth round subtraction products (SP4) of *Monocymbium ceresiiforme* genomic DNA representations at a 200000:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide, where the third subtraction products were either digested with *Mse*I (A), or were undigested (B) before amplification. In both (A) and (B), lanes 1 and 2 represent the subtraction of DG grass representation with the SG grass representation, where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and the SG grass representation was used as driver and the DG grass representation as tester (lane 2). In both (A) and (B) lanes 3 and 4 represent the subtraction of DG grass representation and HG grass representation where the HG grass representation was used as tester and the DG grass was used as driver (lane 3), and reciprocally in lane 4 the DG grass representation was used as tester and the HG grass representation was used as driver. Lane M represents a 100 bp DNA ladder (Roche, Switzerland). Lane 5 (B) represents 2 µg of herring sperm DNA for DNA quantification.

4.4.5 Fifth subtraction products

A fifth round of subtractive hybridization with an increased ratio of driver to tester of 2000000:1 was finally applied. Only fourth round *Mse*I-digested representations were used in the fifth round of subtraction. For the fifth round of subtraction 10 µg of driver representation and 5 pg of tester representation were used. A single band (subtraction product) was observed on an agarose gel in three out of four, after the fifth round of subtraction (Figure 4.6). No subtraction product could be detected after subtraction between the DG grass representation and the SG grass representation, where the DG grass representation was used as tester and the SG grass representation was used as driver. Subtraction products were obtained between the DG grass representation and SG grass representation when the DG grass representation was used as driver and the SG grass representation as tester, and between the DG grass representation and HG grass representation. The size of the different subtraction products was about 240 bp.

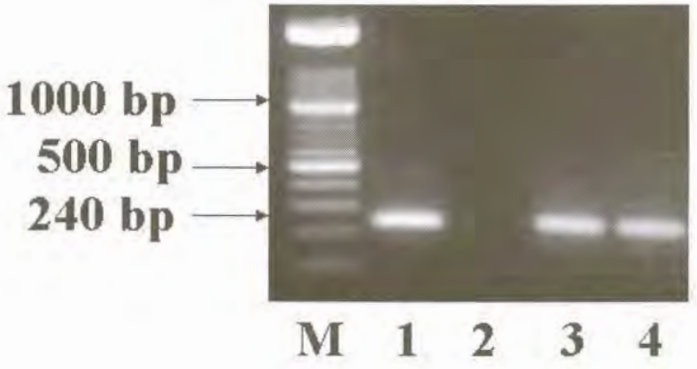
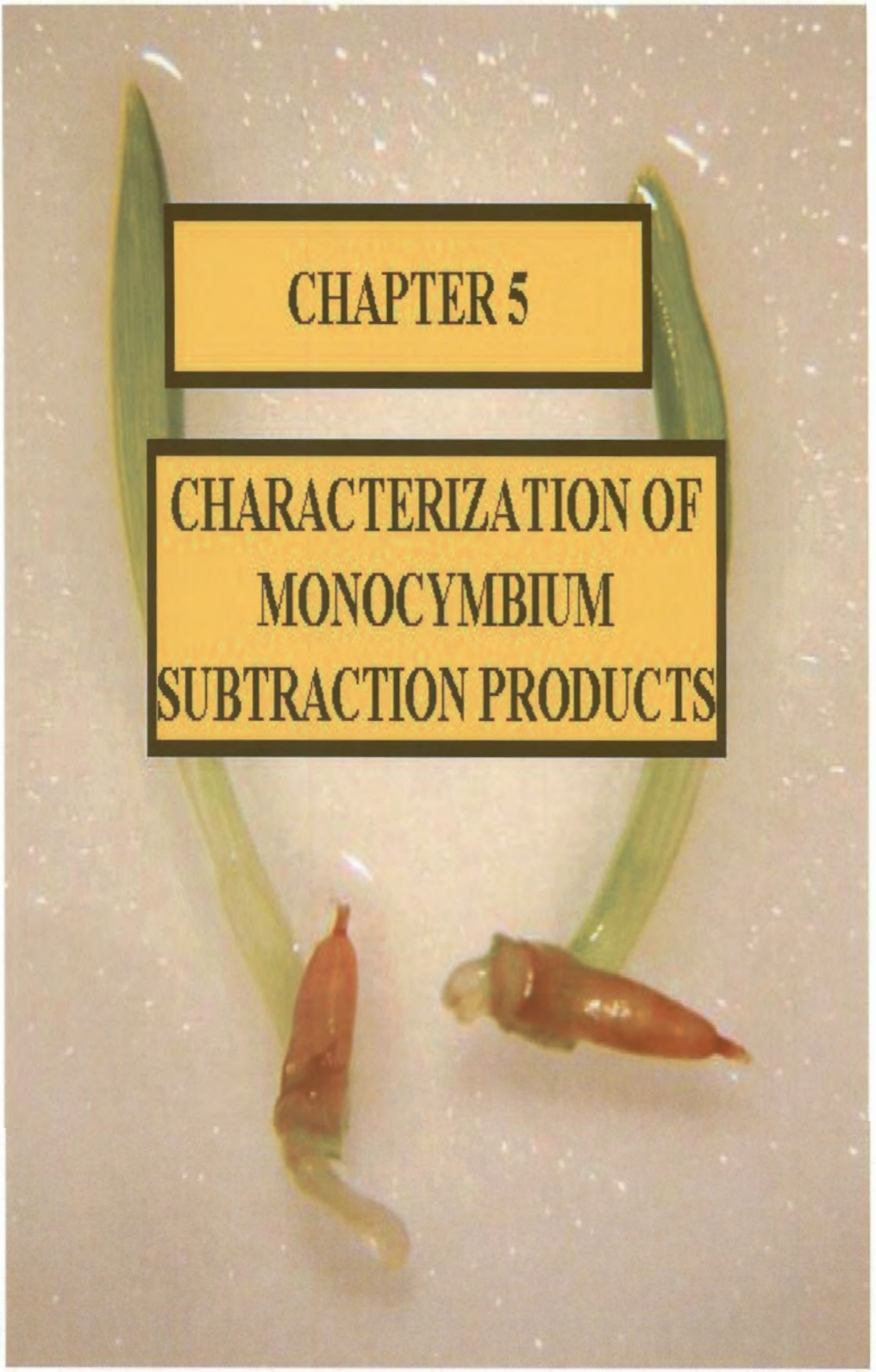


Figure 4.6 Fifth round subtraction products (SP5) of *Monocymbium ceresiiforme* genomic DNA representations at a 2000000:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the subtraction of the DG grass representation and the SG grass representation where the DG grass representation was used as driver and the SG grass representation as tester (lane 1), and

the DG grass representation was used as tester and the SG grass representation as driver (lane 2). Lanes 3 and 4 represent the subtraction of the DG grass representation and the HG grass representation, where the HG grass representation was used as tester and the DG grass representation as driver (lane 3), and the DG grass representation was used as tester and the HG grass representation as driver (lane 4). Lane M represents a 100 bp DNA ladder (Roche, Switzerland).



CHAPTER 5

CHARACTERIZATION OF MONOCYMBIUM SUBTRACTION PRODUCTS

5.1 Abstract

Subtraction products were characterized using bio-informatics tools. Five rounds of subtractive hybridization were performed. A subtraction product that was obtained after two rounds of subtractive hybridization showed homology to a known sequence of maize (LH82 transposon Ins2). This subtraction product was not unique to any of the grasses investigated and the product had further a high copy number in the plant genome. A fifth-round subtraction product DP510 obtained from the SG grass was partially homologous to a *Bacillus* genomic DNA sequence and also to genomic DNA sequences of rice and *Arabidopsis* but failed to hybridize with isolated grass genomic DNA.

5.2 Objective

The objective of this part of the study was to analyze with bio-informatics tools RDA-derived subtraction products from genomic DNAs of *Monocymbium ceresiiforme* collected from three different locations in South Africa (SG, DG, and HG).

5.3 Results

5.3.1 Analysis of second set of subtraction products

After cloning of subtraction products obtained after two rounds of subtractive hybridization into the plasmid p*MosBlue*, forty colonies were screened and five plasmids were found to have an insert, namely S3cl2, S3cl12 S3cl18, S4cl27, and S4cl39 (Figure 5.1). S3 indicates subtraction with the DG grass representation as driver, and the SG representation as tester, and S4 indicates the reciprocal hybridization. One insert (S3cl2) was further analyzed because the other inserts (S3cl12, S3cl18, S4cl27 and S4cl39) could not be successfully sequenced.

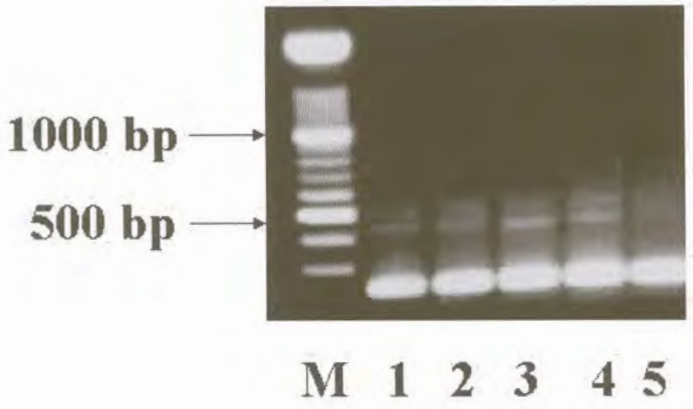


Figure 5.1: Screening of plasmid p*MosBlue* for insertion of subtraction products obtained after two rounds of subtractive hybridization with *Monocymbium ceresiiforme* representations. Lanes 1-5 represent respectively fragments S3cl2, S3cl12, S3cl18, S4cl27, and S4cl39. Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

The 320-base pairs subtraction product obtained after two rounds of subtractive hybridization (S3cl2) was sequenced and used in a homology search using the BLAST online database (Altschul *et al.*, 1990) (Figure 5.2).

```
       10          20          30          40          50          60
AGGCAGCTC GGTATCGAGG GAGAAGCCCN GTGAGATTGA GTCCTACTCA AAGCCAAACT
       70          80          90         100         110         120
AGCACATGGC TGTCAGAAGC CTTCCCGCCG GCCCGGAGCA CATACGNCAT ACTCCCTGGT
      130         140         150         160         170         180
CTCTTCATGA GTCCACTAGA GATCACCCAA ATCTCATAGA TTGCGACGCT CAACAATCTA
      190         200         210         220         230         240
AGCTCATATA GGTATGCTCT TTCAAGAATG CTCTGCAGGA CAGCATCTTC GCTNGTGAAA
      250         260         270         280         290         300
GCCAATANAG CATATTAAGG CATTCTAGCC AACCTACCAT ANAGNGCCTC ACCACTCCNG
      310         320
AGAGNANNGC ANCCNNACA
```

Figure 5.2: Nucleotide sequence of the subtraction product S3cl2 indicating the sequences used as primers for amplification of SG, DG and HG grass genomic DNAs (underlined.

Sequence of the subtraction product had homology to a sequence of maize (*Zea mays* line LH82 transposon Ins2) (Accession number: gi|20502801|gb|AF434192.1|; with an E-value of 8e-25 as well as various *Oryza sativa* (japonica cultivar-group) genomic DNA sequences.

```
A: 94    ggagcacatacgncatactccctggtctcttcatgagtccactagagatcacccaaatct 153
         |||||| ||||| | ||||||  ||||||||||||||  ||||||||||||||||||||
B: 10485 ggagcatatacgtcgtactccaaggtctcttcatgagtttactagagatcacccaaatct 10426

A: 154   catagattgcgacgctcaacaatctaagctcatataggtatgctctttc-aagaatgctc 212
         |||||| ||||||| |  ||| ||||| |||||||||||| || |||||| ||||||| ||
B: 10425 catagactgcgacgttagacagtctaa-ctcatataggtgtgttctttcaaagaatgttc 10367

A: 213   tgcaggacagcatcttcgct 232
         |||||||||  ||||||||||
B: 10366 tgcaggacaacatcttcgct 10347
```

Figure 5.3: Subtraction product S3cl2 sequence (A) and *Zea mays* line LH82 transposon Ins2 sequence (B).

5.3.2 Genomic DNA amplification with primers designed from subtraction product S3cl2

Primers were designed from the obtained sequence of the subtraction product S3cl2, as outlined in Materials and Methods. These primers, S3cl2L and S3cl2R, amplified fragments out of genomic DNA isolated from *Monocymbium ceresiiforme* from all three areas (DG, SG, and HG), generating a strong band of approximately 165 bp, and a weak band of approximately 200 bp (Figure 5.4). These sizes are smaller than the expected 320 bp of the original subtraction product S3cl2. A Less stringent annealing temperature of 55°C was used, because no amplification could be obtained at a higher more stringent annealing temperature of 60°C or 65°C.

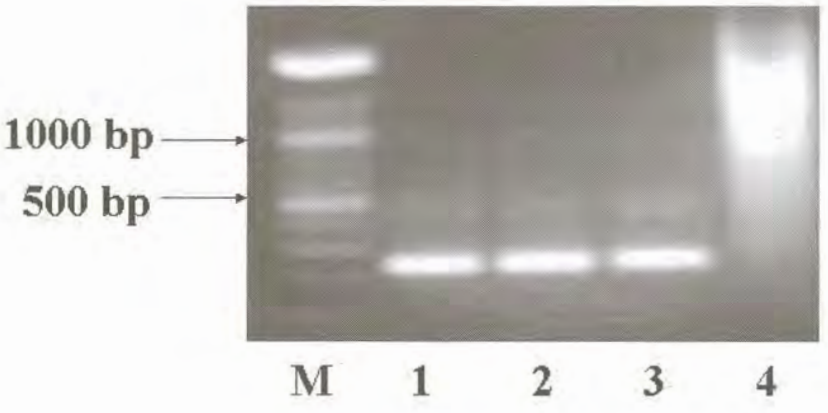


Figure 5.4: PCR amplification of grass genomic DNA at 55°C with primers S3cl2L and S3cl2R visualized on an agarose gel stained with ethidium bromide. Lanes 1, 2, and 3 represent the amplification products from SG, DG, and HG grass genomic DNA, respectively. Lane 4 represents 2 µg of herring sperm DNA for DNA quantification. Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

Sequence alignment of the amplified products H36, D22, D25, S40 and S45 derived from grasses collected at three different locations (HG-H36); [DG- (D22, D25)], and (SG-S40, S45), obtained by amplification using the primers designed from the S3cl2 sequence, showed a high level of homology and consensus between the sequences with several

conserved and variable regions between the amplification products (Figure 5.5). However, variability was not only found for sequences of different types of grass but also for an identical type of grass (Figure 5.5).

```
                10         20         30         40         50         60
S3CL2  GATTGAGTCC TACTCAAAGC CAAACTAGCA CATGGCTGTC AGAAGCGATC ACCCA-ACTT
S40    -ANTG---CC TTAGCCAA-C TAGTCTTGG- CGTG-CTCNN NTTCCTTTNA AN-G--ACTT
D22    CAATG---C- TTAAGCAA-C TAGTCTTGG- CGTGG-TNNN NNNTTTCNN- A-----ACTT
D25    --GGGA-TGC NTCCCNN-TC TNNANGAG-A NA-GAGANG- -G-AGCGT-- AT ---ACTC
S45    C---GAACTC ATCGCCA-TC TTNAAAAACA TA-G--TTCN NNNC-C-T-- ------ACTT
H36    C---GAACTC ATCGCCA-TC TTTACAAAAA CATAGATTNN NNNTTC---- ------ACTN
Cons   CA•TGA••CC TTCGCCAATC TA•ACTAG•A CATGG•T•NN NNN••C•T•• A•••••ACTT
                70         80         90        100        110        120
S3CL2  CCCGCCGGCC CGGAGCACAT ACGC-CATAC TC-CCTGGTC TCTTCATGAG TCCACTAG-A
S40    GGGGT----- -GGAGCATAT TCNCNCATAC TC-CTTANTC TCTTCATGAG CTTACTAG-A
D22    CCNCTCAANG GGGAGCATAT TCTCTCCTAC TCTTAGAAAT CATTCATGAG CTTACTAG-A
D25    CNATAGAACA CTTCATGAGC TCCCTTATAC TCTTAGAGAT CACCCATAAT CTTACTAG-A
S45    CGGGTGAAAA TCCTNTGGAG AGGCC-ATAC TNTCCGGTAT AAGCCCAGAC ACTACTAATA
H36    GCGGTCAAAA TTCTGCGAAG -CGCNTAT-C TCTNNNNTAC AAGCACCAAG ACTACTAGTA
Cons   CCGGTCAA•A •GGAGCA•AT TCGC•CATAC TC•C•GA•AC •ATTCATGAG CTTACTAG-A
               130        140        150        160        170        180
S3CL2  ---------- ATCTCATAGA TTGCGACGCT CAACAATCTA A-GCTCATAT ATAGGT-ATGC TC
S40    GATCACCCAA ATCTCATAGA CTGCGACGTT AAACAGTCTA AC-CTCATAT GATGNC-ATGC CC
D22    GATCACCCAA ATCTCATAGA TTGCGACGTT AAACAGTCTA AC-CTCATAT GCTG-T-ATGC CC
D25    CATCATAGTA ATCTCATAGA CTGCGACGTT AAAGCAGCTA AT-CTGCAAT GATG-T-ATGC CC
S45    CACCATCGCA TTCCTGTTGA CTCCTTCGAG CAACGACCTA.ACGCTCACAA GAAC-TCA-GC GAAA
H36    CACCATCGCA TTCCTGTTGA -TCCTTCGAG CAACGACCAT A-GCTCATAA GACC-TCA-GC GAAA
Cons   CATCATCG•A ATCTCATAGA CTGCGACGTT CAACAATCTA ACGCTCATAT GATG•T-ATGC CC••
```

Figure 5.5: Sequence alignment of the isolated subtraction product S3CL2 and the amplification products clones D25, D22, S40, S45, and H36 showing a consensus sequence with 50% similarity in black, 75% similarity in blue and 100% similarity in red. All black dots (•) showed the regions with no similarities. Empty spaces indicate base pairs deletions (-). Cons = consensus.

5.3.3 Analysis of the fifth subtraction products

In order to eliminate detection of possible highly repetitive elements (retro-transposon) found as a subtraction product after two rounds of subtractive hybridization, further subtractions were done. The third subtraction products were also digested with the frequent-cutter *Mse*I to eliminate most of the repetitive DNA. A fourth round of subtractive hybridization was done at a driver to tester ratio of 200 000:1. No analysis was done for the fourth subtraction products. A fifth round of subtractive hybridization was done at a driver to tester ratio of 2 000 000:1. From the fifth cloned subtraction products, subtraction product-insertions of five individual clones were sequenced. After analysis of the five isolated sequences (DP56, DP57, DP58, DP59 and DP510; Figure 5.6), all of them were found to be almost identical and insertions had similar size of about 240 bp. All sequences showed significant homology to *Bacillus* DNA including the extremophile bacteria species *Bacillus halodurans* genomic DNA (73%) (Accession number AP004602), *Bacillus iheyensis* DNA (74%) (Accession number AP001511), and to DNA of the soil bacterium *Bacillus subtilis* complete genome (65%) (accession number BSUB0006, Table 5.1). All sequences further showed a partial homology in short sequence regions to rice DNA (E-value = 0.27 and accession number AP003825) and to *Arabidopsis* genomic DNA (E-value = 0.11 and accession number AB010068) including an almost perfect identity to a section of *Arabidopsis* DNA encoding a putative auxin-induced protein (95%) (accession number AF361098, Figure 5.7). With the E-Value-14 deemed as a lower significant homology, these homologies were found significant (Table 5.1).

```
                10         20         30         40         50
DP56        ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACAA
DP57        ACCGACGTCG ACTATCCGTG AACAACGAGT AGATGGCCGC GCTTGAACAA
DP58        ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACGG
DP59        ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACAA
DP510       ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACAA

                60         70         80         90        100
DP56        AGTATTGAAG AGACGACACA ATTGGCCAGG G--ATTTCGG GCTCGATTTT
DP57        AGTATTGAAG AGACGACACG ATTGGCCAGG G--ATTTCGG GCTCGATTTT
DP58        AGTATTGAAG AGACGACACG ATTGGCCAGG G--ATTTCGG GCTCGATTTT
DP59        AGTATTGAAG AGACGACACA ATTGGCCAGG G--ATTTCGG GCTCGATTTT
DP510       AGTATTGAAG AGACGACACA ATTGGCCAGG GGAATTTCGG GCTCGATTTT

               110        120        130        140        150
DP56        TATCCGATGC GTTATGAGGT TTGTCCGTCG GATGCCATTT ACACGTTTGG
DP57        TATCCGATGC GTTATGAGGT TTGTCCGTCG GATGTCATTT ACACGTTTGG
DP58        TATCCGATGC GTTATGAGGT TTGTCCGTCG GATGTCATTT ACACGTTTGG
DP59        TATCCGATGC GTTATGAGGT TTGTCCGTCG GATGTCATTT ACACGTTTGG
DP510       TATCCGATGC GTTATGAGGT TTGTCCGTCG GAT-TCATTT ACACGTTTGG

               160        170        180        190        200
DP56        GGCCTACGGG ATGCCGACGC GCTTTTCACA TTGGAGTTTT GGGAAATCAT
DP57        GGCCTACGGG ATGCCGACGC GCTTTTCACA TTGGAGTTTT GGGAAATCAT
DP58        GGCCTACGGG ATGCCGACGC GCTTTTCACA TTGGAGTTTT GGGAAATCAT
DP59        GGCCTACGGG ATGCCGACGC GCTTTTCACA TTGGAGTTTT GGGAAATCAT
DP510       GGCCTACGGG ATGCCGACGC GCTTTTCACA TTGGAGTTTT GGGAAATCAT

               210        220        230        240
DP56        TCC-TAGGAT GAAGCTTGTT CATGGATAGG CGACGTCGGT A
DP57        TCC-TAGGAT GAAGCTTGTT CATGGATAGT CGACGTCGGT A
DP58        TCC-TAGGAT GAAGCTTGTT CATGGATAGT CGACGTCGGT A
DP59        TCC-TAGGAT GAAGCTTGTT CATGGATAGT CGACGTCGGT A
DP510       TCCATAGGAT GAAGCTTGTT CATGGATAGT CGACGTCGGT A
```

Figure 5.6: Sequence alignment of the five sequences obtained as subtraction products after five rounds of subtractive hybridization by using representation DNA from SG grass as tester and representation DNA from DG grass as driver. Nucleotide changes are indicated in bold, and nucleotide deletions with a dash (-).

A

```
        10          20          30          40          50          60
ACGAAGAGAT GGCCGCGCTT GAACAAAGTA TTGAAGAGAC GACACAATTG GCCAGGGGAA

        70          80          90         100         110         120
TTTCGGGCTC GATTTTTATC CGATGCGTTA TGAGGTTTGT CCGTCGGATT CATTTACACG

       130         140         150         160         170         180
TTTGGGGCCT ACGGGATGCC GACGCGCTTT TCACATTGGA GTTTTGGGAA ATCATTCCAT

       190
AGGATGAAGC T
```

B

```
        10          20          30          40          50          60
ACGAAGAGAT GGCCGCGCTT GAACAAAGTA TTGAAGAGAC GACACAATTG GCCAGGGGAA

        70          80          90         100         110         120
TTTCGGGCTC GATTTTTATC CGATGCGTTA TGAGGTTTGT CCGTCGGATT CATTTACACG

       130         140         150         160         170         180
TTTGGGGCCT ACGGGATGCC GACGCGCTTT TCACATTGGA GTTTTGGGAA ATCATTCCAT

       190
AGGATGAAGC T
```

C

```
        10          20          30          40          50          60
ACGAAGAGAT GGCCGCGCTT GAACAAAGTA TTGAAGAGAC GACACAATTG GCCAGGGGAA

        70          80          90         100         110         120
TTTCGGGCTC GATTTTTATC CGATGCGTTA TGAGGTTTGT CCGTCGGATT CATTTACACG

       130         140         150         160         170         180
TTTGGGGCCT ACGGGATGCC GACGCGCTTT TCACATTGGA GTTTTGGGAA ATCATTCCAT

       190
AGGATGAAGC T
```

D

```
        10          20          30          40          50          60
ACGAAGAGAT GGCCGCGCTT GAACAAAGTA TTGAAGAGAC GACACAATTG GCCAGGGGAA

        70          80          90         100         110         120
TTTCGGGCTC GATTTTTATC CGATGCGTTA TGAGGTTTGT CCGTCGGATT CATTTACACG

       130         140         150         160         170         180
TTTGGGGCCT ACGGGATGCC GACGCGCTTT TCACATTGGA GTTTTGGGAA ATCATTCCAT

       190
AGGATGAAGC T
```

Figure 5.7: Sequence analysis of DP510 (A, B, C, and D). Identified hyper-variable DNA region indicated in green. Perfect homology to *Arabidopsis* sequence indicated in pink. Partial homology to *Arabidopsis* genomic DNA indicated in red. Significant

homology to *Bacillus* DNA indicated in blue. E-value-14 was deemed as a lower significant homology.

Table 5.1: Sequence analysis of DP 510

Homology	Overlap (bp)	E-value	% Identity
Bacillus halodurans genomic DNA	179 bp	7.2e-20	73%
Bacillus iheyensis genome	162 bp	1.3e-19	75%
Bacillus subtilis complete genome	194 bp	2.3e-14	65%

To determine whether DP510 is a unique sequence in the genome of the SG grass, DP510 was hybridized to subtraction products obtained after first to fifth subtractions (Figure 5.8). DP510 hybridized weakly to the first subtraction products when HG grass representations were used as tester as well as no hybridization whereever DG grass representations were used as tester. Strong hybridization was obtained to all the other subtraction products, except when the DG grass representation was used as tester after the fifth subtraction (Figure 5.8E). This demonstrates the kinetic enrichment of the subtraction product after individual subtraction steps followed by PCR amplification. Also the DP510 sequence was not unique to the genome of one of the grasses investigated in the study (Figure 5.8). Although the product hybridized to the different subtraction products obtained by RDA, the product failed, despite several attempts, to hybridize by Southern blotting to *Eco*RI-digested genomic DNA isolated from the different types of grasses. When a slot blot technique was used with total genomic DNA, only a very weak hybridization by Southern blotting with DP510 was found (Figure 5.9). So far, the ultimate proof of plant origin of DP510 by Southern blot analysis could not be made. The *Bacillus* genomic DNA was not used for hybridization by Southern blotting analysis with DP510.

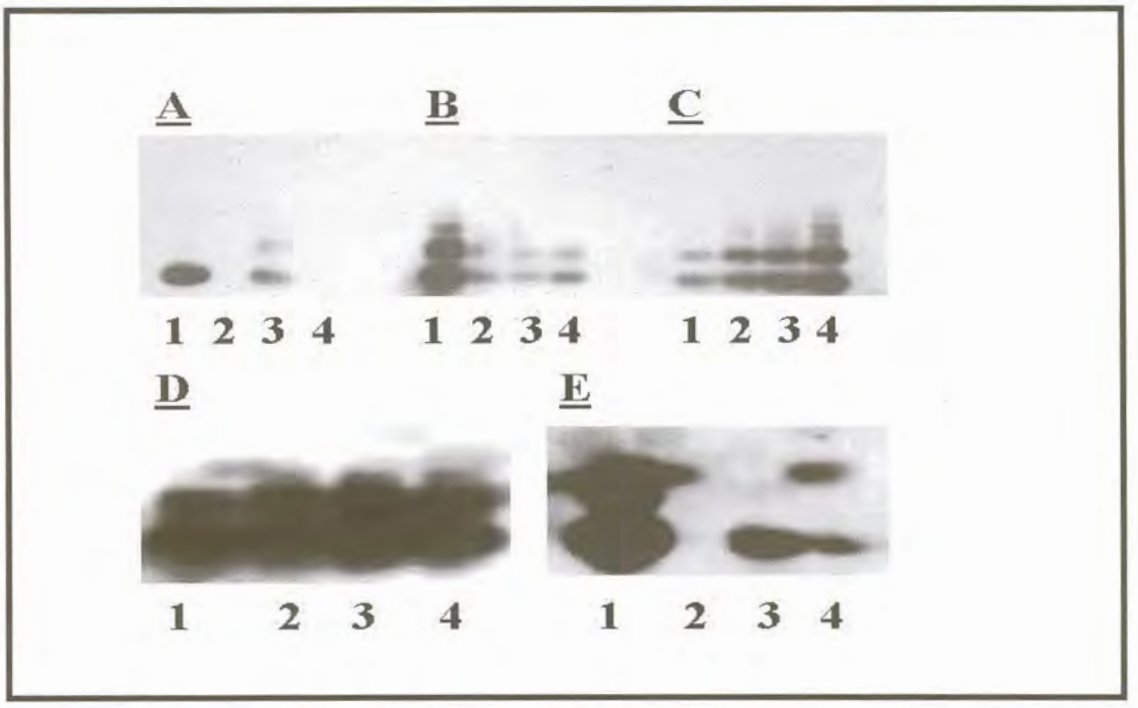


Figure 5.8: Hybridization of first subtraction products (A); second subtraction products (B); third subtraction products (C); fourth subtraction products (D) and fifth subtraction products (E) with labeled DP510 fragment. Lanes 1 and 2 represent the subtraction of SG and DG grass representations, where the SG grass representation was either used as tester (lane 1) or the DG grass representation was used as tester (lane 2). Lanes 3 and 4 represent the subtraction of HG and the DG grass representations, where the HG grass representation was either used as tester (lane 3) or the DG grass representation was used as tester (lane 4).

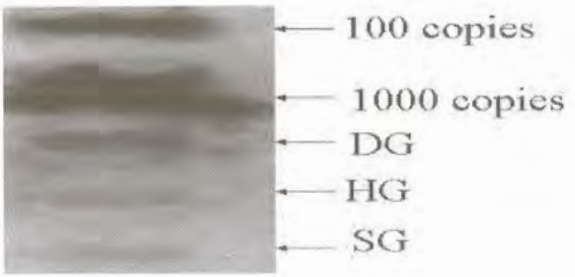


Figure 5.9: Copy number determination in Southern-blotting of EcoRI-digested genomic DNA from DG, HG, and SG grass with labeled DP510. DG, HG, and SG represent genomic DNA from the three different types of grass. Reference bands indicate the intensity of 100 and 1000 copies of DP510 (top two rows).

5.3.4 Genomic DNA amplification with primers designed from DP510

Primers Bhal5A and Bhal3A were designed from the sequence of DP510 (Figure 5.10). These primers were used in a PCR reaction using genomic DNA from the SG grass to amplify DP510 fragment and to identify a possible sequence variation in DP510. For that, amplified products were cloned into the vector p*GEM-T* Easy and transformed colonies were selected by colony hybridization for DP510 inserts with labelled DP510 (Figure 5.11). Nine bacterial colonies that gave the lowest hybridization signal indicating a lower homology to DP510 (probably due to less DNA bound to the membrane) were used for sequencing. But, only 3 sequences gave a satisfactory result in sequence analysis (Bh31, Bh48, and Bh79). All three cloned sequences Bh31, Bh48, and Bh79 had an identical size, and showed significant alignment with all the DP products including DP510 with differences being some base pairs substitutions and deletions (Figure 5.12).

```
ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACAA AGTATTGAAG

        70          80          90         100         110         120
AGACGACACA ATTGGCCAGG GGAATTTCGG GCTCGATTTT TATCCGATGC GTTATGAGGT

       130         140         150         160         170         180
TTGTCCGTCG GATTCATTTA CACGTTTGGG GCCTACGGGA TGCCGACGCG CTTTTCACAT

       190         200         210         220         230
TGGAGTTTTG GGA AATCATT CCATAGGATG AAGCTTGTTC ATGGATAGTC GACGTCGGT
```

Figure 5.10: Sequence of DP510, indicating the sequences used as primers for amplification of SG grass genomic DNA (underlined)

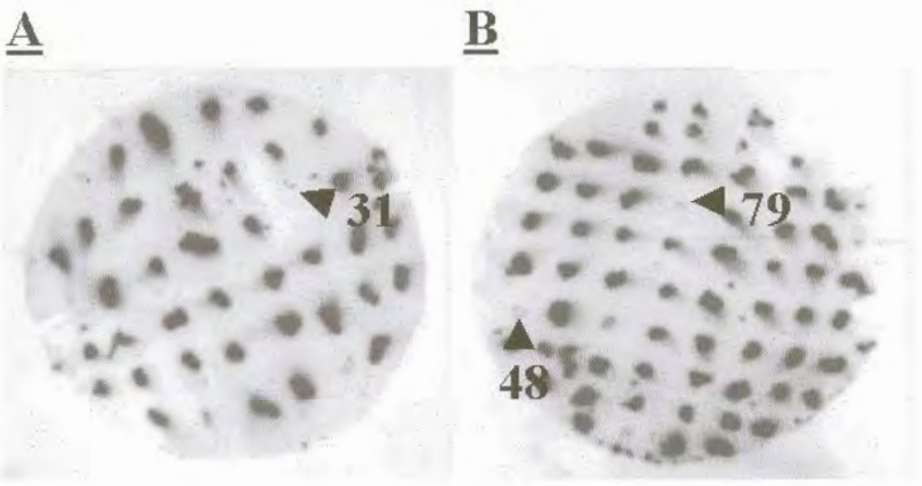


Figure 5.11: Colony blot of *E. coli* colonies carrying cloned DP510 amplification products from genomic DNA of SG grass and using labeled DP510 as a probe for hybridization. Colonies Bh31, Bh48, and Bh79 gave the lowest hybridization signal, which were used for further characterization.

```
                10         20         30         40         50
DP56      ACGAAGAGAT GGCCGCGCTT GAACAA-AGT ATTGAAGAGA CGACACAATT
DP57      ACGAGTAGAT GGCCGCGCTT GAACAA-AGT ATTGAAGAGA CGACACGATT
DP58      ACGAAGAGAT GGCCGCGCTT GAACGG-AGT ATTGAAGAGA CGACACGATT
DP59      ACGAAGAGAT GGCCGCGCTT GAACAA-AGT ATTGAAGAGA CGACACAATT
Bh31      ACGAAGAGAT GGCCGCGCTT -GAACAAAGT ATTGAAGAGA CGACACAATT
Bh48      ACGAAGAGAT GGCCGCGCTT TGAACAAAGT ATTGAAGAGA CGACACAATT
Bh79      ACGAAGAGAT GGCCGCGCTT -GAACAAAGT ATTGAAGAGA CGACACAATT
DP510     ACGAAGAGAT GGCCGCGCTT -GAACAAAGT ATTGAAGAGA CGACACAATT
                60         70         80         90        100
DP56      GGCCAGGG-A -TTTCGGGCT CGATTT-TAT CCGATGCGTT ATGAGGTTTG
DP57      GGCCAGGG-A -TTTCGGGCT CGATTT-TAT CCGATGCGTT ATGAGGTTTG
DP58      GGCCAGGG-A -TTTCGGGCT CGATTT-TAT CCGATGCGTT ATGAGGTTTG
DP59      GGCCAGGG-A -TTTCGGGCT CGATTT-TAT CCGATGCGTT ATGAGGTTTG
Bh31      GGCCAGGG-A -TTTCGGGCT CGATTTTTAT CCGATGCGTT ATGAGGTTTG
Bh48      GGCCAGGC-A -TTTCGGGCT CGATTTTTAT CCGATGCGTT ATGAGGTTTG
Bh79      GGCCAGGG-A -TTTCGGGCT CGATTTTTAT CCGATGCGTT ATGAAGTTTG
DP510     GGCCAGGGGA ATTTCGGGCT CGATTTTTAT CCGATGCGTT ATGAGGTTTG
                110        120        130        140        150
DP56      TCCGTCGGAT GCCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
DP57      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
DP58      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
DP59      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
Bh31      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
Bh48      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
Bh79      TCCGTCGGAT GTCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
DP510     TCCGTCGGAT -TCATTTACA CGTTTGGGGC CTACGGGATG CCGACGCGCT
                160        170        180        190
DP56      TTTCACATTG GAGTTTTGGG AAATCATTCC -TAGGATGAA GC
DP57      TTTCACATTG GAGTTTTGGG AAATCATTCC -TAGGATGAA GC
DP58      TTTCACATTG GAGTTTTGGG AAATCATTCC -TAGGATGAA GC
DP59      TTTCACATTG GAGTTTTGGG AAATCATTCC -TAGGATGAA GC
Bh31      TTTCACATTG GAGTTTTGGG AAATCATTCC ATAGGATGAA GC
Bh48      TTTCACATTG GAGTTTTGGG AAATCATTCC ATAGGATGAA GC
Bh79      TTTCACATTG GAGTTTTGGG AAATCATTCC ATAGGATGAA GC
DP510     TTTCACATTG GAGTTTTGGG AAATCATTCC ATAGGATGAA GC
```

Figure 5.12: Sequence alignment of difference products obtained after fifth round of subtractive hybridisation (DP56, DP57, DP58, DP59, and DP510), with amplified and selected DNA fragments from SG grass genomic DNA using primers Bhal5A and

Bhal3A designed from DP510 (Bh31, Bh48, and Bh79). Detected nucleotide differences are indicated in bold.

5.3.5 Analysis of fragments amplified with internal DP510 primers

Primers Bhal5R and Bhal3L designed for amplifying an internal part of DP510 (Figure 5.13) were used in a further step to amplify fragments from genomic DNA of all three grasses (SG, DG, and HG) (Figure 5.14). The amplified fragments (SG3, and DG1) differed in size to HG4. SG3 and DG1 giving a 157 bp fragment, and HG4 giving a 160 bp fragment. Each of these fragments were sequenced and aligned (Figure 5.15). All the sequences aligned significantly with only few mutations points such as deletions of one to four nucleotides and substitutions of G to C or T to C (Figure 5.15).

```
        10         20         30         40         50         60
ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAACAA AGTATTGAAG
        70         80         90        100        110        120
AGACGACACA ATTGGCCAGG GGAATTTCGG GCTCGATTTT TATCCGATGC GTTATGAGGT
       130        140        150        160        170        180
TTGTCCGTCG GATTCATTTA CACGTTTGGG GCCTACGGGA TGCCGACGCG CTTTTCACAT
       190        200        210        220        230
TGGAGTTTTG GGA AATCATT CCATAGGATG AAGCTTGTTC ATGGATAGTC GACGTCGGT
```

Figure 5.13: Sequence of DP510 with primers Bhal5R and Bhal3L (underlined) designed to amplify an internal segment of sequence in grass genomic DNA.

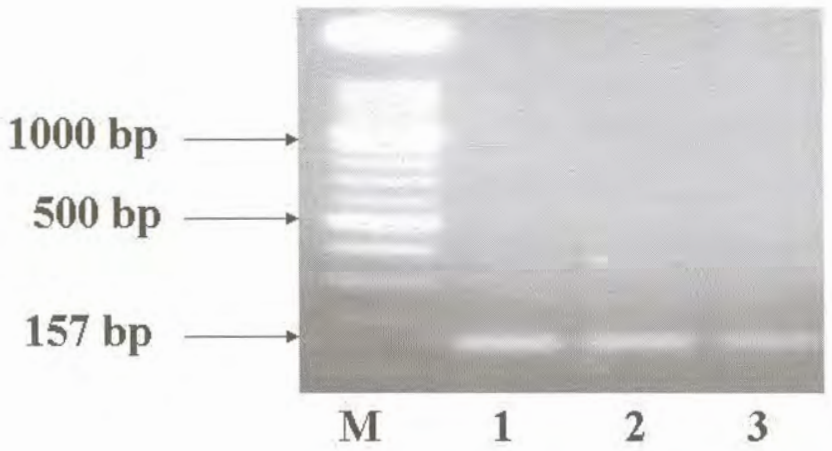


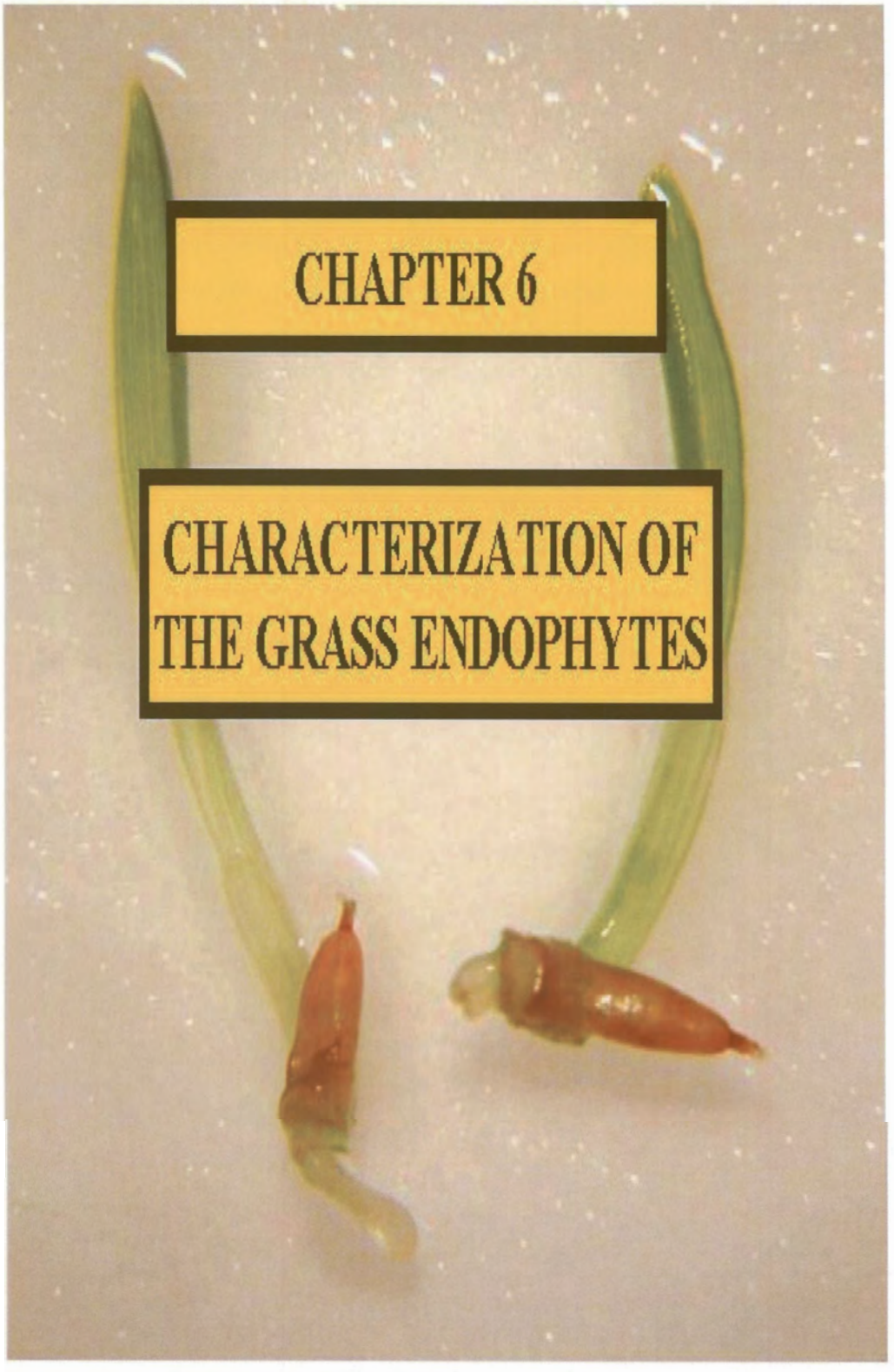
Figure 5.14: Amplification products of grass genomic DNA amplified with primers designed from internal part of DP510, visualized on an agarose gel stained with ethidium bromide. Lane 1 represents the amplified fragment from DG grass genomic DNA; lane 2 the amplified fragment from SG grass genomic DNA and lane 3 represents the amplified fragment from HG grass genomic DNA. Lane M represents 100 bp DNA ladder (Roche, Switzerland).

```
             10         20         30         40         50
DP510 CCGCGCTTGA ACAAAGTATT ----GAAGAG AC-GACACAA TTGGCCAGGG
SG3   CCGCGCTTGA ACAAAGTATT ----GAAGAG AC-GACACAA TTGGCCAGGG
DG1   CCGCGCTTGA ACAAAGTATT ----GAAGAG AC-GACACAA TTGGCCAGGG
HG4   CCGCGCTTGA ACAAAGTATT ATTCGAAGAG ACTGACACAA TTGGCCAGGC

             60         70         80         90        100
DP510 GAATTTCGGG CTCGATTTTT ATCCGATGCG TTATGAGGTT TGTCCGTCGG
SG3   --ATTTCGGG CTCGATTTTT ATCCGATGCG TTATGAGGTT TGTCCGTCGG
DG1   --ATTTCGGG CTCGATTTTT ATCCGATGCG TTATGAGGTT TGTCCGTCGG
HG4   --ATTTCGGG CTCGATTTTT ATCCGATGCG TTATGAGGTT TGTCCGTCGG

            110        120        130        140        150
DP510 AT-TCCATTT ACACGTTTGG GGCCTACGGG ATGCCGACGC GCTTTTCACA
SG3   ATG-CCATTT ACACGTTTGG GGCCTACGGG ATTCCGACGC GCTTTTCACA
DG1   ATGTC-ATTT ACACGTTTGG GGCCTACGGG ATGCCGACGC GCTTTTCACA
HG4   ATGTC-ATTT ACACGCTTGG GGCCTACGGG ATGCCGACGC --TTTTCACA
            160
DP510 TTGGAGTTTT GGGA
SG3   TTGGAGTTTT GGGA
DG1   TTGGAGTTTT GGGA
HG4   TTGGAGTTTT GGGA
```

Figure 5.15: Sequence data of the amplified fragments obtained after amplification of genomic DNA with primers designed for amplification of the internal part of DP510. Sequences of amplified fragments obtained from the SG grass, DG grass and HG grass genomic DNA was indicated by SG3, DG1, and HG4, respectively.


CHAPTER 6
CHARACTERIZATION OF THE GRASS ENDOPHYTES

6.1 Abstract

In this chapter any plant DNA contamination with endophytic bacterial DNA was investigated. In particular it was attempted to amplify by PCR from plant genomic DNA of the three types of grass a 595 bp fragment corresponding to an internal DNA portion of the *Bacillus subtilis* 16S rRNA region. An expected 595 bp DNA fragment was amplified from the isolated *Bacillus subtilis* DNA. However, no amplification product derived from plant genomic DNA. In a further experiment aimed at detecting *Bacillus subtilis* as an endophytic bacterium in the grass, several endophytic bacteria were isolated and characterized but none of them was identified as *Bacillus subtilis*.

6.2 Objective

The objective of this part of the study was to determine if plant genomic DNA carrying a *Bacillus* DNA sequence was contaminated with any bacterial DNA allowing the amplification of *Bacillus* DNA as identified as part of the subtraction product DP510.

6.3 Results

6.3.1 PCR amplification of DNA of the 16S rRNA region

Figure 6.1 shows the PCR amplification product (595 bp) from DNA of the 16S rRNA region using for DNA amplification the primers Bsub3R and Bsub5F and *Bacillus subtilis* DNA. No amplification product was observed from any of the plant genomic DNAs derived from the three types of grass.

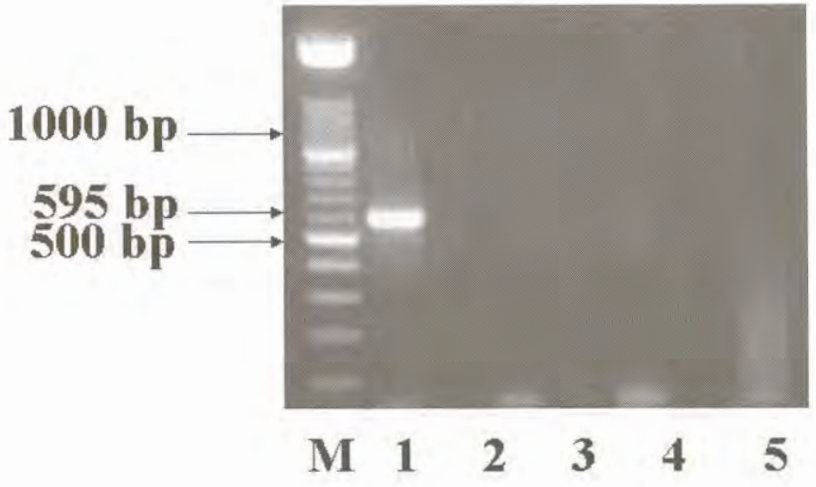


Figure 6.1: Amplification of *Bacillus subtilis* 16S rRNA region with specific primers designed for the 16S rRNA region and products visualized on a 1.5% agarose gel stained with ethidium bromide. Lane 1 represents an amplification product from *Bacillus subtilis* DNA. Lanes 2, 3, 4 represent isolated DG, HG, and SG grass genomic DNA derived from grass flowers. Lane 5 represents a H_2O control. Lane 6 represents genomic DNA isolated from a mixed bacterial culture derived from crushed SG seeds. Lane M represents 100 bp DNA ladder (Roche, Switzerland).

6.3.2 PCR amplification with DP 510 primers

Figure 6.2 shows amplification of a predicted 157 bp DNA fragment using primers Bhal5R and Bhal3L designed for the DP510 fragment covering the region with homology to *Bacillus* DNA. The DNA fragment with the predicted size was amplified from all three types of grass genomic DNAs and also from *Bacillus subtilis* DNA.

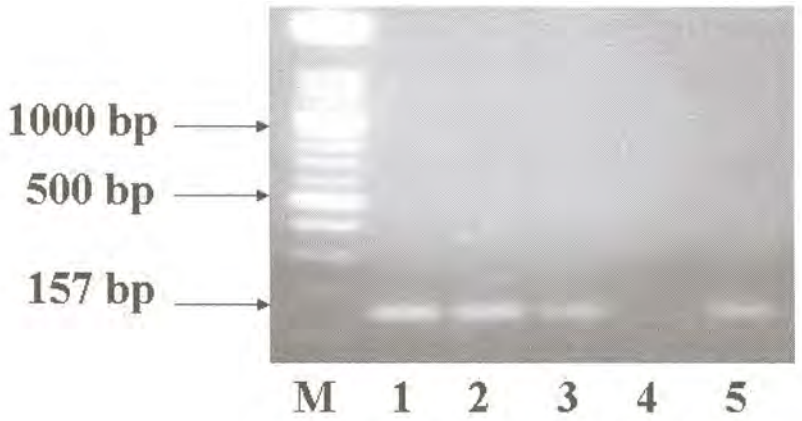
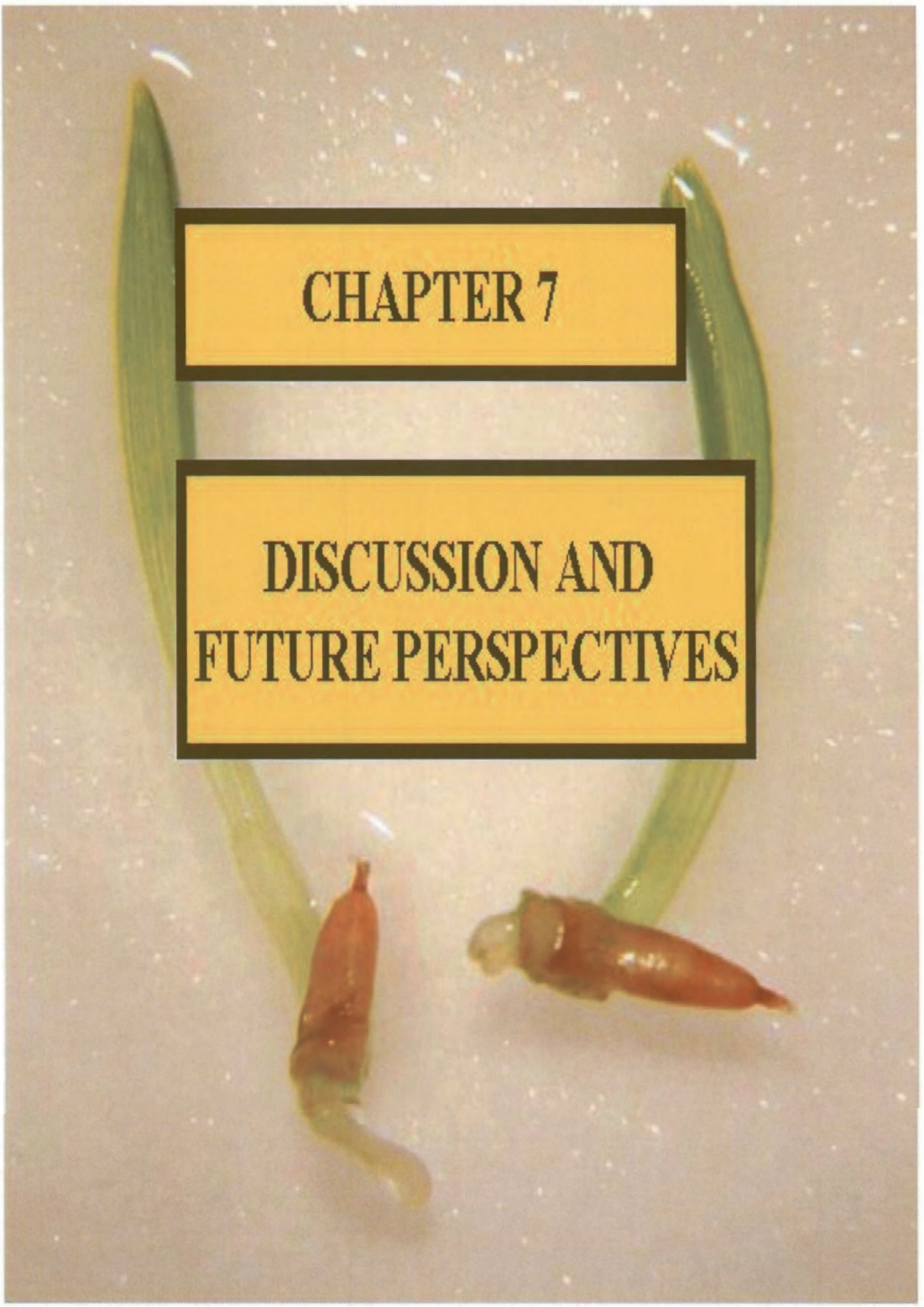


Figure 6.2: Amplification of DP510 fragment covering the region with homology to the Bacillus DNA from grass genomic DNAs and the *Bacillus subtilis* DNA visualized on a 1.5% agarose gel stained with ethidium bromide. Lane 1 represents genomic DNA from DG grass, lane 2 HG grass and lane 3 SG grass. Lane 4 represents H_2O control without DNA addition and lane 5 represents amplification of the predicted 157 bp DP510 fragment from isolated *Bacillus subtilis* DNA. Lane M represents a 100 bp DNA ladder (Roche, Switzerland).

6.3.3 Colonies identification

Streaking of *Bacillus subtilis* cells onto an agar plate resulted in irregular colonies with lobate margins that were dry and flat. This behavior was different to streaked endophytic bacteria isolated from grass seed extracts. These were chromogenic with lobate margins that were dry and flat but with a golden–brown color. By using Gram stain and Hugh Leifson oxidative/fermentative test the bacterium obtained was shown to be a member of the Enterobacteriaceae, either Erwinia spp; Enterobacter spp; or Pantoea spp. All Three genera are known as endophytes of plant species.



CHAPTER 7

DISCUSSION AND FUTURE PERSPECTIVES

7.1 Discussion

One of the most challenging tasks in plant science is the understanding of genome changes in plants that might occur as a response to growth under extreme environmental conditions. For this study, it was therefore considered important to obtain more detailed information about susceptible DNA regions in a grass that grows under different environmental conditions but shows no obvious change in the morphological characteristics. This study accomplished as a first new result the application of the RDA technique to identify and isolate possible putative genome changes from a genome of an inland grass species. RDA also allowed the detailed characterization of these changes with bio-informatic tools and extending the application of the RDA technique to a further plant species.

Execution of the RDA resulted in the isolation of two putative altered *Monocymbium* DNA sequences, which both belong to the class of repetitive DNA and both carried DNA sequences with homology to plant DNA, which was determined by bio-informatics tools. In comparison, the ITS technique only detected fungus-related sequences that had no homology to any plant sequences. Executing the RDA technology and isolating genomic DNA from grass seeds did not result in any interference by fungal DNA, which is difficult to eliminate in any field material. One of the isolated RDA subtraction products had a homology to the LTR region of a maize retro-transposon and seemingly represents a region where a transposition occurred. Such transposons are present in all members of the Poaceae family predominantly in gene-poor regions or non-transcribed repeated DNA (Bennetzen *et al.*, 1998). Using comparative analysis of closely related grass genomes, amplification of transposable elements have been found to be responsible for a large fraction of the variation in the genome size, at both inter-specific and intra-specific levels (San Miguel *et al.*, 1998). Vicient (1999) further reported a large variation in the copy number of the BARE-1 retro-transposons among and within *Hordeum* species. By correlating this variation with intra-specific variation in genome size and with local changes in environmental conditions, it has been suggested that alterations in transposable elements might be adaptive. Massive transposable element activity might

further be the reason why active LTR retro-tranposons can be routinely isolated from plants. Identification of a retro-transposon sequence in *Monocymbium* is consistent with earlier observations that RDA can be used to isolate such families of repetitive sequences.

None of the RDA-derived DNA sequences in this study could so far be clearly linked to detection of grasses derived from a specific location in South Africa. Failure might possibly be due to the very high degree of genomic identity between grasses although growing at environmentally different locations. They might differ, if at all, only in a very small portion of the genome. Possible genome differences might be consequently too small, possibly consisting only of point mutations, to be easily detected by RDA. One should also be aware that a subtractive technology, such as RDA, is inherently subject to several sources of bias. The representation of the genome is based on digestion of the genomic DNA with single restriction enzymes. Further, tester/driver ratios used for subtractive hybridization are critical for the elimination of common regions and enrichment of specific sequences. Also, the initial representation is influenced by the size of the restriction enzyme-digested fragments from total genomic DNA, where larger fragments amplify less efficient by PCR than smaller fragments. In addition, a single or very low copy number DNA difference sequence might also not have been efficiently amplified and enriched by the RDA protocol applied in this study.

Actions to overcome current failure of clear grass identification, if possible at all, might involve the usage of a greater range of different restriction enzymes for genome digestion and also size fractionation of subtraction products after the first round of subtraction. This will allow limiting the genome bias and the selectivity in the genome digestion step. By using different restriction enzymes several representations of the same genome can be scanned in each subtraction. So far the RDA technology has been developed only for four different restriction enzymes namely *Hind*III, *Bgl*II, *Bam*HI and *Hpa*II. And only the *Hind*III digestion was used in this study to avoid any interference with DNA methylation caused by the methylation-sensitive restriction enzyme *Bam*HI and *Hpa*II. In plants, methylation is mainly restricted to the nuclear genome, where methyl-cytosine is specially concentrated in repeated sequences (Finnegan *et al.*, 1998). The increasing

levels of methylation in carrot cells cultures have been observed with a high auxin concentration during embryogenesis (Lo-Schiavo *et al.*, 1989). Arnholdt-Schmitt *et al.* (1991) had also reduced the levels of methylation in carrot cultures when grown on a cytokinin-containing medium.

DNA hyper-methylation has been investigated in genetically modified plants by two antibiotics such as kanamycin and cefotaxime (Schmitt *et al.*, 1997). Schmitt *et al.* (1997) observed also an increasing of methylation in repetitive DNA. However, many more restriction enzymes could possibly be used for genomic digestion, if they generate the same staggered ends, thereby allowing the use of already developed adaptors. Further, a more stringent elimination of highly repetitive DNA sequences with unequal copy numbers, which seemingly are controlled by stresses, in plants might improve the discovery and enrichment of very unique induced genome changes. In this study, the *Mse*I- digestion was applied at a later subtraction step and indeed allowed the isolation of a low copy number repetitive DNA sequence. *Mse*I is known to digest DNA quite frequently in retro-transposons (personal communication, M. van der Merwe) and might possibly be used at the genomic DNA level together with size fractionation of digested DNA fragments in eliminating highly repeated DNA sequences before the production of RDA representations.

A further very interesting new result in this study was the detection of a subtraction product harboring a Bacillus DNA sequence. Although the subtraction product DP510 could be amplified by PCR from genomic DNA of the grass, detection of the DP510 in the genomic DNA by using Southern blotting failed or gave a very weak signal when a slot blot technique was used. So far, the origin of this subtraction product is unclear. The possibility cannot be excluded that the product might either represent a single copy insert, difficult to detect by Southern blotting using chemiluminescence, or simply represents a bacterial contamination. In general, plants play host to a wide variety of microorganisms, including bacteria. The relationships between the bacteria and their host plants are diverse and include pathogenicity, symbiotic root nodule formation, disease suppression and nitrogen fixation, plant growth promotion, interactions such DNA transfer and

probably other, as-yet-undiscovered. Two of the best-studied interactions between plant-hosts and bacteria include the root nodule-inhabiting *Rhizobium* spp and gall-forming *Agrobacterium tumefaciens*. The study of these systems led to the discovery that plants and bacteria communicate by using chemical signals that determine the outcome of the relationships between the organisms (Stacey *et al.*, 1995). Many researches have revealed that compounds mediate relationships between plants and gram-negative bacteria, which influence bacterial gene expression (Rainey, 1999). Although DNA of endophytes might easily interfere in the RDA technique, and the DP510 might represent a contamination by bacterial DNA, there is some evidence that DP510 is a true plant genome-derived subtraction product DNA and has not solely derived from an existing endophyte. Evidence includes (i) the existence of DNA flanking sequences with homology to plant DNA, (ii) failure to amplify a selected, specific *Bacillus subtilis* DNA sequence and (iii) failure of staining a *Bacillus* endophyte in seed material.

It might further be speculated, that detection of bacterial DNA in grass genomic DNA might further indicate that the plant genome shares homology with bacterial DNA or that bacterial DNA has invaded the plant genome by horizontal gene transfer (HGT). Although such transfer has been recently intensively studied in prokaryotic systems (Brown, 2003), there are currently no known naturally occurring vectors, such as plasmids, phages, or transposable elements that could be responsible for inter-domain gene transfer. So far, only a direct transfer of plastidic DNA into nuclear DNA has been found for plants (Bushman, 2002). There is also little evidence that eukaryotic cells are naturally competent, although this is a well-known characteristic of many bacterial species. The closest example of such a promiscuous vector is the Tumor-inducing (T) plasmid of the *Agrobacterium* genus. This bacterium is a soil-borne, bacterial plant pathogens that facilitate the transfer of bacterial DNA to plant somatic cells and the expression of this DNA (Kondo *et al.*, 2002).

In general HGT and inclusion of foreign DNA, possibly into labile regions of the plant genome, can be a mutational process for transfer of nucleotides between organisms. The increasing availability of gene and whole genome sequences has provided clear

indications that such HGT has been a major factor in the evolution of especially prokaryotic species. While there is no doubt about the occurrence of HGT, its importance in adaptive evolution is still debated especially in eukaryotes. It is widely accepted today that HGT heavily relies on mobile genetic elements (MGEs), which collectively form the so-called 'horizontal gene pool' (HGP). These elements can be transferable plasmids, transposons, integrons, genomic islands, or phage, which are able to move within and between genomes, thus allowing 'evolution in quantum leaps'. Among those MGEs, broad host-range (BHR) plasmids can be exchanged between phylogenetically distinct bacteria in various natural ecosystems. In the same way, it is proposed that major evolutionary leaps in eukaryotes (most clearly in unicellular eukaryotes, but possibly also in multi-cellular organisms) are produced by the traffic of mobile elements that operate in the same way as bacterial mobile elements. The eukaryotic mobile genetic elements are the transposons, viruses and bacteria that thrive among them. Initial support for this hypothesis comes from the now widely accepted notion that eukaryotes are, in essence and origin, an assemblage of the components of prokaryotic cells. The same HGT mechanisms that produce speciation in bacteria should, in principle, operate in these assemblages.

It is further thought that most of the genes were transferred from the endosymbiont bacterium to the host nucleus during transition of endosymbiont to organelles (Martin *et al.*, 2002), but many genes of prokaryotic origin remain in the eukaryotic nucleus. An example of eukaryote-to-bacteria transfer is glutamyl-tRNA synthetase, which is found throughout eukaryotes but only in Proteobacteria (Brown and Doolittle, 1999). The origin of mitochondria and plastids from different bacterial endosymbiots has been widely accepted for several decades (Margulis, 1970). However, the extent of additional gene transfer from bacteria to eukaryotic genomes is still being discovered. Brinkman, (2002) found for example that 65% of bacterial proteins in the bacterial genera *Chlamidia* and *Synechocystis* had a highest similarity to eukaryotic protein seemingly due to their close evolutionary relationship with cyanobacteria (ancestral blue green algae) and chloroplasts. The estimated 4500 cyanobacteria genes in the *Arabidopsis* genome are approximately 1000 more genes than being present in the *Synechocystis* genome

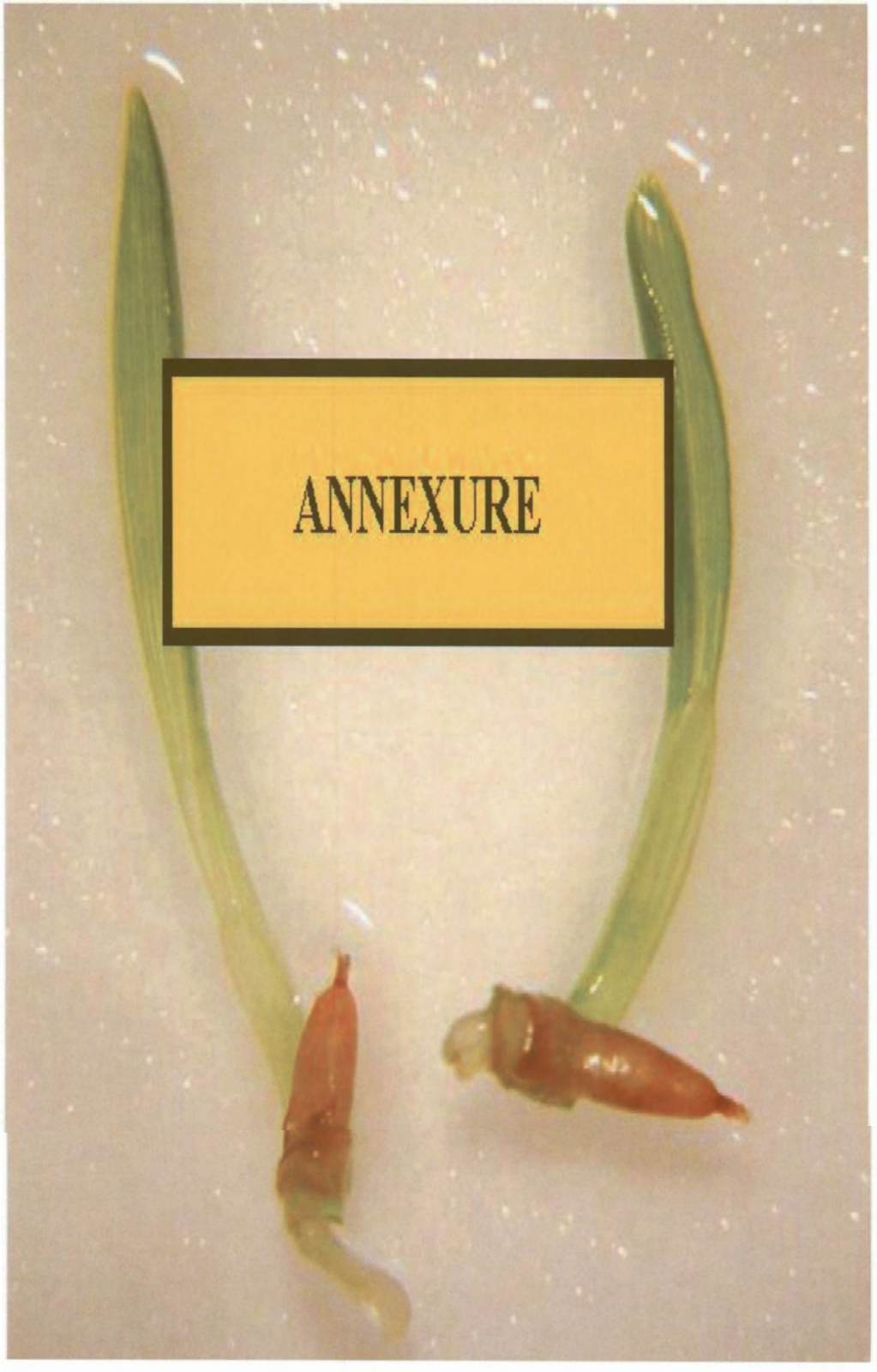
(Archebal and Keeling, 2003). Further evidence for the early integration of bacterial genes into the eukaryotic genome originates from studies of proteins from simple protists such as *Giardia intestinalis* and *Trichomonas vaginalis*, lacking mitochondria. However, molecular studies showed that these amitochondrial protists have genes for several proteins that are not only targeted to the mitochondria in higher eukaryotes but that are also of bacterial origin (Clark and Roger, 1995; Roger *et al.*, 1996). It has also recently been found that the eight-carbon acid sugar 3-deoxy-D-manno-octulosonate (KDO) which is an essential component of the endotoxic lipopolysaccharide or, occasionally capsular polysaccharide and identified on the outer surface of gram-negative bacteria, is also present in the cell wall polysaccharides of green algae and is a pectin component of the cell walls of higher plants (Royo *et al.*, 2000). Numerous conserved eukaryotic genes are closely related to orthologous in species of bacteria other than plastids or mitochondrial endosymbionts. These genes might have become integrated into the eukaryotic genome either by a series of "failed relationships" in which the bacterial "house guests' left behind remnant genes in the nucleus (Doolitte, 1998) or as suggested by Martin *et al*, (2002) or that the fluidity of the genomes of all bacteria, including putative endosymbionts, could have been the factor. Moreover, HGT between bacterial species combined with gene mutation or deletion, results in a high turnover of genes in a bacterial genome over time. If endosymbiosis is transient and frequent, there could be genes from various sources introduced into the eukaryotic genome. However, the extent of gene transfer from bacteria to eukaryotic genomes has still to be determined in greater detail.

7.2 Future perspectives

From this study, a valuable new aspect for a future investigation would be to study a possible HGT between Bacillus and plants using for example RDA to identify labile genome regions and characterize any insertions with bio-informatics tools. However, such study would also include determining the origin of any subtraction product identified by RDA with an efficient DNA amplification process for DNA contaminations.

A second aspect, which could be investigated, is the characterization of flanking sequences of microbial insertions into nuclear plant DNA. This might allow determining whether there are conserved elements at the insertion sites and if integration has occurred only at specific variable regions of the plant genome.

A third new aspect as an outcome of this study would be to conduct searches between completed microbial genome sequences and completed plant nuclear genome sequences to identify by computer-based subtractions any bacterial integration into plant genomes especially at labile regions of the plant genome. This would provide evidence for HGT as an evolutionary process. Types of integrations might also allow studying specially the occurrence and extent of these sequences in plant genomes and the possible function of these insertions. Work on comparative genomics might also uncover relationships between model organisms and facilitate the exploitation of conservation of synteny. In addition, identification and characterization of genome insertion sites might finally allow studying evolutionary plant adaptation processes and the development of trait markers.


ANNEXURE

Buffers for DNA work

Plant genomic DNA

Extraction buffer

100 mM Tris-HCl (pH 8); 1.4 mM sodium chloride (NaCl); 20 mM ethylenediamine tetra-acetic acid, disodium salt ($Na_2EDTA.\ 2H_2O$); 1% 2-mercaptoethanol and 3% hexadecyltrimethyl-ammoniumbromide (CTAB).

To make 500 ml buffer solution the following was added: 6.05 g Tris-base; 810 mg of NaCl; 2.92 g of $Na_2EDTA.\ 2H_2O$ and 15 g of CTAB was added to 400 ml of dsH_2O. The pH was adjusted to pH 8 with 10 N NaOH. 2-mercaptoethanol was diluted from a 50 mM stock solution on the day of use. 2-mercaptoethanol (350 µl) was added to 100 ml dsH_2O and 1.75 ml of the stock solution was added to 400 ml of the buffer solution and finally dsH_2O was added to make up a 500 ml extraction buffer.

Table A.1: DNA extraction buffer (pH 8)

Chemical	Concentration	Mass/volume
Tris-HCl	100 mM	6.05 g/500 ml buffer
NaCl	1.4 mM	0.81 g/500 ml buffer
EDTA	20 mM	2.92 g/500 ml buffer
ß-mercaptoethanol	50 mM	1.75 ml/500 ml buffer
CTAB	3%	15 g/500 ml buffer
Total		500 ml

The buffer was pre-heated on the day of use in a water bath at 60°C.

Bacterial DNA isolation

Resuspension buffer

50 mM glucose; 25 mM Tris-HCl, pH 8; 10 mM $Na_2EDTA.2H_2O$

Glucose (4.5 g), $Na_2EDTA.2H_2O$ (1.46 g) and ice-cold Tris-base (7.5 ml) were dissolved in dsH_2O (400 ml). The pH was set up with 10 N NaOH to pH 8 and the final volume was made up with dsH_2O to 500 ml. The mixture was autoclaved for 20 minutes and 100 µg/ml of RNAase A was added after cooling down to room temperature and the buffer was stored at 4°C.

Alkaline lysis buffer

200 mM NaOH; 1% SDS

NaOH (8.0 g) pellets were dissolved into dsH_2O (950 ml) and 25 ml of a 10% sodium dodecyl sulphate (SDS; sodium lauryl sulfate) solution was added.

10% SDS stock solution

SDS (10%) was made up the day before use by dissolving SDS (100 g) into dsH_2O (900 ml) using a protection shield to avoid breathing the dust. The mixture was heated to 68°C to assist the dissolution. The pH was adjusted to 7.2 by adding a few drops of concentrated HCl and the volume was adjusted to 1 l with dsH_2O. The 10% SDS solution was not further sterilized.

Neutralization buffer

3 M potassium acetate (pH 5.5)

Potassium acetate (294.5 g) was dissolved in sdH_2O (500 ml). The pH was adjusted to 5.5 with glacial acetic acid (~110 ml) and the volume adjusted to 1 l with dsH_2O.

Low TE buffer

10 mM Tris; 0.1 mM Na_2 EDTA. $2H_2O$ (pH 8)

Tris-base (18 mg) and Na_2EDTA. $2H_2O$ (121 mg) were added to dsH_2O (75 ml), mixed well and 10 N of NaOH was used to set the pH to 8 and then dsH_2O was added to 100 ml. The buffer was made up the day before use.

Precipitation solution

3 M sodium acetate (NaAc) (pH 4.8)

Sodium acetate (40.8 g) was dissolved first in 90 ml dsH_2O, the pH was adjusted to 6.8 with acetic acid and then dsH_2O was added to a final volume of 100 ml.

Buffers for Southern blotting

Denaturation solution for DNA transfer

1.5 M NaCl; 0.5 M NaOH

Sodium chloride (43.83 g, NaCl) and sodium hydroxide (10 g, NaOH) were dissolved in dsH_2O (400 ml) and made up to a final volume of 500 ml by adding dsH_2O. The solution was sterilized by autoclaving.

Neutralization solution for DNA transfer

1.4 M NaCl; 0.5 M Tris-HCl

Sodium chloride (43.83 g) and Tris-base (30.27 g) were dissolved in dsH_2O (400 ml). The pH was adjusted to 7.5 by adding concentrated HCl slowly and carefully under stirring. Finally the volume was made up to 500 ml with dsH_2O and the mixture was autoclaved.

Hybridization buffer

5×SSC; 0.1% (w/v SDS); Dextran sulfate sodium salt; liquid block (Amersham life science, UK)

Into dsH_2O (800 ml), NaCl (175.3 g) and Na_3-citrate $2H_2O$ (88.2 g) were dissolved to produce a 20×SSC stock solution. The pH was adjusted to 7.0 with a few drops of 10 **N** NaOH and dsH_2O was added to a final volume of 1 l. Aliquots were sterilized by autoclaving.

For the preparation of the hybridization buffer, dsH_2O (26.6 ml), 20×SSC (10 ml), 10% SDS (0.4 ml) and liquid block (2 ml) were mixed to make up a hybridization stock solution in which dextran sulfate sodium salt (2 g) was dissolved at 60°C in a total volume of 40 ml.

Wash buffer

Buffer 1

1×SSC; 0.1% (w/v) SDS

To make up 1×SSC, 20×SSC (20 ml) and 10% SDS (4 ml) was added to 376 ml of dsH_2O for a total volume of 400 ml.

Buffer 2

0.5×SSC; 0.1% (w/V) SDS

20×SSC (10 ml) (stock solution) and 10% SDS (4 ml) (stock solution) were added to dsH_2O (386 ml) for a total volume of 400 ml. Both buffers were autoclaved for 20 minutes at 105 kPa to avoid any contamination.

Incubation and blocking buffer

100 mM Tris-HCl; 300 mM NaCl (pH 9.5)

NaCl (58.76 g) and Tris-base (6.05 g) were dissolved in dsH_2O (400 ml). The pH was adjusted to 9.5 with concentrated HCl and dsH_2O was added to a total volume of 500 ml and autoclaved in a 1 l bottle for 20 minutes at 105 kPa.

Table A.2: Buffers/Southern blotting

Required solutions	**Description**	**Concentration**	**Mass/Volume**
HCl	250 mM	250 mM	73 ml/500 ml H_2O
dsH_2O	Distilled, sterile water		
Denaturation buffer	0.5 N NaOH 1.5 M NaCl	0.5 N 1.5 M	10 g/500 ml buffer 43.83 g/500 ml buffer
Neutralization buffer	0.5 M Tris-HCl pH 7.5 1.5 M NaCl	0.5 M 1.5 M	30.27 g/500 ml buffer 43.83 g/500 ml buffer

20×SSC buffer	3 M NaCl 300 mM sodium citrate, pH 7.0	3 M 300 mM	97.66 g/1000 ml buffer 88.2 g/1000 ml buffer
5×SSC	750 mM NaCl 75 mM sodium citrate, pH 7.0	750 mM 75 mM	43.83 g/1000 ml buffer 22.05 g/1000 ml buffer
10%SDS	High SDS buffer	10%	100 g/900 ml H_2O
Hybridization buffer	5×SSC 0.1% (w/v SDS) Dextran sulfate liquid block		500 µl/40 ml buffer 400 µl/40 ml buffer 2 g/40 ml buffer 2 ml/40 ml buffer
2×wash buffer 0.5×wash buffer	2×SSC 0.1% SDS 0.5×SSC 0.1% SDS	 0.1% 0.1%	

Buffers for DNA colony hybridization

Denaturation buffer

0.5 N NaOH; 1.5 M NaCl

NaOH (10 g) pellets and NaCl (43.8 g) were dissolved in dsH_2O (500 ml) and the mixture was sterilized by autoclaving for 20 minutes.

Neutralization buffer

10 M Tris-HCl (pH 7.5); 1.5 M NaCl

Tris-base (60.5 g) and NaCl (43.8 g) were dissolved in dsH_2O (400 ml). The pH was adjusted by adding concentrated HCl slowly and carefully with stirring to 7.5. Finally, the volume was made up to 500 ml by adding dsH_2O and autoclaved for sterilization.

Buffers for DNA electrophoresis

Tris -acetate (TAE) buffer

Stock solution (50×TAE)

50×TAE: contained 2 M Tris-base; 0.5M $Na_2EDTA.2H_2O$ and glacial acetic acid (pH 8). To make up a 1 l stock solution, Tris-base (242 g) was added to 0.5 M $Na_2EDTA.2H_2O$ solution (100 ml) and dsH_2O (800 ml) were added. The pH was adjusted to 8 with 57.1 ml of glacial acetic acid and the volume made up to 1 l with dsH_2O. The stock solution was stored at room temperature in a glass bottle after autoclaving.

1×TAE buffer for DNA electrophoresis

50×TAE (200 ml) was diluted with dsH_2O (9.8 l) in a total volume of 10 l for a final concentration of 0.04 M Tris-acetate and 1 mM $Na_2EDTA.2H_2O$ .The buffer was stored at room temperature, away from light, for further uses.

Table A.3: TAE buffer (50×stock)

Chemical	Concentration	Mass/Volume
Tris	2 M	242 g
EDTA disodium salt	0.5 M	37.2 g
Glacial acetic acid	5.71% (w/v)	57.1 ml
Total volume		1000 ml

DNA loading buffer (agarose gel)

Table A.4: DNA loading buffer

Chemical	Concentration	Mass/Volume
Glycerol	50%	5 ml
TAE buffer	1×	200 µl of 50×stock
Bromophenol blue	1%	0.1 g
Xylene cyanol	1%	0.1%

Agarose gel composition

Agarose gels had the following composition:

Table A.5: Agarose gel composition

Tray	1×TAE buffer	1% agarose	1.5% agarose
7×10 cm	50 ml	0.50 g	0.75 g
15×10 cm	100 ml	1 g	1.50 g
15×15 cm	150 ml	1.5 g	2.25 g

Polymerase chain reaction buffer

10×PCR buffer

PCR reaction buffer consisted of a 10×PCR buffer (Takara, Japan) containing 500 mM KCl; 25 mM $MgCl_2$; 100 mM Tris-HCl (pH 8.3).

Buffers for RDA technique

Ligation buffer

10×ligase buffer contained 66 mM Tris-HCl (pH 7.6); 6.6 mM $MgCl_2$; 10 mM dithiothreitol (DDT) and 66 mM ATP. The supplier of ligase supplied the ligation buffer (Amersham, UK).

Elution TE-buffer

10 mM Tris-HCl (pH 8); 0.1 mM $Na_2EDTA.\ 2H_2O$

Tris- base (605 mg) and $Na_2EDTA\ 2H_2O$ (9 mg) were dissolved in dsH_2O (400 ml). The pH was adjusted with 10 N HCl to 8 and the volume was adjusted to 500 ml with dsH_2O.

Buffer for subtractive hybridization

30 mM EPPS [(N-[2-hydroxyethyl] piperazine)-N'-(3-propane sulfonic acid; HEPPS)] (pH 8.0) at 20°C; 3 mM $Na_2EDTA.\ 2H_2O$; 5 M NaCl.

EPPS (1.51 g), $Na_2EDTA\ 2H_2O$ (220 mg) and NaCl (58.43 g) were dissolved in dsH_2O (150 ml). The pH was adjusted to 8 by stirring the solution at 20°C and the total volume was set up to 200 ml with dsH_2O.

10 M Ammonium acetate

Ammonium acetate (770 g) was dissolved in dsH_2O (800 ml). The volume was adjusted with dsH_2O to 1 l and the mixture sterilized by filtration.

10×Mung bean nuclease buffer

50 mM Tris-HCl (pH 8.9)

Tris-base (300 mg) was dissolved in dsH_2O (90 ml) and the pH adjusted to 8.9 dsH_2O was added to a total volume of 100 ml.

Cloning reagents

Isopropyl-β-D-thiogalactopyranoside; IPTG (0.1 M stock solution)

IPTG (1.2 g) was dissolved in dsH_2O (50 ml) and the stock mixture was filter-sterilized and stored at -20°C.

5-bromo-4-chloro-3-indolyl- β-D-galactoside; X-gal (2% stock solution)

X-gal (20 mg) was dissolved in dimethylformamide (DMF) (1 ml). The stock mixture was covered with aluminum foil and stored at -20°C.

Ampicillin50

Ampicillin (50 mg) was dissolved in dsH_2O (1 ml). The mixture was filter-sterilized and stored at -20°C.

Growth media

Bacteria growth medium (Luria–Bertani broth)

To dsH_2O (1 l), Tryptone (10 g); Yeast extract (5 g) and NaCl (10 g) were added and the pH was adjusted to 7.4 with NaOH. For a solid medium LB medium, agar (15 g) was added. Both media were sterilized by autoclaving to avoid any contamination.

LB plates with ampicillin

LB-agar medium (30-35 ml) was poured into 85 mm petri dishes before adding ampicillin (20 or 40 µl/plate) to a final concentration of 50 µg or 100 µg/ml. The medium was allowed to cool down to 50°C. After agar hardened, the plates were stored at 4°C for up to one month or at room temperature for up to one week.

LB plates with ampicillin/ IPTG/X-gal

LB plates containing ampicillin were produced as outlined above but then supplemented with IPTG (20-100 µl of 0.1 M stock solution) and X-gal (20-35 µl of a 2% stock solution). The LB agar was mixed with the reagents and plates were dried for 30 minutes at room temperature.

Table A.6: Growth medium composition

Reagents	Concentration	Quantity/plates	Mass/volume
IPTG	100 mM	20-100 µl	1.2 g/50 ml dsH_2O
X-gal	2%	20-35 µl	0.02 g/1 ml DMF
Ampicillin	50-100 µg/ml	20-35 µl	50 mg/1 ml dsH_2O
Tryptone			10 g in 1 l dsH_2O
Yeast extract			5 g/1 in1 l dsH_2O
NaCl			10 g/1 in 1 l dsH_2O
Agar			15 g/1 in 1 l LB broth
LB medium		25-35 ml	

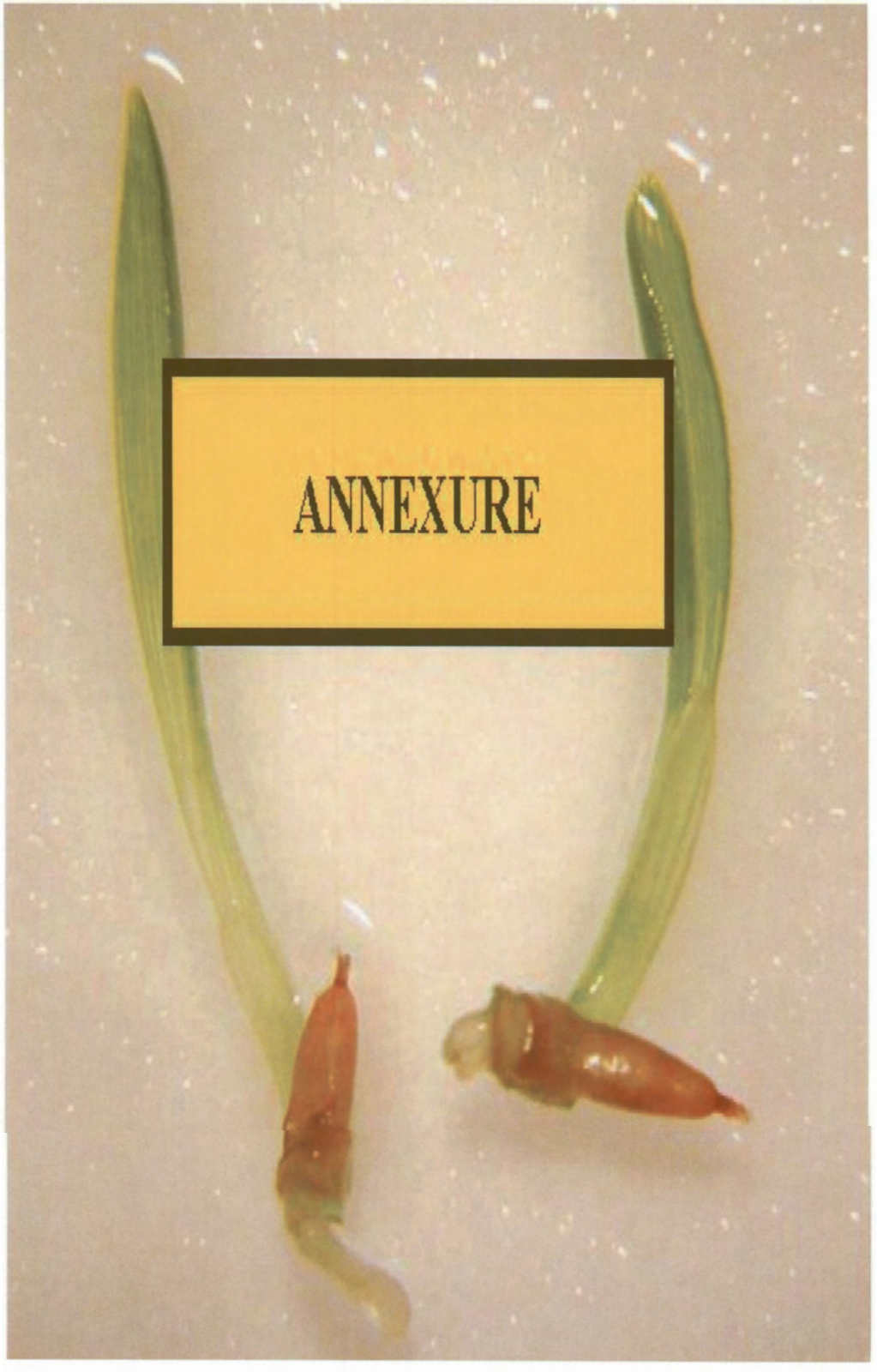
Sequence of primers used in experiments

Table A.7: Sequences of all the primers used for amplification of the DNA regions." ITS" represents the sequence of the primer set used to amplify the internally transcribed spacer sequence (ITS region); "NTS" primer set used to amplify the non-transcribed spacer (NTS region); "Retrotransposon" primer used to amplify a retrotransposon like region from the grass species *Monocymbium ceresiiforme*. "DP510" primer set used to amplify a fragment with homology to *Bacillus halodurans* region and "*Bacillus subtilis*" the primer set used to amplify the *Bacillus subtilis* 16s rRNA region.

DNA regions	Primer	Primer sequence
ITS	ITS1	5'-TCCGTAGGTGAACCTGCGGG-3'
	ITS4	5'-GCTGCGTTCTTCATCGATGC-3'
NTS	NTS1	5'-TTTAGTGCTGGTATGATCGC-3'
	NTS2	5'-TTGGAAGTCCTCGTGTTGCA-3'
Retrotransposon	S3Cl2L	5'-CTCGGTATCGAGGGAGA-3'
	S3CL2R	5'-TTTCAAGAATGCTCTGCAGG-3'
DP 510	Bhal5R	5'-CCGCGCTTGAACAAAGTATT-3'
	Bhal3L	5'-TTCACATTGGAGTTTTGGGA-3'
	Bhal5A	5'-ACCGACGTCGACTATCCATGAACAA-3'
	Bhal3A	5'-AAGCTTGTTCATGGATAGTCGACGTCGGT-3'
Bacillus subtilis	Bsub3R	5'-CCAGTTTCCATTGACCCTCCCC-3'
	Bsub5F	5'-AAGTCGAGCGGACAGATGG-3'

Table A.8: Sequence of the three adaptor sets used for execution of the RDA.

RDA adaptor sets	Adaptor sequence
Set 1	
RHind12	5'-AGCTTCGGGTGA-3'
RHind24	5'-AGCACTCTCCAGCCTCTCACCGCA-3'
Set 2	
JHind12	5'-AGCTTGTTCATG-3
JHind24	5'-ACCGACGTCGACTATCCATGAACA-3'
Set 3	
NHind12	5'-AGCTTCTCCCTC-3'
NHind24	5'-AGGCAGCTGTGGTATCGAGGGAGA-3'


ANNEXURE

Buffers for DNA work

Plant genomic DNA

Extraction buffer

100 mM Tris-HCl (pH 8); 1.4 mM sodium chloride (NaCl); 20 mM ethylenediamine tetra-acetic acid, disodium salt ($Na_2EDTA.\ 2H_2O$); 1% 2-mercaptoethanol and 3% hexadecyltrimethyl-ammoniumbromide (CTAB).

To make 500 ml buffer solution the following was added: 6.05 g Tris-base; 810 mg of NaCl; 2.92 g of $Na_2EDTA.\ 2H_2O$ and 15 g of CTAB was added to 400 ml of dsH_2O. The pH was adjusted to pH 8 with 10 N NaOH. 2-mercaptoethanol was diluted from a 50 mM stock solution on the day of use. 2-mercaptoethanol (350 µl) was added to 100 ml dsH_2O and 1.75 ml of the stock solution was added to 400 ml of the buffer solution and finally dsH_2O was added to make up a 500 ml extraction buffer.

Table A.1: DNA extraction buffer (pH 8)

Chemical	Concentration	Mass/volume
Tris-HCl	100 mM	6.05 g/500 ml buffer
NaCl	1.4 mM	0.81 g/500 ml buffer
EDTA	20 mM	2.92 g/500 ml buffer
ß-mercaptoethanol	50 mM	1.75 ml/500 ml buffer
CTAB	3%	15 g/500 ml buffer
Total		500 ml

The buffer was pre-heated on the day of use in a water bath at 60°C.

Bacterial DNA isolation

Resuspension buffer

50 mM glucose; 25 mM Tris-HCl, pH 8; 10 mM $Na_2EDTA.2H_2O$

Glucose (4.5 g), $Na_2EDTA.2H_2O$ (1.46 g) and ice-cold Tris-base (7.5 ml) were dissolved in dsH_2O (400 ml). The pH was set up with 10 N NaOH to pH 8 and the final volume was made up with dsH_2O to 500 ml. The mixture was autoclaved for 20 minutes and 100 µg/ml of RNAase A was added after cooling down to room temperature and the buffer was stored at 4°C.

Alkaline lysis buffer

200 mM NaOH; 1% SDS

NaOH (8.0 g) pellets were dissolved into dsH_2O (950 ml) and 25 ml of a 10% sodium dodecyl sulphate (SDS; sodium lauryl sulfate) solution was added.

10% SDS stock solution

SDS (10%) was made up the day before use by dissolving SDS (100 g) into dsH_2O (900 ml) using a protection shield to avoid breathing the dust. The mixture was heated to 68°C to assist the dissolution. The pH was adjusted to 7.2 by adding a few drops of concentrated HCl and the volume was adjusted to 1 l with dsH_2O. The 10% SDS solution was not further sterilized.

Neutralization buffer

3 M potassium acetate (pH 5.5)

Potassium acetate (294.5 g) was dissolved in sdH_2O (500 ml). The pH was adjusted to 5.5 with glacial acetic acid (~110 ml) and the volume adjusted to 1 l with dsH_2O.

Low TE buffer

10 mM Tris; 0.1 mM Na_2 EDTA. $2H_2O$ (pH 8)

Tris-base (18 mg) and Na_2EDTA. $2H_2O$ (121 mg) were added to dsH_2O (75 ml), mixed well and 10 N of NaOH was used to set the pH to 8 and then dsH_2O was added to 100 ml. The buffer was made up the day before use.

Precipitation solution

3 M sodium acetate (NaAc) (pH 4.8)

Sodium acetate (40.8 g) was dissolved first in 90 ml dsH_2O, the pH was adjusted to 6.8 with acetic acid and then dsH_2O was added to a final volume of 100 ml.

Buffers for Southern blotting

Denaturation solution for DNA transfer

1.5 M NaCl; 0.5 M NaOH

Sodium chloride (43.83 g, NaCl) and sodium hydroxide (10 g, NaOH) were dissolved in dsH_2O (400 ml) and made up to a final volume of 500 ml by adding dsH_2O. The solution was sterilized by autoclaving.

Neutralization solution for DNA transfer

1.4 M NaCl; 0.5 M Tris-HCl

Sodium chloride (43.83 g) and Tris-base (30.27 g) were dissolved in dsH_2O (400 ml). The pH was adjusted to 7.5 by adding concentrated HCl slowly and carefully under stirring. Finally the volume was made up to 500 ml with dsH_2O and the mixture was autoclaved.

Hybridization buffer

5×SSC; 0.1% (w/v SDS); Dextran sulfate sodium salt; liquid block (Amersham life science, UK)

Into dsH_2O (800 ml), NaCl (175.3 g) and Na_3-citrate $2H_2O$ (88.2 g) were dissolved to produce a 20×SSC stock solution. The pH was adjusted to 7.0 with a few drops of 10 **N** NaOH and dsH_2O was added to a final volume of 1 l. Aliquots were sterilized by autoclaving.

For the preparation of the hybridization buffer, dsH_2O (26.6 ml), 20×SSC (10 ml), 10% SDS (0.4 ml) and liquid block (2 ml) were mixed to make up a hybridization stock solution in which dextran sulfate sodium salt (2 g) was dissolved at 60°C in a total volume of 40 ml.

Wash buffer

Buffer 1

1×SSC; 0.1% (w/v) SDS

To make up 1×SSC, 20×SSC (20 ml) and 10% SDS (4 ml) was added to 376 ml of dsH_2O for a total volume of 400 ml.

Buffer 2

0.5×SSC; 0.1% (w/V) SDS

20×SSC (10 ml) (stock solution) and 10% SDS (4 ml) (stock solution) were added to dsH_2O (386 ml) for a total volume of 400 ml. Both buffers were autoclaved for 20 minutes at 105 kPa to avoid any contamination.

Incubation and blocking buffer

100 mM Tris-HCl; 300 mM NaCl (pH 9.5)

NaCl (58.76 g) and Tris-base (6.05 g) were dissolved in dsH_2O (400 ml). The pH was adjusted to 9.5 with concentrated HCl and dsH_2O was added to a total volume of 500 ml and autoclaved in a 1 l bottle for 20 minutes at 105 kPa.

Table A.2: Buffers/Southern blotting

Required solutions	**Description**	**Concentration**	**Mass/Volume**
HCl	250 mM	250 mM	73 ml/500 ml H_2O
dsH_2O	Distilled, sterile water		
Denaturation buffer	0.5 N NaOH 1.5 M NaCl	0.5 N 1.5 M	10 g/500 ml buffer 43.83 g/500 ml buffer
Neutralization buffer	0.5 M Tris-HCl pH 7.5 1.5 M NaCl	0.5 M 1.5 M	30.27 g/500 ml buffer 43.83 g/500 ml buffer

20×SSC buffer	3 M NaCl 300 mM sodium citrate, pH 7.0	3 M 300 mM	97.66 g/1000 ml buffer 88.2 g/1000 ml buffer
5×SSC	750 mM NaCl 75 mM sodium citrate, pH 7.0	750 mM 75 mM	43.83 g/1000 ml buffer 22.05 g/1000 ml buffer
10%SDS	High SDS buffer	10%	100 g/900 ml H_2O
Hybridization buffer	5×SSC 0.1% (w/v SDS) Dextran sulfate liquid block		500 µl/40 ml buffer 400 µl/40 ml buffer 2 g/40 ml buffer 2 ml/40 ml buffer
2×wash buffer 0.5×wash buffer	2×SSC 0.1% SDS 0.5×SSC 0.1% SDS	 0.1% 0.1%	

Buffers for DNA colony hybridization

Denaturation buffer

0.5 N NaOH; 1.5 M NaCl

NaOH (10 g) pellets and NaCl (43.8 g) were dissolved in dsH_2O (500 ml) and the mixture was sterilized by autoclaving for 20 minutes.

Neutralization buffer

10 M Tris-HCl (pH 7.5); 1.5 M NaCl

Tris-base (60.5 g) and NaCl (43.8 g) were dissolved in dsH_2O (400 ml). The pH was adjusted by adding concentrated HCl slowly and carefully with stirring to 7.5. Finally, the volume was made up to 500 ml by adding dsH_2O and autoclaved for sterilization.

Buffers for DNA electrophoresis

Tris -acetate (TAE) buffer

Stock solution (50×TAE)

50×TAE: contained 2 M Tris-base; 0.5M $Na_2EDTA.2H_2O$ and glacial acetic acid (pH 8). To make up a 1 l stock solution, Tris-base (242 g) was added to 0.5 M $Na_2EDTA.2H_2O$ solution (100 ml) and dsH_2O (800 ml) were added. The pH was adjusted to 8 with 57.1 ml of glacial acetic acid and the volume made up to 1 l with dsH_2O. The stock solution was stored at room temperature in a glass bottle after autoclaving.

1×TAE buffer for DNA electrophoresis

50×TAE (200 ml) was diluted with dsH_2O (9.8 l) in a total volume of 10 l for a final concentration of 0.04 M Tris-acetate and 1 mM $Na_2EDTA.2H_2O$ .The buffer was stored at room temperature, away from light, for further uses.

Table A.3: TAE buffer (50×stock)

Chemical	Concentration	Mass/Volume
Tris	2 M	242 g
EDTA disodium salt	0.5 M	37.2 g
Glacial acetic acid	5.71% (w/v)	57.1 ml
Total volume		1000 ml

DNA loading buffer (agarose gel)

Table A.4: DNA loading buffer

Chemical	Concentration	Mass/Volume
Glycerol	50%	5 ml
TAE buffer	1×	200 µl of 50×stock
Bromophenol blue	1%	0.1 g
Xylene cyanol	1%	0.1%

Agarose gel composition

Agarose gels had the following composition:

Table A.5: Agarose gel composition

Tray	1×TAE buffer	1% agarose	1.5% agarose
7×10 cm	50 ml	0.50 g	0.75 g
15×10 cm	100 ml	1 g	1.50 g
15×15 cm	150 ml	1.5 g	2.25 g

Polymerase chain reaction buffer

10×PCR buffer

PCR reaction buffer consisted of a 10×PCR buffer (Takara, Japan) containing 500 mM KCl; 25 mM $MgCl_2$; 100 mM Tris-HCl (pH 8.3).

Buffers for RDA technique

Ligation buffer

10×ligase buffer contained 66 mM Tris-HCl (pH 7.6); 6.6 mM $MgCl_2$; 10 mM dithiothreitol (DDT) and 66 mM ATP. The supplier of ligase supplied the ligation buffer (Amersham, UK).

Elution TE-buffer

10 mM Tris-HCl (pH 8); 0.1 mM $Na_2EDTA.\ 2H_2O$

Tris- base (605 mg) and $Na_2EDTA\ 2H_2O$ (9 mg) were dissolved in dsH_2O (400 ml). The pH was adjusted with 10 N HCl to 8 and the volume was adjusted to 500 ml with dsH_2O.

Buffer for subtractive hybridization

30 mM EPPS [(N-[2-hydroxyethyl] piperazine)-N'-(3-propane sulfonic acid; HEPPS)] (pH 8.0) at 20°C; 3 mM $Na_2EDTA.\ 2H_2O$; 5 M NaCl.

EPPS (1.51 g), $Na_2EDTA\ 2H_2O$ (220 mg) and NaCl (58.43 g) were dissolved in dsH_2O (150 ml). The pH was adjusted to 8 by stirring the solution at 20°C and the total volume was set up to 200 ml with dsH_2O.

10 M Ammonium acetate

Ammonium acetate (770 g) was dissolved in dsH_2O (800 ml). The volume was adjusted with dsH_2O to 1 l and the mixture sterilized by filtration.

10×Mung bean nuclease buffer

50 mM Tris-HCl (pH 8.9)

Tris-base (300 mg) was dissolved in dsH_2O (90 ml) and the pH adjusted to 8.9 dsH_2O was added to a total volume of 100 ml.

Cloning reagents

Isopropyl-β-D-thiogalactopyranoside; IPTG (0.1 M stock solution)

IPTG (1.2 g) was dissolved in dsH_2O (50 ml) and the stock mixture was filter-sterilized and stored at -20°C.

5-bromo-4-chloro-3-indolyl- β-D-galactoside; X-gal (2% stock solution)

X-gal (20 mg) was dissolved in dimethylformamide (DMF) (1 ml). The stock mixture was covered with aluminum foil and stored at -20°C.

Ampicillin50

Ampicillin (50 mg) was dissolved in dsH_2O (1 ml). The mixture was filter-sterilized and stored at -20°C.

Growth media

Bacteria growth medium (Luria–Bertani broth)

To dsH_2O (1 l), Tryptone (10 g); Yeast extract (5 g) and NaCl (10 g) were added and the pH was adjusted to 7.4 with NaOH. For a solid medium LB medium, agar (15 g) was added. Both media were sterilized by autoclaving to avoid any contamination.

LB plates with ampicillin

LB-agar medium (30-35 ml) was poured into 85 mm petri dishes before adding ampicillin (20 or 40 µl/plate) to a final concentration of 50 µg or 100 µg/ml. The medium was allowed to cool down to 50°C. After agar hardened, the plates were stored at 4°C for up to one month or at room temperature for up to one week.

LB plates with ampicillin/ IPTG/X-gal

LB plates containing ampicillin were produced as outlined above but then supplemented with IPTG (20-100 µl of 0.1 M stock solution) and X-gal (20-35 µl of a 2% stock solution). The LB agar was mixed with the reagents and plates were dried for 30 minutes at room temperature.

Table A.6: Growth medium composition

Reagents	Concentration	Quantity/plates	Mass/volume
IPTG	100 mM	20-100 µl	1.2 g/50 ml dsH_2O
X-gal	2%	20-35 µl	0.02 g/1 ml DMF
Ampicillin	50-100 µg/ml	20-35 µl	50 mg/1 ml dsH_2O
Tryptone			10 g in 1 l dsH_2O
Yeast extract			5 g/1 in1 l dsH_2O
NaCl			10 g/1 in 1 l dsH_2O
Agar			15 g/1 in 1 l LB broth
LB medium		25-35 ml	

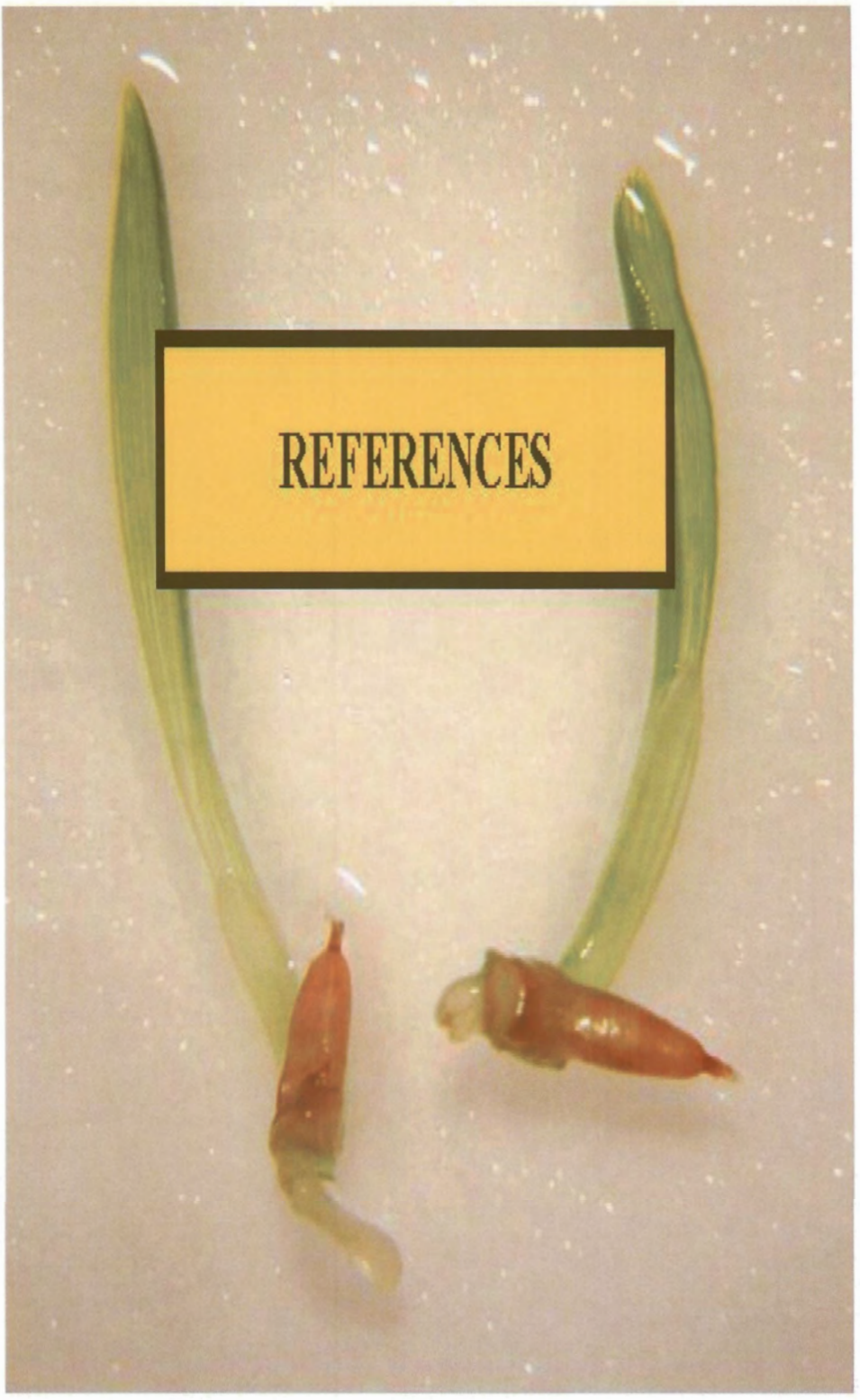
Sequence of primers used in experiments

Table A.7: Sequences of all the primers used for amplification of the DNA regions." ITS" represents the sequence of the primer set used to amplify the internally transcribed spacer sequence (ITS region); "NTS" primer set used to amplify the non-transcribed spacer (NTS region); "Retrotransposon" primer used to amplify a retrotransposon like region from the grass species *Monocymbium ceresiiforme*. "DP510" primer set used to amplify a fragment with homology to *Bacillus halodurans* region and "*Bacillus subtilis*" the primer set used to amplify the *Bacillus subtilis* 16s rRNA region.

DNA regions	Primer	Primer sequence
ITS	ITS1	5'-TCCGTAGGTGAACCTGCGGG-3'
	ITS4	5'-GCTGCGTTCTTCATCGATGC-3'
NTS	NTS1	5'-TTTAGTGCTGGTATGATCGC-3'
	NTS2	5'-TTGGAAGTCCTCGTGTTGCA-3'
Retrotransposon	S3Cl2L	5'-CTCGGTATCGAGGGAGA-3'
	S3CL2R	5'-TTTCAAGAATGCTCTGCAGG-3'
DP 510	Bhal5R	5'-CCGCGCTTGAACAAAGTATT-3'
	Bhal3L	5'-TTCACATTGGAGTTTTGGGA-3'
	Bhal5A	5'-ACCGACGTCGACTATCCATGAACAA-3'
	Bhal3A	5'-AAGCTTGTTCATGGATAGTCGACGTCGGT-3'
Bacillus subtilis	Bsub3R	5'-CCAGTTTCCATTGACCCTCCCC-3'
	Bsub5F	5'-AAGTCGAGCGGACAGATGG-3'

Table A.8: Sequence of the three adaptor sets used for execution of the RDA.

RDA adaptor sets	Adaptor sequence
Set 1	
RHind12	5'-AGCTTCGGGTGA-3'
RHind24	5'-AGCACTCTCCAGCCTCTCACCGCA-3'
Set 2	
JHind12	5'-AGCTTGTTCATG-3
JHind24	5'-ACCGACGTCGACTATCCATGAACA-3'
Set 3	
NHind12	5'-AGCTTCTCCCTC-3'
NHind24	5'-AGGCAGCTGTGGTATCGAGGGAGA-3'



REFERENCES

Abe T., Li N., Togashi A. and Sasahara T. 2002. Large deletions in chloroplast DNA of rice calli after long-term culture. Journal of Plant Physiology **159:** 918-923.

Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology **215:** 403-410.

Anh S. N., Anderson J. A., Sorrells M. E. and Tanskley S. D. 1993: Homeologous relationships of rice, wheat and maize. Molecular and General Genetics **241:** 483-490.

Archibal J. M. and Keeling P. J. 2003. Comparative genomics: Plant genomes: Cyanobacterial genes revealed. Heredity **90:** 2-3.

Arnholdt-Smitt B., Holzapfel B., Schillinger A. and Neumann K. H. 1991. Variable methylation and differential replication of genomic DNA in cultured carrot root explants during growth induction as influenced by hormonal treatments. Theoretical and Applied Genetics **82:** 283-288.

Bennetzen J. L. and Freeling M. 1993. Grasses as a single genetic system: Genome composition, co linearity and compatibility. Trends in Genetics **9:** 259-261.

Bennetzen J. L., San Miguel P., Chen M., Tikhonov A., Francki M. and Avramova Z. 1998. Grass genomes. Proceedings of the National Academic of Science of the U.S.A. **95:** 1975-1978.

Bennetzen J. L. 2000a. Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions. Plant Cell **12:** 1021-1029.

Binelli G., Giangranceschi L., Pe M. E., Taramino G., Busso C., Stenhouse J. and Ottaviano E. 1992. Similarity between maize and sorghum genomes as revealed by maize RFLP probes. Theoretical and Applied Genetics **84:** 10-16.

Blundy K. S., Cullis C. A. and Hepburn A. G. 1987. Ribosomal DNA methylation in a flax genotroph and a crown gall tumor. Plant Molecular Biology **8:** 217-225.

Brinkman F. S. 2002. Evidence that plant-like genes in Chlamidia species reflect an ancestral relationship between Chlamidiaceae, Cyanobacteria, and the chloroplast. Genome Research **12:** 1159-1167.

Brown J. R. and Doolittle W. F. 1999. Gene descent, duplication, and horizontal transfer in the evolution of glutamyl-and glutamyl-tRNA synthetases. Journal of Molecular Biology and Evolution **49:** 485-495.

Brown R. B. 2003. Ancient horizontal transfer. Nature Reviews **4:** 121-132.

Bushman F. 2002. Lateral DNA transfers (Cold Spring Harbor Laboratory Press, New York.

Burr B., Evola S. V., Burr F. A. and Beckman J. S. 1983. The application of restriction fragment length polymorphisms to plant breeding. In: Estlow J. K, Hollaender. A, (Eds). Genetic Engineering Principles and Methods, Volume 5. New York: Plenum, Pp 45-59.

Capy P. 1998. A plastic genome. Nature **396:** 522-523.

Cassels A. C. and Curry R. F. 2001. Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micro propagators and genetic engineers. Plant Cell, Tissue and Organs **64:** 145-157.

Chase M. W. and Palmer. J. D. 1989. Chloroplast DNA systemic of lilioid monocots: resources, feasibility, and an example from the Orchidaceae. American Journal of Botany **76:** 1720-1730.

Clark C. G and Roger A. J. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolityca*. Proceedings of the National Academic of Science of the U.S.A. **92:** 6518-6521.

Clayton W. D. 1972. Gramineae. In F. N Hepper: Flora of West Tropical Africa Volume 3, 2. Crown Agents, London.

Clayton W. D. and Renvoize S. A. 1986. Genera graminum: Grasses of the world. Kew Bulletin Additional Series XIII. London, Royal Botanical Gardens, Kew, Her Majesty's stationary office, London.

Clay O. and Bernadi G. 2001. The isochrones in human chromosomes 21 and 22. Biochemistry and Biophysical Research in Communication **285:** 855-856.

Copenhaver G. P. and Preuss D. 1999. Centromeres in the genomic era: unraveling paradoxes. Current Opinion in Plant Biology **2:** 104-108.

Cox A., Bennett M. D., and Dyer T. A. 1992: Use of the polymerase chain reaction to detect spacer size heterogeneity in plant 5S-rRNA gene cluster and to locate such clusters in wheat (*Triticum aestivum* l.). Theoretical and Applied Genetics **83:** 684-690.

Cullis C. A. 1979. Quantitative variation of ribosomal RNA genes in flax genotrops. Heredity **42:** 237-246.

Cullis C. A. 1999. Environmental Stress: a generator of adaptative variation? In: Lerner. H. R. (eds.), Plant Adaptations to stress Environments. Marcel Dekker, New York, Pp 149-160.

Cullis C. A. and Kunert K. J. 1999. Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. In: A. Altman, M. Ziv and S. Izhar (eds).

Plant Biotechnology and In Vitro Biology in the 21st century, Kluwer Academic Press, The Netherlands, Pp. 245-248.

Cullis C. A. and Kunert K. J. 2000. Isolation of tissue culture-induced polymorphism in bananas by representational difference analysis. Acta Horticulturae **530:** 421-428.

Davis G. L., McMullen M. D., Baysdorf C., Musket T., Grant D., Staebell M., Xua G., Polacco M., Kosterd L. and Meliahancock S. 1999. A maize map standard with sequenced core markers, grass genome reference points and 932 expressed sequence tagged sites (ESTs) in a 1736-locus map. Genetics **152:** 1137-1172.

Dean C. and Schmidt R. 1995. Plant genomes: a current description. Annual Review in Plant Physiology and Plant Molecular Biology **46**:395-418.

De Boer S. H. and Kelman A. 2001 II. Gram-negative B-2 Erwinia Soft Rot Group. In: Laboratory Guide for the identification of plant pathogenic bacteria. Third Edition. Editors: N. W Schaad, J. B Jones. and W Chun. APS Press, St. Paul Minnesota, USA. Pp. 56-71.

Devos K. M., Chao S., Li Q. Y., Simonetti M. C. and Gale M. D. 1994. Relationships between chromosomes 9 of maize and wheat homologous group 7 chromosomes. Genetics **138:** 1287-1292.

Devos K. M. and Gale M. D. 2000. Genome relationships: the grass model in current research. Plant Cell **12:** 637-646.

Doolittle W. F. 1998. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends in Genetics **14:** 307-311.

Ducusvsky J., Ramakrishna W., SanMiguel P., Busso C., Yan L., Bryan A., Shiloff L. and Bennetzen J. 2001. Comparative sequence analysis of collinear barley and rice-bacterial artificial chromosomes. Plant Physiology **125:** 1342-1353.

Dufour P., Grivet L., D'Hont A., Deu M., Paulet F., Bouet A., Lanaud D., Glaszmann J. C. and Hamon P. 1997. Construction of a composite sorghum genome map for comparison with sugarcane, a related complex polyploid. Theoretical and Applied Genetics **94:** 409-418.

Dvorák J., Zhang H. B., Kota R. S. and Lassner M. 1989. Organization and evolution of the 5S ribosomal RNA gene family in wheat and related species. Genome **32:** 1003-1016.

Fernandez I. Á., Aguillar J. F., Panero J. L. and Feliner G. N. 2001. A phylogenetic analysis of Doronicum (Asteraceae, Senecioneae) based on morphological, nuclear ribosomal (ITS), and chloroplast (trnL-F) evidence. Molecular Phylogenetics and Evolution **20:** 41-64.

Finnegan E. J., Brettell R. I. S. and Dennis E. S. 1993. The role of DNA methylation in the regulation of plant gene expression. DNA Methylation: Molecular Biology and Biological Significance. Birkhauser Verlag Bassel.

Flavell R. B., Bennett M. D., Smith J. B. and Smith J. B. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochemistry and. Genetics **12:** 257-269.

Flavell R. B., Rimpau, J. and Smith D. B. 1977. Repeated sequence relationship in four cereal genomes. Chromosoma **63:** 205-222.

Franklin A. E. and Cande W. Z. 1999. Nuclear organization and chromosome segregation. Plant Cell **11:** 523-534.

Freeling M. 2001. Grasses as a single system. Reassessment. Plant Physiology **125:** 1191-1197.

Frisch P. W. 2001. Phylogeny and biogeography of the flowering plant genus *Styrax* (Styracaceae) based on chloroplast DNA restriction sites and DNA sequences of the internal transcribed spacer region. Molecular Phylogenetics and Evolution **19:** 387-408.

Funk V. A. and Robinson H. 2001. A bully new genus from the Andes (Compositeae: Liabeae). Systemic in Botany **26:** 216-225.

Gale M. D. and Devos K. M. 1998. Comparative genetics in the grasses. Proceedings of the National Academic of Science of the U.S.A. **95:** 1971-1974.

Gale M., Moore G. and Devos K. 2001. Rice: the pivotal genome in cereal comparative genetics. Novartis Foundation Symposium **236:** 46-53.

Gawel N. J. and Jarret R. L. 1991. A modified CTAB DNA extraction procedure for Musa and Ipomoea. Plant Molecular Biology Reporter **9:** 262-266.

Haberer G. and Fisher T. C. 1996. Mapping of the nucleolus organizer on chromosome 4 in *Arabidopsis thaliana*. Molecular and General Genetics **250:** 123-128.

Heldt H. W. 1997. Plant Biochemistry and Molecular Biology. Oxford University Press New York.

Henry R. J. 1997. Practical Applications of Plant Molecular Biology. Chapman and Hall, London.

Jansen R. K. and Palmer J. D. 1987. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene localization, and characterization of a large inversion. Current Genetics **11:** 553-564.

Jeffreys A. J., Wilson V. and Theins S. L. 1985. Hyper-variable "minisatellite" regions in Human DNA. Nature **314:** 67-73

Kaeppler S. M., Kaeppler H. F. and Rhee Y. 2000. Epigenetic aspects of somaclonal variation in plants. Plants Molecular Biology **43:** 179-188.

Kanazin V., Ananiev E. and Blake T. 1993. The genetics of 5S rRNA encoding multigene families in barley. Genome **36:** 1023-1028.

Karp A., Seberg O. and Buiatti M. 1996. Molecular techniques in the assessment of Botanical Diversity. Annals of Botany **78:** 143- 149.

Karp A and Oleseberg M. B. 1997. Molecular technologies for biodiversity evaluation: opportunities and challenges. Natural Biotechnology **15:** 625-628.

Keller B. and Feuillet C. 2000. Colinearity and gene density in grass genomes. Trends in Plant Science **5:** 246-251.

Kellogg E. A. 1998. Relationships of cereal crops and other grasses. Proceedings of the National Academic of Science of the U.S.A. **97:** 9121-9126.

Ko H. L., Henry R. J., Graham G. C., Fox G. P., Chadbone D. A. and Haak I. C. 1994. Identification of cereals using polymerase chain reaction. Journal of Cereal Science **19:** 101–106.

Kondo N., Nikoh N., Ijichi N., Shimada M. and Fukatsu T. 2002. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. Proceedings of the National Academy of Science of the U.S.A. **99:** 14280-14285.

Kumar A. and Bennetzen J. 1999. Plant retrotransposons. Annual Review of Genetics **33:** 479-532.

Kunert K. J., Voster J., Bester C. and Cullis C. A. 2002. DNA microchip technology in the plant tissue culture industry. In K Rajaseran, T. J. Jacks and J. W. Finley (eds). Crop Biotechnology, Symposium Series N° 829. American Chemical Society, Washington D. C (in press).

Litti M. and Luty J. A. 1989. A hyper variable micro satellite revealed by in vitro amplification of dinucleotide repeats within the cardiac muscle actin gene. American Journal of Human Genetics **44:** 397-401.

Lisitsyn N. A., Lisitsyn N. M. and Wigler M. 1993. Cloning the differences between two complex genomes. Science **259:** 946-951.

Lo-Schiavo F., Pitto L., Giuliano G., Torti G., Nuti-Roncchi V., Marazziti D., Vergara r., Orselli S. and Terzi M. 1989. DNA methylation of embryogenic carrot cell cultures and its variations as caused by mutation, differentiation, hormones and hypomethylating drugs. Theoretical and Applied Genetics **77:** 325-331.

Long E. O. and David I. B. 1980. Repeated genes in eukaryotes. Annual Review of Biochemistry **49:** 727-764.

Lu Z., Kunnimalaiyaan M. and Nielsen B. L. 1996. Characterization of replication origins flanking the 23S rRNA gene in tobacco chloroplast DNA. Plant Molecular Biology **32:** 693-706.

Maheswaran M., Subudhi P. K., Nandi S., Xu J. C., Parco A., Yang D. C. and Huang N. 1997. Polymorphism, distribution, and segregation of AFLP markers in a doubled haploid rice population. Theoretical and Applied Genetics **94:** 39-45.

Marillonnet S. and Wessler S. R. 1998. Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics **150:** 1245-1256.

Martin W., Rujan T., Richly E., Hansen A., Cornelson S., Lins T., Leister D., Stoebe B., Hasegawa M. and Penny D. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proceedings of the National Academic of Science of the U.S.A. **99:** 12246-12251.

McCouch S. 1998. Toward a plant genomic initiative: Thoughts on the value of cross-species and cross-genera comparisons in the grasses. Proceedings of the National Academic of Science of the U.S.A. **95:** 1983-1985.

Moore G., Devos K. M., Wang Z. and Gale M. D. 1991. The conservation of dinucleotide micro satellite among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics **10:** 654-660.

Moore G., Devos K. M., Wang Z. and Gale M. D. 1995. Grasses, line up and form a circle. Current Biology **5:** 737-739.

Moynihan J. and Watson L. E. 2001. Phylogeography, generic allies, and nomenclature of Caribbean endemic genus Neolaugeria (Rubiaceae) based on internal transcribed spacer sequences. International Journal of Plant Sciences **162:** 393-401.

Murray M. G. and Thompson W. F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acid Research **8:** 4321-4325.

Nagamura Y., Takana T., Nozawa H., Kaidai H., Kasuga S. and Sasaki T. 1998. Syntenic regions between rice and sorghum genomes. National Grassland Research Institute Report (Japan) **9:** 97-103.

Nagl W. and Rucker W. 1976. Effect of phytohormones on thermal denaturation profiles of Cymbidium DNA: Indication of differential DNA replication. Nucleic Acids Research **3:** 2033-2039.

Nukretenko A., Makova K. D. and Baker R. J. 2000. Isolation of a binary species-specific PCR- based markers and their value for diagnostic applications. Gene **249:** 47-51.

Nierras C. R., Liebman S. W. and Warner J. R. 1997. Does Saccarhomyces need an organized nucleolus? Chromosoma **105:** 4007-4021.

O'Hanlon P. C., Peakall R. and Briese D. T. 2000. A review of new PCR-based genetic markers and their utility to weed ecology. Weed Research **40:** 239-254.

Pardue M. L., Danilevskaya O. N., Traverse K. L. and Lowenhaupt K. 1997. Evolution links between telomeres and transposable elements. Genetica **100:** 73-84.

Pluhar S. A., Erickson L. and Pauls K. P. 2001. Effects of tissue culture on a highly repetitive DNA sequence (E180 satellite) in *Medigo sativa*. Plant Cell Tissue and Organ Culture **67:** 195-199.

Pruitt R E. and Meyerowitz E. M. 1986. Characterization of the genome of *Arabidopsis thaliana*. Journal of Molecular Biology **187:** 169-328.

Rainey P. B. 1999. Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. Environmental Microbiology **1:** 243-257.

Roger A. J., Clark C. G. and Doolittle W. F. 1996. A possible mitochondria gene in the early branching amitochondriate protists *Trichomonas vaginalis*. Proceedings of the National Academy of Science of the U.S.A. **93:** 14618-14622.

Royo J., Gimez E. and Hueros G. 2000. CMP-KDO synthetase: a plant gene borrowed from gram-negative eubacteria. Trends in Genetics **16:** 432-433.

Saghai-Maroof M. A., Soliman K. M., Jorgesen R. A. and Allard R. W. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal

location, and population dynamics. Proceedings of the National Academy of Science of the U.S.A. **81:** 8014-8018.

Sambrook J., Fritsch E. F. and Maniatis T. 1989. Molecular cloning: a laboratory Manual, second edition. Cold Spring Harbor Laboratory Press, New York, U.S.A.

Sandhu D. and Gill K. S. 2002. Gene-containing regions of wheat and other grass genomes. Plant Physiology Online **128:** 803-811.

Sanger F., Nicklen S. and Coulson A. R. 1977. DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science of the U.S.A. **74:** 5463-5467.

San Miguel P., Tikhonov A., Jin Y K., Motchoulskala N., Zakharov D., Melake-Berhan A., Springer P. S., Edwards K. J., Avromova Z. and Bennetzen J. L. 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765-768.

Stacey G., Sanjuan J., Luka S., Dockendorff T. and Carlson R. W. 1995. Signal exchange in the *Bradyrhizobium soycan* symbiosis. Soil Biology and Biochemistry **27:** 473-483.

Schmidt T. and Heslop-Harrisson J. S. 1998. Genomes, genes and junk: the large scale organization of plant chromosomes. Trends in Plant Science **3:** 195-199.

Schmitt F., Oakeley E. J. and Jost J. P. 1997. Antibiotics induce genome-wide hypermethylation in cultured *Nicotiana tabaccum* plants. Journal of Biological Chemistry **272:** 1534-1540.

Schuster W. and Brennicke A. 1994. The plant mitochondria genome: physical structure, information content, and gene migration to the nucleus. Annual Review in Plant Physiology and Plant Molecular Biology **45:** 61-78.

Scribner K. T. and Pearce J. M. 2000. Microsatellites: Evolutionary and methodological background and empirical applications at an individual, population and phylogenetic levels. In Molecular Methods in Ecology, Baker A (eds), Blackwell London Pp. 235-271.

Senior M. L. and Heun M. 1993. Mapping maize micro satellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome **36:** 884-889.

Shantz H. L. 1954. The place of grasslands in the earth's cover of the vegetation. Ecology **35:**143-145.

Shirasu K., Schulman A. H., Lahaye T. and Schulze-Lefert P. 2000. A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Research **10:** 908-915.

Sperveslage J., Stack Brandt E., Lembke F. W. and Koch C. 1996. Detection of bacterial contamination, including *Bacillus* spores, in dry growth media and milk by identification of their 16S rDNA by polymerase chain reaction. Journal of Microbiology Methods **26:** 219-224.

Song K. M., Osborn T. C. and Williams P. H. 1988a. Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs). 1. Genome evolution of diploid and amphidiploids species. Theoretical and Applied Genetics **75:** 784-794.

Song K. M., Osborn T. C. and Williams P. H. 1988b. Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 2. Preliminary analysis of subspecies within *Brassica rapa* and *Brassica oleracea*. Theoretical and Applied Genetics **76:** 593-600.

Song K. M., Osborn T. C. and Williams P. H. 1990. Brassica taxonomy based on nuclear restriction fragment length polymorphisms (RFLPs) 3.Genome relationships in *Brassica*

and related genera and the origin of *Brassica. rapa* and *Brassica oleracea*. Theoretical and Applied Genetics **79:** 503-517.

Sugiura M. 1992. The chloroplast genome. Plant Molecular Biology **19:** 149-168.

Sun X., Walhlstrom J. and Karpen G. 1997. Molecular structure of a functional Drosophila centromere. Cell **91:** 1007-1019.

Tao Y. Z., Jordan D. R., Henzell R. G. and Mcintyre C. L. 1998. Construction of a genetic map in a sorghum recombinant inbred line-using probes from different sources and its comparison with other sorghum maps. Australian Journal of Agricultural Research **49:** 729-736.

Taramino G. and Tingey S. 1996. Simple sequence repeats for germplasm analysis and mapping in maize. Genome **39:** 277-287.

Tarchini R., Biddle P., Wineland R., Tingey S. and Rafalski A. 2000. The complete sequence of 340kb of DNA around the rice Adh-1-adh2 region reveals interrupted co linearity with maize chromosome 4. Plant Cell **12:** 381-391.

Ting I. P. 1982. Plant Physiology. Addison-Wesley Publishing Company, California.

Ushijima T., Morimura K., Hosoya Y., Okonogi H., Tatematsu M., Sugimura T. and Nagao M. 1997. Establishment of methylation sensitive by representational difference analysis and isolation of hypo-and hyper-methylated genomic fragments in mouse liver tumors. Proceedings of the National Academic of Science of the U.S.A. **94:** 2284-2289.

Van der Vyver C. 2003. Stress-induced genome changes in plants. PHD Thesis. Faculty of Natural and Agricultural Sciences. University of Pretoria.

Van Deynze A. E., Dubcovsky J., Gill K. S., Nelson J. C., Sorrells M. E., Dvorak J., Gill B. S., Lagudah E. S., McCouch S. R. and Appels R. 1995a. Molecular genetic maps for group 1 chromosome of Triticeae species and their relation to chromosomes in rice and oat. Genome **38:** 45-59.

Vicient C. M., Jaaskelainen M. J., Kalendar R. and Schulman A. H. 1999. Retrotransposons BARE-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell **11:** 1769-1784.

Vodkin M. and Kattermann F. R. H. 1971. Divergence of ribosomal RNA sequences within angiospermae. Genetics **69:** 435-451.

Vos P., Hogers R. and Bleeker M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research **23:** 4407-4414.

Voster B. J., Kunert K. J. and Cullis C. A. 2002. Use representational difference analysis for the characterization of sequences between date palm varieties. Plant Cell Report **21:** 271-275.

Voystas D. F. 1996. retro elements in genome organization. Science **274:** 737-738.

Walbot V. and Cullis C. A. 1983. The plasticity of the plant genome-is it a requirement for success? Plant Molecular Biology Report **1:** 3-11.

Walbot V. and Cullis C. A. 1985. Rapid genomic changes in higher plants. Annual Review of Plant Physiology **36:** 367-396.

Watson L. and Dallwitz M. L. 1991. Grass genera of the world: an INTKEY package for an automated identification and information retrieval of data including synonyms, morphology, anatomy, physiology, cytology, classification, pathogens, world and local distribution, and references. 2nd edition. Flora Online 22.

Watson L. and Dallwitz M. J.1992. The grass genera of the world. 2nd edition, 1081pp. (CAB international: Wallingford).

Webb D. M. and Knapp S. J. 1990. DNA extraction from a previously recalcitrant plant genus. Plant Molecular Biology **8:** 180-185.

Whitkus R., Doebley J. and Lee M. 1992. Comparative genome mapping of sorghum and maize. Genetics **132:** 1119-1130.

Wicker S. N., Albar L., Feuillet C., Schlagenhauf E. and Keller B. 2001. Analysis of a contiguous 211kb sequence in diploid wheat (*Triticum monoccocum* L.) reveals multiple mechanisms of genome evolution. Plant Journal **26:** 307-316.

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

Wright D. A., Ke N., Smalle J., Haugr B. M., Goodman H. M. and Voytas D. F. 1996. Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics **142:** 569-578.

Xie Z. W., Ge S. and Hong D. Y. 1999. Preparation of silica gel dried mini-amount of leaves of *Oryza rufipogon* Griff for RAPD study and total DNA bank construction. Acta Botany **41:** 884-890.

Zheng K. L., Castiglione S., Biasini M. G., Biroli A., Morandi C. and Sala F. 1987. Nuclear DNA amplification in cultured cells of *Oryza sativa* L. Theoretical and Applied Genetics **74:** 65-70.

SOMMAIRE

Le génome des graminées diffère en taille; degré de ploïdy; et nombre de chromosome. Depuis le siècle dernier, les methodes d´ identification et de charactérisation des génomes ont dramaticallement changées dans la reproduction des plantes. Elles sont passées des croisements simples aux croisements retour jusqu'aux techniques moleculaires actuelles. L'analyse des différences représentatives de deux génomes qui en est une des techniques moleculaires, a été appliquée sur l'avoine sauvage collectée á différents endroits en Afrique du Sud pour isoler une unique fraction de son génome. Cinq séries d'hybridation soustractive ont été appliquée. Après la deuxième série, un produit différentiel obtenu etait homologue à une séquence connue de 'rétrotransposon'du maïs et aussi à une région chromosomique du riz. Ce produit de soustraction n'était pas unique à une seule des échantillons testées. En plus, ce produit avait aussi un nombre élevé de copies dans le genome de la plante. La troisième, quatrième et cinquième tour d'hybridation soustractive ont été aussi appliquées. La cinquième étape d'hybridation soustractive a été appliquée sur un quatrième produit ayant subi une digestion enzymatique au *Mse*I reconnue active pour couper l'ADN repetitif. Ce cinquième produit de soustraction analysée etait homologue a une séquence de l'ADN bactérienne, ainsi qu'a une séquence partielle d'ADN de riz et de mil. L'homologie du produit de différence génomique à une séquence d'ADN bactérienne nous a fait penser à la contamination de l'ADN de départ par une bactérie endophyte de la plante. Pour s'assurer de la pureté de notre matériel de départ, l'ADN isoler des plantes a été utilisée pour amplifier un fragment de 595 bp caractéristique de la région 16S de l'ADN ribosomal du *Bacillus subtilis*. Cette réaction a été négative. De même l'identification du *Bacillus subtilis* comme endophyte spécific de la même plante a donné plutôt lieu à d'autres espèces bactériennes.