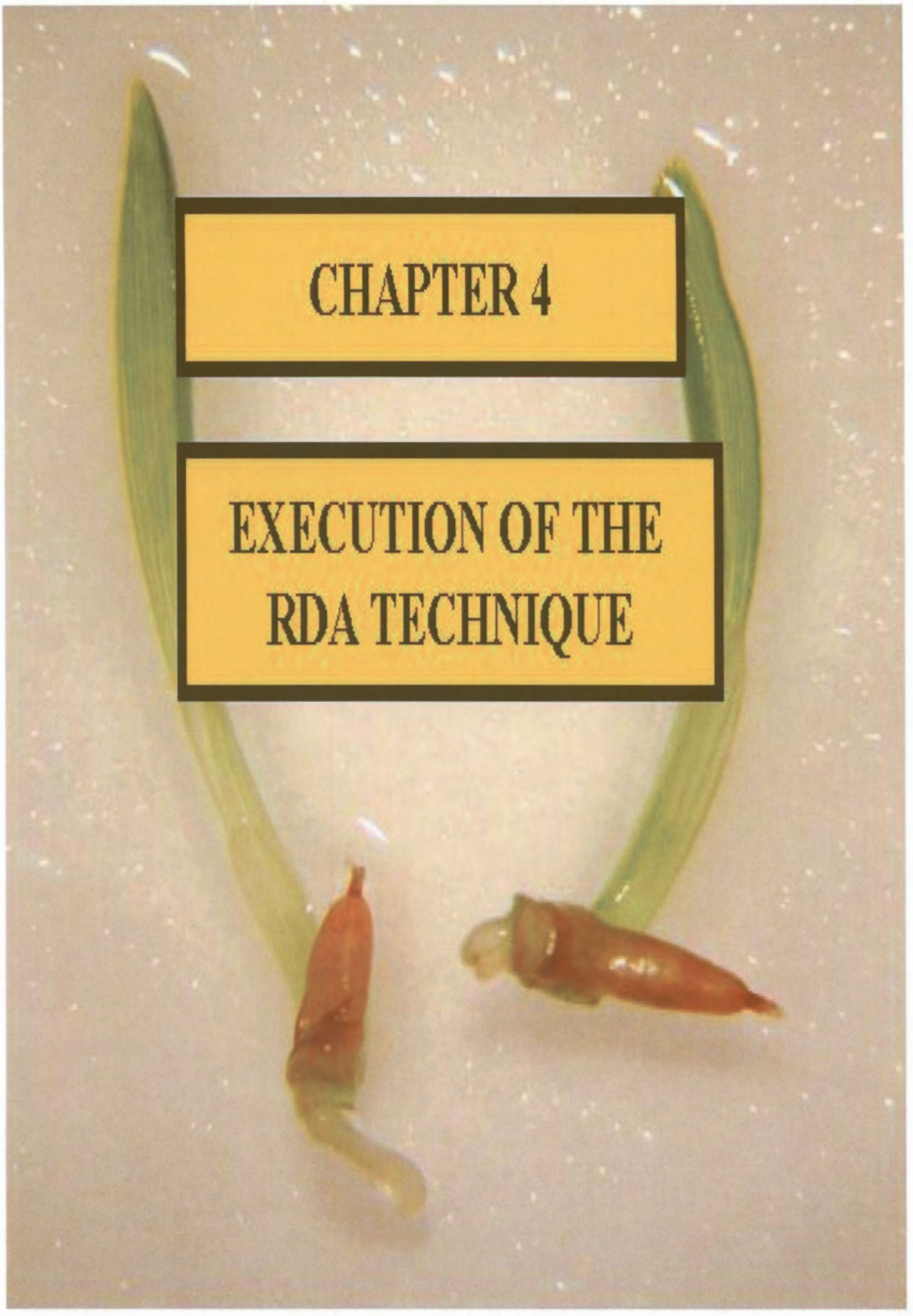

CHAPTER 4
EXECUTION OF THE RDA TECHNIQUE

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.

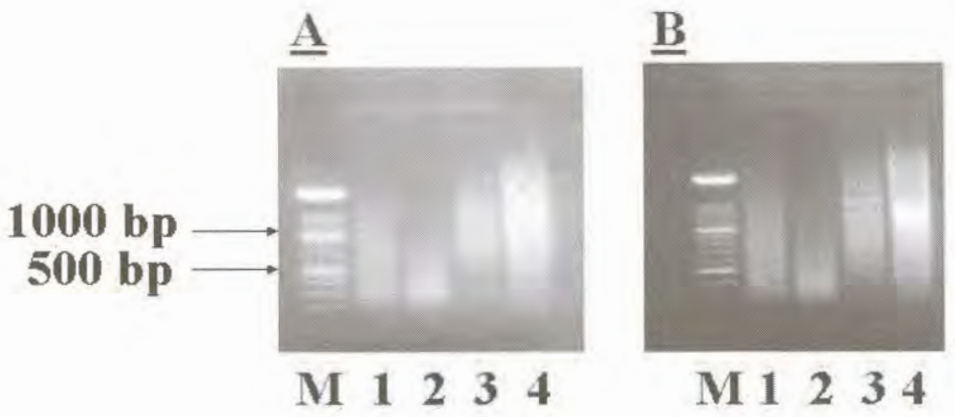


Figure 4.1: Digestion of genomic DNA isolated from *Monocymbium* collected from three different areas with restriction enzyme *Hin*dIII (A) and amplicons obtained from genomic DNA after adaptor ligation and amplification with adaptor pair RHind 12 and 24 (B), visualized on an agarose gel after staining 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).

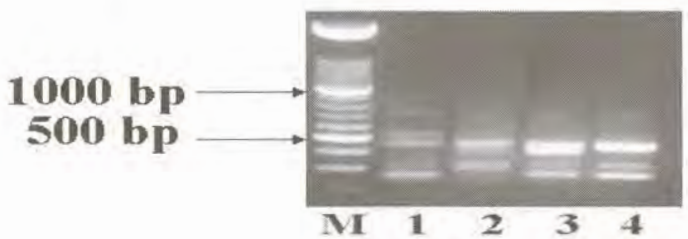


Figure 4.2: First subtraction products (SP1) of *Monocymbium ceresiiforme* genomic DNA representations visualized on an agarose gel stained with ethidium bromide. Lane M represents a 100 bp DNA ladder (Roche, Switzerland). Lanes 1 and 2 represent the subtraction products resulting from the subtraction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

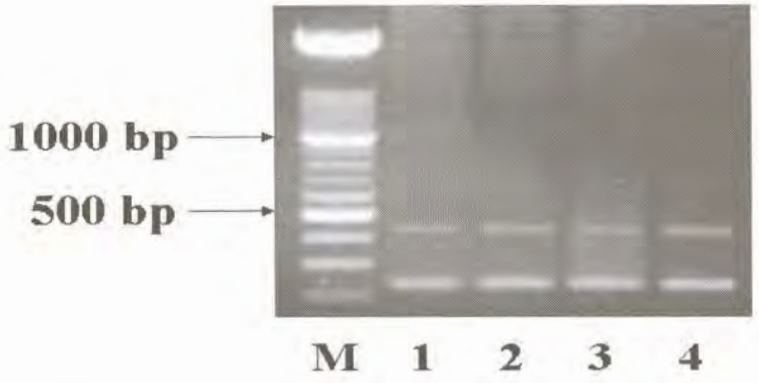


Figure 4.3: Second subtraction products (SP2) of *Monocymbium ceresiiforme* genomic DNA representations at a 150:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the resulting products from the subtraction of DG grass representation and SG 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.

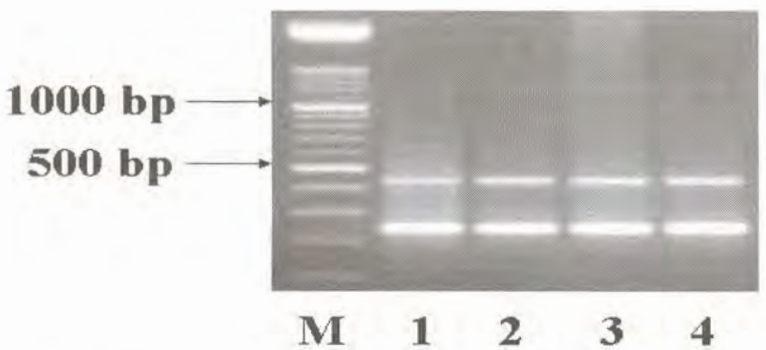


Figure 4.4: Third subtraction products (SP3) of *Monocymbium ceresiiforme* genomic DNA representations at a 1500:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the resulting products from the subtraction of DG grass representation and SG 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.

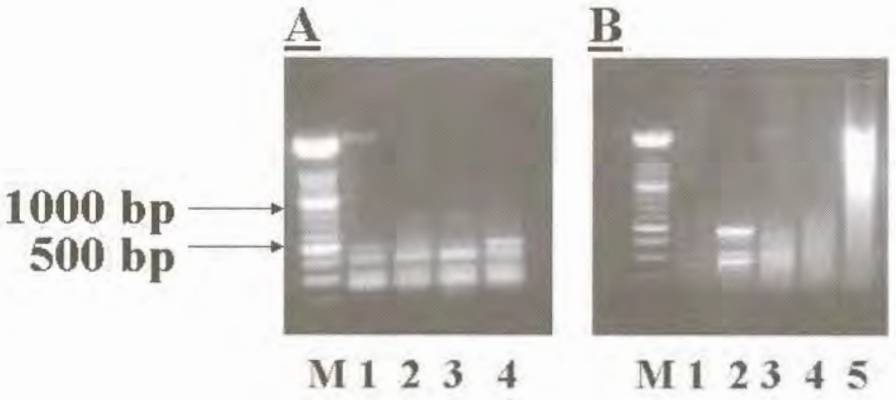


Figure 4.5 Fourth round subtraction products (SP4) of *Monocymbium ceresiiforme* genomic DNA representations at a 200000:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide, where the third subtraction products were either digested with *Mse*I (A), or were undigested (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.

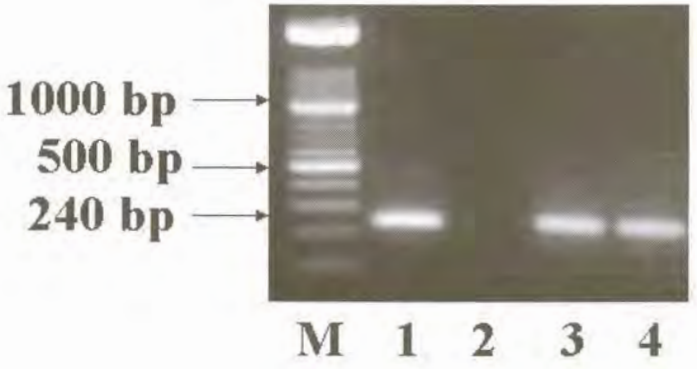
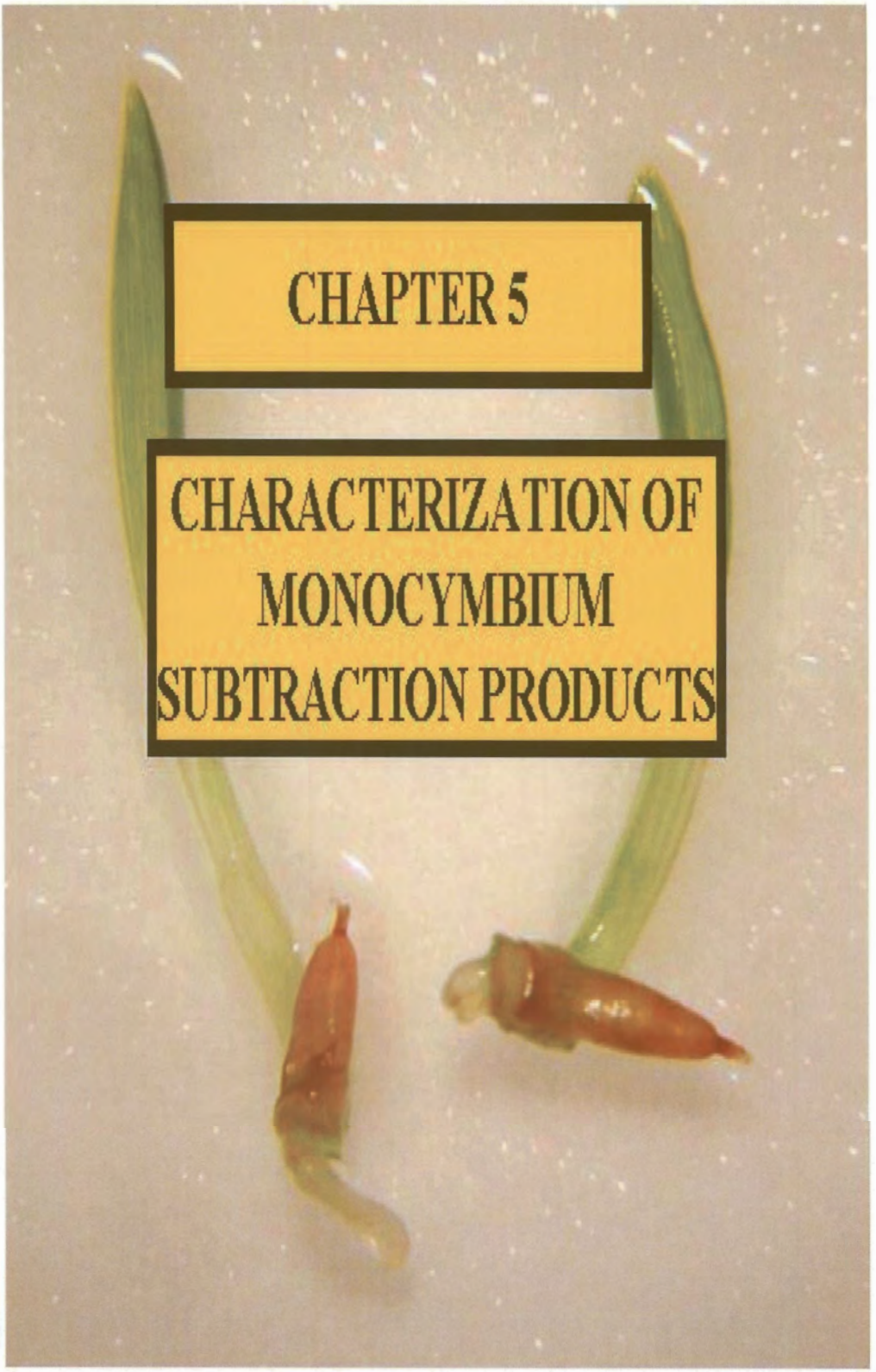


Figure 4.6 Fifth round subtraction products (SP5) of *Monocymbium ceresiiforme* genomic DNA representations at a 2000000:1 driver to tester ratio, visualized on an agarose gel stained with ethidium bromide. Lanes 1 and 2 represent the subtraction of the DG grass representation and the SG grass representation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

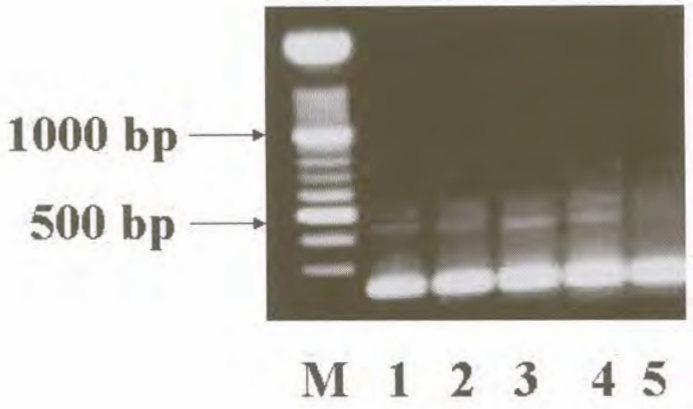


Figure 5.1: Screening of plasmid p*MosBlue* for insertion of subtraction products obtained after two rounds of subtractive hybridization with *Monocymbium ceresiiforme* representations. Lanes 1-5 represent respectively fragments S3cl2, S3cl12, S3cl18, S4cl27, and S4cl39. Lane M represents a 100 bp DNA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

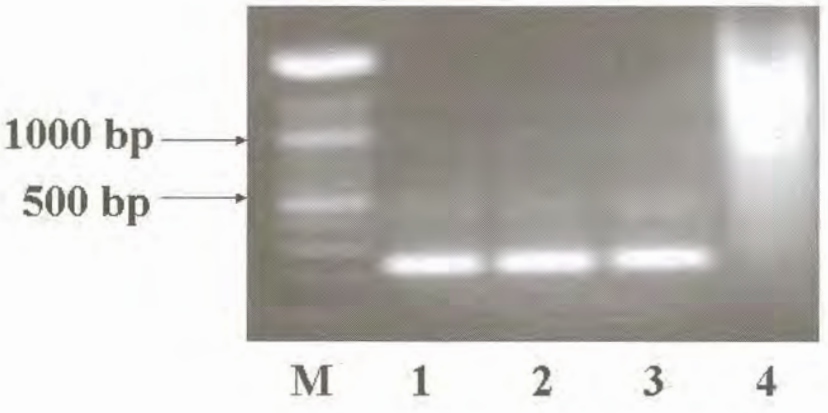


Figure 5.4: PCR amplification of grass genomic DNA at 55°C with primers S3cl2L and S3cl2R visualized on an agarose gel stained with ethidium bromide. Lanes 1, 2, and 3 represent the amplification products from SG, DG, and HG grass genomic DNA, respectively. Lane 4 represents 2 µg of herring sperm DNA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

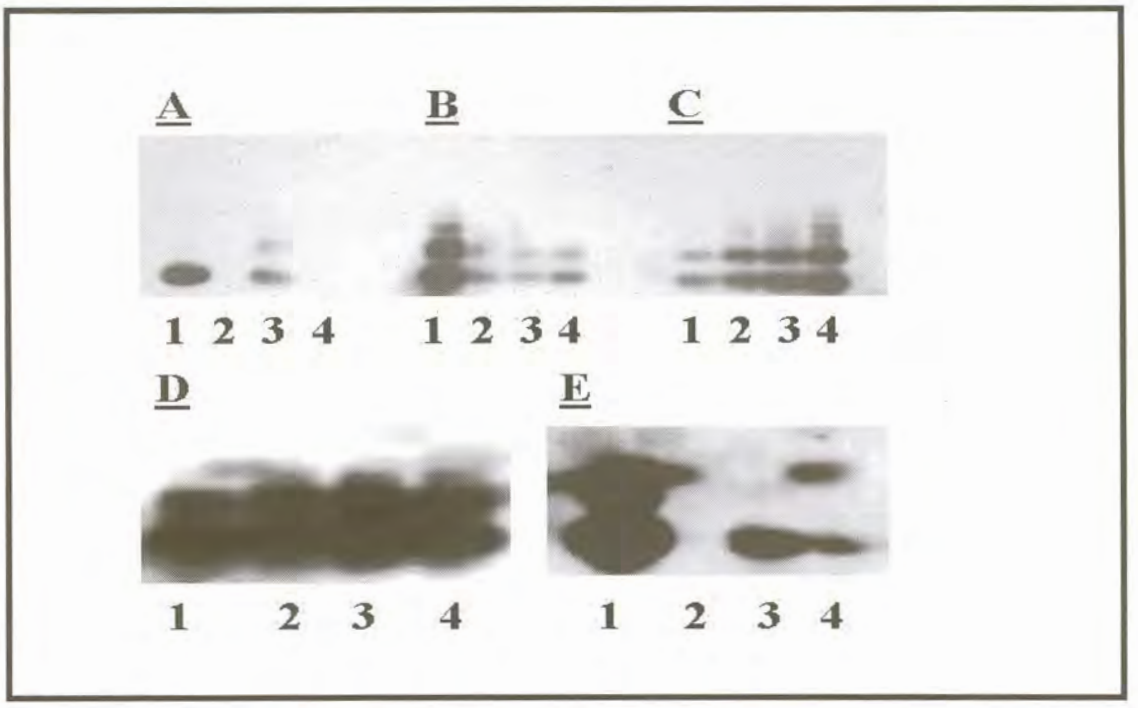


Figure 5.8: Hybridization of first subtraction products (A); second subtraction products (B); third subtraction products (C); fourth subtraction products (D) and fifth subtraction products (E) with labeled DP510 fragment. Lanes 1 and 2 represent the subtraction of SG and DG grass representations, where 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).

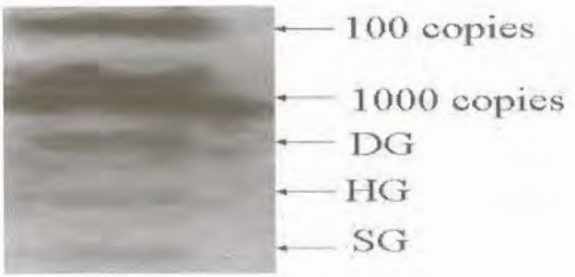


Figure 5.9: Copy number determination in Southern-blotting of EcoRI-digested genomic DNA from DG, HG, and SG grass with labeled DP510. DG, HG, and SG represent genomic DNA from the three different types of grass. Reference bands indicate the intensity of 100 and 1000 copies of DP510 (top two rows).

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)

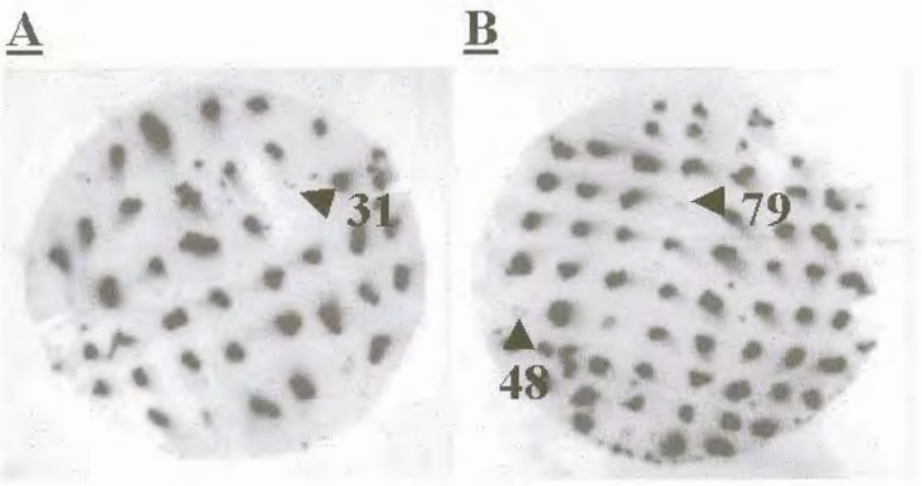


Figure 5.11: Colony blot of *E. coli* colonies carrying cloned DP510 amplification products from genomic DNA of SG grass and using labeled DP510 as a probe for hybridization. Colonies Bh31, Bh48, and Bh79 gave the lowest hybridization signal, which were used for further characterization.

```
                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.

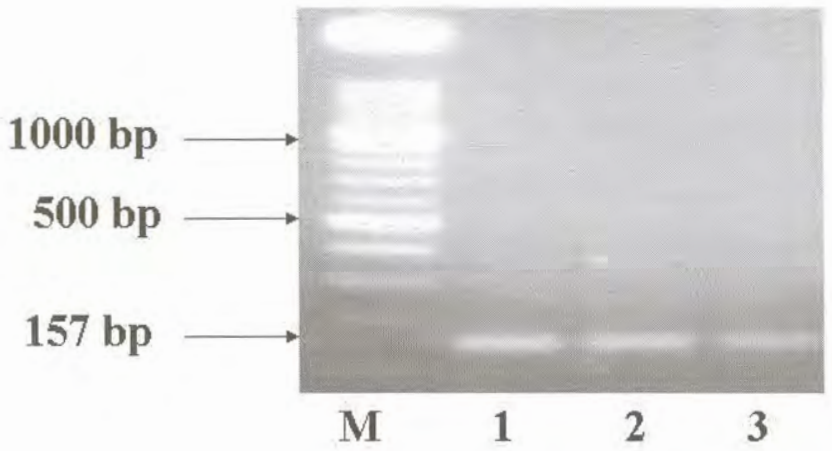


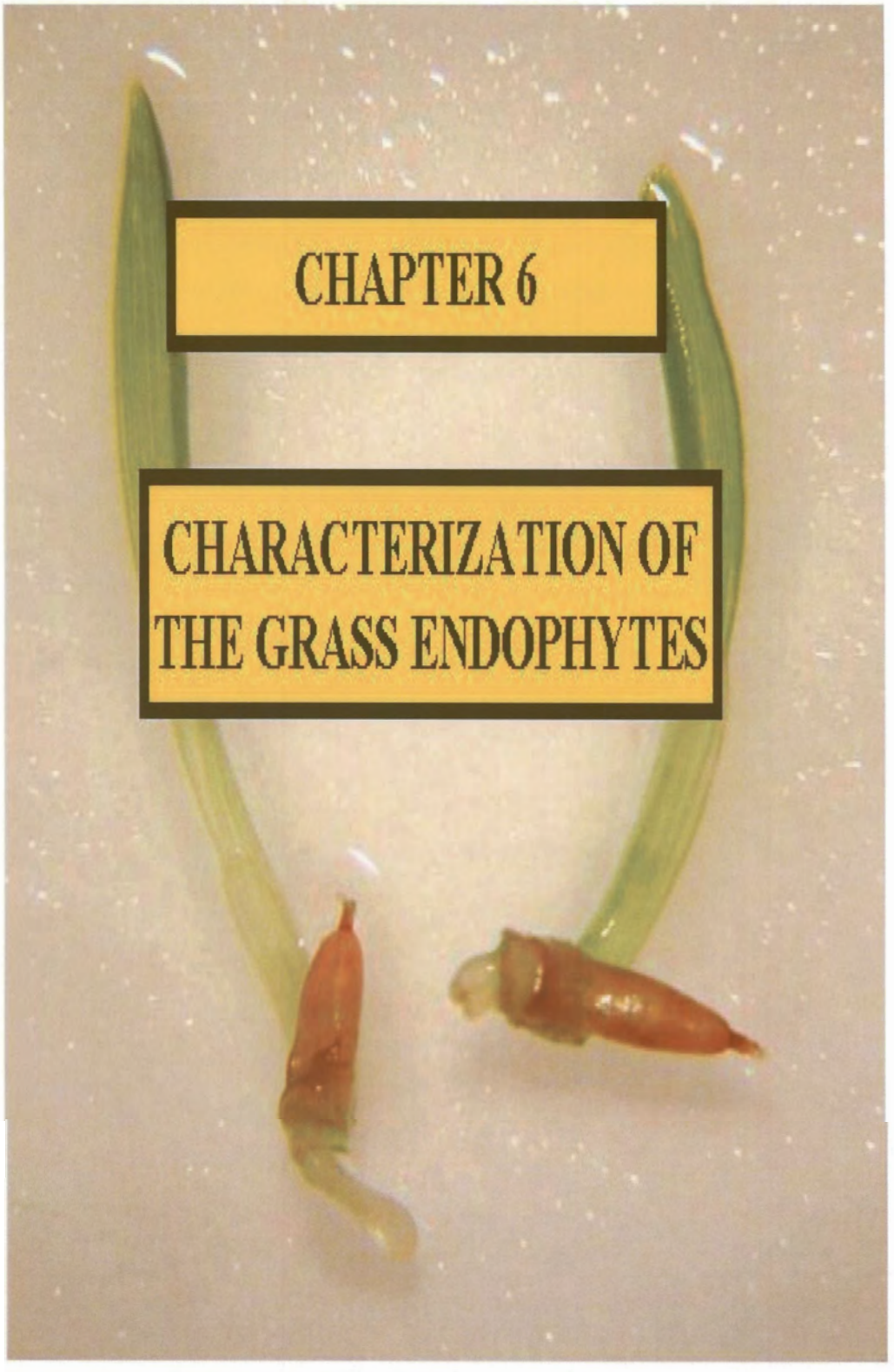
Figure 5.14: Amplification products of grass genomic DNA amplified with primers designed from internal part of DP510, visualized on an agarose gel stained with ethidium bromide. Lane 1 represents the amplified fragment from DG grass genomic DNA; lane 2 the amplified fragment from SG grass genomic DNA 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.

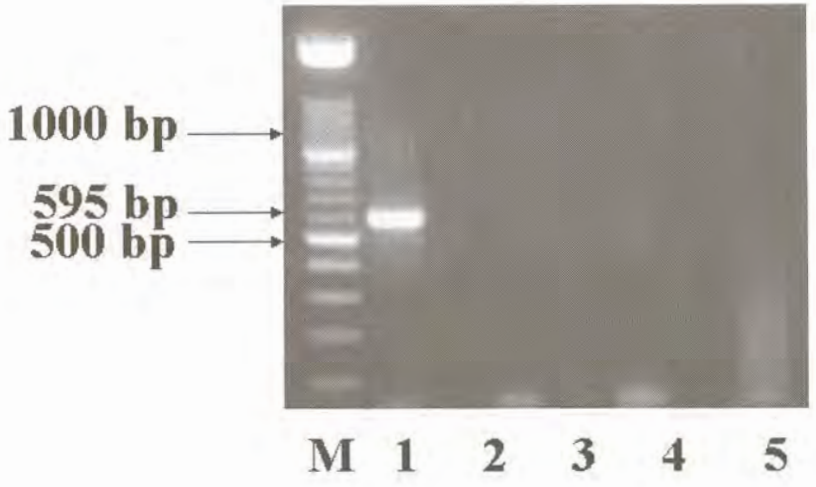


Figure 6.1: Amplification of *Bacillus subtilis* 16S rRNA region with specific primers designed for the 16S rRNA region and products visualized on a 1.5% agarose gel stained with ethidium bromide. Lane 1 represents an amplification product from *Bacillus subtilis* DNA. Lanes 2, 3, 4 represent isolated 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.

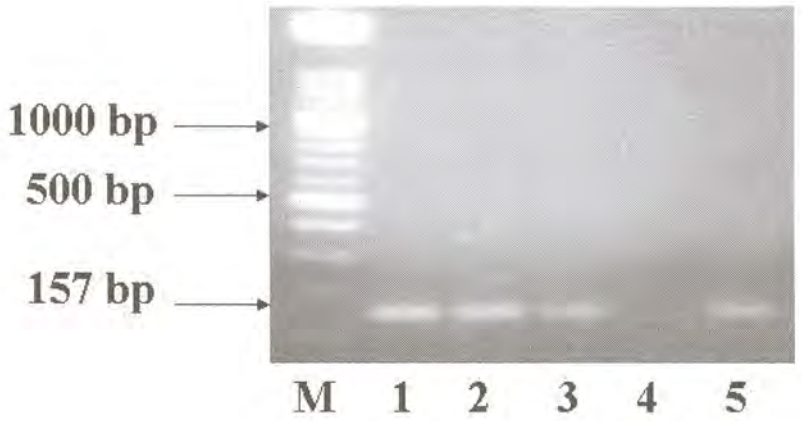
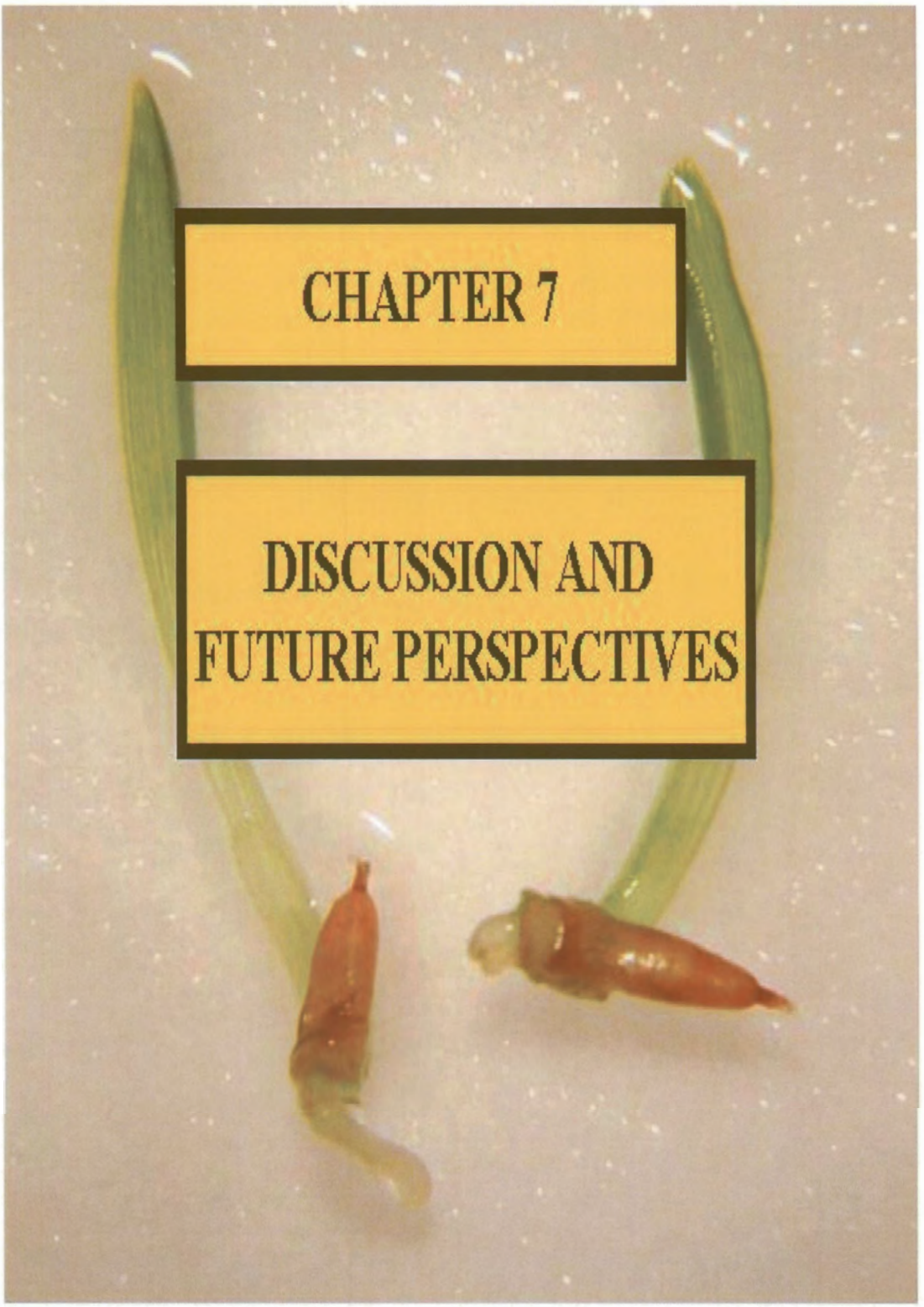


Figure 6.2: Amplification of DP510 fragment covering the region with homology to the Bacillus DNA from grass genomic DNAs and the *Bacillus subtilis* DNA visualized on a 1.5% agarose gel stained with ethidium bromide. Lane 1 represents genomic DNA from DG grass, lane 2 HG grass and lane 3 SG grass. Lane 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

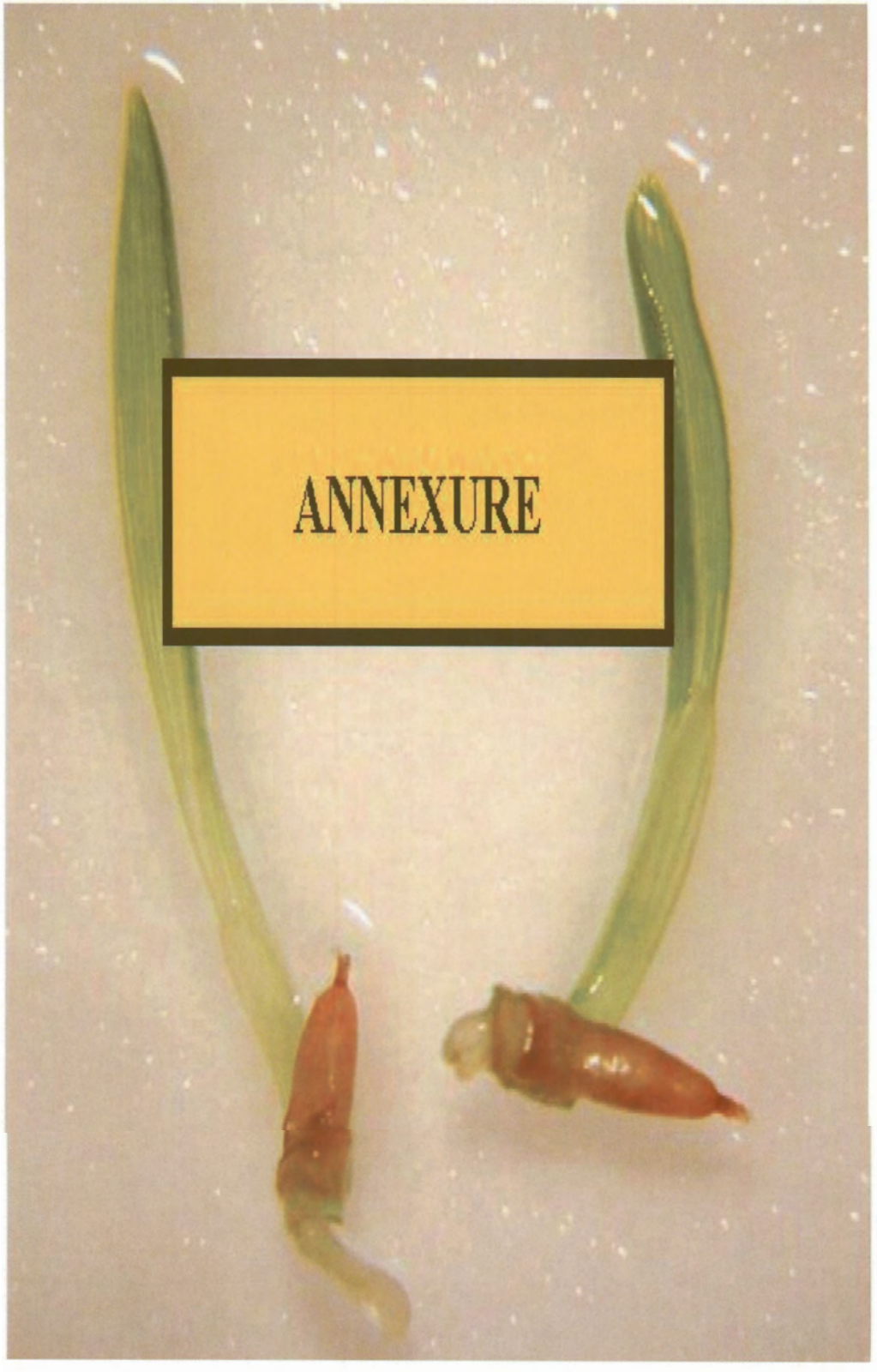
(Archebal and Keeling, 2003). Further evidence for the early integration of bacterial genes into the eukaryotic genome originates from studies of proteins from simple protists such as *Giardia intestinalis* and *Trichomonas vaginalis*, lacking mitochondria. However, molecular studies showed that these 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'