# 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.