The application of representational difference analysis and plant differentiation

By

BJ VORSTER

Dissertation submitted in partial fulfillment of the requirements 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. K.J. KUNERT
CO-SUPERVISOR: PROF. A-M. OBERHOLSTER

November 2003

DECLARATION

I, the undersigned, hereby declare that the thesis submitted herewith for the degree Magister Scientiae to the University of Pretoria, contains my own independent work and has not been submitted for any other degree at any other university.

BJ Vorster

November 2003

TABLE OF CONTENTS

ABSTRACT

The methods used in analyzing plant genomes have changed dramatically in the last century, from extensive breeding experiments using crosses and backcrosses to techniques that identify molecular differences in the DNA sequence of different genomes and studying these differences much more intensely. Representational Difference Analysis (RDA) is one such molecular technique. Originally developed in cancer research, it has recently also been applied to study plant genomes. RDA uses subtractive hybridization to isolate genomic regions that differ between two genomes, allowing the isolation of unique genetic sequences from one genome. Since RDA has only been used on a very limited scale in plant genome analysis, the aim of this project was to test the potential of this technique on the level of plant identification and the application of the technique on large and small plant genomes. Eleven subtraction products were isolated from date palm using the restriction enzyme *Bam*HI for genome differentiation and production of genome representations. These subtraction products were shown to be dispersed repetitive sequences. It was show that out of these subtraction products, Dp41, was much more abundant in the genome of one date palm cultivar compared to the other cultivar used in this study. Analysis of this element in a number of date palm plants indicates that Dp41 possibly represents a hot-spot for stress-induced mutations. In addition three additional subtraction products were isolated from date palm using the methylation sensitive restriction enzyme *Hpa*II for the creation of genome representations. Two of these showed homology to rRNA genes. In a third application subtraction products were isolated from the very large genome of *Pinus strobus*. Due to the size of the pine genome the complexity and number of the obtained subtraction products could not be used for analysis. A difference in genome methylation between the different lines could however be demonstrated.

OPSOMMING

Die analise van plantgenome het drasties gedurende die afgelope eeu verander. Uitgebreide telingseksperimente, wat gebruik maak van kruising en terugkruisings word aangevul met tegnieke wat molekulêre verskille in die nukleotiedvolgorde van verskeie genome identifiseer en in diepte ondersoek. Verteenwoordigende verskilanaliese (RDA) is een van hierdie tegnieke. Alhoewel oorspronklik ontwikkel vir kankernavorsing, word dit suksesvol in die bestudeering van plantgenome aangewend. RDA isoleer unieke geenvolgordes van een genoom deur die verskille tussen twee genome te identifiseer deur middel van substraksie-hibidisasie. Die doel van hierdie projek is om die potensiaal van RDA in die identifiseering van plante en die toepassing daarvan op groot en klein plantgenome te toets, aangesien die gebruik van RDA op die huidige oomblik baie beperk is in plantnavorsing. Elf substraksie-produkte is uit datelpalms geïsoleer deur die snydingsensiem *Bam*HI te gebruik vir genoom-differensiasie. Hierdie substraksie-produkte is verspreide herhalende nukleotiedvolgordes waarvan een, Dp41, baie meer algemeen voorgekom het in die genoom van een van die dadel-kultivars wat bestudeer is, in vergelyking met die ander kultivar. Verdere analise van hierdie element in 'n aantal dadelpalms het getoon dat Dp41 'n moontelike teikengebied vir stress-geïnduseerde mutasies mag wees. Die gebruik van die snydingsensiem *Hpa*II, wat sensetief vir DNS-metilering is, het nog drie substraksie-produkte opgelewer, waarvan twee ooreenkomste met rRNS gene getoon het. Ten derde is daar ook geslaag om substraksie-produkte uit die baie groot genoom van *Pinus strobus* te isoleer. As gevolg van die grootte van die genoom was die hoeveelheid en kompleksiteit van die subtraksie-produkte te groot om geanaliseer te word. 'n Verskil in genoom-metileering tussen die versikilinde lyne kon egter gedemonstreer word.

Research Objectives

There are many molecular techniques available that can be used to do genome analysis, each with their own advantages and disadvantages. In this MSc study one such technique was evaluated for its suitability in plant genome analysis. Representational difference analysis (RDA), which was developed in human cancer research to isolate differences between cancerous and non-cancerous cells, have recently been applied to study plant genomes, but very little is still known about the potential of RDA in plant genome analysis. The research objectives were therefore (1) to evaluate its potential in plant identification on the variety level by using methylation insensitive and sensitive restriction enzymes for the generation of genome representations, and (2) the evaluation of the RDA technique when applied to large genomes such as the pine genome. The experiments carried out in this MSc project focused in particular on (1) the isolation and characterization of possible genomic variations between closely related date palm varieties (2) the evaluation of the usefulness of these isolated subtraction products to be used as markers for cultivar identification and differentiation, (3) the isolation and characterization of differences due to methylation changes in the date palm genome, and (4) the ability of the RDA technique to subtract large genomes with a high complexity.

Thesis Composition

The following theses composes of seven chapters and discuss the application of representational difference analysis (RDA) to genome analysis in plants. Since (RDA) is a technique developed to isolate differences between cancerous and non-cancerous cells in humans, the hypothesis was that it could be used to isolate differences between closely related cultivars or cell-lines to isolate differences that could be useful in the creation of molecular markers. In this study RDA was used to isolate subtraction products from the genomes of two date palm cultivars as well as from the genome of white pine. It is also used to isolate subtraction products using methylation sensitive and non-sensitive restriction enzymes. A breakdown of the various chapters is provided below.

Chapter 1 of this thesis presents an introduction into molecular biology regarding the theories and philosophy forming the basis of this field of study and how it developed. It also gives an overview as to how plant genomes are organized and the processes involved in creating genome variation and evolution. It also deals with some of the most popular techniques used today in plant genome analysis, random amplified polymorphic DNAs (RAPD), and discusses the advantages and disadvantages of each technique focusing to a larger extent on representational difference analysis. **Chapter 2** presents the results obtained using RAPD to differentiate between two date palm cultivars, *'Medjool'* and *'Barhee'*. This includes two different methods for genomic DNA isolation, RAPD analysis, the cloning of a polymorphic DNA fragment, its sequence and the design of Sequence characterized amplified region (SCAR) primers for this fragment. **Chapter 3** focuses on the application of RDA on date palm. This chapter outlines the results obtained from application of this subtractive technique executed on genomic DNA digested with restriction enzymes including the isolation and cloning of subtraction products. **Chapter 4** focusus on the characterization of the RDA subtraction products using bioinformatic tools to conduct sequence homology searches and to do sequence alignments. The results obtained from the detailed characterization of one subtraction product DP41 is also presented in this chapter. **Chapter 5** deals with genome methylation. It consists of two parts, firstly detecting genome methylation in date palm using RDA, and secondly using RDA to detect genome methylation in the much larger and complex genome of *Pinus strobus*. It also includes a comparison of genome

methylation in *P. strobus* embryogenic cultures grown on tissue culture medium containing either a low or high concentration of 2,4-D, a synthetic plant growth regulator known to influence genome methylation during tissue culture. In **Chapter 6** the results of the various experiments are discussed and **Chapter 7** the **Achievements and conclusion** outlines the scientific achievements made by this thesis and also the failures are mentioned and discussed, and an overview about possible future research activities is provided. **References** lists all the literature used and sited in this thesis and finally, in the **Annex** the methods, composition of the tissue culture media as well as plasmid maps, and general buffers and solutions used in the various protocols, and lastly an article that was published from this results.

Acknowledgements

ABBREVIATIONS, DEFINITIONS AND SYMBOLS

°C	Degrees Celsius
%	Percentage
µg	microgram
µl	microlitre
A	adenine
ABI	Applied Biosystems (Perkin Elmer)
AFLP	Amplified fragment length polymorphism
Amp	Ampicillin
App.	Appendix
BA	Benzyladenine
bp	base pair
C	cytosine
CH_3COONH_4	Ammonium acetate
CpG	Cytosine and guanine dinucleotide
CTAB	Cetyltrimethylammonium bromide
2,4-D	2,4-dichloropenxyacetic acid
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetra acetic acid
EPPS	*N*-(2-hydroxyethy)piperazine-*N'*-(3-propanesulfonic acid)
g	Grams
G	guanine
H	Hour
H_2O	Water

IPTG	Isopropyl-[beta]-D-thiogalactopyranoside
L	Litre
LB	Luria Bertani
LTR	Long terminal repeat
M	Molar
mer	Oligomer
mg	Milligrams
$MgCl_2$	Magnesium cloride
ml	Millilitres
mM	Millimolar
mRNA	messenger Ribonucleic Acid
MS – RDA	Methylation sensitive Representational Difference Analysis
NaCl	Sodium chloride
ng	nano gram
nmol	nano mole
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
oligo	Oligomer
PCR	Polymerase Chain Reaction
pH	Log hydrogen ion concentration
QTL	Quantitative Trait Locus
RAPD	Random Amplified Polymorphic DNA
RDA	Representational Difference Analysis
RDA-WEEC	RDA-With Elimination of Excessive Clones
rDNA	Ribosomal DNA
RE	Restriction Endonuclease
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
RNaseA	Ribonuclease A
rpm	revolutions per minute
SDS	Sodium Dodecyl Sulphate
sec	Second(s)
ss	single stranded
SCAR	Sequence Characterized Amplified Regions

SSC	Saline sodium citrate
T	Thymine
T_a	Annealing Temperature
TAE	Tris-acetate EDTA
Taq	*Thermus aquaticus* DNA polymerase
TE	Tris EDTA
T_e	Extension Temperature
Tet	Tetracyclin
T_m	Melting Temperature
Tris	2-amino-2-(hydromethyl) propane-1, 3 diol
tRNA	transfer RNA
U	Unit
UV	Ultraviolet
X-Gal	5-Bromo-4-chloro-indoyl-[beta]-D galactoside

Term	**Definition**
Adenine (A)	purine base that pairs with thymine (T) in DNA.
Agarose	the neutral gelling fraction of agar commonly used in gel electrophoresis.
Algorithm	a step by step process for solving a problem.
Alignment	the juxtaposition of amino acids or nucleotides in homologous molecules that are assumed to contain residues that are all derived from a single common ancestral residue.
Aliquot	a subsample of a reagent; to divide into several subsamples.
Allele	one of a series of possible alternative forms of a given gene differing in DNA sequence and affecting the structure and/or function of a single product (RNA and/or protein).
Alu family	a short (300 bp), interspersed DNA sequence repeated about 500,000 times in the human genome, and characterized by containing a distinctive *Alu*I restriction site.

Bacterial colony	a clone of bacterial cells.
Base pair	a pair of hydrogen-bonded nucleotides that join the two strands of a DNA double helix. In a double-stranded DNA molecule, adenine (A) forms a base pair with thymine (T), and guanine (G) pairs with cytosine (C).
Base sequence	order of bases in a DNA molecule.
C-value	a measure of haploid DNA content per cell.
Cathode	the negative electrode in an electrolytic cell (such as an electrophoresis chamber) toward which cations migrate.
cDNA	complementary DNA, often refering to a cDNA library made with mRNA and the enzyme reverse transcriptase.
cDNA clone	double-stranded DNA sequence that is complementary to a specific RNA and inserted in a cloning vector such as a plasmid.
Chromatid	the eukaryotic chromosome prior to replication, or one of the two longitudinal subunits of a chromosome after replication, joined by a centromere.
Chromosome	structure containing DNA and proteins in the cell nucleus.
Clone	1.*verb*. to insert a piece of DNA into a vector for subsequent amplification and isolation of that specific piece; 2.*noun*. a piece of DNA composed of a vector and its insert.
Cloning vector	biological carriers such as plasmids, bacteriophage, or cosmids used to amplify an inserted DNA sequence.
Coding strand	DNA strand that is transcribed into mRNA.
Colony - hybridizatrion	method for detecting bacteria that carry a vector with a desired inserted sequence.
Complementary sequence	a sequence of nucleotides related by the base-pairing rules. For example, in DNA a sequence A-G-T in one strand is complementary to T-C-A in the other strand. A given sequence defines the complementary sequence.

DNA reannealing   double-stranded DNA separates into single strands when heated which reanneal back into double strands when temperature is lowered.

DNase   deoxyribonuclease; enzyme that cleaves phosphodiester bonds in DNA, to break the molecule into pieces.

Dot (slot)-blot   a DNA analysis system where sample DNA is directly pipetted onto a membrane, as opposed to the Southern blot procedure of enzymatic digestion, electrophoresis, and Southern transfer.

Double-stranded helix   three-dimensional shape exhibited by two complementary base-paired DNA strands.

Electrophoresis   the separation of macromolecules in the presence of an electric current. Electrophoresis is routinely used to separate both proteins and DNA fragments; allozymes are separated based on differences in net charge, whereas DNA fragments are separated based on differences in size.

Epigenetic   all processes relating to the expression and interaction of genes.

Fingerprinting   separation of the DNA of an individual into defined fragments the lengths of which are determined by the spacing of given restriction of enzyme sites. Numbers and lengths of fragments form a unique 'DNA fingerprint' for an individual.

Gene   a sequence of DNA that functions as a unit (e.g., coding for a specific protein).

Gene family   set of very similar genes derived by duplication of an ancestral gene and subsequent minor alteration in each gene in the family.

Genome   the sum total of all the DNA on a haploid set of chromosomes in the nucleus of an individual, including both coding and non-coding sequences.

Genotype   genetic consititution of an individual organism.

Guanine (G)   purine base that pairs with cytosine in DNA.

Haploid	chromosome number in the gametes of a species, symbolized by "n".
Heterochromatin	chromosomal segments or whole chromosomes that generally exhibit a condensed state throughout interphase and late replication.
Homology	common ancestry of two or more genes or gene products (or portions thereof).
Hotspot	region in DNA where mutations occur at exceptionally high frequency.
Hybridization	formation of a double-stranded nucleic acid molecule from complementary single-stranded molecules.
Hybridization stringencies	the fidelity with which single strands of DNA reanneal depends on the stringency of hybridization determined by temperature and ionic conditions.
Hybridize	to induce the pairing of complementary DNA strands, often from different individuals or species, to form a DNA-DNA hybrid molecule.
Hypervariability	extreme genetic variations between individuals in certain genomic sequences.
Hyper variable region	a segment of a chromosome characterized by considerable variation in the number of tandem repeats at one or more loci.
In vitro	means "in glass" and refers to a biological process carried out in the laboratory separate from an organism.
Insert	the DNA of interest that has been cloned, or inserted, into a vector.
Insertion	placement of additional nucleotide pairs in a specific site in DNA.
Intergenic	nucleotide sequences located between genes.

Interspersed repetitive DNA	short segments of DNA that have been found in hundreds of thousands of copies interspersed throughout the genome, rather than being serially repeated like satellite DNAs.
Inversion	a reversed chromosome segment.
Inverted repeats	symmetrical nucleotide sequence of DNA that is repeated in opposite orientations on same molecule.
Isochizomer	restriction endonuclease with the same recogniton sequence as another restriction endonuclease.
Label	to chemically "tag" a piece of DNA so that it can be visualized in some manner. Most commonly, DNA is labeled with a radioactive element, enabling detection of its presence using x-ray film. Labeling may also be achieved by colorimetric methods or by chemiluminesence. Visualization of larger amounts of DNA requires less sensitivity and is often possible using stains such as ethidium bromide.
Ligase	an enzyme that covalently joins two single stranded DNA molecules when annealed end to end on a DNA template.
Ligation	enzymatically catalyzed formation of a phosphodiester bond that links two DNA molecules.
Locus	a specific position on a chromosome.
Long-terminal repeats (LTR)	nucleotide sequence that is repeated at the end of a DNA molecule.
Melt	the process of disrupting the hydrogen bonds linking complementary DNA strands.
Melting temperature (Tm)	midpoint of the heat denaturation curve for double-stranded DNA.
messenger RNA (mRNA)	single-stranded template RNA that contains information for amino acid sequence of the protein.
Methylation	the chemical process of adding a methyl group to a molecule.

Methylation (me)	one form of methylation, the most common in mammals, invloves the conversion of cytosine to 5-methyl cytosine. Methylation can prevent cleavage of DNA at a restriction enzyme recognition site, for example, Hpa II cleaves at C^CG G
Microsatellite	a class of repetitive DNA. Microsatellites are simple sequence repeats two to eight nucleotides in length. For example the repeat unit can be simply "CA", and might exist in a tandem array (CACACACACA...) be highly polymorphic.
Minisatellite	tandem array of from 10 to 50 copies of a non-coding length (typically 10 to 100 nucleotides) of DNA. Arrays on different chromosomes are usually with different numbers of repeated copies giving rise to unique individual DNA fingerprints.
Mismatch	bases that do not match in "complementary" DNA strands. Depending in the blot wash stringency conditions, some mismatch can be tolerated between hybridized sample and probe DNA complementary
Mobile elements	lengths of DNA that can move from one position to another in the genome.
Multigene family	a collection of identical or near identical genes in the genome. The numbers of gene copies and their distribution amongst chromosomes varies widely between species depending on the gene family in question.
Mutagen	an agent that causes changes in the nucleotide sequence of DNA.
Mutant	organism that carries a modified inherited gene.
Mutation	change in the nucleotide sequence of DNA that is inherited.
Nuclear genome	the portion of the genome contained in the nucleus of eukaryotes, i.e., the chromosomes.
Nuclease	enzyme that cleaves phophordiester bonds in nucleic acids.

Nucleolar organizer region	a region on a chromosome that contains the ribosomal RNA genes and associated spacers.
Nucleotide	one of the monomeric units from which DNA molecules are constructed, consisting of a purine or pyrimidine base, a pentose sugar, and a phosphoric acid group. The nucleotides of DNA are deoxyadenylic acid, thymidylic acid, deoxygualinic acid, and deoxycytidylic acid. The nucleotides are often referred to interchangeably with their corresponding nitrogenous base, i.e., the nucleotide deoxyadenylic acid is often referred to as adenine (represented in a sequence by "A").
ORF	open reading frame; start and stop codons are present around a DNA sequence.
Pellet	the button of particulate material formed after a suspension has been centrifuged.
Phenotype	the physical make-up of an individual as define by genetic and non-genetic factors.
Plasmid	a self-replicating extrachromosomal genetic element found in a variety of bacterial species that generally confers some advantage to the host cell (i.e., resistance to antibiotics, etc.). Plasmids are double-stranded, circular DNA molecules ranging in size from 1 to 200 thousand base pairs commonly used as vectors for cloning.
point mutation	a mutation involving a single nucleotide substitution.
Polymerase	an enzyme that assembles the subunits of macromolecules. DNA polymerases have the ability to synthesize the complementary strand of a single stranded DNA template. Synthesis only extends from existing double-stranded sequence across single-stranded template; thus synthesis can be controlled in reactions such as PCR or sequencing using oligonucleotide primers that will anneal to form short double stranded sections contiguous to specific regions of interest.

Polymerase Chain Reaction (PCR)   a series of thermal cycles of denaturation, annealing of primers, and primer extension catalyzed by a thermostable DNA polymerase, in which a target DNA fragment is amplified exponentially; primers that have nucleotide sequences complementary to the DNA that flanks the target region are added to sample DNA along with a heat-stable DNA polymerase. The DNA is heated to separate the complementary strands and then cooled to let the primers bind to the flanking sequences. The polymerase initiates synthesis of complementary DNA. The reaction is allowed to proceed for a series of replication cycles. Twenty cycles will yield a millionfold amplification; thirty cycles will yield an amplification factor of one billion.

Polymorphism   intraspecific variation. On the DNA level, this refers to differences in base pair sequence between two individuals.

Primers   short pieces of single stranded DNA (10-30 bp) annealed to the 5' end of a DNA template used to initiate synthesis of the complementary strand of the template piece of DNA. Primers can be designed so that they will bind only to a very specific region of the DNA, and will thus initiate synthesis of a targeted sequence (as in PCR or DNA sequencing).

Probe   in molecular biology, any biochemical molecule that can be used to distinguish a specific molecule of interest apart from others of its kind. A DNA probe is a segment of DNA with a sequence homologous to the DNA of interest. If the probe is labeled, the sequence can be visualized independently from the rest of the DNA in the sample.

Purine   a heterocyclic ring molecule with various side chains. (Adenine and guanine are purines found in DNA and RNA.)

Pyrimidine   a heterocyclic ring molecule with six carbons and various side chains. (Cytosine and uracil are pyrimidines found in RNA, while cytosine and thymine are found in DNA.)

RAPD	(random amplified polymorphic DNA) An analysis of PCR products produced by short non-specific primers. High levels of polymorphism are often observed using this technique.
RDNA	ribosomal DNA; the genes for several classes of ribosomal RNA molecules that go into the construction of ribosomes, usually in long tandem arrays in the chromosomes.
Repetitive DNA	nucleotide sequences occurring repeatedly in chromosomal DNA. Repetitive DNA can belong to the highly repetitive or middle repetitive categories. The highly repetitive fraction contains sequences of several nucleotides repeated millions of times. Middle repetitive DNA consists of segments 1 to 500 base pairs in length repeated 100 to 10,000 times each.
Restriction enzyme (endonuclease)	an enzyme that cleaves double-stranded DNA. Type I restriction endonucleases are not sequence-specific; type II restriction endonucleases cleave DNA at particular recognition sequences (typically 4-6 bp palindromes).The enzymes are named by an acronym that indicates the bacterial species from which they were isolated, followed by a Roman numeral that gives the chronological order of discovery when more than one enzyme came from the same source. DNA fragments produced by certain enzymes, such as *Eco*RI, can anneal with any other fragment produced by that enzyme. This property allows splicing of foreign genes into *E. coli* plasmids or bacteriophage vectors.
Restriction fragment length polymorphism (RFLP)	a polymorphism in an individual, population, or species defined by restriction fragments of a distinctive length. Usually caused by gain or loss of a restriction site, but may result from an insertion or deletion of a fragment of DNA between two conserved restriction sites.
restriction sites	short motifs of DNA capable of being recognized by a restriction enzyme leading to the cutting of the DNA molecule into separate fragments. Each restriction enzyme has a unique cutting site.

RNA **(ribonucleic acid)**	polynucleotide that contains ribose sugar.
RNAse	an enzyme capable of degrading RNA.
RRNA	ribosomal RNA, the nucleic acid component of ribosomes, which functions in translation of proteins from mRNA.
Single copy DNA	usually refers to sequences that appear only once in the entire genome. Specific genes or sequences that are single copy are much more difficult to isolate because they represent such a small percentage of the total DNA of an organism.
Single-copy gene	genes for which only two alleles exist (one from each parent) in a diploid cell.
Size marker	DNA fragments of known molecular weight and base pair length, such as l-phage digested with the restriction enzyme *Hind*III, run on electrophoresis gels for the determination of DNA sample fragment sizes.
Siippage	a mechanism of DNA turnover by which gains-and losses occur of short motifs (usually less than 10 nucleotides) in DNA helix leading to pure and cryptic DNA simplicity.
Southern blot	a membrane onto which DNA has been transferred directly from an electrophoretic gel.
Stringency	in DNA-DNA hybridization or DNA-RNA hybridization, the conditions of the hybridization (such as temperature and concentration of chemical additives) that determine the degree of similarity that will result in formation of hybrid molecules.
Tandem array	multiple copies of a sequence of DNA that are arranged one after another in series. Repeat units can be short nucleotide sequences or entire sets of genes.
***Taq* polymerase**	a thermostable DNA polymerase from *Thermus aquaticus*, thermophilic bacterium. Used for amplification via the polymerase chain reaction.

Thymine	pyrimide base found in DNA.
thymine dimer	adjacent thymine residues in DNA that have been chemically linked, usually by the action of ultraviolet irradiation.
Transform	to cause bacterial cells to take up a plasmid host.
Translocation	change in the position of a chromosome segment within a genome.
Transposable element	a genomic element that can move from site to site in the genome of an organism, either through direct DNA copying (at least in prokaryotes) or reverse transcription from an RNA intermediate (probably the usual mechanism in eukaryotes).
Transposition	see mobile elements. Duplicative transposition occurs when a given DNA region replicates and the extra copy moves to another position in the genome. Non-duplicative transposition occurs when the DNA region moves from one position to another: no extra copies are involved.
Transposon	a segment of DNA flanked by transposable elements that is capable of moving its location in the genome.
Uracil	pyrimidine base in RNA that appears in place of the thymine found in DNA.
Vector	a self-replicating DNA molecule that exists with, but is separate from the genome of the host cell. Many different vectors have been identified and genetically engineered for use in molecular biology. DNA inserted into a vector will be replicated along with the vector. In this manner, DNA of interest can be obtained in large quantities i.e., cloned. For example, the human insulin gene can be cloned into the plasmid vector pBR 322 which in turn will replicate in *E. coli* cultures.

References

Freifelder D. and Malacinski GM. 1993. Essentials of Molecular Biology 2nd Ed. Jones and Bartlett, Boston, Massachusetts.

Hillis DM. and Morits C. 1990. Molecular Systematics. Sinauer, Sunderland, Massachusetts.

Hoelzel AR. and Dover GA. 1991. Molecular Genetic Ecology. Oxord University Press, New York. D. Rickwood, ed.

Kirby LT. 1990. DNA fingerprinting: An introduction. Stockton Press, New York.

INDEX OF TABLES

Chapter 1

The history of molecular biology and detection of genetic variation

1.1   Abstract

The past century has witnessed a breathtaking array of discoveries in the biological sciences. This is in particular in the general area of molecular biology, the scientific discipline that seeks to fully understand the molecular basis of heredity, genetic variation, and the expression patterns of individual units of heredity called genes. To fully appreciate the current status of molecular biology and where it is heading it is important to understand the early principals and the theories that gave rise to it. In this chapter a brief history of the development of the field of molecular biology and the basic philosophy on which our understanding of biological systems is based is presented. Furthermore an overview of the plant genome and its composition is given. The mechanisms that is responsible for creating variation in genomes and the most popular techniques molecular biologists use to find, study and utilize these changes are presented.

1.2   Biology: then and now

Building on observations about natural plant variation in time and space, early geneticists, such as Darwin, Mendel, and Vavilov, posed fundamental questions regarding the origin, structure, and evolution of genetic diversity. They postulated that an underlying reservoir of innate and heritable genetic possibilities delineated the options for growth, development, and reproduction of organisms at both the individual and population level. The field of plant genetics today continues to address many of the same questions while integrating new developments in molecular biology and bioinformatics. Over the last 15 to 20 years, new, highly automated tools have created unprecedented opportunities for generating and analysing large biological data sets. The systematic processing of nucleic acid and protein sequence information from many different organisms has further fundamentally changed the way that biologists approach the study of living things (McCouch, 2001).

In retrospect, the 6th decade of the 19th century was truly remarkable with respect to the development of the science of biology. By the end of those years all of the pieces were in place for the maturation of what had been a purely observational discipline into one with a theoretical basis. The result, the field of molecular biology and its attendant sub-disciplines, is grounded philosophically in a mechanistic, deterministic and reductionistic view that derives from the logical empiristic setting in which it was

born (Judson, 1996). In November 1859, Charles Darwin first published his "*On the Origin of Species*" and in 1866 Fr. Gregor Mendel his *"First and Second Laws of Heredity"* as well as his conceptualisation of the gene as the unit of inheritance. By 1869, the chemist JF Miescher isolated from human pus the substance he named nuclein, which was later called nucleic acid. At the time of these events, biology could hardly be compared to the so-called 'hard' sciences such as physics and chemistry, which rested upon strong theoretical platforms. Biology had essentially been an exercise in observation and classification. But with Darwin's theory and Mendel's laws, biology had for the first time a potential theoretical basis of it's own.

Modern biology has had great success in representing the reality of living systems in a form that yields a great deal of both theoretical and practical information. Our current understanding of all of the mechanisms by which these macromolecules act and interact as a part of the life functions of an organism derive from the reductionist paradigm and the techniques it has produced. And yet, there are clearly aspects of living systems that have not yielded to this analysis. Among these features are included the organization of the human brain and, more importantly, the origin of the mind and consciousness. It has become increasingly difficult to model these kinds of natural phenomenon in terms of the reductionist paradigm extent in much of biological thinking.

A major part of the difficulty appears to be the framework within which the natural world is viewed in modern biology. As an inheritor of the logical 'empiricists' position, the biologist believes that the only aspects of the world that are observable and open to rational investigation. In fact, relevant are those that could be, in Aristotelian terms, material cause (the set of objective possibilities) and formal cause (the shape or form of the substance). However, modern biology mostly does not recognise nor does it incorporate into theoretical considerations either the final cause (action of will) or the efficient cause (projection into reality of this act of will) (Ayala, 1970). The problem is a science that embraces Hume's assertion that the efficient and final causes of a thing can never be known. But consider for a moment the nature of chance or random events and the mistake becomes evident in this chain of reasoning. The neo-Darwinist assumes that the random nature of mutational changes in DNA eliminates the causality from the consideration. In fact, the mechanisms of mutation are well understood and proceed by quite specific steps. In a world where no system is entirely closed, that is, isolated from its surroundings, blind chance cannot exist: there are interconnections that provide a complex, if sometimes subtle, input into all

events (Laszlo, 1996). The overall result can be seen as having both an efficient cause (the agent of mutation) and a final cause (the consequence of the mutated gene on the function of the organism). Although the biologist cannot predict which gene will be mutated in a particular organism, the dependence of the functioning of that organism on that change can be observed (Ayala, 1970).

Explanation by design, or teleology, is "the use of design, purpose, or utility as an explanation of any natural phenomenon" (Webster's Third New International Dictionary, 1966). An object or behaviour is said to be teleological when it gives evidence of design or appears to be directed towards certain ends. Teleological explanations account for the existence of a certain feature in a system by demonstrating the feature's contribution to a specific property or state in the system. The idea of teleological thinking is viewed as a slur when applied to any kind of biological model or conclusion. Nonetheless, purpose and plan are obvious in living systems. Some time ago Francisco Ayala pointed out three types of teleological explanations that are appropriate to biological systems: (1) conscious anticipation of an end-state or goal, (2) self-regulating systems and (3) structures designed to perform a specific function. As Barbara McClintock (1984) expressed it: There are 'shocks' that a genome must face repeatedly, and for which it is prepared to respond in a programmed manner. An example is the 'heat shock' response in eukaryotic organisms. Each of these initiates a highly programmed sequence of events within the cell that serves to cushion the effects of the shock. Some sensing mechanisms must be present in these instances to alert the cell to imminent danger, and to set in motion the orderly sequence of events that will mitigate the danger. But there are also responses to the genome to unanticipated challenges that are not so precisely programmed. The genome is unprepared for these shocks. Nevertheless, they are sensed, and the genome responds in a discernible but initially unforeseen manner.

And as the co-holder of the 1945 Nobel prize for physiology and medicine, Sir Ernest Chain put it: "To postulate that the development and survival of the fittest is entirely a consequence of chance mutation seems to me a hypothesis based on no evidence and irreconcilable with the facts. These classical evolutionary theories are a gross over-simplification of an immensely complex and intricate mass of facts, and it amazes me that they are swallowed so uncritically and readily and for such a long time, by so many scientists without a murmur of protest." Teleology asserts that a casual non-physical agent as postulated by vitalism is purposeful and that there is purpose and design in nature.

1.3   The plant genome: size and organisation

DNA re-association kinetics studies showed that non-transcribing repeat (NTR)-DNA is an integral part of most plant genomes and its amount is proportional to the genome size (Flavell *et al.*, 1974). Most plant genomes are large and complex and NTR-DNA is primarily composed of retro-transposons (Bennetzen *et al.*, 1998). The composition of plant NTR-DNA seems to be a result of multiple invasions by different retro-transposons. Replication of retro-transposons then occurred followed by their inactivation by transposition and/or hetero-chromatization (Sandhu and Gill, 2002). The NTR-DNA is unevenly distributed in the plant genomes. Paucity of genes observed from physical maps (Sandhu *et al.*, 2001) and an abundance of hetrochromatin visualized as C-bands (Curtis and Lukaszewski, 1991; Gill *et al.*, 1991; Jiang *et al.*, 1996) allegorise that repetitive DNA is especially abundant around the centromeric regions (Copenhaver and Preuss, 1999). Regions present between two gene-rich regions are also composed of NTR-DNA as well as regions present near the tip of chromosome arms deficient in genes (Sandhu and Gill, 2002). In addition to retro-transposons, psuedo-genes also seem to be an important part of the non-transcribed regions.

Whether transposons, retro-transposons, and other repetitive elements accumulate extensively in a given evolutionary lineage may depend on several factors. Among them are the efficiency or repressive mechanisms and the rate at which the sequences undergo mutational and deletional decay. For example, methylation of C residues enhances the mutability of CG base pairs hence methylation accelerates the divergence rate of newly arising duplications. This happens at an extreme form in Neurospora, in which many methylated CGs are mutated in the span of a single generation, and at more measured rates in plants and mammals (SanMiguel *et al.*, 1998; Wang *et al.*, 1998). If multiple copies of a DNA sequence are present in a genome we can think of each sequence as a single 'species' evolving on it's own 'line of decent' because each repeat will be mutated at random.

Retro-transposons display a high degree of sequence variability (Casacuberta *et al.*, 1995; Marillonnet and Wessler, 1998) and in most cases they represent elements that have lost the ability to transpose. It is also known that both the *Ac* and *Spm* transposons of maize frequently give rise to internally deleted elements, and *Ac* ends

are much more abundant in the maize genome than are full-length elements, suggesting deletional decay of transposon sequences (Fredoroff *et al.*, 1983,1984; Masson *et al.*, 1987).
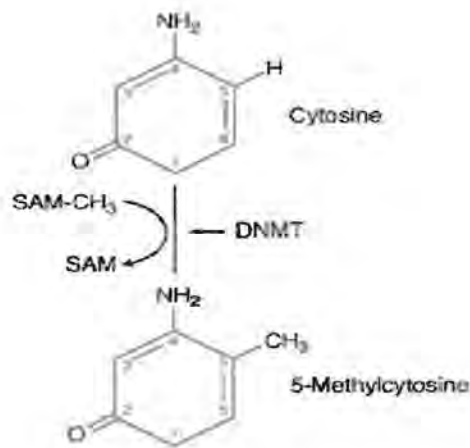
Figure 1.1. Mechanism of DNA methylation. 5-Methylcytosine is produced by the action of the DNA methyltransferases (DNMT), which catalyse the transfer of a methyl group (CH_3) from *S*-adenosylmethionine (SAM) to the carbon-5 position of cytosine (Strathdee and Brown, 2002).

1.3.1 Functional and non-functional sequences

Sequences within the genome can be classified according to a number of criteria. The most important of these is functionality and the largest class of functional DNA elements (also known as euchromatin) consists of coding sequences within transcription units. For the most part, transcription units correspond one-to-one with Mendelian genes. They usually function on behalf of the organism within which they are contained. The functional class of DNA elements also includes a number of specialized sequences that play roles in chromosome structure and transmission. The best-characterized structural elements are associated with the centromeres and telomers (Figure 1.2) (Fitzgerald-Hayes *et al.*, 1982; Bloom and Carbon, 1982; Sun *et al.*, 1997; Wright *et al.*, 1996; Pardue *et al.*, 1996; Zakian 1996).

Most of the genome appears to consist of DNA sequences that have no apparent function. This non-functional class includes psuedo-genes that derive from specific

6

genes but are not themselves functional with a lack of transcription or translation. For the most part, however, non-functional DNA, also known as heterocrhomatin, is present in the context of long lengths of apparently random sequence, and repetitive elements with origins that have long since become indecipherable as a consequence of constant genetic drift.
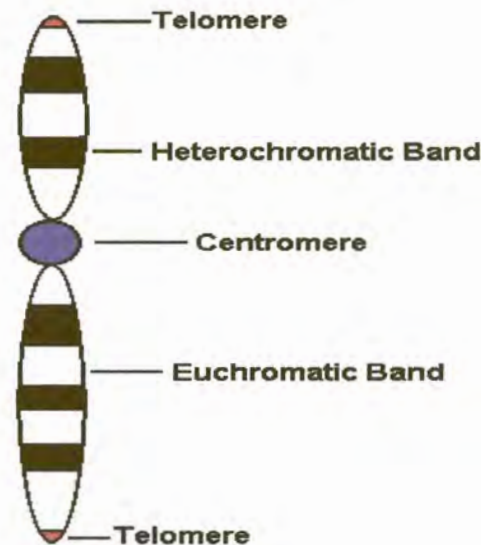
Figure 1.2. Diagram indicating different structural elements found on the chromosome (Brown, 1992).

Both functional and non-functional sequences can be distinguished by a second criterion – copy number. 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, were recombination is suppressed. In contrast, repeats occurring in euchromatin 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. The two major families, the long and short interspersed nucleotide elements (LINEs and SINEs), have significant roles in genome function and evolution.

1.4. Sources of variation: Genomic elements, genome variation and evolution

Towards the end of the sixties of the last century it became clear that the genome of eucaryotes contain, in contrast to the prokaryotic genome, a high percentage of non-coding, usually repetitive nucleotide sequences. Repetitive DNA consists of a repeated sequence of a certain size (the repeat unit) with a given copy number organized in a particular manner in space. Repeat units can be organized in three ways: (1) tandem repeats have no spaces between individual repeat units (Figure 1.2), (2) hyphenated repeats are separated by small gaps but are still grouped together and (3) dispersed repeats are spread throughout the genome. There is obviously nothing remarkable about dispersed repeats of one, or even two or three nucleotides. Hence, dispersed repeats are only significant when they involve reasonably long DNA sequences, which occur substantially more frequently than would be expected by chance. Some genome-wide dispersed repetitive DNA corresponds to members of multigene families, comprising both functional genes and pseudo-genes. Otherwise, it may represent motifs that function at the DNA level. Most dispersed repetitive DNA, however, corresponds to either functional transposable elements or their remains. Today, we know that transposons constitute a large fraction – even a majority – of the DNA in some species of plants and animals, among them mice, humans, and agriculturally important plants such as maize and wheat.
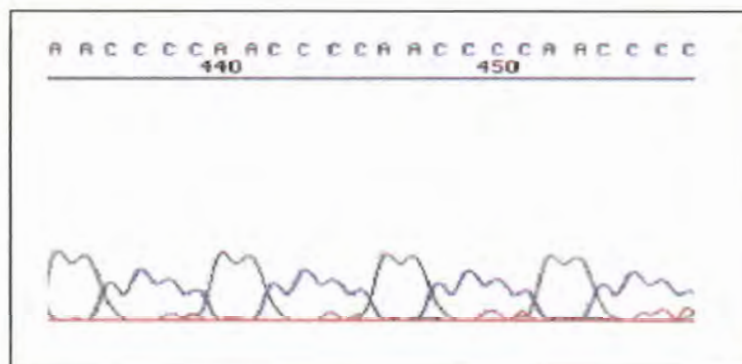
Figure 1.3. DNA sequence showing a tandem AACCC repeat

1.4.1 Retro-elements and genome variation

Retro-elements have been found in the genomes of all plant species that have been examined. But they seem to be highly abundant only 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). Bennetzen and Kellogg (1997) raised the question if retro-transposons are the largest single component of many flowering plant genomes, because not all plant genomes have expanded with the amplification of these elements. Retro-elements transpose without excision and their mobility will always increase their copy number and thereby increase the genome size (Figure 1.4). Therefore, the question has to be asked does continuous or episodic retro-element amplification mean that all plants are on the road to larger genomes, or is there an active process for removing these interspersed repetitive DNAs from plant genomes?
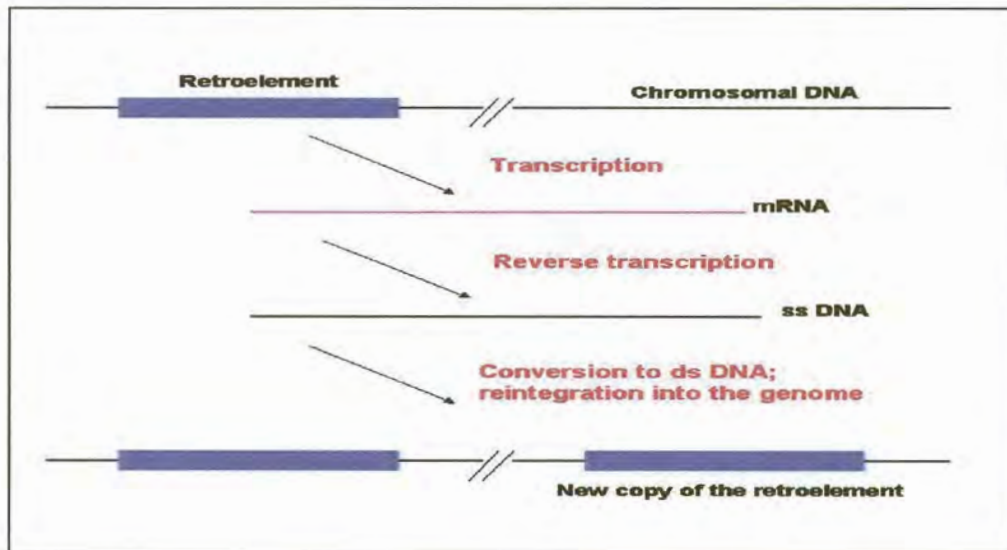
Figure 1.4. Duplication of retro-transposons in the genome *via* a RNA intermediate (Brown, 1992).

Logic dictates that there should be an upper limit to the level of transposition that genomes can endure. It is therefore quite surprising to discover the magnitude of transposable elements in genomes of plants, such as maize, which together make up more than 50% of the genome (SanMiguel *et al.*, 1996). The presence of multiple low-copy number families further indicates that many hundreds, if not thousands, of

9

distinct retro-element families exist in maize (Voytas, 1996). But how can a genome function with such a large burden of retro-elements? It is of course, in the element's best interest to minimize genetic damage caused by integration, because the host's survival is necessary for persistence of the element. In the yeast *Saccharomyces cerivisiae*, in which retro-transposons have been studied extensively, it appears that the five families of retro-transposons found within the *S. cerevisiae* genome, have a strong bias for sites in the genome where they integrate. Retro-elements are found particularly upstream from tRNA genes or at the telomeres. Regions targeted by yeast retro-transposons are typically devoid of open reading frames (Voytas, 1996).

The organisation of retro-elements in the interspacer regions of maize is reminiscent of the findings in yeast. Targeted integration, as opposed to amplification by recombination, is suggested by the overall structural integrity of the retro-elements and the presence of intact target-site duplications flanking most insertions. The under-representation of the most highly abundant retro-element families in the maize DNA sequence databases further suggests that these elements specifically avoid coding regions or that their presence near genes has been strongly selected against. Intergenic regions are hyper-methylated relative to gene sequences (Bennetzen *et al.*, 1994). By extrapolating from the yeast model, one might predict that some such unique chromatin feature serves as a homing device for maize retro-elements during integration. Hyper-methylated arrays of retro-transposons within retro-transposons have also been observed in the slime mold *Physarum polycephalum* (Rothnie *et al.*, 1991). This suggests that targeted integration may be a widespread strategy adopted by retro-elements to proliferate within host genomes (Voytas, 1996). According to this model, the increased number of retro-transposons in large-genome lineages may be due part to an increase in the number of possible non-deleterious insertion sites in the genome.

The fundamental issue is the likelihood of changes in genome size over evolutionary time and, in particular, the likelihood of decreases in genome size versus increases. Although increases - via amplification of retro-transposons - are clearly possible and apparently easy, decreases may be more difficult and/or may occur less frequently. The striking variation in genome size observed between closely related species has been termed the C-value paradox (Thomas, 1971), meaning that it is paradoxical that genomic complexity (i.e., size) does not correlate with biological complexity of the organism. Some genome-size variation is due to the polyploidy commonly found in the angiosperms or to tandemly repeated satellite sequences, but most is associated

with ill-defined classes of interspersed highly repetitive and middle repetitive DNAs (Flavell *et al.*, 1974). Recent studies have indicated that the majority of this reiterated DNA is composed of retro-elements (Moor *et al.*, 1991; SanMiguel *et al.*, 1996; Smyth *et al.*, 1989)

The question still stands as to whether all plant genomes are destined for genome obesity o if an active process for removing these interspersed repetitive DNAs from plant genomes exists? Studies on spontaneous mutations have shown that deletions are more frequent and longer than insertions. For example, in mammals, deletions are three to seven-times more frequent than are insertions and are, on average, somewhat longer (Graur *et al.*, 1989). In *Drosophila*, the difference is even more profound – deletions are almost ten-times more frequent and almost seven-times longer than are insertions (Petrov *et al.*, 1996). This biases in mutation frequency and size will lead to the progressive elimination of nonessential sequences. Admittedly, this process is very slow.

It is also known that both the *Ac* and *Spm* transposons of maize frequently give rise to internally deleted elements, and *Ac* ends are more abundant in the maize genome than are full-length elements, suggesting deletional decay of transposon sequences (Fedoroff *et al.*, 1983,1984; Masson *et al.*, 1987; Schwarz-Sommer *et al.*, 1985). It would not be surprising to find mechanisms that preferentially eliminate sequences, as has been found in wheat for the preferential loss of non-redundant sequences early after polyploidization (Feldman *et al.*, 1997). Whether transposons, retro-transposons, and other repetitive elements accumulate in a given evolutionary lineage may depend on several factors, among them the efficiency of repressive mechanisms and the rate at which the sequence undergo mutational and deletional decay (Fedoroff, 2000).

1.4.2   Stress and genome variation

Growth conditions are seldom optimal and when the environment changes an organism must be able to adapt in order to survive or die. Any environmental change that results in a response of an organism that is less than optimal might be considered as stressful (Levitt, 1972). Stress factors can be either biotic, imposed by other organisms, abiotic, arising from an excess or deficit in the physical or chemical

environment or of a genetic nature for example the introduction of foreign genetic material and viral infections. When plants are grown in artificial conditions additional stress might be introduced by the synthetic growth medium or additional growth regulators and antibiotics added to the plants' environment (Buchanan *et al.*, 2000).

Many of the plant retro-transposons studied to date are transcriptionally activated by various biotic and abiotic stress factors (Grandbastien, 1998). The expression of tobacco *Tnt*1 and *Tto*1 retrotransposons is greatly increased by several abiotic stresses such as cell culture, wounding, methyl jasmonate, $CuCl_2$ and salicylic acid (Hirochika, 1993; Mhiri *et al.*, 1997; Moreau-Mhiri *et al.*, 1996; Pouteau *et al.*, 1994; Takeda *et al.*, 1998, 1999). Similarly, various biotic stress factors, such as infection with various viral, bacterial or fungal pathogens (Mhiri *et al.*, 1999; Pouteau *et al.*, 1994), have been shown to activate transcription of these retrotransposons. In contrast to *Tnt*1 and *Tto*1, transcription of *Tos*17 is induced only by tissue culture.

Many factors determine how plants react to these stresses. The genetic make-up of the plant, its developmental state, the duration and severity of the stress, the number of times the plant is subjected to stress and any synergistic effects of multiple stresses influence the plants' response. In response to stress some genes are expressed more strongly, whereas the expression of others are repressed. Several mechanisms, such as quantitative modification of repetitive DNA, DNA methylation, excision and insertion of transposable elements, gene amplification or deletion and histone acetylation have been suggested as points of control for these changes (Capy, 1998; Cullis 1990; Johnson *et al.*, 1996).

1.4.3 Somaclonal variation

Somaclonal variation is a widespread phenomenon in tissue culture plants. The term, somaclonal variation, was first defined by Larkin and Scowcroft (1981) as the genetic variation in plants regenerated from the tissue culture process. In all organisms, spontaneous mutations occur from generation to generation, but somaclonal variation describes the additional mutations in plants produced via tissue culture (Bouman and De Klerk, 1997). This unexpected source of variability was once hailed as a "novel source of variation for crop improvement", but due largely to its unpredictability as a breeding tool, enthusiasm for this application has diminished and somaclonal variation has lost much popularity in recent years (Karp, 1993).

Somaclonal variation and its causes are not well understood and have not been elucidated. Plants are generated by a series of cell divisions in meristematic tissues. During the early stages of embryogenesis the apical meristem is formed from the zygote. The axillary meristem in turn originates from the apical meristem. However, new apical meristems may also originate from non-zygotic cells, in particular from somatic cells from callus or cell suspension cultures. Plants generated from these adventitious meristems are often genetically different from the mother plant (Bouman and De Klerk, 1997). Somaclonal variations are therefore often associated with callus formation (Skirvin and Janick, 1976) and the use of growth regulators, such as 2,4-dichloropenxyacetic acid (2,4-D) and benzyladenine (BA) (Figure 1.5) have been reported to play an important role in the induction of variability (Evans, 1988). Linacero et al. (2000) found hot spots of DNA instability in rye plants generated from immature embryos. At least 40% of the studied rye plants showed at least one variation, and the number of mutations per plant was quite high, ranging from 2 to 12. In 2001, Leroy et al. found in a study done on cauliflower calli using microsatellite primers, only six calli out of a total of 224 with stable original DNA patterns. Many more examples of somaclonal variation have been found in plants grown in vitro including beet plants (Sabir et al., 1992), red clover (Nelke et al., 1993), and oilseed rape (Poulsen et al., 1993).
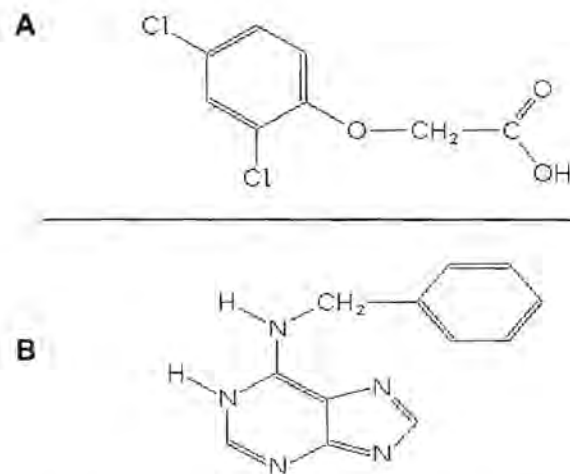
Figure 1.5. (A) The molecular structure of 2,4-dichloropenxyacetic acid (2,4-D) and (B) the structure of benzyladenine (BA).

1.4.4 Genomic re-arrangements and somaclonal variation

Many of the mechanisms by which the genome is re-organized due to stress have been observed in cells in tissue culture or in plants regenerated from such cultures. Reviews discuss a number of reasons why the tissue culture environment can be stressful to plant cells. They focus especially on the physical nature of the medium, such as high salt, water stress, mineral deficiency, an excess of metal ions and overexposure to auxins (Phillips *et al.*, 1994; Skirvin *et al.*, 1994; Cullis 1999; Arnault and Dufournel, 1994). Genomic changes in tissue culture can result in changes in ploidy level, such as aneuploidy, chromosomal rearrangements (Figure 1.6), activation of transposable elements and other genes (Figure 1.7), point mutations, genome re-arrangements, methylation changes and even altered copy number of sequences (Cullis, 1990; Peschke *et al.*, 1987; Hirochika, 1993; Phillips *et al.*, 1994). Blundy *et al.* (1987) found an almost three-fold reduction in the ribosomal RNA genes in callus cultures of flax. Lee and Phillips (1986) detected chromosomal instabilities in *in vitro* grown maize plants. The extent of these chromosome abnormalities was found to be dependent on the time the cells had been in culture (Chandler *et al.*, 1986). Translocations are a commonly observed chromosome abnormality with inversions, insertions and deletions occurring in the DNA sequence. Repetitive DNA sequences are especially sensitive to stress-related DNA changes and account for a large portion of variation in sequence copy numbers. When cultured *Cymbidium* protocorms were exposed to auxin, the amplification of AT-rich satellite DNA was observed, while exposure to gibberellic acid increased GT-rich regions (Nagl and Rucker, 1976). Zheng *et al.* (1987) found that in rice suspension cultures, highly repeated sequences were amplified up to 75-fold. Another representative of highly repetitive sequences, the ribosomal RNA DNA sequences (rDNA) are also part of the variable component and a decrease in ribosomal RNA genes has been reported in flax callus cultures (Blundy *et al.*, 1987).
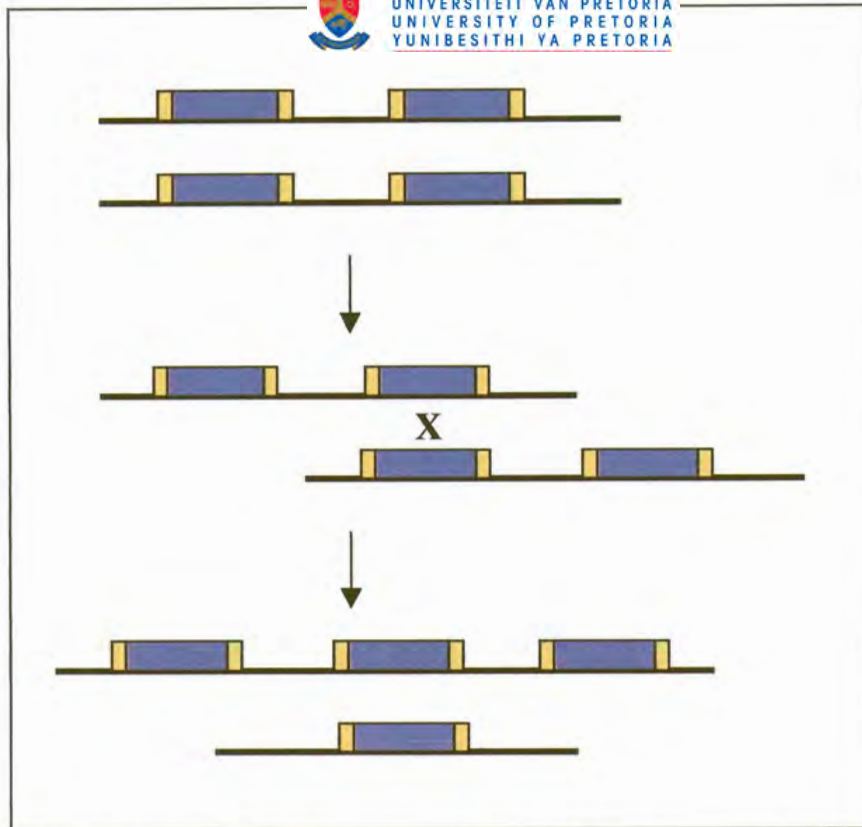
Figure 1.6. Chromosome rearrangements caused by unequal crossing over of transposable elements.
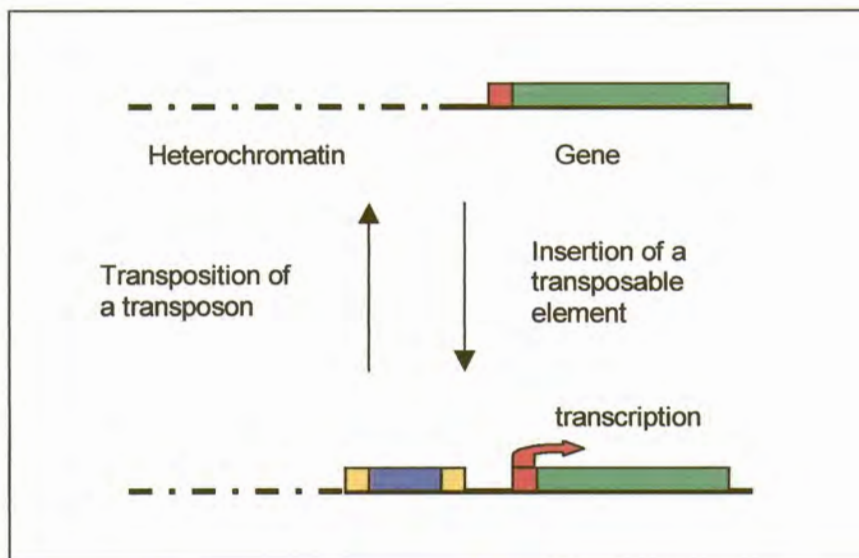
Figure 1.7. Gene activation/deactivation by transposition. Genes located near heterochromatin may be inactive due to the DNA structure. Insertion of a transposable element may move the gene further away from heterochromatin and thereby activating it. Transposition of such an element may again render the gene inactive.

1.4.5 Ribosomal RNA and their genes

Two sizes of ribosomes, 70S and 80S, are known in higher plants, with the 80S ribosomes located in the cytoplasm and the 70S located in the chloroplast and mitochondria. Each of the small subunits of the ribosomes are themselves composed of two or more smaller rRNAs and are repeated and arranged in one or more tandem arrays, termed nucleolar organizer regions (NOR) (Nierras *et al.*, 1997). With the exception of some legumes, almost all plant chloroplast genomes contain two copies of a large inverted repeat which contains the 16S, 23S and 5S rRNA, genes as well as some tRNA and ribosomal protein genes, situated in the opposite orientation and separated by a large single-copy (LSC) and small single-copy (SSC) region (Lu *et al.*, 1996). In contrast, the rRNA unit in the cytosol consist out of the 18S, 5.8S and 25S rRNA coding regions with their non-coding spacers (Haberer and Fischer, 1996), while the mitochondrion rRNA are made up by the 18S, 5S and 26S coding units and non-coding spacers (Heldt, 1997). Copy numbers of rRNA genes are high in most plants with about 570 repeats per haploid genome (Pruitt and Meyerowitz, 1986).
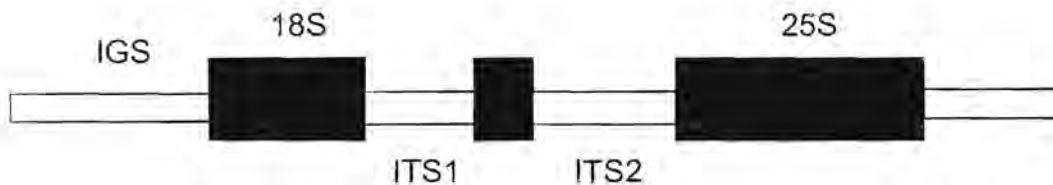
Figure 1.8. Ribosomal RNA genes in the cytosol: arrangement of the 18S-5,8S-25S RNA gene complex. IGS = intergenic spacer; ITS = internal transcribed spacer (Henry, 1997).

1.4.6 Control of retro-transposons

Because retro-transposons have the potential to dramatically alter gene function and host genome structure, it is not surprising that their transposional activities are regulated both by retro-transposon- and host-encoded factors, possibly to avoid deleterious effects on host and retro-transposon survival. The intimate relationship between retro-transposons and their plant hosts has existed for many millions of years. We are just beginning to understand how retro-transposons and their hosts'

genomes have co-evolved mechanisms to regulate transposition, insertion specificities, and mutational outcomes in order to optimise each other's survival.

Because retro-transposons cannot transpose without the presence of an RNA template available for reverse transcription, the simplest way to control their activity would be via the regulation of transcriptional initiation. Many retro-transposons show unique patterns of developmental and/or environmental regulation. A correlation between transcription and transposition of retro-transposons has been demonstrated for the tobacco *Tto*1 and rice *Tos*17 retro-transposons (Hirochika, 1993; Hirochika *et al.*, 1996). For example, transposition of *Tto*1 and *Tos*17 was associated with an increase in the levels of their RNAs, suggesting that transposition of these retro-transposons is regulated mainly at the transcriptional level. However, this is not the situation for several other retro-transposons. The *BARE*-1 of barley is highly transcribed in leaves, but its transposition has not been observed (Suoniemi *et al.*, 1996).

1.4.7 Repetitive DNA and DNA methylation

In plants and filamentous fungi, genomic methylation is restricted mostly to transposons and other repeats (Rabinowicz *et al.*, 1999; Colot and Rossignol, 1999). Most of our current knowledge concerning possible roles for methylation in eukaryotes derives from the study of organisms amenable to genome manipulation. Among these, two filamentous fungi, *Neurospora crassa* and *Ascobolus immerses*, came up with a surprise. Both fungi were found to be endowed with the ability to methylate *de novo* (and concurrently in *Neurospora*, to mutate) all gene-size duplications, and to maintain this methylation vegetatively and sexually (Colot and Rossignol, 1999).

Cytosine methylation is associated with two effects that can serve as defence mechanisms against mobile repetitive elements. First, cytosine methylation, cause a loss of RNA-polymerase-II-dependent transcription in the methylated region, either by preventing transcription initiation or by impending transcript elongation. Therefore, methylation of transposable element sequences can silence the expression of transposon-encoded genes and prevent their amplification and transposon-mediated DNA rearrangements. Secondly, cytosine methylation correlates with reduced homologous recombination between methylated regions (Eggleston *et al.*, 1995). Therefore, methylation of repetitive sequences might suppress recombination

between repeats in different genomic positions – which otherwise would lead to translocations and other chromosomal rearrangements (Bender, 1998). Striking examples of differences in methylation that correlate with differences in repeat content can be found in plants such as *Arabidiopsis thaliana* and maize (Leutwiler *et al.*, 1984; Bennetzen *et al.*, 1994).

How are repeat sequences distinguished from the rest of the genome and targeted for methylation? Evidence suggests that at least some repeated sequences are detected by a DNA-DNA pairing mechanism; unique features of the paired region mark it for methylation along the lengths of the interacting repeats (Selker *et al.*, 1987; Selker and Garrett, 1988; Vongs *et al.*, 1993). But are targeted methylation of repetitive DNA and transposons only a mechanism of silencing or is it also a mechanism to induce targeted mutations and thereby inactivate these transposons? It is known that spontaneous de-amination of 5-methylcyosine (5meC) in dividing cells causes hot spots of CG → TA mutations in *Escherichia coli,* as well as human cells (Lieb and Rehmat, 1997), hence methylation accelerates the divergence rate of newly arising duplications.

1.4.8 Retro-transposon regulation as a form of host defence

Some retro-transposons might have beneficial effects on a plant, through mutations that provide new regulatory properties to a gene or centromere function (Miller *et al.*, 1998). However, many of the expression, mutation, and insertion properties of these elements suggest that their effects are being minimized. With any highly adapted host/parasite interaction, the parasite will contribute as little as possible to the decreased host fitness. The host should also evolve such minimization/defence processes, and several seem to be acting on plant retro-transposons.

Most of the known retro-transposons appear to be defective in their ability to encode all necessary transcription functions, owing to insertions, deletions, and other mutations (Flavell *et al.*, 1992a; Flavell *et al.*, 1992b; Hu *et al.*, 1995; Jin and Bennetzen, 1989; Pearce *et al.*, 1996; Voytas *et al.*, 1992; White *et al.*, 1994). The epigenetic regulation of plant retro-transposons associated with DNA methylation and presumed hetrochromatization may also be involved in keeping retro-transposon transcription at a low level (Yoder *et al.* 1997). This DNA methylation is associated with a two- to threefold higher transition mutation rate in these elements, thus causing them to decay to a non-functioning form mare rapidly than other sequences

1.4.9 Mutations

Mutations, if they have an observable effect, are almost always harmful. Most changes however do not take place in the genes but in the great bulk of the so-called 'junk' DNA, most of which has no known function. Most mutations have very minor effects, if any, but that does not mean that they are unimportant (Crow, 1997). Spontaneous deamination of 5-methylcytosine ($5^{me}C$) causes hot spots of CG → TA mutations. Deamination of 5meC produces thymine, which is not recognised by uracil glycosylase and consequently can result in C → T mutations (Figure 1.9). It is likely that mutation hot spots at 5meC in rapidly dividing cells are attributable to insufficient time for T to G correction in the interval between deamination of 5meC and subsequent DNA replication (Lieb and Rehmat, 1997). Because cells divide rapidly during the tissue culture process as a result of hormones, such as auxins, that are added to the growth medium, and the alteration in methylation patterns observed during tissue culture, this kind of mutations can be expected to be especially higher in their cells.
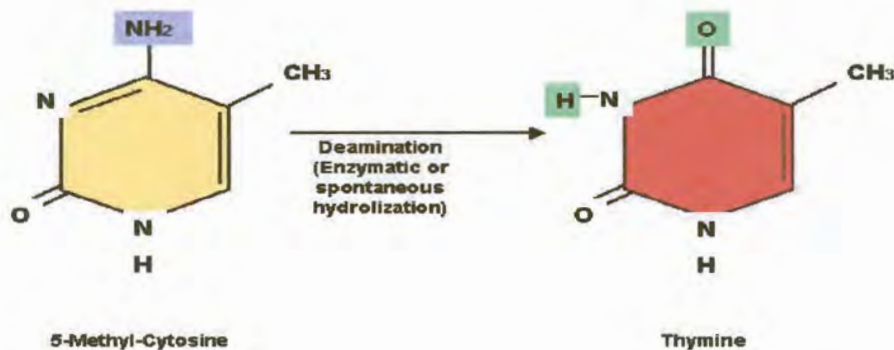
Figure 1.9. Deamination of 5-Methyl-Cytosine converts it into Thymine.

1.5 Detecting genome variation

Molecular markers and marker mapping are part of the intrusive 'new genetics' that is finding its way into all areas of modern biology, from genomics to breeding, from systematics to ecology and also into plant physiology. What are molecular markers? Molecular markers reveal neutral sites of variation at the DNA sequence level. By 'neutral' is meant that these variations do not show themselves in the phenotype, and might be nothing more than a single nucleotide difference in a gene or piece of repetitive DNA (Jones *et al.*, 1997).

1.5.1 Molecular biology and molecular markers

In April 1983, Kary Mullis chanced the course of molecular biology when he conceived the Polymerase Chain Reaction (PCR). The introduction of PCR to molecular biology has enabled the development of powerful genetic markers for the measurement of genotypic variation. By measuring genotype, rather than phenotype, genetic markers avoid complicating environmental effects and provide ideal tools for assessing genetic variation, identification and defining genetic relationships (O'Hanlon *et al.*, 2000).

PCR is an *in vitro* method whereby defined sequences of DNA are enzymatically synthesized. The reaction uses two oligonucleotide primers that hybridise to the opposite DNA strands and flank the target DNA sequence that is to be amplified. The elongation of the primers is catalysed by a heat-stable DNA polymerase. A repetitive series of cycles, involving template denaturation, primer annealing, and extension of the annealed primers by the polymerase result in exponential accumulation of a specific DNA fragment.

The PCR technique is so pervasive in molecular biology that it is difficult to think of life without it. Because of PCR, 'insufficient nucleic acid' is no longer a limitation in molecular biology. More importantly innovative research is continuously updating the definition of 'PCR applications' thereby increasing the usefulness and scope of the technique.

1.5.2 Non-PCR fingerprinting techniques

1.5.2.1 Restriction fragment length polymorphysim (RFLP)

If two DNA molecules are essentially the same, but nevertheless have one or more small differences in their nucleotide sequence, then the fact that they are not identical may become apparent by means of RFLPs (Buscot et al., 1996). The difference between two genomes in the size of the restriction fragments at a defined genetic locus is thus termed a restriction fragment length polymorphism. RFLP analysis is predominantly used to compare closely related species by the comparison of slower evolving regions of their genomes and to assess the diversity within populations (Bruns et al., 1991).

Restriction patterns are generated by the cleavage of DNA with restriction enzymes and separating the bands using electrophoresis. One of two approaches to be followed is the PCR-RFLP, which involves the restriction of PCR amplicons of a selected portion of the genome usually with four- or six-base restriction enzymes. A second approach would be the restriction of the entire genome with restriction enzymes followed by Southern-blotting using selected probes (Maclean et al., 1993).
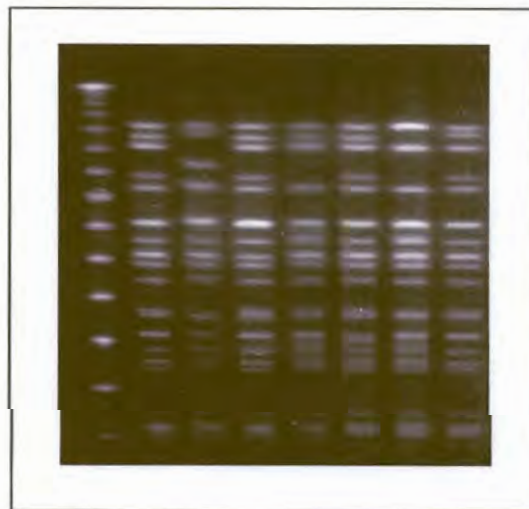
Figure1.10: A typical PCR-RFLP gel profile

There are disadvantages with RFLP analysis, as often a number of probes and enzyme combinations have to be tested in order to generate significant numbers of RFLPs. This method generally uses radioactive probes, and requires a large amount of target DNA. In fact, RFLPs detect about 1 in 10000 polymorphic nucleotides in the human genome (Soller and Beckmann, 1986). The problem with searching for direct associations between RFLP and traits of economic value is the low likelihood of finding one; at best 1 : 200, but probably 1 : 20000 (Soller and Beckmann, 1986). RFLP markers are also difficult to transfer between different linkage maps, as what is polymorphic in one population may well not be in another. This, however, is true for any randomly selected type of marker, but RFLPs do generate a reasonable number of markers, which can be easily placed onto a map.

1.5.3   PCR-based fingerprinting techniques

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. A range of different approaches is used to try and detect these genetic differences. Molecular techniques based on the polymerase chain reaction (PCR) have been used very successfully as a tool in genetic mapping, molecular taxonomy, and evolutionary studies. Among the techniques that are being used in the differentiation of plants are the analyses of r-DNA intergenic regions (Scribner and Pearce 2000), simple sequence repeats (SSRs), which are also known as microsatellites, restriction fragment length polymorphic DNA (RFLP), random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphic DNA (AFLP). The two most widely used molecular techniques to detect plant variation are Random Amplified Polymorphic DNA (RAPD) analysis, which detects DNA polymorphisms amplified by arbitrary primers (Williams *et al.*, 1990) and Amplified Fragment Length Polymorphisms (AFLPs) (O'Hanlon *et al.*, 2000).

1.5.3.1 *Random amplified polymorphic DNA (RAPD)*

This technique was developed by Williams *et al.* (1990) and Welsh and McClelland (1990) to utilize the decreased specificity in binding at low annealing temperatures for short oligo-nucleotides (usually 10 nucleotides). Large numbers of fragments are amplified by this method, some of which are polymorphic. RAPD analysis can be a

very useful tool for characterizing genetic variability among different cultivars and varieties as also shown in this study. It is very simple and sensitive and provides a PCR fingerprint for related organisms based on the genome rather than individual genes (Foster *et al.*, 1993). This technique scans the DNA for short inverted repeat sequences and amplify inverting DNA segments (Hadrys *et al.*, 1992). A PCR is carried out by the use of a single primer or primer set, usually 9-10 nucleotides in length (Foster *et al.*, 1993). These primers find homology on the template DNA and generate random amplified polymorphic DNA by initiation and extension. The different band sizes can be analysed by electrophoresis to generate specific banding patterns.

The RAPD technique offers several advantages. It can produce more polymorphisms than the 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).

RAPD markers have been used in many species for a variety of investigations: gene cloning, medical diagnostics and trait intro-gression in breeding programs (Williams *et al.*, 1990). Levin *et al.* (1993) used RAPD markers to generate new markers on the Z chromosome of the chicken in order to identify sex-linked traits. RAPD markers have also been useful in determining phylogenetic relationships between species as demonstrated by Barral *et al.* (1993) with the *Shistosoma* genome. Recently, the RAPD technique has been applied to identify date palm varieties (Corniquel and Mercier 1994; Sedra *et al.*, 1998) which has also been a subject of this study.

Because of the unreliable nature of RAPDs certain modifications and improvements on the technique can be made to produce a more reliable differentiation system. Paran and Michelmore (1993) developed sequence-characterized amplified regions (SCARs), also done in this study. SCARS are derived from RAPD markers by developing longer primers. After the RAPD fragment is cloned and sequenced, a pair of primers are designed and synthesized. These SCAR primers are then used to amplify the specific regions of DNA. SCAR markers are advantageous over RAPD markers because they usually detect only a single locus, their PCR amplification is less sensitive to reaction conditions, and they are more likely to be co-dominant markers. SCARs have been used for mapping genes of interest, map-based cloning,

and marker-assisted selection (MAS) on several plants including lettuce (Paran and Michelmore, 1993), common bean (Gu *et al.*, 1995), oak (Bodennes *et al.*, 1997), and citrus (Deng *et al.*, 1997).

RAPDs are widely used as an easier alternative for RFLPs, because less information about the genome is required, it is less expensive and is much faster than RFLPs (Foster *et al.*, 1993). This technique provides the advantages of simplicity of minute amounts of sample DNA and a universal set of primers can be used for all living organisms, constructed without any knowledge of the organism's genome. Some disadvantages might be that this technique requires a lot of standardization, is unable to distinguish between homo- and heterozygotes and differences in band intensity. It is also difficult to obtain the same results repeatedly, which makes it sometimes difficult to draw solid conclusions. RAPD markers are the least informative of all fingerprinting techniques, but they are detected much more easily than RFLPs (Foster *et al.*, 1993).
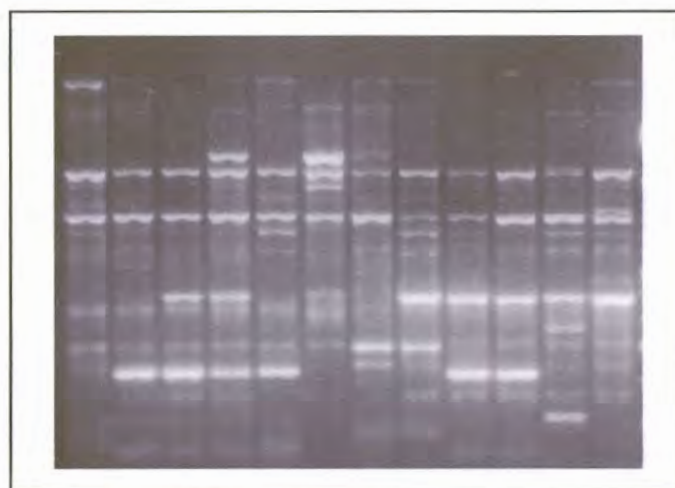
Figure 1.11: A typical RAPD gel profile.

1.5.3.2 *Amplified fragment length polymorphism (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. Adaptors are very important in this technique, because the restriction site sequence at the end of the fragments is insufficient for primer design (Karp *et al.*, 1997).

If these restriction 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 pair/s, 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. 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.

Radio-labelled primers can be used to visualize the amplified products with exposure to X-ray film, but the cost and danger involved make non-radiolabelled 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).

Figure 1.12. A typical AFLP gel profile (Russel 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.

1.5.3.3 *Representational difference analysis (RDA)*

The techniques described above rely on patterns consisting of the presence or absence of DNA fragments rather than DNA sequence variation. Understanding DNA sequence variations should allow us to understand the genetic basis of evolution, the genetic control of development, as well as the physiological abnormalities and variation. It is well known that genetic variation can result from many phenomena such as genomic rearrangements, gene duplication, viral insertions, deletions, or simple base pair changes. Detecting these sorts of variations, however, has been very difficult due to the complexity of the genomes being analysed and the small size of the differences.

In the past, subtractive hybridisation has been used with some success to identify large differences between two genomes, such as insertions or deletions. Subtractive hybridisation involves hybridising DNA containing the sequence of interest with large amounts of the DNA lacking these sequences. Hybridisation is usually followed by the physical separation of the undesired sequences from the target sequences, using methods such as chromatography. However, the larger and more complex the genomes or the smaller the difference between them, the more difficult this separation becomes. Successful subtractive hybridisation usually results in only a 10- to 100-fold enrichment of the target sequences and therefore must be followed by the

time-consuming process of sorting through large numbers of DNA sequences to find the one of interest. A new technique has recently been described that eliminates the classical problems of subtractive hybridisation and will allow direct isolation of small differences between two complex genomes without having to sort through excess products. Lisitsyn *et al.* (1993) have developed a technique called Representational Difference Analysis (RDA). RDA couples subtractive hybridisation with polymerase chain reaction (PCR) to amplify exponentially only the target DNA sequences after repeated rounds of subtractive hybridisation.

RDA is a two-step technique. In the first step a representation of the genome is created by digesting the genome with an appropriate restriction enzyme and ligating adaptors to the restriction fragments, which is then amplified by a PCR step to form amplicons. (Figure 1.13). In the second step new adapters are ligated to the amplicons of the representation of the tester genome and the amplicons of the tester and driver are subtracted from each other. This results in the amplification of unique DNA sequences found in the tester (Figure 1.14)

RDA allows one to target regions of the genome that differs in more than single nucleotide differences between two genomes. These differences may represent unique sequences, genome rearrangements differences in copy number or sequences with a high mutation rate. This allows one to find regions of genetic variation in organisms were genome variation might otherwise be low or hard to find.
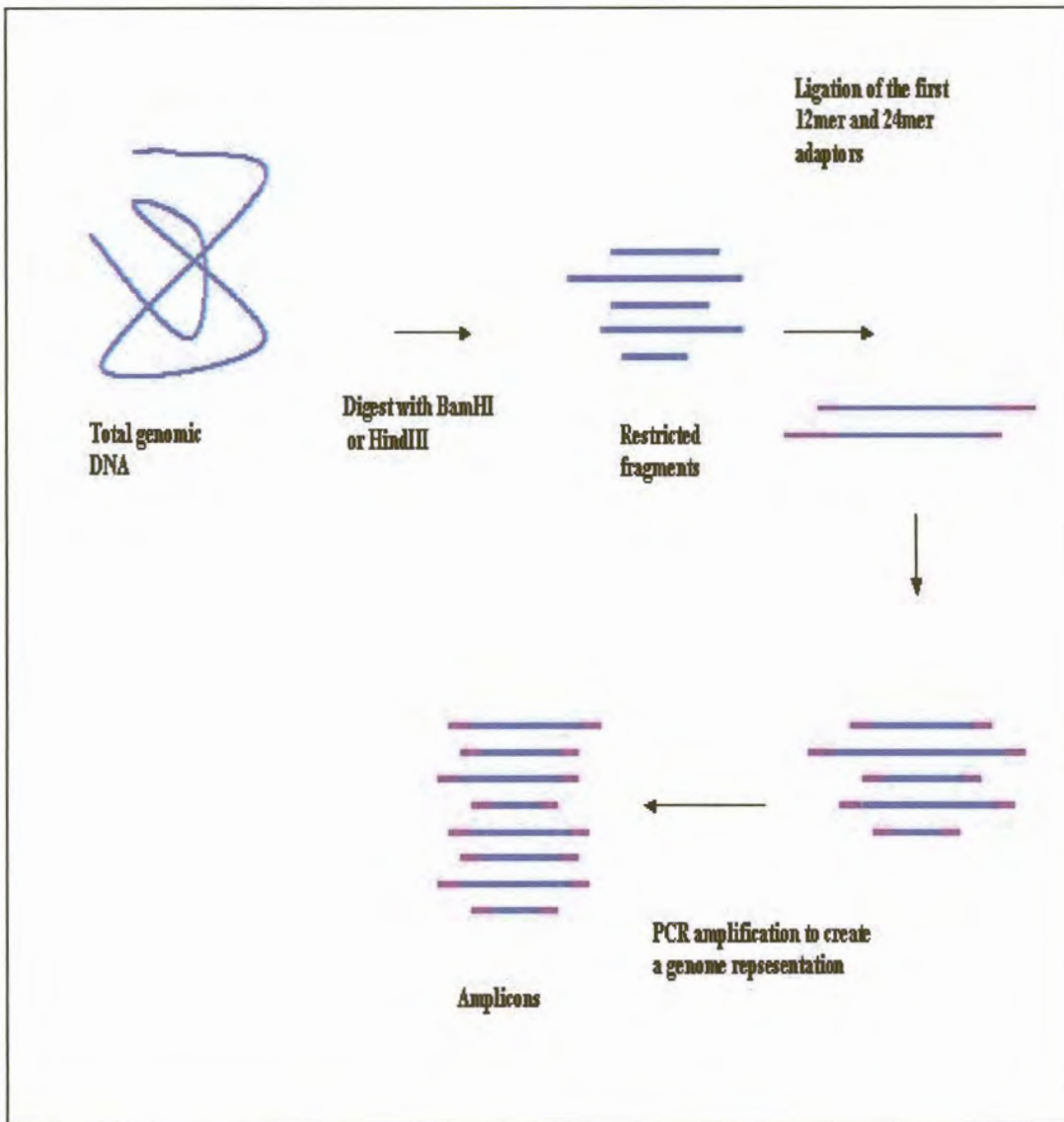
Figure 1.13. Generation of RDA amplicons from total genomic DNA
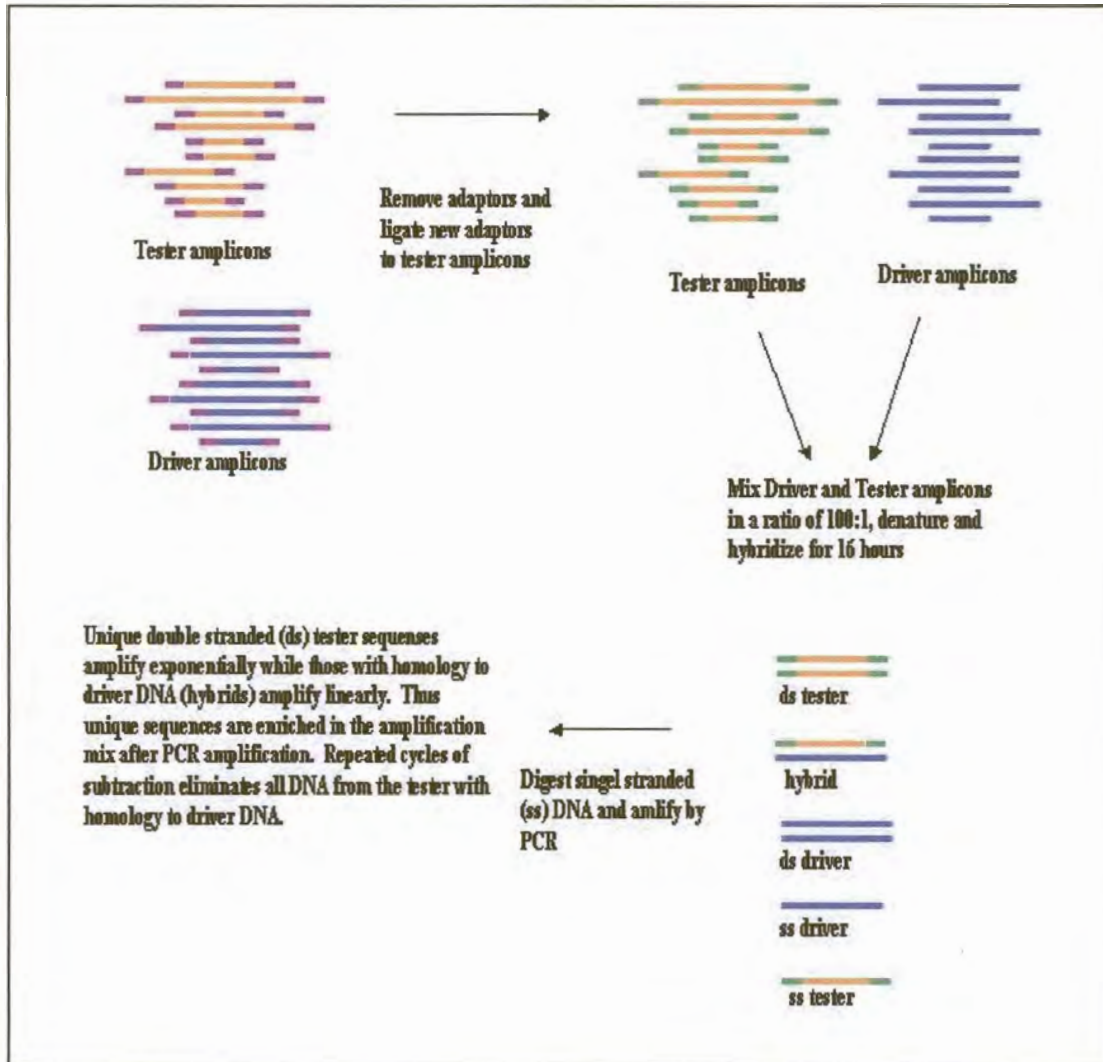
Figure 1.14. Subtractive hybridisation of tester and driver amplicons and kinetic enrichment of differences.

1.5.4 First trials of RDA

Several uses and successful trials of RDA as well as the technique were described by Lisitsyn *et al.* (1993). The originality of this approach relies on the fact that subtraction is applied to a fraction amplified by PCR (a representation) rather than the entire genome, thus making genomic subtraction applicable to complex genomes. The use of different restriction enzymes can provide several representations of the genome, if necessary. To show that RDA could detect viral DNA in the human genome, the authors used human genomic DNA, mixed with lambda phage DNA or adenovirus, as the tester. They used the same human

genomic DNA without the viral DNA as driver. These two populations, therefore, were identical except for a few additional phage or viral sequences in the tester. They found that the small restriction fragments of the adenovirus and the lambda phage were the only difference products after 3 rounds of subtractive hybridization and PCR amplification. The large restriction fragments present in the added viral DNA were not isolated by RDA, consistent with the hypothesis that only small restriction fragments would amplify during PCR.

RDA was next shown to identify RFLPs between two closely related individuals. DNA from two sisters of an Amish family with an established pedigree was analyzed using RDA. A complex, yet clear, pattern of difference products was seen on an agarose electrophoresis gel after 3 rounds of subtractive-hybridization and amplification. Five distinct difference products were isolated and all were used in Southern blot analysis to identify unique RFLPs between the sisters. The inheritance of these RFLPs was analyzed in other relatives and each identified a distinct genetic locus with two different alleles that were inherited according to Mendelian genetics. This application of RDA could be useful in identifying loci linked to inherited disorders by analyzing individuals in families segregating the disorder.

Lisitsyn et al. (1993) proposed several future applications of RDA (1) to detect genetic abnormalities that result in cancer, (2) to generate RFLP markers that exist between related species or individuals to be used in genetic mapping, and (3) to identify loci linked to genetic diseases that result from spontaneous mutations or rearrangements in the fertilized egg. In addition, RDA was proposed as a method of detecting RFLPs linked to a mutant gene of unknown location in any organisms that can be bred. This last possibility could greatly benefit the field of molecular biology by speeding up the process of mapping and cloning genes.

Recently, RDA was used to isolate sequences that are unique to the tumorogenic tissue of patients with AIDS-associated Kaposi's sarcoma (Chang et al., 1994). RDA has also been used to isolate families of repetitive sequences present in only one of compared genomes (Navin, 1996). Further, Nekrutenko et al. (2000) used RDA to create a species-specific marker for voles and Toder et al. (2001) have applied RDA in evolutionary genomics to search for overall genome differences between humans and the great apes. RDA has also been used to determine differences between two distantly related oak species where similarities of isolated RDA fragments with known retro-transposons were found (Zoldos et al., 2001). In addition, Donnison et al. (1996) applied RDA to identify male-specific restriction fragments in the dioecious

plant *Silene latifolia*. RDA has also been used to identify polymorphisms in banana lines that are a result of genomic rearrangements during *in vitro* propagation resulting in markers useful for the detection of early variation in the initiation of tissue culture plants (Cullis and Kunert, 2000). One of the specific advantages of RDA is that subtractions between pooled DNA samples can be performed in order to identify specific polymorphisms only present in a particular individual rather than relying on Identification based on a particular pattern of polymorphic bands.

In the following chapters results are presented regarding the outcome of experiments using the RDA technique to subtract genomes from date palm varieties as well as the outcome of methylation sensitive RDA subtractions done on date palm and pine clones.

Chapter 2

Differentiation of date palm varieties with Random Amplified Polymorphic DNAs (RAPDs)
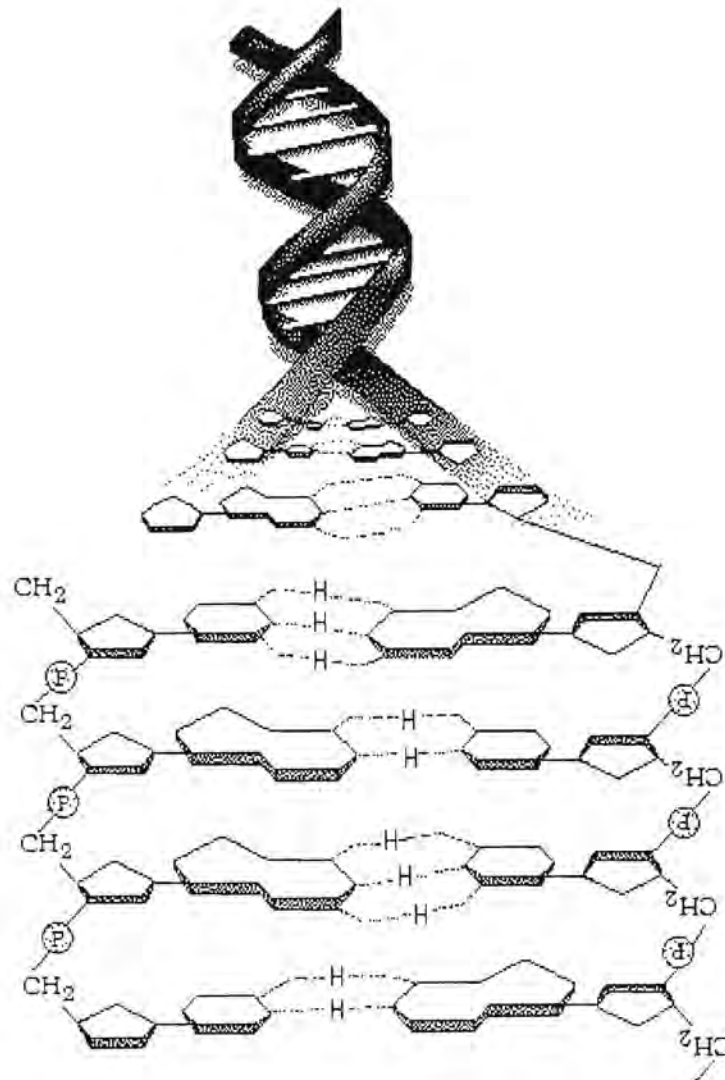
2.1 Abstract

Identification of plant cultivars where no early morphological differences are visible is crucial for many plant producers. Random Amplified Polymorphic DNA or RAPD is among the commonly used techniques for plant differentiation. To test the applicability of this technique for date palm, genomic DNA from tissue culture-derived date palm plants of the cultivars 'Barhee' and 'Medjool' were amplified for RAPD analysis using the commercially available DNA primers OPE-01 and OPE-06. Both primers could differentiate the two cultivars. In an attempt to produce a more robust amplification, sequence information of a 700 bp fragment amplified from 'Medjool' was used to create a SCAR primer pair. However, this SCAR primer pair amplified an dentical DNA fragment from both tested genomes.

2.2 Objective

The first objective of this part of the study was to confirm a reported differentiation system for date palm using the RAPD technique. The second objective was to develop a SCAR primer pair from one of the amplification products to obtain a more robust test system. In particular (1) two reported RAPD primers, OPE-01 and OPE-06, were applied in this RAPD study (Corniquel and Mercier 1994) to differentiate the two date palm cultivars 'Medjool' and 'Barhee' and (2) an attempt was made to develop a SCAR primer pair from one of the sequenced amplification products.

2.3 Results

2.3.1 DNA extraction

Two methods were used for genomic DNA extraction from date palm. The CTAB (cetyltrimethylammonium bromide) method resulted in an amount of about 320 µg genomic DNA isolated per gram fresh plant leaf material. In comparison, 5 µg genomic DNA per gram of fresh material were obtained using the Nucleon Phytopure Plant DNA extraction kit (Figures 2.1 and 2.2). DNA isolated according to the CTAB method was, however, of a much lower quality when compared to the Nucleon

33

Phytopure system when used for DNA amplification. After repeated attempts to optimize the PCR reaction, amplification of genomic DNA isolated with the CTAB method resulted in either smears on the agarose gel or complete failure of amplification by PCR. The shift in 'Barhee' DNA band in figure 2.2 lane 5 may be due to the larger genome size as a result of the greater number of repetitive elements found as shown in this study.

    1   2   3   4   5   6
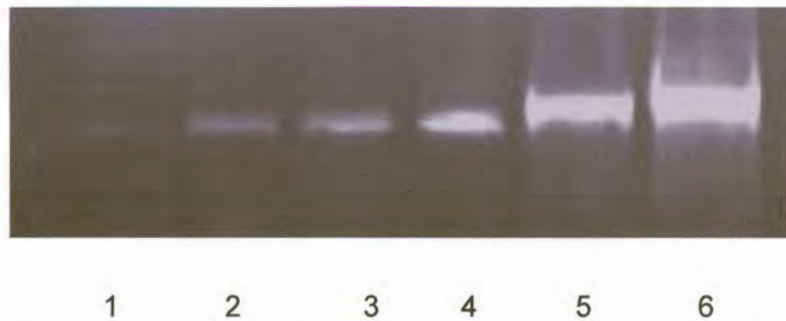
Figure 2.1. DNA extraction from date palm using the CTAB method. Lanes 1-4: 25 μg (1), 50 μg (2), 100 μg (3) and 250 μg λ-phage DNA (4) as standards to determine the amount of genomic DNA isolated; lane 5: 2 μl 'Barhee' DNA; lane 6: 2 μl 'Medjool' DNA

    1   2   3   4   5   6
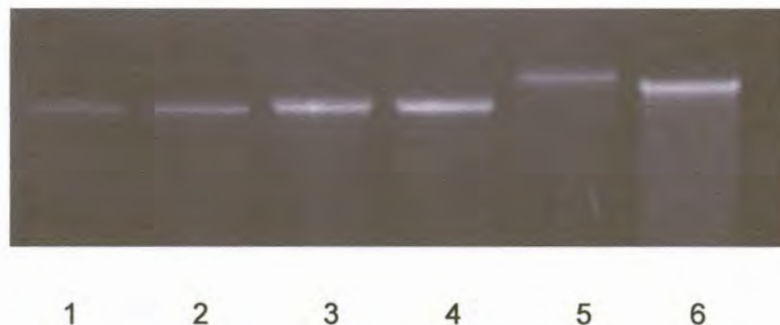
Figure 2.2. DNA extraction from date palm using the Nucleon Phytopure Plant DNA extraction kit. Lanes 1-4: 25 μg (1), 50 μg (2), 100 μg (3) and 250 μg λ-phage DNA (4) as standards to determine the amount of genomic DNA isolated; lane 5: 2 μl 'Barhee' DNA; lane 6: 2 μl 'Medjool' DNA

2.3.2 RAPD analysis

Application of primers OPE-01 and OPE-06 differentiated the two cultivars. OPE-01 amplified a distinct DNA fragment from 'Barhee' genomic DNA of about 1700 bp but not from 'Medjool' genomic DNA (Figure 2.3). Primer OPE-06 amplified a fragment of about 700 bp from 'Medjool' but not from 'Barhee' genomic DNA. A control test with all PCR reagents except any template DNA resulted in no amplification of any product (Figure 3.3). In addition to the two primers OPE-01 and OPE-06, the primer OPB-07 was also used randomly. However, this primer did not distinguish between the two cultivars under the conditions used in this study.
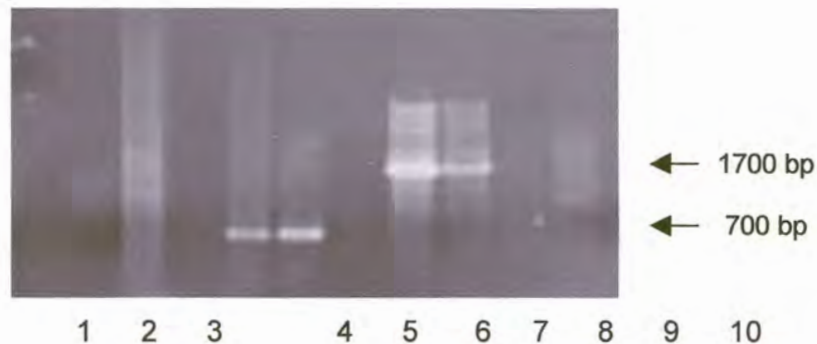
Figure 2.3. RAPD analysis of the date palm cultivars 'Medjool' and 'Barhee'. Lane 1 represents a control without genomic DNA but with primer OPE-06; lanes 2 and 3: amplification of 'Barhee' DNA with OPE-06; lanes 4 and 5 amplification of 'Medjool' DNA with OPE-01 producing a 700 bp fragment; lane 6: control without template DNA but with primer OPE-01; lanes 7 and 8: amplification of 'Barhee' DNA with primer OPE-01 producing a 1500 bp fragment; lanes 9 and 10: amplification of 'Medjool' DNA with primer OPE-01.

2.3.3 Characterization of amplification product

The 700 bp DNA fragment amplified from 'Medjool' genomic DNA was cloned into the vector p*MosBlue* and then sequenced. Sequencing of the cloned fragment was done in both directions in order to obtain a complete sequence of the DNA fragment. Analysis of the fragment including the OPE-06 primer (Figure 1.4) revealed, however,

no significant homology to any known plant sequences when a Blast search was performed.

```
  1  AAGACCCCTCCATGCTGGATTTTATTCTGATTCTACNTCAGTCCCCTCCA
 51  CTTGAATACAACTCGTCCTCGAGTTCTCATCTGGATATAAATGTTCCTCA
101  GGCTATGGTATTCATAGATATCAGCGACATTAAAAGTCTTTGAAATTCTC
151  ATATCATCTTTGTAATAACCCCAAAATATTTTTTTATATAAAAAGGGAT
201  AGAATAGTCCTTTTCAATTAATATAAGGAGTAAAATAGGAAAGAATCAAA
251  AGATAATGGCATAGTTGTAGATGCATTGAAGTGGTAAGGGTAAAATTGGA
301  AGGAATCAAAAGTTAATGGCATAACGGTAGATATGTTGAAGTGTTAGGGG
351  CAAAATGGAAATTTAATAATATTAAACAAAGGGTGAATAGTGTTTTGATG
401  TGATTTAATTAGGGGGTTTGCTGTGGCTTTTGGATGGAAACCCGAGAGGG
451  AGTCTCAGACGGACAAAGAGGAAGAAAAAGAAGGAGAGGAAAGGAAAGGA
501  AGAAAGGAAGAGGAAAGGAAAGGAAGAAAGGAAGAGGAAGGAGATCCCGG
551  TTGCATTTCTAGCTGGAGGGAAAAACGTAGATCTCCTTCTCCTCTACACA
601  AGAACCTCGATTTCCAAAAAGAAAAAGGTAAATATCCATCCCTATCTAGT
651  CTTTTGATGACATTANGCTATACNAANGGTGGATATAGTCTANAAGGGTC
```

Figure 2.4. Sequence of the 700 bp fragment amplified with primer OPE-06 from 'Medjool' genomic DNA. OPE-06 primer site (underlined) and sites used to design the SCAR primers DpSL and DpSR are underlined.

2.3.4 SCAR primer design and testing

The SCAR primers DpSL and DpSR were constructed from the sequence obtained from the polymorphic band amplified from 'Medjool' with OPE-06 (Figure 2.4). The two 20-mer primers (Table 2.1) were designed using a standard design program (Expasy, Switzerland). The primer pair amplifying a 128 bp DNA fragment was used in a PCR reaction with 'Barhee' and 'Medjool' DNA as template at various annealing temperatures to optimize the PCR reaction. Fragments of the same size (about 128 bp) were, however, amplified from both types of genomic DNA (Figure 2.5) and therefore SCAR primers failed to differentiate between the two date palm cultivars. A further 400bp fragment was also amplified from genomic DNA of one of the 'Medjool' plants, but could not be amplified in other genomic 'Medjool' DNAs tested.
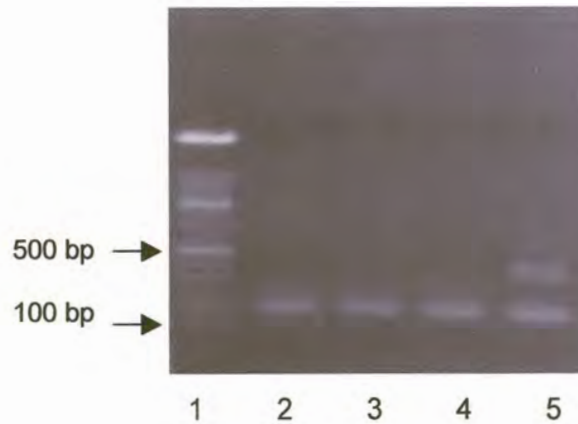
Figure 2.5. SCAR analysis of 'Barhee' and 'Medjool' genomic DNA. Lane 1 represents a 100 bp molecular marker ladder (Roche, Switzerland); lanes 2 and 3: amplification of 'Barhee' genomic DNA with SCAR primers; lanes 4 and 5: amplification of 'Medjool' genomic DNA with SCAR primers.

Table 2.1. Primer sequences used for RAPD and SCAR analysis.

Primer	Sequence
OPE-01	5'-CCCAAGGTCC-3'
OPE-06	5'-AAGACCCCTC-3'
DpSL	5'- GTGTTAGGGGCAAAATGGAA–3'
DpSR	5'- TTGTCCGTCTGAGACTCCCT–3'

2.4   Materials and methods

2.4.1   Plant material and DNA extraction

Tissue culture plants of 'Barhee' and 'Medjool' were used for genomic DNA extraction. The *in vitro* plants used were 'Medjool' derived from explant material collected in California, and 'Barhee' derived from explant material collected in the United Arabic Emirates. Total cellular DNA was isolated in a first method from the entire plant (1 g) using the Nucleon Phytopure Plant DNA extraction kit (Amersham Life Science, UK) according to the manufacturer's instructions.

In a second method for genomic DNA extraction, the protocol of Ait-Chitt *et al.* (1993) was used. This method is based on CTAB precipitation of the DNA. For DNA isolation leave tissue (1 g) from tissue culture plants was used. Leaf tissue was ground to a fine powder with a mortar and pestle in liquid nitrogen. Grinding was assisted with acid-washed sand. The powder was then homogenized in 7ml of DNA extraction buffer (1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, 3% w/v CTAB, 1% v/v 2-mercaptoetanol) and incubated at 65°C for 30 min. This mixture was then extracted with an equal volume of chloroform-isoamyl alcohol (24 : 1), and the DNA in the aqueous phase precipitated with an equal volume of isopropanol. The DNA was collected by centrifugation, washed with 70% (v/v) ethanol and dissolved in TE buffer. The DNA was then treated with RNase.

2.4.2 RAPD analysis

For RAPD analysis, oligo-nucleotide primers OPE-01 and OPE-06 from Operon Technologies Inc. (Alameda, CA) were used. For detailed PCR reaction conditions see Annex.

2.4.3 DNA isolation from agarose gels and cloning

Amplified band (700 bp) from genomic 'Medjool' DNA was cut out of the gel with a scalpel and purified with a Sephaglas™ BrandPrep kit (Pharmacia Biotech Inc, USA) using the recommended protocol of the supplier for DNA purification from agarose gels. The isolated fragments were cloned into the vector pMos*Blue*. For detailed cloning procedure see Annex.

2.4.4 Sequence analysis

Sequencing reactions were carried out as described under Annex.

Chapter 3

Isolation of RDA difference products from date palm

3.1 Abstract

Several research groups have applied RDA (Representational Difference Analysis) especially in medical sciences and the technique has been predominantly used on the cDNA level to study differences in gene expression. Application of RDA on the genomic level and in plant research is still very limited, which is very likely due to its complexity in execution on large plant genomes. RDA was used in this part of the study to demonstrate its applicability in genome differentiation and to isolate differences in the genome sequence of two closely related date palm cultivars. Several RDA subtraction products with an approximately same size were isolated from 'Barhee' genomic DNA after *Bam*HI digestion of genomic date palm DNA. RDA subtraction products were cloned into an appropriate cloning vector allowing sequence analysis. Subtraction products could not be detected after digestion of genomic date palm DNA with *Hind*III followed by the subtraction of genomes from the two investigated date palm genomes.

3.2 Objective

The objective of this chapter of the study was to evaluate the applicability of the RDA for date palm genome analysis to isolate possible genomic sequence differences between two closely related date palm cultivars. In particular, the genomes of the two date palm cultivars 'Barhee' and 'Medjool' were subtracted from each other after digestion of genomic DNA of both cultivars with either the restriction enzyme *Bam*HI or *Hind*III.

3.3 Results

3.3.1 Isolation and digestion of genomic DNA

Genomic DNA of the two date palm cultivars 'Barhee' and 'Medjool' digested with either the the restriction enzyme *Bam*HI or *Hind*III revealed no obvious differences in the DNA patterns of the two cultivars after agarose gel electrophoresis and staining of DNA with ethidium bromide (Figure 3.1).

1 2 3 4

Figure 3.1. Restriction enzyme digestion of total genomic DNA from two date palm cultivars separated on a 1% agarose gel and stained with ethidium bromide. Lane 1 represents 'Barhee' genomic DNA cut with *Bam*HI cut; lane 2 'Barhee' DNA cut with *Hind*III, lane 3 'Medjool' DNA cut with *Bam*HI cut and lane 4 'Medjool' DNA cut with *Hind*III.

3.3.2 Amplification and subtraction of genomic DNA

After genomic DNA digestion, adaptor sequences were ligated to digested genomic DNA to allow amplification of digested DNA by the PCR reaction. By executing a PCR reaction using adaptor sequences as PCR primers, DNA representations (amplicons) of each date palm genome were produced using a PCR reaction. Figure 3.2 shows the amplicons produced from *Bam*HI digested genomic DNA after agarose gel electrophoresis.

After amplicon production, RDA subtractions were performed between 'Barhee' and 'Medjool' amplicons using either 'Barhee' DNA as tester and 'Medjool' as driver or with 'Medjool' DNA as tester and 'Barhee' DNA as driver. Following a single round of subtraction using a tester to driver ratio of 1 to 100 only one of the two subtractions produced subtraction products of approximately 150 bp (Figure 3.3). This was when 'Barhee' amplicon DNA was used as tester and 'Medjool' DNA as driver but not when 'Medjool' amplicon DNA was used as tester and 'Barhee' DNA as driver.

Figure 3.2. First round amplicons from *Bam*HI digested genomic DNA of two date palm cultivars after separation on a 1% agarose gel and staining with ethidium bromide. Lane 1 represents 'Barhee' amplicon and lane 2 'Medjool' amplicon.
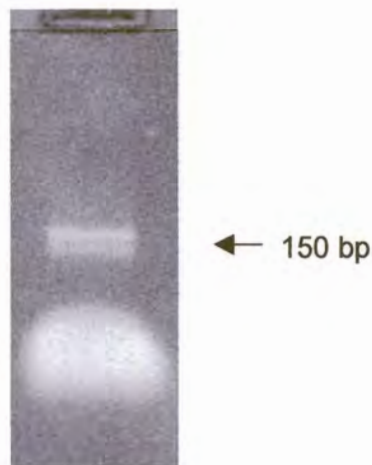
Figure 3.3. RDA subtraction product separated on a 1% agarose gel and staining with ethidium bromide after subtracting 'Barhee' amplicon DNA from 'Medjool' amplicon DNA.

3.3.3 Cloning and hybridization of subtraction products

After extraction of the region DNA from the agarose gel, the purified total extracted DNA was cloned into the unique *Bam*HI site of vector p*BlueScript*, which was used to

transform *Escherichia coli* cells of the strain JM109. After selection on an ampicillin-containing medium, fifty *Escherichia coli* (*E. coli*) colonies containing the cloned difference product were hybridized with randomly labeled 'Barhee' amplicon DNA (Figure 3.4) or 'Medjool' amplicon DNA (data not shown). Both sets of labeled amplicons hybridized to all colonies with different intensity indicating that the isolated subtraction products were not unique to 'Barhee' genomic DNA.
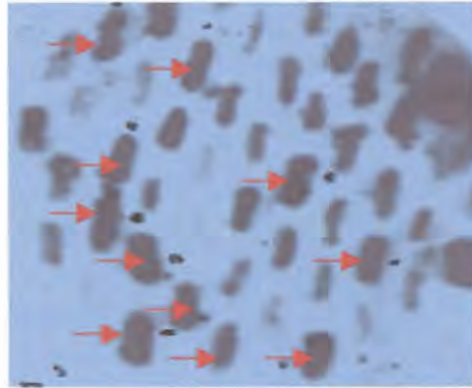
Figure 3.4. Hybridisation of *E.coli* colonies carrying plasmid DNA with subtraction product inserts to randomly labelled 'Barhee' amplicon DNA. Red arrows indicate the *E. coli* colonies that were selected.

To select for the cloned inserts with the strongest homology to 'Barhee' genomic DNA, ten *E. coli* colonies with the strongest hybridisation signal after probing with labelled 'Barhee' amplicon DNA were selected. The plasmid DNA was then isolated from colony and the cloned subtraction product inserts released after *Bam*HI digestion. Figure 3.5 shows that two different sizes of subtraction product inserts had been cloned into the plasmid.

← 140 bp

Figure 3.5. Subtraction product inserts after separation on a 1% agarose gel and staining with ethidium bromide after release from plasmid p*BlueScript* digested with *Bam*HI. (*) indicates the two inserts Dp41 and Dp50 with a different size.

3.4 Materials and methods

3.4.1 Plant material and DNA extraction

'Medjool' and 'Barhee' plant material originated from 14 plants grown in California that have been propagated solely via off-shoots from mother plant material imported in the beginning of the last century from Northern Africa (Nixon 1950). The *in vitro* plants used were 'Medjool' derived from explant material collected in California, and 'Barhee' derived from explant material collected in the United Arabic Emirates. Total cellular DNA was isolated from leaves (1 g) using the Nucleon Phytopure Plant DNA extraction kit (Amersham Life Science, UK) according to the manufacturer's instructions.

3.4.2 Representational Difference Analysis

3.4.2.1 Preparation of RDA amplicons

RDA was performed following the general outline described by Lisitsyn et al. (1993) as described in the annex. Total genomic DNA was digested with the restriction enzymes *Bam*HI and *Hind*III. Amplicons were prepared by ligating either the adaptor pair Rbam 12 & Rbam 24 or Rhind 12 & Rhind 24 (see appendix for sequences) to the digested DNA. After ligation the DNA was amplified in eight 100μl volumes using a Perkin-Elmer GeneAmp 9600 thermocycler as described in the annex.

3.4.2.2 Removal of the adaptors from amplicons

Of the amplicons that was to be used as driver 150 µg and 10 µg of the tester amplicons were digested with the appropriate enzyme. The driver and tester DNAs were both redissolved at approximately 400 µg/ml. Both the digested driver and tester amplicons were run on a 1.5% TAE agarose gel alongside an equal aliquot of undigested amplicons to check the completeness of the digestion. On the same gel, standard lambda phage DNA was run so that the final concentrations of driver and tester DNAs could be estimated and adjusted if necessary.

3.4.2.3 Change of adaptors on tester amplicons

Tester DNA was prepared by adding a second adaptor pair JBam 12 and 24 (sequences in annex) for *Bam*HI digested DNA or JHind 12 and 24 (see annex for sequences) for *Hind*III digested DNA to 1 µg of the first round amplicons. An aliquot of the ligate was amplified for 20 cycles in a reaction volume of 20 µl to check that the newly ligated adaptors would support amplification with the new primer.

3.4.2.4 Subtractive hybridisation and kinetic enrichment

The hybridisation reaction was set up by mixing the driver and tester amplicons at a ratio of driver : tester of 100 : 1. The hybridisation mixture was then precipitated and the pellet redissolved in 4 µl 3X EE (30mM EPPS, 3 mM EDTA; pH 8) buffer. The solution was collected at the bottom of the tube and overlaid with light mineral oil so that the spherical droplet could be seen to be completely covered by oil. The DNA was then denatured at 98°C for five minutes and one µl of 5 M sodium chloride solution was added. Hybridization was done overnight at 67°C. Following hybridisation the reaction mixture was amplified by PCR for 10 cycles after which the amplicons were digested with mung bean nuclease (Amersham Life Science, UK) remove all single stranded DNA. Following the mung bean nuclease treatment the hybridization mixture was once again submitted to PCR amplification.

3.4.3 Cloning of the difference products

Two µg of these subtraction products were digested with the 50 units of the appropriate restriction enzyme and cloned into the p*Bluescript*II vector (Strategene, USA) that was used to transform competent XL1*Blue* cells. Fifty plasmid-containing colonies carrying an insert were selected and probed with either the 'Barhee' or 'Medjool' labeled driver amplicons using the *Gene Images* random prime-labelling module (Amersham Life Sciences UK). Ten colonies that showed a much stronger signal after hybridisation with the 'Barhee' amplicons than with the 'Medjool' amplicons were selected for plasmid isolation and determination of the insert sequence and size.

Chapter 4

Characterization of RDA subtraction products using bioinformatic tools
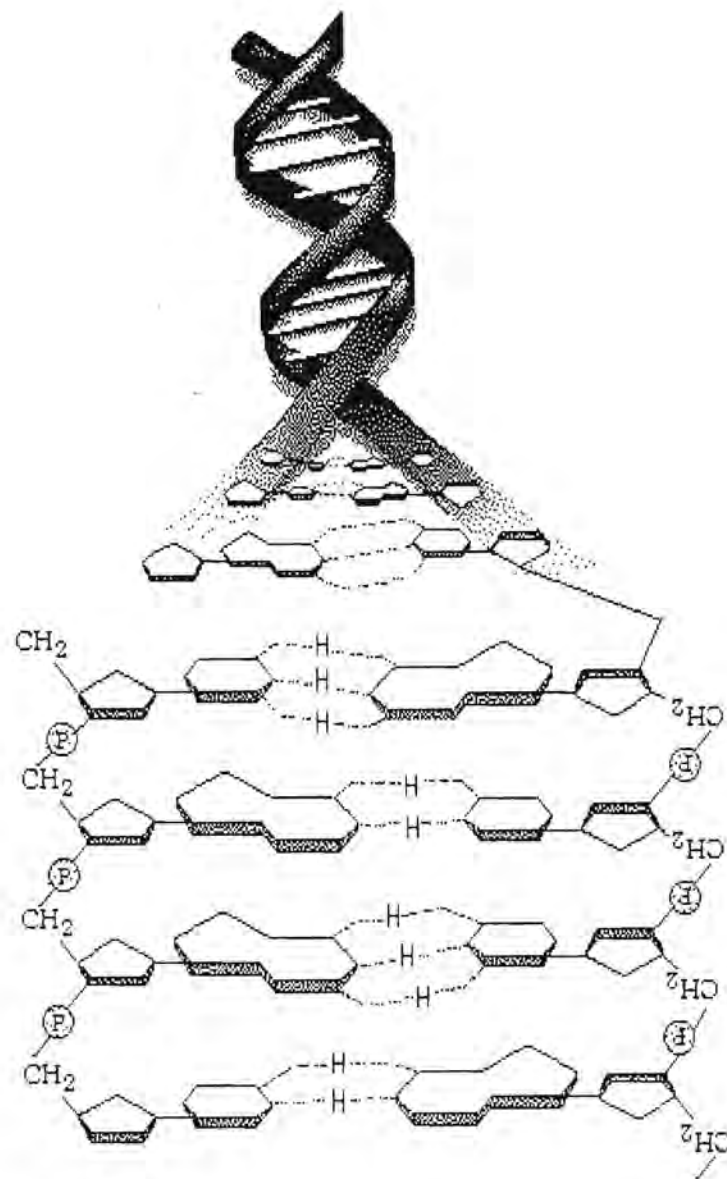
4.1 Abstract

Bioinformatics has advanced genetic research through the analysis of DNA sequences, the computerized processing of sequence data and comparison of unknown sequences with known DNA sequences through database searches. In this chapter subtraction products obtained by subtracting the two date palm genomes derived from the varieties 'Barhee' and 'Medjool' using the RDA technique were characterized with various bioinformatic tools. Results obtained showed that isolated subtraction products belong to the group of plant repetitive elements with a variety of sequence differences between the different copies of the Dp41 repetitive element, which was analysed in greater detail.

4.2 Objectives

The objective of this chapter was to characterize isolated subtraction products from the two date palm varieties 'Barhee' and 'Medjool' using bioinformatic tools to obtain detailed information about the DNA sequence of products and possible homologies to known DNA sequences available in DNA databases.

4.3 Results

4.3.1 DNA sequence analysis

Sequence analysis of the cloned subtraction products was carried out by using a commercial sequencing kit (Roche Molecular Biochemicals, Mannheim, Germany) using T7 and SP6 primers (see annex for sequence). DNA sequence analysis of 11 cloned subtraction products, revealed that these products consisted of at least three types of sequences with lengths of 141 bp (Dp41), 147 bp (Dp36), and 156 bp (Dp2) (Figure 4.1). When these three DNA sequences were aligned with all the DNA sequences determined, most of the sequenced subtraction products were homologous to Dp2, with only slight base differences occurring (Figure 4.2).

48

```
Dp2     CCTATCGAAC CCATTCATAC AGAGCCAGTA TTCAATGTCC CTCAACCATC  50
Dp2     GCGCGGATCT AGTAGGGTCT CCCATCCTCC CGATAGATAC TTAGGTATTC 100
Dp2     TAGAAGAGGA TACCGAGGAA ATGTTCCTAG TGGGAGATAG GGATCACATA 150
Dp2     CAGGAT                                                 156

Dp36    CCTATCGACG ACAGGCTGAC ATGGCAATTG TGCCGCACCA ATCATGCTCG  50
Dp36    GATAGGAAAG AGTCGACCTC GACGAAAGCG GCTCGGTAAA GCCCCGGTAT 150
Dp36    ACTCCAACAA GTCCGGGTCA ATCCACGGT ATCTCCTCGC GCTGGAT     147

Dp41    CCTTCTCCCG GTAGGATCCG GCCTCACCGC AAATCCTGCA AGTATGACTG  50
Dp41    AGGGGGAAGA AGAAGGAGGG GACTCCGGAC CTGCCGTCGC GTCGTGGGGA 100
Dp41    CACCGTAGAT GGCTCGGTAG GTTGCCTTTC CTCCGTTGGA T         141
```

Figure 4.1. The three main RDA subtraction products isolated.

```
Dp2     CCTATCGAACCCATTCATACAGAGCCAGTATTCAATGTCCCTCAACCA  TCG      C
Dp3     -----A-------------A----------CA-G-------------CC---TAGAT-
Dp8     --------G------A-------------CA-G-------------CC---TAGAT-
Dp12    ------------A-------------CA-G-------------CC---CAGAT-
Dp18    ------------T-------AG--------CA-G-------------CC---TAGAT-
Dp26    -----T------------------------CA-G-----------G-GGGCC---TAGAT-
Dp33    ---------------A---------------A--------------GG------CAGAT-
Dp36    -----G-CGA--GGCTGAC-TGGCA-T-GTG-CGCAC-AA---TG-    TCG  GAT
Dp41    ---TCTCCC-GT-G-A-CCGGCCT--CCGCAAATC-TG-AAGTAT-ACTGAG-GGGAAG
Dp47    ---------------T-----------C--CA-G-------------CC---CAGAT-
Dp50    ---------------A-------------CA-G-------------CC---TAGAT-

Dp2     GCGGATCTAGTAGGGTCTCCCATCCTCCCGATAGATACTTAGGTATTCTAGAAGAGGAT
Dp3     ----------------------------------------------A------------
Dp8     A-----------------------------------------A------------
Dp12    -----------T----------------------------A------------
Dp18    ------------------------------------------A------------
DP26    -------T----------------G-------------A------------
Dp33    A-CGACC--GA-GA-AG-GG-TCGGTA-AG-CCC------ACTCC-ACA --
Dp36    AGGAA--A---GA-- --GA-GAA-GC-GCT-GGTAAAGCC-CG-T-TAC   -
Dp41    AAG-AGGAGGGG--TCCG-A-GTGCCTC-CGTCGT--GGACACC-T--ATGGC
Dp47    A-----------------------------------A---------
Dp50    -----------------------------------------A------------

Dp2     ACCGAGGAAATGTT CCTAGTGGGAGATAGGGATCACATACAGGAT
Dp3     ------A-------- ----------------A-A-------------
Dp8     -------------- ----------------------CG------
Dp12    -------A-------T----------------A-----------
Dp18    -------------- ----------------------------
DP26    -------------T----------------------------
Dp33    ----G-TC---CC-TT------------------G-G------
Dp36    --A-CA-GTCCGG GTCAA-CC--CGGTATCTC-TCGCGCT----

Dp41    TCGGTAGGTTGCC-TTCCTCC-TT          ----
Dp47                                     ----
Dp50    -------------- ----------------------------
```

Figure 4.2. Alignment of the different date palm RDA subtraction product sequences

49

4.3.2 Bioinformatic sequence analysis

Sequences were first analysed using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990). The standard nucleotide-nucleotide BLAST (*blastn*) option was used. This allows for the unknown sequence to be compared to all nucleotide sequences in the database. Using this approach no significant homology was found between any of the isolated difference products with any known plant DNA sequences in available databases. Secondly the sequence data was subjected to the FASTA algorithm. Using FASTA version 3.4t (Pearson and Lipman, 1988), alignment algorithm provided by the European Bioinformatics Institute (http://www.ebi.ac.uk/fasta33), the Dp41 sequence showed homology against many known *Oryza sativa* (rice) genomic DNA sequences. And lastly homology searches were preformed using the Smith-Waterman Algorithm. When the Smith-Waterman algorithm (Smith and Waterman. 1981) was used to produce local alignments between the Dp41 sequence and database sequences the best homology (75%) found, was between genomic DNA from chromosome 1 of *Oryza sativa* (accession number AP002902) and Dp41 (Figure 4.3). No homology with known DNA sequences was found for Dp2 or Dp36.

4.3.3 Primer design

From the sequence information obtained by sequencing the different cloned subtraction products, three primer pairs were designed to represent the three different groups of sequences obtained. Primer pairs DP36L and DP36R; DP41L and DP41R; DP50L and DP50R (Table 4.1) were designed from the sequence information of Dp41, Dp36 and Dp50. When these primer pairs were tested on tissue culture-derived date palm plants fragments with the expected sizes were amplified from all plants regardless if 'Medjool' or 'Barhee' template was used (Figure 4.4). When genomic DNA from non-tissue culture derived plants were used as template DNA, amplification with different primer pairs varied and genomic DNA from two 'Barhee' plants did not produce any amplification product regardless which primer pair was used for amplification (Figure 4.5).

```
Dp41        6        TCCCCGTAGTAACCGGCCTCCC..CGCAAATCTT..GCAAGTA.. 44
                     ::}::::  :  :::  ::::::: :::   ::::  :: ::::
AP002902   89189     tccccgacgccacctccctcccggcgc...tcttcagcgagtagc
89230

Dp41        45       .....TCACTGAGGGGGA...AGAAGGAGGAGGGGCCTCCGCACG 81
                         ::  :  : ::  :::: ::  ::::  :::::  :  :::::::
AP002902   89231     tcacctcgcagatcgggatggagtgggagtagggg.ttgcgcacg
89274

Dp41        82       TGCCTCACGTC........GTGGGGACACCG.............. 104
                     :  :::  :  ::  :         ::::  :  :  :::
AP002902   89275     t.ccacgcgccccgcgagggtggcgccgccgcggcccacgctctg
89318

Dp41       105       ..TAGATGGCTCGG.TAGGTTGCCTTTCCTCCG            134
                       ::  :::::::  ::  :  ::    :::::::
AP002902   89319     gaacgacagctcggccagct..cc..gcctccg
89347



Local: Dp41 vs AP002902                     %similarity = 75.83
Score: 201.00
```

Figure 4.3 Local alignment between Dp41 and *Oryza sativa* chromosome 1 DNA, accession number AP002902, using the Smith-Waterman algorithm.

Figure 4.4 Amplification products obtained with the three different primers pairs to amplify RDA subtraction products DP36 (100 bp), DP41 (100 bp) and DP50 (120 bp). The products were amplified from seven individual tissue culture-derived 'Medjool' plants (DP36 and DP41) and six individual plants (DP50). Amplification of subtraction products using 'Barhee' DNA from tissue culture plants as template DNA resulted in identical products.
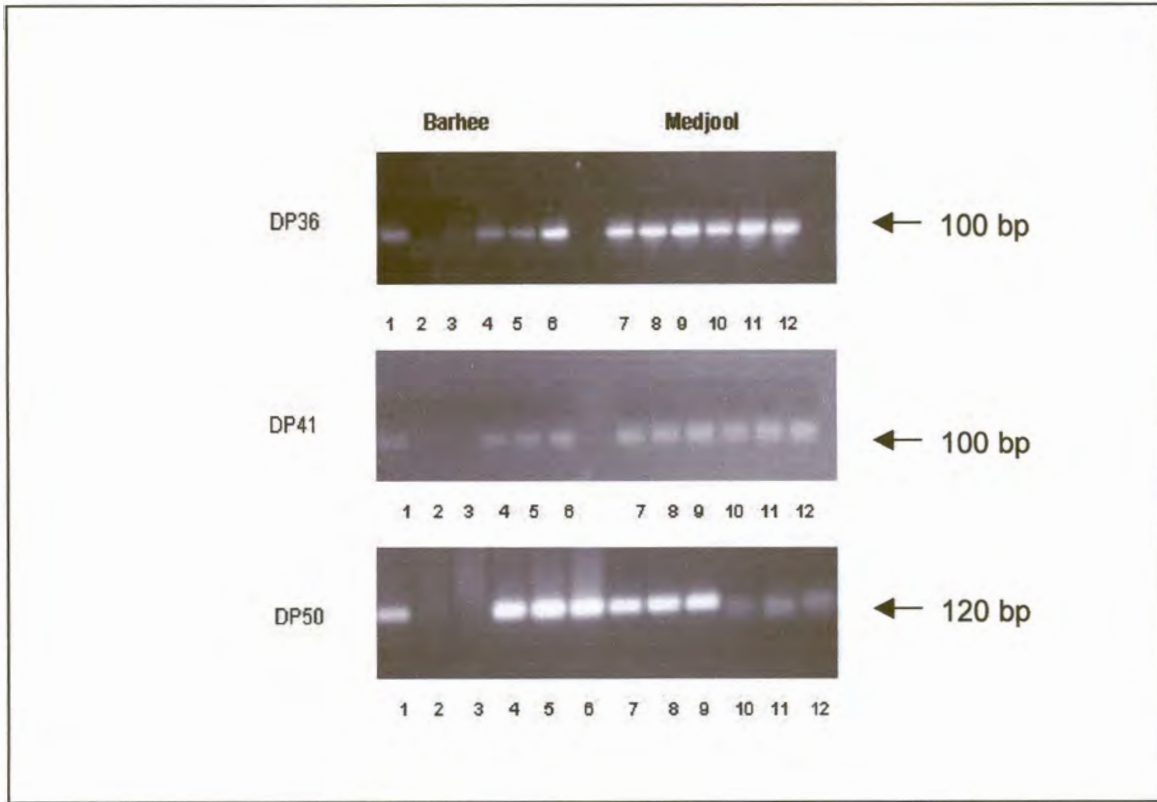
Figure 4.5 Variation in amplification of product from genomic DNA of different non-tissue culture plants. Lanes 1- 6: Amplification of DNA from 6 non-tissue culture-derived 'Barhee' plants; lanes 7- 12: Amplification of DNA from 6 non-tissue culture-derived 'Medjool' plants.
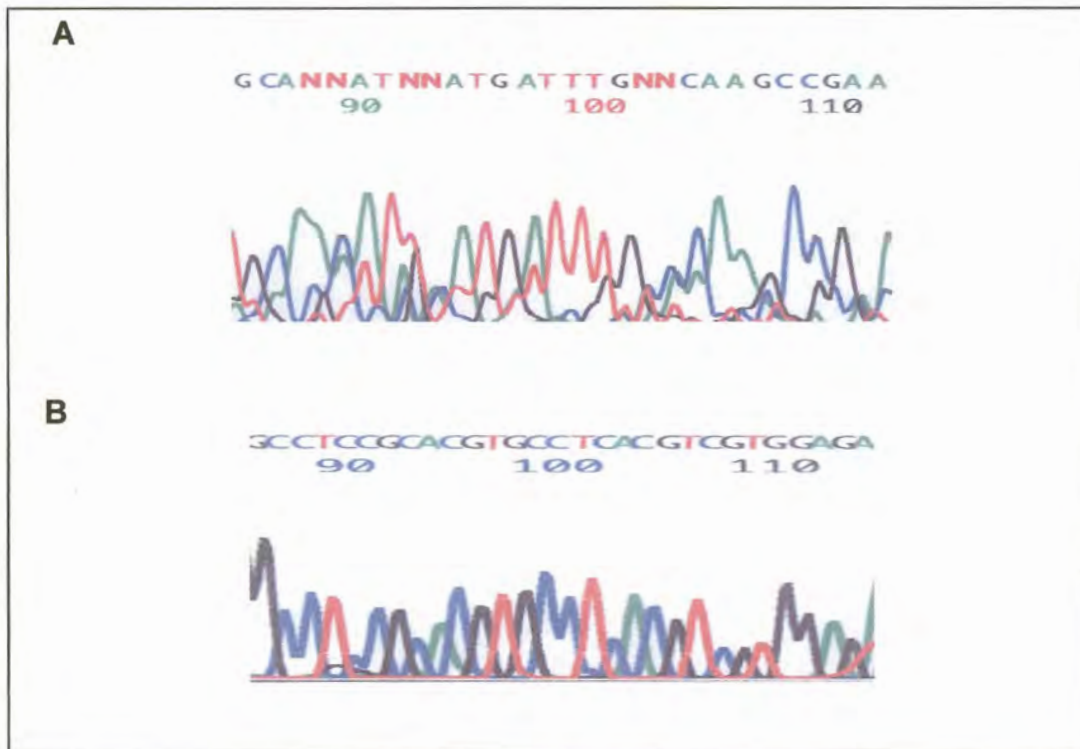
Figure 4.6 Electropherogram data from direct sequencing (A) with unacceptable high background noise, and from sequencing of cloned product (B) with acceptable background noise.

4.3.4 Detailed characterization of product Dp41

The Dp41 amplification product was characterized in more detail. From each of the six tissue culture-derived 'Barhee' and 'Medjool' plants, three independent clones of the cloned Dp41 amplification product were sequenced. As a first approach direct sequencing was applied. Using this approach, a high background noise was obtained in the sequence data and the obtained sequences were of very low quality not suitable for sequence analysis (Figure 4.6). Therefore the amplification products were first cloned and then sequence. The second approach resulted in good sequence quality.

Thirty-six sequenced clones from different 'Medjool' and 'Barhee' plants revealed a high degree of homology (above 95% sequence simalarity) with the original Dp41 subtraction product. Only those DNA sequences showing differences in comparison to Dp41 are shown in Figure 4.7. Detected differences were single base pair changes / deletions occurring mainly in a variable 45 bp region of the fragment (Figure 4.1, sequence in red). In general, more changes were observed between the individual 'Barhee' sequences than in the 'Medjool' sequences. Changes detected were specifically single base deletions in this variable region. From the 18 different sequences analyzed for each variety, 6 of the 'Medjool' sequences and 8 of the 'Barhee' sequences were identical to Dp41. This indicates that the region of the Dp41 sequence used for primer design is common to both genome types. Among the sequence variants found within the two genome- types, two, namely M6 and B9, were identical. Two sequence variants, M1 and M4, the latter with a six base deletion in its sequence, were unique to 'Medjool' and one variant (B11) was unique to 'Barhee'.

```
DP41   TCTTGCAAGTATCACTGAGGGGGAAGAAGGAGGAGGGGCCTCCGC
M1     --------------T--A-----A---G------------------T
M2     --------------------------G----------C-------
M3     --------------T--A---------G-----------------
M4     -T-----------A------------G---        T-G------
M5     -G-----------T---------T--------------------
M6     ---A-----A---------------------------------
M7     ---------------------------------------C-------
M10    ---------------------------------------G------
M11    ------------------------------------------AA--
M12    -------------------------------------------T--
M13    -------------------------------------------T--
M16    --------------TT----------------------------

DP41   TCTTGCAAGTATCACTGAGGGGGAAGAAGGAGGAGGGGCCTCCGC
B1     ----  -------A-----------------------G------
B2     ------------A-----------------------T-G------
B5     ------------A-----------------------T-G------
B6     ---------A--A-----A--------------------------
B9     ---A-----A----------------------------------
B10    ----------------T---------------------------
B11    ------CGCA---TTGC-A-TATC--T-----------------
B12    -------------------------   ---------------
B14    ----  --------------   ----  --  --  ----  -------
B15    ---C------------------   ----------------  -------
```
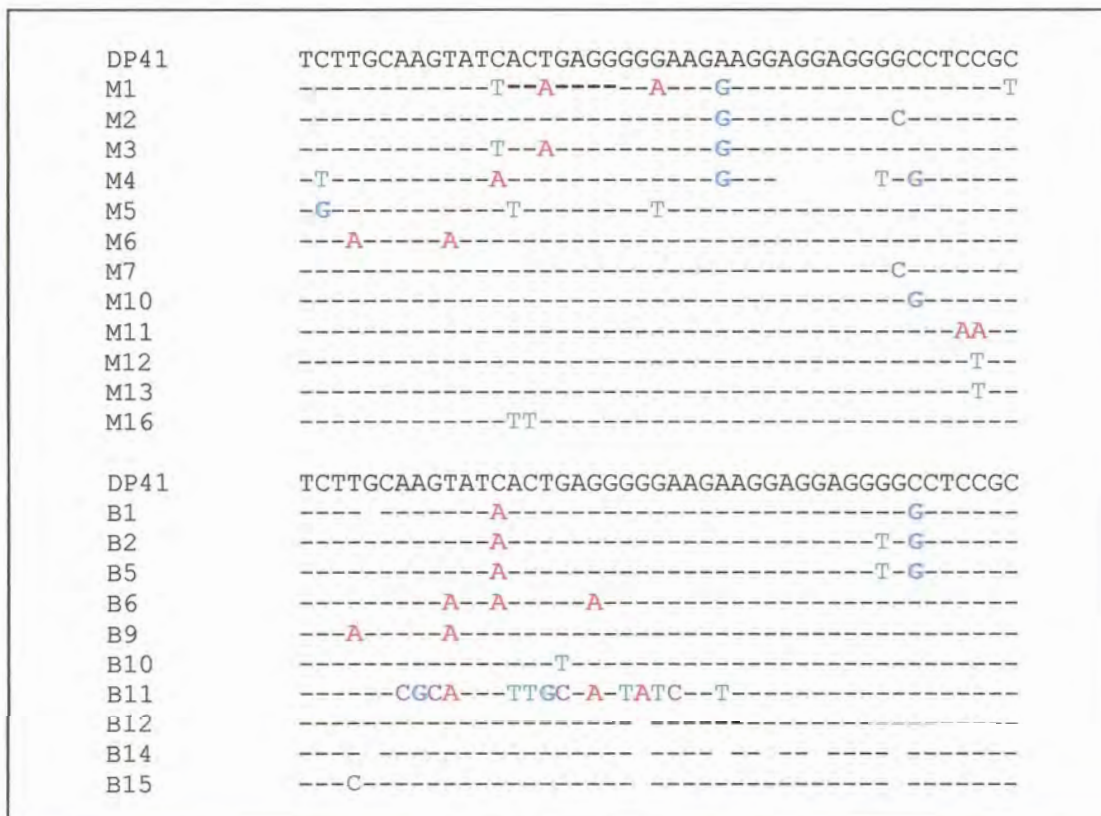
Figure 4.7. Differences found within the DNA sequence of amplified Dp41 fragment from different individual plants.

Primers PLM1, PLM4 and PLB11 (Table 4.2) were therefore designed from sequence M1, M4 and B11, to cover the variable portion of these sequences (Figure 4.6) when used in conjunction with DP41R. All 6 tissue culture-derived 'Medjool' plants and also all tissue-culture-derived 'Barhee' plants, which originated from a single mother plant, amplified a PCR product with the expected size with all three primers used. However, primer PLB11 at an optimal annealing temperature of 65°C only amplified a PCR product with the expected size of about 110 bp from 2 of 7 non-tissue culture-derived 'Barhee' and 6 of 7 'Medjool' plants. An identical result was observed with primer PLM1 at an optimal annealing temperature of 60°C. Primer PLM4 (at 65°C annealing temperature), which covered a unique 6 base pairs deletion, amplified a PCR product from DNA of all 7 non-tissue culture-derived 'Medjool' plants (Figure 4.2) but only from one 'Barhee' plant.

4.4 Materials and methods

4.4.1 Sequence analysis

Sequence analysis was done as outlined in the Annex.

4.4.2 Bioinformatic sequence analysis

For the BLAST (Altschul *et al.*, 1990) analysis the standard nucleotide-nucleotide BLAST or BlastN option was used. The 'nr' database option was used allowing the algorithm to search all GenBank, EMBL, DDBJ, PDB sequences. A low complexity filter was chosen and the 'expect' or the statistical significance threshold were set at 10 and the word size at 11. FASTA version 4.3t was used (Pearson and Lipman, 1988). From the FASTA search submission form the following options were picked. The program chosen was Fasta3. The data base option the following to parameters were chosen: (1) Nucleic acid and (2) EMLB. Gap penalties: Open -16 and Residue -4. Scores and Alignment were both set at 50. The KTUP value was set at 2 and both DNA strands were scored. The E-value threshold was set at 10. For the Smith-Waterman analysis (Smith and Waterman, 1981) the following parameters were used. Matrix: NUC4X4HB; Gap open penalty: -10 and Gap extention penalty: -5.

4.4.3 Primer design and testing

Pairs of primers were designed using a standard design program (Expasy, Switzerland) and the sequence information obtained for 4 of the different subtraction products. The primer pairs were used in a PCR reaction using 'Barhee' and 'Medjool' DNA as template at various annealing temperatures to optimize the PCR reaction. The amplification products were separated on a 1.5 % agarose gel, stained with ethidium bromide and visualized under UV light. See Annex for details on primer sequences.

Table 4.1. Primers designed for amplification of subtraction products.

Primer	Sequence
DP36L	5'-CTATCGACGACAGGCTGACA -3'
DP36R	5'-GACCCGGACTTGTTGGAGTA-3'
DP41L	5'-CCTTCTCCCCGTAGTAACCG-3'
DP41R	5'-AGGAAAGGCAACCTACCGAG-3'
DP50L	5'-TACACGATGTCCCTCAACCA-3'
DP50R	5'-GGAACATTTCCTCGGTATCC-3'

Table 4.2 Primers designed to include changes identified within the different amplified Dp41 products.

Primer	Sequence
PLM 1	5'-TTACAGAGGGGAAAGGAGGA-3'
PLM 4	5'-GGAAGGAGGTGGCTCCG-3'
PLB11	5'-CGCAATCTTGCAAGTATCAGT-3'

Chapter 5

Detecting methylation changes in the genome
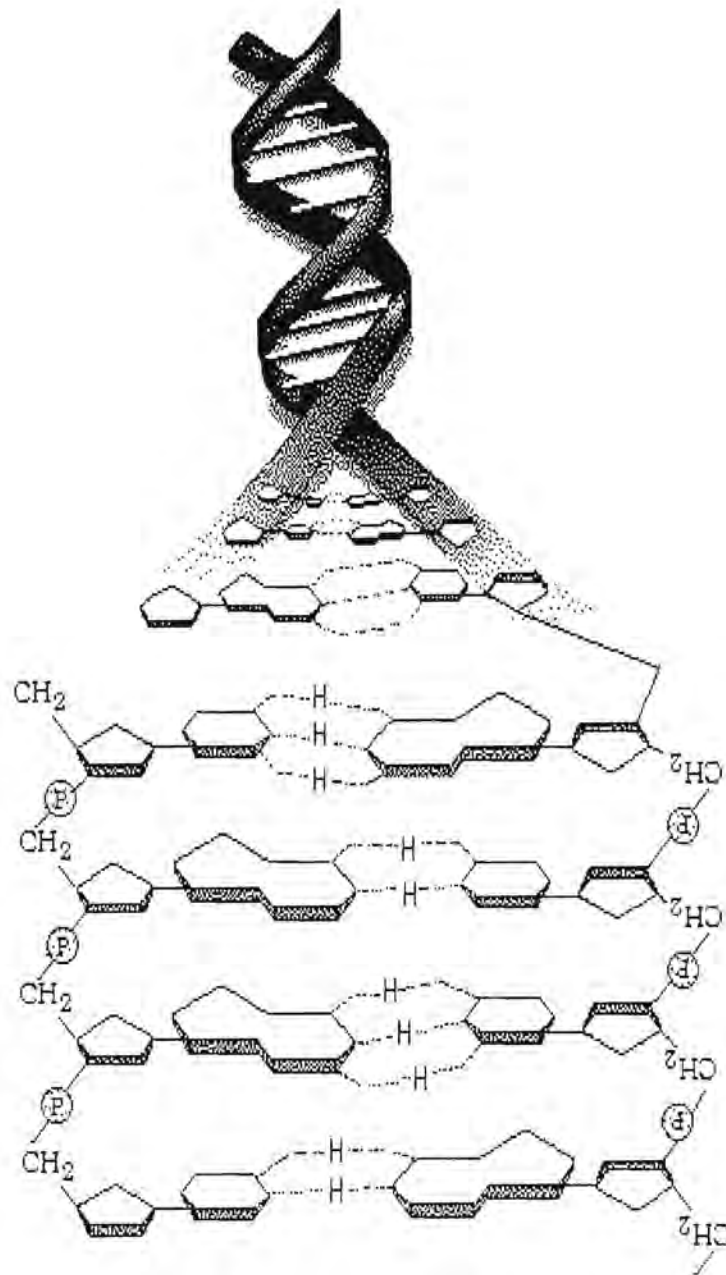
5.1 Abstract

In an extention of the project the RDA technique was applied to see whether changes in genomic methylation between two date palm cultivars and also two types of embryogenic tissues from *Pinus strobus* could be isolated. From date palm a repetitive sequence was isolated as a subtraction product and further products with homology to rRNA genes. These genes also represent repetitive sequences in the plant genome, and as such are a target for methylation. Embryogenic tissue of *Pinus strobes* was used to provide information whether the RDA protocol could also be applied to study large genomes and to isolate subtraction products that could be correlated to differences observed during somatic embryogenesis of tissue grown on different 2,4-D concentrations. Although a difference in the methylation levels of genomic DNA of these cells receiving the different 2,4-D treatments was found, no subtraction products could successfully be isolated, which was very likely due to the size and complexity of the pine genome.

5.2 Objectives

The objectives of this study were to determine whether the RDA protocol could be applied to isolate DNA sequences that are differently methylated between two genomes (MS-RDA). In particular, tissue culture plants from two date palm cultivars were compared ('Barhee' and 'Medjool') as a first objective. They were chosen because of the differences in the occurrence of somaclonal variation that is observed between the two cultivars when regenerated by tissue culture. Further, in a second objective, embryogenic material from *Pinus strobus* (*P. strobus*) that have been grown *in vitro* and exposed to different concentrations of 2,4-D were compared. This experiment was aimed at determining the capability of the standard RDA protocol to compare large genomes and complex genomes such as that of *P. strobus*. *Phoenix dactylifera* (date palm) has a 1C content of 931 Mbp (Olszewska & Osiecka, 1982) whereas *P. strobus* has a IC content of 25137 Mbp (O'Brien *et al.*, 1996)

5.3   Results

5.3.1   Date palm

5.3.1.1 *Comparison of methylation in date palm*

Comparison of methylation was done using the methylation-sensitive restriction enzyme *Hpa*II and the less sensitive enzyme *Msp*I. Using these two enzymes the relative methylation status of the genomic DNA of date palm plants in tissue culture was compared to those that did not derive from a tissue culture process. The results obtained indicate that the date palm DNA obtained from the tissue culture plants were hyper-methylated when compared to that of non-tissue culture plants (Figure 5.1). The genomic DNA of the 'Barhee' cultivar also contained a higher percentage of methylated DNA compared to that of 'Medjool' regardless of whether the plants were exposed to a tissue culture process.

Figure 5.1. Comparison of methylation of tissue culture and non-tissue culture date palm plants. Lanes 1 and 2 represent digested DNA of 'Barhee' non-tissue culture plants digested with with *Hpa*II (1) and *Msp*I (2); lanes 3 and 4 digested DNA of 'Medjool' non-tissue culture plants with *Hpa*II (3) and *Msp*I (4); lanes 5 and 6: digested DNA of 'Barhee' tissue culture plants with HpaII (5) and MspI (6); lanes 7 and 8: digested DNA of 'Medjool' tissue culture plants with HpaII (7) and MspI (8). Lane 9 represents molecular weight marker III.

5.3.1.2 *First MS-RDA subtraction on date palm*

Two MS-RDA subtractions were performed with either 'Medjool' as tester and 'Barhee' as driver or *vice versa*. The subtraction using 'Barhee' DNA as tester yielded no subtraction products. Using 'Medjool' as tester yielded a number of fragments ranging from 200 bp to 500 bp (Figure 5.2 A). Cross hybridisation of the subtraction products revealed a shared homology between all the different cloned fragments representing a common element. This homology was confirmed by subsequent sequence analysis.

Southern blot analysis probing the subtraction products against the first amplicons of both 'Barhee' and 'Medjool' showed that the element was present in both genomes, but a stronger hybridisation signal was observed with the 'Medjool' amplicons. This indicates that a possible higher abundance of this sequence in the 'Medjool' genome. Southern blot analysis probing the subtraction product against genomic DNA gave a smear providing evidence for a dispersed repetitive element. Southern blot hybridisation using a labelled subtraction product hybridised against the other subtraction products revealed homology between all the isolated subtraction products (Figure 5.2 B)

Sequence analysis of the fragments showed a 27 bp repeat within each fragment. The first 24 bp of the repeat was homologous to the 24 bp of the RHpa24 adapter that was used to create the first round amplicons (Figure 5.3). A BlastN homology search revealed some homology of the total sequence with microsattelites from *Hippoglossus hippoglossus* (accession number: AF133244) (E-value: 0.002) and *Phytophthora cinnamomi* (accession number: AF312885) (E-value: 0.12), but no plant-related sequences.

A

500 bp →

200 bp →

B
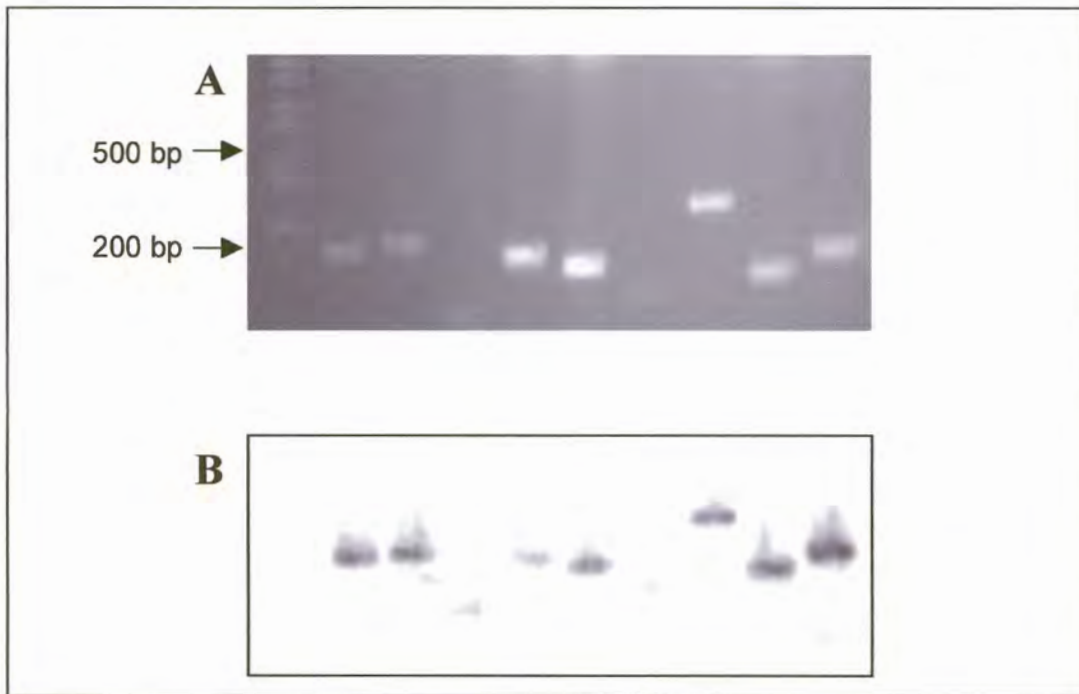
Figure 5.2. Cloned subtraction products (A) cut from plasmids with appropriate restriction enzymes showing various sizes and (B) Southern blot analysis showing homology between the subtraction products.

```
5′GGACCTTGCAAGCCCCTCGCCGAACGCCGTGGTTGT*AGCACTC
TCCAGCCTCTCACCGACCG*AGCACTCTCCAGCCCCTCACCGACCG
C*AGCACTCTCCAGCCTCTCACCGACCGC*GCACTCTCCAGCCTCT
CACCGACCGT*AGCACTCTCCAGCCTCTCACCGACCGT*AGCACTC
CCCAGCCTCTCACCGACCGT*AGCACTCTCCAGCCTCTCACCGACC
GT*AGGCTTCTCCAGCCTCTCACCGACCGT*AGCACTCTTCAGCCT
CTCACCGACCGT*AGCACTCTCCAGCCTCTCA 3′

RHpa 24    5′AGCACTCTCCAGCCTCTCACCGAC 3′
```

Figure 5.3. Sequence of a 298 bp subtraction product obtained from the first MS-RDA subtraction of date palm. Repeats are separated by a red asterix (*) and the RHpa24 primer sequence is shown for comparison.

5.3.1.3 *Second MS-RDA subtraction on date palm*

Because of the sequence homology found between the first subtraction products and the RHpa24 adapter, the *Hpa*II subtraction was repeated, using the SHpa12 and 24 adapter pair instead of the RHpa adapter pair. The second subtraction resulted in the isolation of three fragments from 'Medjool' as tester (Figure 5.4). The fragment sizes were 444 bp, 330 bp and 215 bp. Sequence analysis and BLAST search of these fragments revealed a 99% homology between the 444 bp fragment and part of the 18S rDNA of tobacco (*Nicotiana tabacum*) (accession number: AJ236016.1) (E-value: 0.0) as well as that of a number of other plant species. The 330 bp fragment showed homology to 23S and 4.5S rDNA sequences including that of tobacco. The 215 bp fragment showed no significant homology to any plant related sequences.
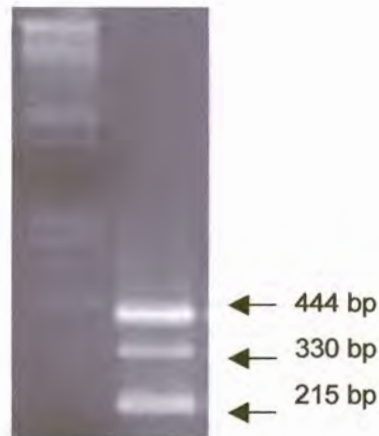
Figure 5.4. Subtraction products obtained from the second MS-RDA subtraction of date palm. Sizes of fragments are indicated.

5.3.2 Pinus strobus

5.3.2.1 *Pinus strobus culture and somatic embryogenesis*

Mega-gametophytes containing zygotic embryos in developmental stages ranging from the cleavage poly-embryony, to the early (pre-cotyledonary) dominance stage were used to initiate embryonal masses (EM) used for somatic embryogenesis. Embryogenic tissue began to grow from explants after the second week of culture, with the majority of explants responding by the 12th week (Figure 5.5A). These

embryonal masses were carefully separated from the mega-gametophytes and transferred to new medium, (Figure 5.5B). Analysis of the tissue under a light microscope confirmed their embryonal nature (Figure 5.5C).

Maintenance of the cultures on the medium used for the initiation of the specific line ensured rapid growth once the line was established after approximately 2 months. However, with a few genotypes an occasional gradual decline in growth was observed. From approximately 300 mg of fresh weight at day 0 the tissue fresh weight increased 6 to 8-fold after a 2-week culture period. This material was used for DNA extraction for MS-RDA, as well as for the maturation experiment to show the differences in the amount of mature embryos produced by the two different 2.4-D treatments.

Embryogenic tissue of *P. strobus* maintained on UL medium produced on average more mature somatic embryos (Figure 5.6) than the ones maintained on Std medium. medium. Figure 5.7 shows differences in the amount of mature embryos produced per plate for two lines (1053 and 1055) when embryogenic tissue maintained on either MLV UL or Std were used for the maturation of embryos. Table 5.1 shows the average number of mature embryos produced per gram of fresh weight for the four lines used.

Table 5.1. Average number of embryos produced per gram of fresh weight used.

Line	Number of embryos per gram of fresh weight MLV UL	Number of embryos per gram of fresh weight MLV Std
1053	319	42
1054	307	13
1055	263	55
1073	282	80
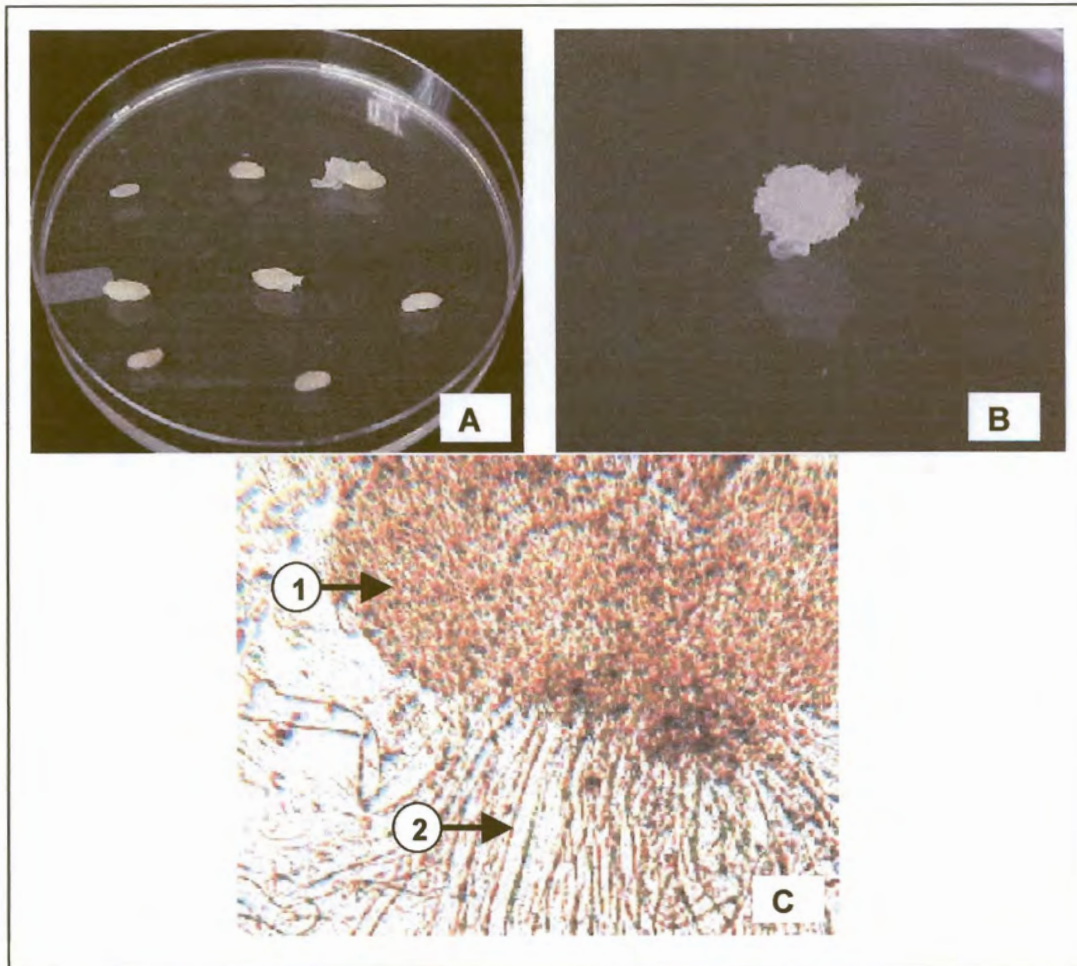
Figure 5.5. (A) Mega-gametophytes on initiation medium. Micropylar extrusion and proliferation of embryonal mass (EM) indicated by arrow. (B) EM on medium and (C) light microscope image of EM showing small embrogenic cells (1) and large suspensor cells (2).

Figure 5.6. Mature somatic embryo of *P. strobus*.
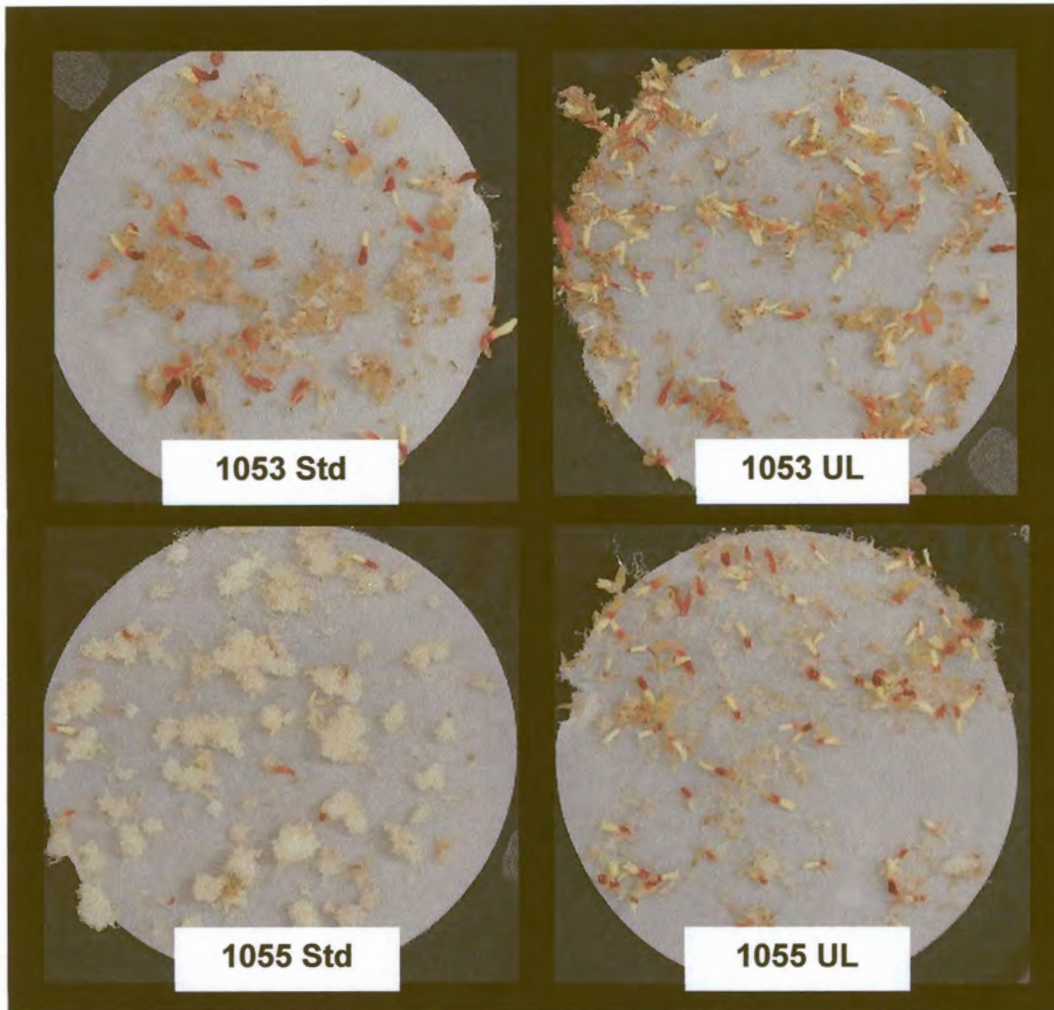
Figure 5.7. Maturation of somatic embryos from tissue maintained on either MLV UL medium (1053 UL and 1055 UL) vs. from tissue maintained on MLV Std medium (1053 Std and 1055 Std).

5.3.2.2 *Comparison of methylation in P. strobus*

Genomic methylation of embryogenic masses of *P. stobus* grown on medium containing different concentrations of 2,4-D was compared. DNA isolated from those cells of *P. strobus* that were grown on high 2,4-D medium also showed a higher level of methylation when compared to those that were grown on a lower 2,4-D medium. This can be seen by the amount of undigested DNA still in the wells of the gel (Figure 5.8; lanes 9-16) and further by the amplification of larger fragments in the *Hpa*II amplicons generated from genomic DNA from cells grown on MLV Std medium *vs.* MLV UL medium (Figure 5.9).

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
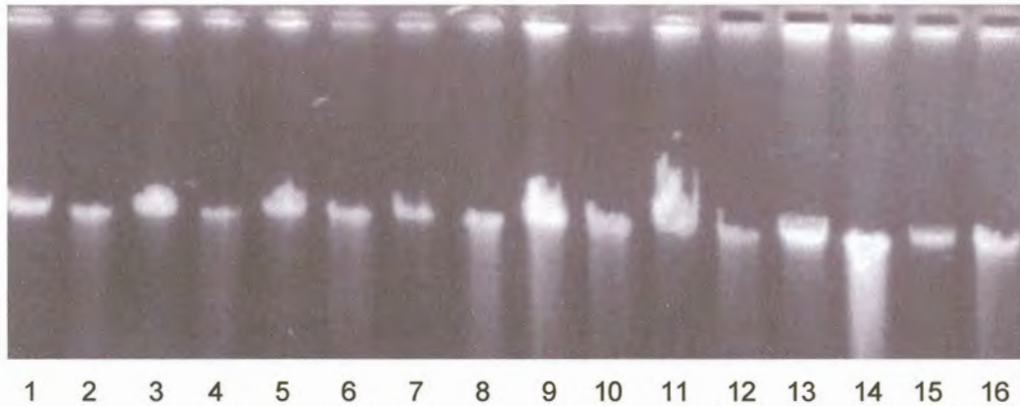
Figure 5.8. Comparison of DNA methylation between cells grown on high and low 2,4-D medium. Lanes 1-8 represent cells grown on low 2,4-D media; lanes 9-16 cells grown on high 2,4-D media. Each DNA sample was cut with *Hpa*II and *Msp*I and the two samples run in adjacent lanes.

A   B

1 2 3 4 5 6 7 8   1 2 3 4 5 6 7 8
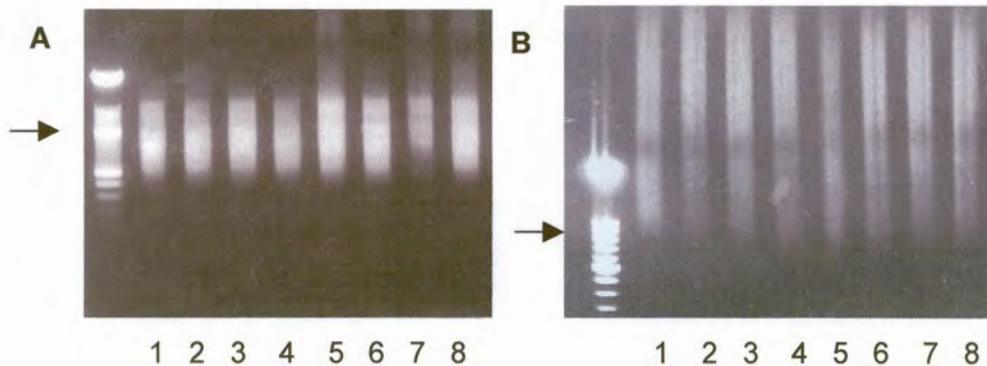
Figure 5.9. Gel profiles of RDA *Hpa*II amplicons generated from genomic DNA of lines grown on (A) MLV UL medium (1.1 μM 2,4-D) and (B) the same lines grown on MLV Std. medium (9.5 μM 2,4-D). The amplification of larger fragments in B indicates higher methylation levels. Arrow indicates the position of the 1000 bp marker.

5.3.2.3 *Pinus strobus MS-RDA subtraction*

Using the tissue grown on MLV UL medium (1.1 μM 2,4-D) as tester and tissue from the same line grown on MLV Std. (9.5 μM 2,4-D) as driver, three rounds of subtractive hybridisation at tester to driver ratios of 1: 1000; 1: 10000 and 1: 100000 were performed. However, the subtractions failed to produce any subtraction

67

products. The profile obtained after separating the amplicons on a 1.5% agarose gel shows only a smear indicating a high complexity of amplified fragments (Figure 5.10).
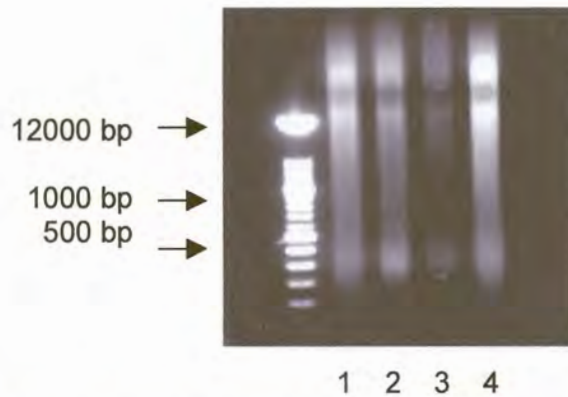
Figure 5.10. Amplification profile of 3rd round subtraction products of 4 *P. strobus* embryogenic lines, 1053 (line 1); 1054 (line 2); 1055 (line 3) and 1073 (line 4).

5.5   Materials and methods

5.5.1   Plant material and DNA extraction

In vitro grown date palm plants from the cultivars 'Medjool' and 'Barhee' and embryogenic tissue of *P. strobus* were used for the experiments. The embryogenic material was obtained from plant material collected from the Nova Scotia Tree Improvement Working Group's (NSTIWG) clonal seed orchard in Debert, Nova Scotia, Canada. Tissue from four cell lines was used (NSTIWG clone numbers 1053, 1054, 1055 and 1073). The embryogenic tissue of each cell line was grown on ML medium containing either 9.5 μM or 1.1 μM of 2,4-D. Total cellular DNA from different types of plant material was isolated from tissue (1g) using a Nucleon Phytopure Plant DNA extraction kit (Amersham Life Science, UK) according to the manufacturer's instructions.

5.5.2 Tissue culture and somatic embryogenesis

5.5.2.1 *Plant material*

For culture initiation, seeds were extracted from developing cones and placed in petri dishes containing moistened filter paper. Within 1 hour of extraction, seeds were surface sterilized in preparation for dissection. Seeds were surface-sterilised by placing them in metal baskets and stirring vigorously for 6 minutes in 200 ml of 3% (v/v) hydrogen peroxide with one drop of Tween 20. Seeds were then rinsed three-times for 2 minutes each with sterile water and transferred to a petri dish containing a moist, sterile filter paper. The seed coat, nucellus, and membrane surrounding the mega-gametophyte were removed under a stereomicroscope, and the mega-gametophyte was placed on initiation medium.

Five to eight mega-gametophytes were cultured in each petri dish containing approximately 25 ml of either MLV UL or Std. medium (see Annex), and dishes were sealed with parafilm. Explants were cultured in darkness at 24°C for 16 weeks and examined every 4 weeks for embryogenic tissue initiation.

5.5.2.2 *Embryogenic tissue proliferation and culture maintenance*

Explants displaying proliferation from the zygotic embryo cells were microscopically selected, then embryogenic tissue was removed, dispersed by vigorous shaking in liquid MLV medium, poured onto a sterile filter paper disc (Whatman No. 2, 5.5 cm) in a Büchner funnel and a short low-pressure pulse (55 kPa) was applied to drain the medium prior to transferring the disc to fresh medium. This approach was also used for maintenance of embryogenic tissue. Maintenance and proliferation cultures were incubated in darkness at 24°C.

5.5.2.3 *Maturation of somatic embryos*

The somatic embryo experiments were performed by combining embryogenic tissue of one line (from several plates), one week after subculture, in a 50 ml test tube, adding liquid MLV media without growth regulators and vigorously shaking the tube to break up the clumps of tissue into a fine suspension. Subsequently 3 ml containing

approximately 0.2 g of suspended embryonal masses were withdrawn with a wide-mouth pipette and placed on a moist filter paper disc (Whatman No. 2 55 mm in diameter) in a Büchner funnel attached to a vacuum pump. A short low-pressure pulse (5 s, 55 kPa) was applied to drain all liquid medium and anchor the embryonal masses to the filter paper as a thin layer. Each disc of filter paper with the embryonal masses were subsequently placed on maturation medium (see Annex) in a 10 X 20 mm petri dish and cultured for up to 10 weeks. The cultures were kept under dim light condition at 1.6 μmol m^{-2}s^{-1} from cool white florescence lamps (Philips F72T12/CW, 65 W) under a 16 hour photoperiod at 24°C.

5.5.3 Comparison of relative methylation

Methylation of total genomic DNA was compared between date palm material derived from tissue culture and non-tissue culture-derived plants (Figure 5.1). For *P. strobus*, DNA methylation was compared between cells grown on high (9.5 μM) and low (1.1 μM) 2,4-D medium (Figure 5.2). The DNA was digested with the restriction enzymes *Hpa*II and *Msp*I both recognizing the same site (C/CGG). The digested DNA was separated by electrophoresis and stained with ethidium bromide and the obtained restriction profiles were compared.

5.5.4 MS-RDA subtraction

5.5.4.1 *Date palm*

RDA was performed following the general outline described by Lisitsyn *et al.* (1993) and as described in the Annex. Total genomic DNA was digested with the methylation sensitive restriction enzyme *Hpa*II. Amplicons were prepared by ligating the adaptor pair RHpa12 and RHpa24 (see Annex) to the digested DNA. After ligation, the DNA was amplified to produce the tester and driver amplicons in eight individual 100 μl volumes using a Perkin-Elmer GeneAmp 9600 thermocycler as described in the Annex.

Driver (150 μg) and tester (10 μg) amplicons were digested with *Hpa*II. Driver and tester DNAs were both re-dissolved at approximately 400 μg/ml. Both the digested

driver and tester amplicons were run on a 1.5% TAE agarose gel alongside an equal aliquot of undigested amplicons to check the completeness of the digestion. On the same gel, standard lambda phage DNA was run so that the final concentrations of driver and tester DNAs could be estimated and adjusted to 400 µg/ml if necessary.

Tester DNA was prepared by adding a second adaptor pair JHpa 12 and 24 to 1 µg of the first round amplicons. An aliquot of tester DNA with ligated adaptors was amplified for 20 cycles in a reaction volume of 20 µl to test that ligated adaptor pair will amplify tester amplicon DNA. A hybridisation reaction was set up by mixing the driver and tester amplicons at a ratio of driver to tester DNA of 100:1. The hybridisation mixture was precipitated and the pellet re-dissolved in 4 µl 3X EE buffer (see Annex). The solution was overlaid with light mineral oil. The DNA was then denatured and sodium chloride solution was added. Hybridisation of DNAs was done overnight at 67°C. Following hybridisation the reaction mixture was amplified by PCR.

The adapters on the amplicons that were obtained after amplification of the first hybridisation reaction, was removed and replaced with a third set of adapters, namely NHpa 12 and 24. A second round of hybridisation was then performed in the same way as the first except that a ratio of driver to tester of 1000:1 was used this time. A second MS-RDA subtraction was also done on the date palm genome that differed from the first in that a different adaptor pair, SHpa 11 and 24 (see Annex), was used instead of the RHpa adapter pair.

5.5.4.2 *P. strobus*

RDA amplicons were prepared as those for date palm using the restriction enzyme *Hpa*II for genomic DNA digestion and the RHpa adaptor pair. The first subtraction was done at a driver to tester ratio of 100:1 followed by a second subtraction of 1000:1 and a third of 10 000:1 using the adaptor pairs (see Annex)

5.4.5   Cloning and sequencing of the difference products

The RDA subtraction products obtained were cloned blunt-ended into the *EcoR*V cloning site of the cloning vector pMos*Blue* (Amersham Life Science, UK). Blue/white selection of *E. coli* colonies and screening for ampicillin-resistant colonies were used to identify recombinant colonies.

Cloned subtraction products were sequenced using an automated sequencer as described in the Annex. Obtained sequences were compared and submitted to the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) to search for homology with known sequences.

Chapter 6

Discussion
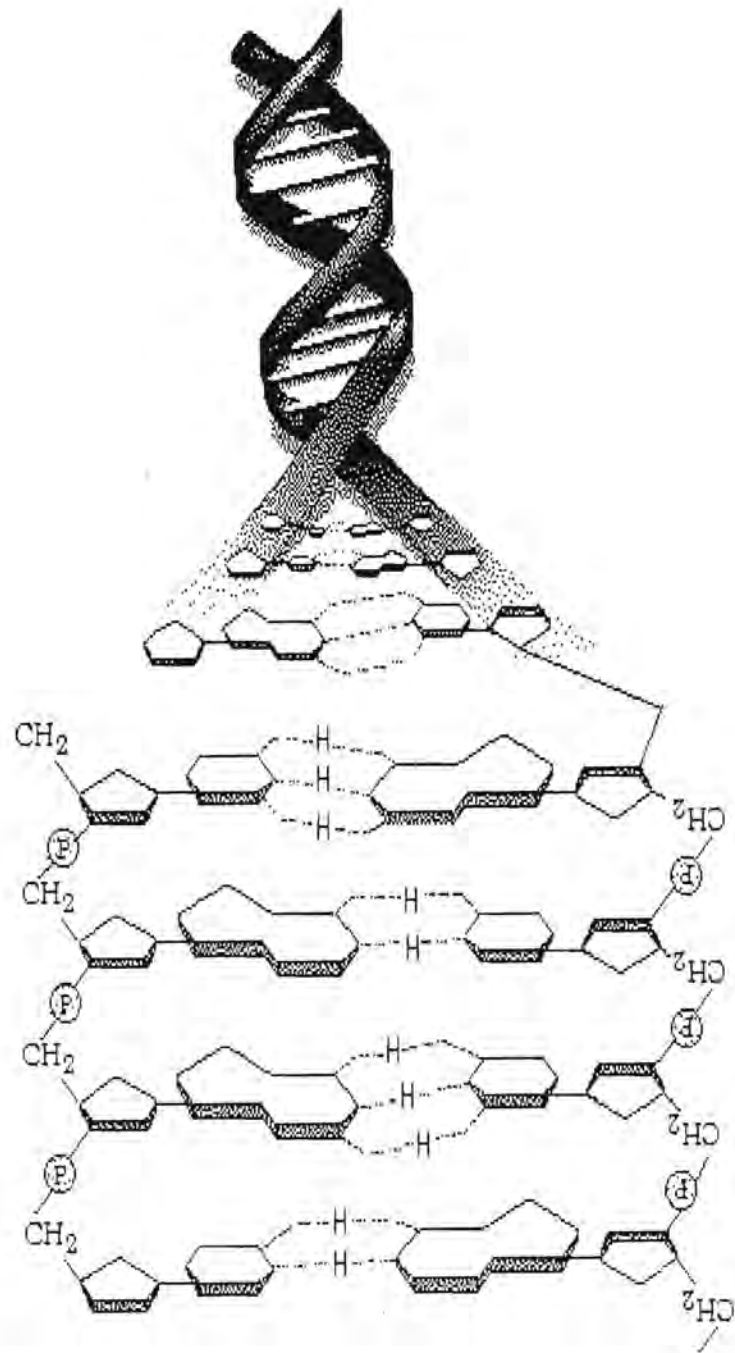
6.1 Date palm RAPD analysis

In this study the two date cultivars 'Barhee' and 'Medjool' could be differentiated with polymorphic DNA bands by the RAPD technique as outlined by Corniquel and Mercier (1994) using the RAPD primers OPE-01 and OPE-06. For optimal amplification, it was however vital in this study to optimize the reaction conditions for PCR amplification. This included optimizing the DNA isolation and purification method, the DNA template concentration and the selection of both enzyme and the DNA annealing temperature. The optimal DNA concentration was found to be approximately 25 ng. More or less DNA in the reaction mixture usually resulted in a failure to amplify a DNA fragment. The enzyme used for amplification was also important and best and consistent amplification results were always obtained with Taq polymerase from TaKaRa (Japan). Amplification profiles obtained by other enzymes, such as Taq polymerase from Amersham or Roche, varied and were not useful due to lack of consistency in amplification when test were repeated 3-times. It was also found that profiles were more consistent when the annealing temperature was lowered from the original 36°C to 33°C.

As a new aspect of this study, the potential of designed SCAR primers was investigated. RAPD reactions are very sensitive and the slightest variation in either temperature, quality of DNA or in any of the reagents used can easily result in failure of amplification. This is due to the use of a short 10 mer DNA primer in the RAPD technique. Since the ability of such primer to bind correctly can be influenced by very small changes in the reaction conditions, SCAR primers (20 mers) were constructed from an amplified polymorphic 'Medjool' DNA fragment (700 bp fragment). Designed SCAR primers amplified a DNA fragment with the expected size of 128 bp irrespectively which genomic DNA of date palm was used in the PCR reaction and therefore did not differentiate between the two types of genomic DNA. This possibly indicates that the original polymorphic fragment used for SCAR primer design had a base pair change in the OPE-06 primer site preventing efficient binding of the primer in the PCR reaction.

Overall, in this part of the study the results reported by Corniquel and Mercier (1994) for date palm differentiation using the RAPD technique could be confirmed. But, SCAR primers designed to obtain a more robust PCR amplification system than RAPD failed.

6.2   Isolation of RDA subtraction products from date palm

Application of RDA successfully allowed a direct one-step cloning of putative genome differences of two date palm cultivars despite carrying out a more complex technical procedure than the RAPD technique applied in chapter 2 of this study. With RDA, two types of subtraction products from 'Barhee' genomic DNA with a different size were isolated after subtracting the two genomes. However, subtraction products were only identified after *BamH*I digestion of genomic DNA and not after *Hind*III digestion. Also cross-hybridization with 'Medjool' amplicon DNA of the 'Barhee' subtraction product revealed that these products are present in both genomes and not unique to the 'Barhee' genome. Date palm has a relatively small genome ($5X10^8$ bp), which is roughly 4.2-times greater than the genome of *Arabidiopsis thaliana* ($1.17X10^8$ bp) but very similar to that of rice ($4.3X10^8$ bp). Generally, such small genomes tend to be more stable than larger genomes, which are present in many cereals or trees. The relatively small date palm genome might therefore not have sufficient detectable variation specifically when digested with the restriction enzyme *Hind*III, which is required to easily detect significant sequence differences between cultivars.

It is also unclear why subtraction products were formed and could be isolated from 'Barhee' when these products are present in both genomes. One possibility is that these subtraction products are in excess of more than 100-times in the 'Barhee' genome than in the 'Medjool' genome. This might result in elimination of the subtractive potential of the driver DNA when the driver DNA is only applied in a 100-times higher concentration than the tester DNA. A greatly uneven copy number of subtraction products in both genomes might also explain why no difference product was obtained when 'Medjool' instead of 'Barhee' amplicon DNA was used as a tester. Also, methylation differences might exist between the genomic DNAs of the two cultivars, with 'Medjool' DNA being more methylated. The restriction enzyme *BamH*I is methylation sensitive and any methylation of a restriction site for this enzyme in the driver DNA might have prevented sufficient PCR amplification of the sequence region detected as a subtraction product when the driver amplicon DNA was produced.

6.3   Bioinformatic analysis of isolated subtraction products

In this part of the study several bioinformatic tools, such as the Smith-Waterman algorithm, FASTA algorithm and BLAST, were applied to characterize one subtraction product sequence (Dp41) in more detail. The Smith-Waterman algorithm is the *de facto* standard for searching databases. It employs the Dynamic Programming (DP) algorithm to each and every database sequence, so it is the most sensitive method for finding related proteins in a database. But, because the DP algorithm effectively makes every possible pair-wise comparison between the query (or reference) sequence and the library (or database) sequences, it is also the slowest method for similarity searches of sequence databases. The Smith-Waterman Algorithm produces local alignment(s) between two sequences. Instead of investigating each sequence in it's entirety it compares segments of all possible lengths, and chooses whichever maximize the similarity measure (Smith and Waterman, 1981). This is useful if some parts of the sequences have undergone any processes of evolution that similarity is undetectable; or if one is only a fragment or only partially overlaps the other. These are all common situations.

The FASTA algorithm is a heuristic approximation to the Smith-Waterman algorithm. In Computer science the term 'heuristic' is by definition: Relating to or using a problem solving technique in which the most appropriate of several found by alternative methods is selected at successive stages of a program for use in the next step in the program. The heuristics used by FASTA allows it to run much faster than the Smith-Waterman algorithm, but at the cost of sensitivity (Pearson and Lipman, 1988). Finally BLAST is a much faster heuristic approximation of the Smith-Waterman algorithm. However BLAST cannot guarantee missing a score as sequences not found by BLAST could still have a significant Smith-Waterman score.

In this study the subtraction product sequence (Dp41) was specifically characterized, using molecular tools, with respect to its variability and presence in individual date palm plants of the two different date palm varieties. A 75% homology was found with rice DNA from chromosome 1 using the Smith-Waterman algorithm but without a known gene function. Sequence analysis of the amplified subtraction products from both date palm varieties, 'Barhee' and 'Medjool', identified a variable region and the existence of several variant members of the repeated sequence. This provides strong evidence that this Dp41 family of DNA sequences represents a variable "genetic hotspot" in the date palm genome as found for rye by Linacero et al. (2000). Variation

detected included changes from C→ T and G→A and deletion of single or several base pairs. In general, deletions occurred at a higher frequency in 'Barhee' than in 'Medjool'. Base pair changes represent the most commonly observed point mutations in plants, and can also be a consequence of a plant tissue culture process where plant growth regulators are used (Phillips et al. 1994). The extent of variation observed between the different plants further indicates that Dp41 might be a rapidly evolving/changing sequence. It might be hypothesized that 'Barhee' plants are generally more susceptible to these variations and that these variations have resulted in a heterogenous 'Barhee' population among non-tissue culture-derived plants including variation in the copy number of this sequence in the different date palm cultivars. However, since all tissue culture-derived 'Barhee' plants originating from a single 'Barhee' mother plant, no conclusions about the behaviour of this family of sequences through tissue culture can be drawn at this stage and might be subject of a further detailed investigation.

Furthermore, PCR amplification patterns were more variable between individual plants collected from a genome bank than from tissue-derived plants, as would be expected because of the clonal nature of the tissue culture plants. Higher variability might also be due to the age of tested plants. Whereas all tissue culture-derived plants were young with a uniform genetic composition, plants collected from the genome bank were mature and fruit bearing. The complexity to which this fragment evolved indicate a genetic hot-spot for genome mutation.

6.4   Detection of methylation changes in the genome using RDA

6.4.1   DNA methylation

In this study, comparison of methylation levels in different types of plants showed higher methylation of genomic DNA in tissue culture plants compared to that of non-tissue culture plants. Further, genomic DNA of non-tissue culture plants digested with the restriction enzymes *Hpa*II and *Msp*I gave similar profiles on an agarose gel. The methylation process of genomic DNA in plants and the effect a tissue culture process has on genomic methylation are both well documented (Arnold-Schmitt *et al.* 1995; Phillips *et al.*, 1994 Fredorff, 2000; Martienssen and Coleot, 2001). However in date palm, the profiles of enzyme-digested 'Medjool' and 'Barhee' DNA differed, with 'Barhee' DNA showing a higher degree of methylation compared to that of 'Medjool'.

In plants and filamentous fungi, genomic methylation is restricted mostly to repetitive DNA elements (Rabinowicz et al., 1999; Colot and Rossignol, 1999). Higher level of methylation might therefore be due to a higher copy number of the highly repetitive subtraction product DP41.

In *P. strobus* increased 2,4-D concentrations in the medium caused higher genomic DNA methylation and higher methylation gave different digestion profiles of DNA with the restriction enzymes *Hpa*II vs. *Msp*I. Both enzymes recognise the same DNA sequence but only methylation at the restriction site produces differences in the digestibility of the DNA, with methylation of DNA rendering the DNA indigestible for *Hpa*II. Furthermore, amplification of different, mostly larger, size fragments for amplicons derived from *Hpa*II digested DNA of cells grown on a high 2,4-D containing medium might be explained by the inability of *Hpa*II to digest methylated DNA to smaller amplifiable fragments.

6.4.2   Date palm

The isolation of repetitive DNA sequences with RDA has been previously reported (Ushijima *et al.*, 1997; Chen *et al.*, 1998). In this study, repeated RHpa 24 adapter sequences were identified in one isolated and characterized subtraction product. Very likely this product does not represent an PCR artefact because identified repeats are interrupted by an extra three base pairs unlikely to be present in an PCR artefact. Also, a similar repeated DNA has been found in other organisms, such as *Hippoglossus* and banana (Cullis, unpublished results). Although an artefact cannot be completely excluded.

Methylation of repeated DNA is a common phenomenon in plant genomes and the amplification or deletion of sequences resulting in alteration of copy number is a known phenomenon occurring during tissue culture (Brown PTH, 1989). Differences in the copy number of subtraction product between the two genomes might also be associated with differences in the levels of methylation in the different genomes, allowing the isolation of subtraction product. The variation found in the fragment length might be due to the RHpa primer binding at different positions during the first round amplification reaction due to the sequence homology found between the primer and the fragment. resulting in amplified fragments with different sizes.

In this study, when RDA was repeated, the elimination of the R series adaptor resulted in the isolation of a new set of difference products. Such isolation of a different set of difference products has also been reported by Ushijima *et al.* (1997) using RDA-WEEC (RDA - With Elimination of Excessive Clones) to amplify nine novel fragments after obtaining two repetitive sequences in the first RDA. Two of the three fragments obtained in this study as subtraction products – a 444 bp fragment and a 330 bp fragment – showed a high level of homology with rRNA genes. These genes, in themselves, represent repetitive DNA sequences and a high copy number of rRNA genes with about 570 repeats per haploid genome have been found in plants (Pruitt and Meyerowitz, 1986). These fragments were obtained only from 'Medjool'. Seemingly same size fragments, which can be efficiently amplified by PCR, were not amplified for 'Barhee' amplicons due to methylation at the HpaII restriction site in genomic 'Barhee' DNA.

6.4.3 *P. strobus*

The cleavage polyembryony to the early (precotyledonary) dominance stage represents the most responsive embryo developmental stage for somatic embryogenesis initiation (Percy *et al.*, 2000). Finger *et al.* (1989) showed that high induction rates could be obtained when precoteledonary embryos, enclosed in the megagametophytes, were used as explants. In this study, 40% of explants produced embryonal masses and different levels of 2,4-D applied for cell growth did not resulted in any noticeable differences in growth of the cell cultures. However, in this study noticeable differences were obtained in the number of somatic embryos produced where tissue grown on MLV UL produced more mature embryos. Higher levels of methylation to be present in the genome of tissue grown on high 2,4-D medium might be correlated with loss of gene expression and subsequent loss of embryogenic potential.

The subtractions experiments that were done using genomic DNA of *P. strobus* digested with *Hpa* II failed to reduce the sample size and complexity sufficiently in order to isolate difference products. Even after subtractions with a ratio of 1: 100 000, amplification of the samples resulted in a smear with fragment sizes varying between 200 bp and 2000 bp when visualized on an ethidium bromide stained agarose gel. Whereas the standard RDA protocol is sufficient in reducing the complexity of small

and medium size genomes, large genomes such as that of conifers and specifically *P. strobus* presents an obstacle.

Chapter 7

Achievements and conclusions
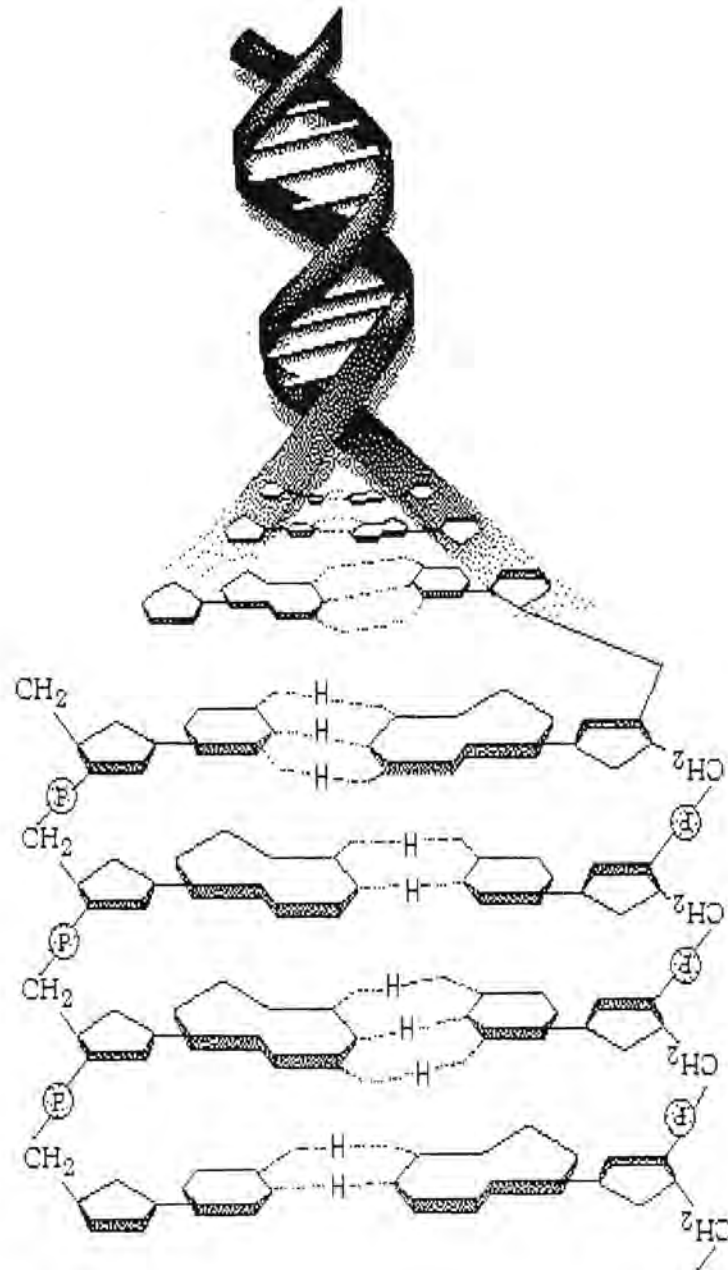
There are always advantages and disadvantages associated with every molecular technique and careful consideration should be given before choosing a specific technique. In this study, the applicability of Representational Difference Analysis (RDA) towards plant genome analysis was investigated as one strategy to identify genome differences in two types of plants. Since RDA was originally developed to isolate genetic differences between cancerous and non-cancerous cell lines, one of challenges in this study was to demonstrate if the RDA technique is at all a useful technique to isolate genetic differences on the genome level between closely related plant cultivars or plant cell lines with different phenotypes, and if any isolated difference might ultimately be useful in establishing a genetic marker that can be used either for cultivar identification or for marker assisted selection of superior cell lines and quality control purposes.

In this study, the well-established technique of Random Amplified Polymorphic DNA (RAPD) was first used to demonstrate if two date palm cultivars 'Barhee' and 'Medjool' could be differentiated at all. RAPD results obtained in this study clearly confirmed the findings of Corniquel and Mercier (1994), and RAPD verified the identity of plant material that was used for RDA. Unfortunately, the more robust technique of Sequence Characterised Amplified Region (SCAR) for plant material verification failed in this study indicating that any change in the date palm genome detected by RAPD is possibly due to only a single nucleotide polymorphism in the original RAPD primer-binding site.

For execution of RDA on date palm material three restriction enzymes were used for amplicon production (*Bam*HI, *Hpa*II and *Hind*III) as outlined in the original paper by Lisitsyn *et al.* (1993) and this study achieved as a first outcome to obtain subtraction products from *Bam*HI and *Hpa*II-derived amplicons. *Bam*HI-derived subtraction products were characterized in this study in much more detail and products represented mainly three related sequences. One of these, Dp41, showed more variability and represented a highly repetitive sequence that was up to a hundred-times more abundant in 'Barhee' compared to 'Medjool'. Amplification of this region from a number of individual plants of both cultivars showed further variation in the sequence between cultivars, as well as between individual plants and between different copies amplified from one plant. Several primers were designed from this sequence data and one showed indeed potential for cultivar identification purposes amplifying the expected fragment consistently only from 'Medjool' plants. Further, the sequence variation that was found showed also certain tendencies such as $C \rightarrow T$

and G→ A changes. The sequence variation occurred, however, to such an extent that it could not be explained by applying the rate of natural random mutation to this sequence. This study therefore achieved as a second outcome to identify a possible genetic mutational hot spot in date palm. It might be speculated, that such hot spots act as a genetic buffer to collect mutations, preventing them from occurring in essential genes or other important regions of the genome.

This study also showed as a third outcome, differences in the genomic DNA methylation status between tissue culture and non-tissue culture plants as well as between plants grown on different 2,4-D concentrations. Degree of DNA methylation was clearly higher in tissue culture-derived date palm plants when compared to non-tissue culture plants. Further, DNA methylation of *P.strobus* cells grown on a high 2,4-D containing media were also higher than of cells grown on a lower 2,4-D containing media, which fits to the theory that 2,4-D leads to hyper-methylation.

This study also focused on application of the RDA technology on large plant genome using cell lines of *Pinus strobus* as a model and testing if possible DNA methylation differences in a genome due to 2,4-D treatment could be detected by the RDA technology. Unfortunately, RDA was inefficient to identify any subtraction products after three rounds of subtractions of *Hpa*II-derived amplicons and subtraction product profile was still too complex representing a mixture of a great number of fragments too difficult to analyze.

In conclusion, the RDA technique could be extended in this study to two further plant species and has shown potential to isolate variable DNA regions, which differ in plant genomes. Using date palm several repetitive elements were isolated. This would be by far more difficult with, for example, RAPD only based on mutations in short DNA primer sites. However, RDA results in plants have still to be interpreted with caution because a subtractive cloning approach is always subject to several sources of bias. First, the representation of the genome in RDA is based on digestion of the genomic DNA with a single restriction enzyme and therefore depends on the sequence of the restriction site. Secondly, the tester to driver ratios used for subtractive hybridization are critical for the selection of specific sequences, so that a bias introduced by a badly chosen ratio could be further amplified during the PCR step (Zoldos *et al.*, 2001).

Although it was demonstrated in this study that RDA could be successfully applied to analyse small plant genomes, more work remains to be done in the future. To fully understand the significance of these products analysis of the flanking sequences need to be done and the position of the sequences in the genome needs to be determined. Using, RDA-WEEC as previously described by Ushijima *et al.* (1997) can be useful to eliminate the repetitive sequences found commonly in plant genomes and isolate other low-copy regions. In terms of larger genomes such as that of *P. stobus* the size and complexity of the genome seems to be an obstacle for the successful isolation of subtraction products, therefore a greater reduction in the genome complexity by the introduction of a second restriction enzyme such as *Msp*I that cuts the sequence A/ATT commonly found within repetitive sequences might be useful. Otherwise cDNA-RDA that is currently been used to determine differences in gene expression between cell lines grown on high and low 2,4-D containing media might be a better approach when dealing with large plant genomes.

The results obtained from this study resulted in two publications Kunert *et al.*, 2000 and Vorster *et al.*, 2002.

References

Ait-Chitt M., Ainsworth CC. and Thangavelu M. 1993. A rapid and efficient method for extraction of total DNA from mature leaves of date palm (*Phoenix dactylifera* l.) Plant Molecular Biology Reports 11: 317-319

Altschul SF., Gish W., Miller W., Myers EW. and Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410

Arnault C. and Dufournel I. 1994. Genome and stresses: reactions against aggressions, behaviour of transposable elements. Genetica 93: 149-160

Arnold-Schmitt B., Herterich S. and Neumann K-H. 1995. Physiological aspects of genome variability in tissue culture. I. Growth phase-dependant differential DNA methylation of the carrot genome (Daucus carota L.) during primary culture. Theoretical and Applied Genetics 91: 809–815

Ayala F. 1970. Teleological explanations in evolutionary biology. Philosophy of Science (March) 1-15

Barakat A., Han DT., Banslimane A-A., Rode A. and Bernardi G. 1999. The gene distribution in the genomes of pea, tomato and date palm. FEBS Letters 463: 139-142

Barral V., This P., Imbert-Establet D., Combes C. and Delseny M. 1993. Genetic variability and evolution of the Schistosoma genome analised by using random amplified polymorphic DNA markers. Molecular Biochemistry and Parasitology 59: 211-222

Bender CM., Pao MM. and Jones PA. 1998. Inhibition of DNA methylation by 5-aza-2'-deoxycytidine suppresses the growth of human tumor cell lines. Cancer Research 58: 95-101

Bennett MD. and Smith JB. (1991). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London* B 334: 309-345

Bennetzen JL. and Kelogg EA. 1997. Do plants have a one-way ticket to genome obesity? Plant Cell 9: 1509-1514

Bennetzen JL., Schrick K., Springer PS., Brown WE. And SanMiguel P. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 37: 565– 576.

Bloom KS. and Carbon J. 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell 29: 305-317

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

Bodens C., Joandet S., Laigret F. and Kremer A. 1997. Detection of genomic regions differentiating two closely related oak species *Quercus petreae* (Matt.) and *Q. robur* (L.). Heredity 78: 433-444

Bouman H. and de Klerk GJ. 1997. Somaclonal variation. In: Biotechnology of Ornamental Plants p. 165-183. Oxford University Press, NY.

Brown PTH. 1989. DNA methylation in plants and its role in tissue culture. Genome 31: 717- 729

Brown TA. 1992. Genetics a molecular approach. Second edition. Chapman and Hall London.

Buchanan BB., Gruissem W. and Jones RL. 2000. Biochemistry and molecular biology of plants. American Society of Plant Physiologists, Rochville, Maryland.

Burns TD., White TJ. and Taylor JW. 1991. Fungal molecular systematics. Annual Review of Ecological Systematics 22: 524- 564

Buscot F., Wipf D., Di Battista C., Munch J-C, Botten B. and Martin F. 1996. DNA polymorphisim in morels: PCR/ RFLP analysis of the ribosomal DNA spacer and microsatellite primed PCR. Mycology Research 100: 63- 71

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

Casacuberta JM., Vernhettes S. and Grandbastien M-A. 1995. Sequence variability within the tobacco retrotransposon *Tnt*1 population. EMBO Journal 14: 2670-2678

Chandler V., Rivin C. and Walbot W. 1986. Stable non-mutator stocks of maize have sequences homologous to the Mu1 transposable element. Genetics 114: 1007-1021

Chang Y., Cesarman E., Pessin M.S., Lee F., Culpepper J., Knowles DM. and Moore PS. 1994. Identification of herpes virus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266: 1865-1869

Chen M., SanMiguel P. and Bennetzen JL. 1998. Sequence organizatin and conservation in sh2/ a1- homologous regions of sorgum and rice. Genetics 148: 435-443.

Colot V. and Rossignol. 1999. Eukaryotic DNA methylation as a evolutionary device. BioEssays 21: 402- 411

Copenhaver GP. and Preuss D. 1999. Centromeres in the genomic era: unravelling paradoxes. Current Opinions in Plant Biology 2: 104- 108

Corniquel B. and Mercier L. 1994. Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. Plant Science 101: 163-172

Crow JF. 1997. The high spontaneous mutation rate: Is it a health risk? Proceedings of the National Academy of Science USA 94: 8380-8386

Cullis CA. 1990. DNA rearrangements in response to environmental stress. Advances in Genetics 28: 73-97

Cullis CA. 1999. Environmental Stress - a generator of adaptive variation? In: H.R. Lerner (ed.). Plant Adaptations to Stress Environments. Marcel Dekker, New York, in press.

Cullis CA. and Kunert KJ. 2000. Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. Acta Horticulturae 530: 421-428

Cullis CA., Rademan S. and Kunert KJ. 1999. Method for finding genetic markers of somaclonal variation. International publication number WO 99/53100.

Curtis CA. and Lukaszewski AJ. 1991. Genetic linkage between C- bands and storage protein genes in chromosome 1B of tetraploid wheat. Theoretical and Applied Genetics 81: 245- 252

Deng Z., Huang S., Xiao S. and Gmitter JR. FG. 1997. Development and characterization of SCAR markers linked to the *Citrus tristeza*. Genome 40: 697-740

Donnison IS., Siroky J., Vyskot B., Saedler H. and Grant SR. 1996. Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. Genetics 144: 1893-1901

Eggleston WB., Alleman M. and Kermicle JL. 1995. Molecular organization and germinal instability of *R- stippled* maize. Genetics 141: 347- 360

Evans DA. 1988. Applications of somaclonal variation. Biotechnology in Agriculture p. 203-223

Fedoroff N. 2000. Transposons and genome evolution in plants. Proceedings of the National Academy Of Science USA 97: 7002-7007

Fedoroff N., Furtek D. and Nelson O. 1984. Cloning of the Bronze locus in maize by a simple and generalizable procedure using the transposable controlling element activator (AC). Proceedings of the National Academy Science USA 81: 3825-3829

Fedoroff N., Wessler S. and Shure M. 1983. Isolation of the transposable maize controlling elements Ac and Ds. Cell 35: 243- 251

Feldman M., Liu B., Segal G., Abbo S., Levy AA. and Vega JM. 1997. Rapid elimination of low- copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of Homoeologous chromosomes. Genetics 147: 1381- 1387.

Fitzgerald-Hayes M., Clarke L. and Carbon J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell 29: 235-244

Flavell AJ., Dunbar E., Anderson R., Pearce SR., Hartly R. and Kumar A. 1992b. Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Research 20: 3639-44

Flavell AJ., Smith DB. and Kumar A. 1992a. Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. Molecular and General Genetics. 231:233-42

Flavell RB., Bennett MD., Smith JB. and Smith D. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochemical Genetics 12: 257-269

Finer JJ., Krebel HB. and Becwar MR. 1989. Initiation of embryogenic callus and suspension cultures of eastern white pine (*Pinus strobus* L.) Plant Cell Reports 8: 203- 206

Foster LM., Kozac KR., Loftus MG., Stevens JJ. and Ross IK. 1993. The polymerase chain reaction and its application to the filamentous fungi. Mycological Research 97: 769-781

Gill BS., Fiebe B. and Endo TR. 1991. Standard karyotype and nomenclature system for description of chromasome bands and structural aberrations in wheat (*Triticum aestivum*). Genome 34 : 830- 839

Grandbastien M-A. 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Science 3: 181-187

Graur D., Shuali Y. and Li WH. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. Journal of Molecular Evolution 28: 279-285

Gu WK., Weeden NF., Yu J. and Wallace DH. 1995. Large-scale, cost-effective screening of PCR products in marker assisted selection application. Theoretical and Applied Genetics 91: 465- 470

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

Hadrys H., Balick M. and Schierwater B. 1992. Applications of RAPD fingerprinting in molecular ecology. Molecular Ecology

Heldt HW. 1997. Plant Biochemistry & Molecular Biology. Oxford University Press, New York.

Henry RJ. 1997. Practical Applications of Plant Molecular Biology. Chapman & Hall, London.

Hirochika H. 1993. Activation of tobacco retrotransposons during tissue-culture. EMBO Journal 12: 2521-2528

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H. and Kanda M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. Proceedings of the National Academy of Science USA 93: 7783-7788

Hu W., Das OP. and Messing J. 1995. *Zeon*-1, a member of a new maize retrotransposon family. Molecular and General Genetics 248: 471-480

Jiang JM., Nasuda S., Dong FG., Scherrer CW. and Woo SS. *et al.*, 1996 A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proceedings of the National Academy of Science USA 93:14210-14213

Jin Y-K. and Bennetzen JL. 1989. Structure and coding properties of *Bs*1, a maize retrovirus-like transposon. Proceedings of the National Academy of Science USA 86:6235-6239

Johnston JS., Jensen A., Czeschin DG. and Price HJ. 1996. Environmentally induced nuclear DNA content instability in *Helianthus annus* (Asteraceae). American Journal of Botany 83: 1113-1120

Jones N., Ougham H. and Thomas H. 1997. Markers and mapping: we are all geneticists now. New Phytologist 137: 165-177

Judson HF. 1996. The Eighth Day of Creation. Cold Spring Harbor Press, New York

Karp A. 1993. Are your plants normal? - Genetic instability in regenerated and transgenic plants. Agro-Food-Industry Hi-Tech: May/June 7-12

Kunert KJ., Vorster J., Bey E. and Cullis C. 2000. Introduction of DNA-based quality assurance procedure for commercially micropropagated date palms in South Africa. Proceedings of the International Date Palm Symposium,Windhoek/Namibia, pp 74-80.

Larkin PJ. and Scowcroft WR. 1981. Somaclonal variation - a novel source of variability from cell cultures for plant improvement. Theoretical and Applied Genetics 60: 197-214

Laszlo E. 1996. The systems view of the world: A holistic vision for our time. (Advances in systems theory, complexity, and the human sciences). Hampton Press, Australia.

Lee M. and Phillips RL. 1986. Genomic rearrangements in maize induced by tissue culture. Genome 29: 122-128

Leroy XJ., Hily JM. and Chaumeil P. 2001. Detection of in vitro culture-induced instability through inter-simple sequence repeat analysis. Theoretical and Applied Genetics 102: 885-891

Leutwiler LS., Hough- Evans BR. and Meyerowitz EM. 1984. The DNA of *Arabidopsis thaliana*. Genetics 194: 15- 23

Levin I., Crittenden LB. and Dogson JB. 1993. Genetic map of the Z chromosome using random amplified polymorphic DNA (RAPD) markers. Genomics 16: 224- 230

Levitt J. 1972. Responses of plants to environmental stresses. Academic Press, New York and London

Lieb M. and Rehmat S. 1997. 5-Methylcytosine is not a mutation hot spot in nondividing *Escherichia coli*. Proceedings of the National Academy of Science USA 94: 940- 945

Linacero R., Alves EF. and Vazquez AM. 2000. Hot spots of DNA instability revealed through the study of somaclonal variation in rye. Theoretical and Applied Genetics 100: 506-511

Lisitsyn N., Lsistsyn N. and Wigler M. 1993. Cloning the differences between two complex genomes. Science 259: 946-950

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

MacLean DJ., Braithwaite KS., Manners JM. and Irwin JAG. 1993. How do we identify and classify fungal pathogens in the era of DNA analysis? In Advances in Plant Pathology Volume 10. Andrews JH. and Tommerup IC. (Ed.) Academic Press London

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

Martiensenn RA. and Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293: 1070- 1074

Masson P., Surosky R., Kingsbury JA. and Fedoroff NV. 1987. Genetics 177: 177-137

McClintock B. 1984. The significance of responses of the genome to challenge. Science 226: 792-801

McCouch SR. 2001. Genomics and synteny. Plant Physiology 125: 152-155

Mhiri C., De Wit PJGM. and Grandbastien M-A. 1999. Activation of the promoter of the *Tnt1* retrotransposon in tomato after inoculation with the fungal pathogen *Cladosporium fulvum*. Molecular Plant-Microbe Interactions 12: 592-603

Mhiri C., Morel JB., Vernhettes S., Casacuberta JM., Lucas H. and Grandbastien MA. 1997. The promoter of the tobacco *Tnt*1 retrotransposon is induced by wounding and by abiotic stress. Plant Molecular Biology 33: 257-266

Miller JT., Dong F., Jackson SA., Song J. and Jiang J. 1998. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150: 1615- 1623.

Moor G., Lucas H., Batty N. and Flavell R. 1991. A family of retrotransposons and associated genomic variation in wheat. Genomics 10: 461- 468.

Moreau-Mhiri C., Morel JB., Audeon C., Ferault M., Grandbastien MA. and Lucas H. 1996. Regulation of expression of the tobacco *Tnt*1 retrotransposon in heterologous species following pathogen related stresses. Plant Journal 9: 409-419

Murray B. (1998). Nuclear DNA amounts in gymnosperms. *Annals of Botany* 82 (Supplement A): 3-15.

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

Navin A, Prekeris R, Lisitsyn NA, Sonti MM, Grieco DA, Narayanswami S, Lander ES, Simpson EM. 1996. Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics 36: 349-353

Nekrutenko A, Makova KD, Baker RJ. 2000. Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene 249: 47-51

Nelke M., Nowak J., Wright J.M. and Mclean N.L. 1993. DNA fingerprinting of red clover (*Trifolium pratens* L.) with Jeffrey's probes: detection of somaclonal variation and other applications. Plant Cell Report 13: 72-78

Nierras CR., Liebman SW. and Warner JR. 1997. Does Saccharomyces need and organized nucleolus? Chromosoma 105: 444-451

Nixon RW. (1950) Imported varieties of dates in the United States. US Department of Agriculture Circular No. 834

O'Brien IEW., Smith DR., Gardner RC. and Murray BG. 1996. Flow cytometric determination of genome size in Pinus. Plant Science 115: 91-99

O'Hanlon PC., Peakall R. and Briese DT. 2000. A review of new PCR- based genetic markers and their utility to weed ecology. Weed Research 40: 239- 254

O'Niell RJW., O'Niell, MJ., and Graves, MA. 1998. Undermethylation associated with retroelements activation and chromosomes remodelling in an interspecific mammalian hybrid. Nature 393: 68-72

Olszewska MJ. and Osiecka R. 1982. The relationship between 2C DNA content, life-cycle type, systematic position, and the level of DNA endoreplication in nuclei of parenchyma cells during growth and differentiation of roots in some monocotyledonous species. Biochemie Und Physiologie Der Pflanzen 177: 319-336

Paran I. and Michelmore, RW. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics 85: 985-993

Pardue ML., Danilevskaya ON., Lowenhaupt K., Slot F. and Traverse KL. 1996. Drosophila telomeres: new views on chromosome evolution. Trends in Genetics 12: 48-52

Pearce SR., Harrison G., Li D., Heslop-Harrison JS., Kumar A. and Flavell AJ. 1996. The Ty1-copia group retrotransposons in Vica species: copy number, sequence heterogeneity and chromosomal location. Molecular and General Genetics 250: 305-315

Pearson WR. and Lipman DJ. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Science USA 85: 2444-2448

Percy RE.,m Klimaszewska K. and Cyr DR. 2000. Evaluation of somatic embryogenesis for clonal propagation of eastern white pine. Canadian Journal of Forest Reasearch 30: 1867- 1876.

Peschke VM., Phillips RL. and Gengenbach BG. 1987. Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science 238: 804-807

Petrov DA., Lozovskaya ER. and Hartl DL. 1996. High intrinsic rate of DNA loss in Drosophila. Nature 384: 346-349

Phillips RL., Kaeppler SM. and Olhoft P. 1994. Genetic instability of plant tissue cultures: Breakdown of normal controls. Proceedings of the National Academy of Science USA 91: 5222-5226

Poulsen GB., Kahl G. and Weising K. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassic napus* L. Theoretical and Applied Genetics 85: 994-1000

Pouteau S., Gandbasien MA., Boccara M. 1994. Microbial elicitors of plant defense response activate transcription of a retrotransposon. Plant Journal 5: 535-542

Pruitt RE., and Meyerowitz EM. 1986. Characterization of the genome of *Arabidopsis thaliana.* Journal of Molecular Biology 187: 169-183

Rabinowicz PD., Schutz K., Dedhia N., Yordan C., Parnell LD., Stein L., McCombie WR. and Martienssen RA. 1999. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genetics 23: 305- 308

Rothnie HM, McCurrach KJ., Clover LA. and Hardman N. 1991. Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum.* Nucleic Acids Research 19: 279- 286

Russell JR., Fuller JD., Macaulay M., Hatz BG., Jahoor A., Powell W. and Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics 95: 714–722

Sabir A., Newbury HJ., Todd G., Catty J. and Fors-Lloud BV. 1992. Detection of genetic stability using isozymes and RFLPs in beet plants regenerated in vitro. Theoretical and Applied Genetics 84: 113-114

Sambrook J., Fritsch EF. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Sandhu D. and Gill KS. 2002. Gene- containing regions of wheat and the other grass genomes. Plant Physiology 128: 803- 811

Sandhu D., Champoux JA., Bondareva SN. and Gill KS. 2001. Identification and physical localization of useful genes and markers to a major gene- rich region on wheat group *1S* chromosomes. Genetics 157: 1735- 1747

SanMiguel P., Gaut, BS., Tikhonov, A., Nakajima, Y. and Bennetzen, JL. 1998. The palaeontology of intergene retrotransposons of maize. Nature Genetics 20: 43-45

SanMiguel P., Tikhonov A., Jin YK., Motchoulskaia N. and Zakharov D, et al., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765-68

Schilling AG. 1996. Polymerase chain reaction-based assays for species-specific detection of *Fusarium culmorum*, *F.graminearum*, and *F.avenaceum*. Phytopathology 86: 515 522

Schwarz-Sommer Z., Gieri A., Cuypers H., Peterson PA. and Saedler H. 1985. Plant transposable elements generate the DNA sequence diversity needed for evolution. EMBO Journal 4: 591-597

Scribner KT. and Pearce JM. 2000. Microsatellites: evolutionary and methodological background and empirical applications at an individual, population and phylogenetic levels. In Molecular methods in ecology, Baker A (ed), Blackwell London pp. 235- 271

Sedra MH., Lashermes P., Trouslot P., Combes MC. and Hamons S. 1998. Identification and genetic diversity analysis of date palm (*Phoenix dactilyfera* L.) varieties from Morocco using RAPD markers. Euphytica 103: 75- 82.

Selker EU. and Garrett PW. 1988. DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. Proceedings of the National Academy of Science USA 85: 6870- 6874.

Selker EU., Cambareri EB., Jenesen BC. and Haack KR. 1987. Rearrangement of duplicated DNA in specialized Neurospora.Cell 51: 741- 752.

Skirvin RM. and Janick J. 1976. Tissue culture-induced variation in scented *Pelargonium* spp. Journal of the American Society for Horticultural Science 101: 281-290

Skirvin RM., McPheeters KD. and Norton M. 1994. Sources and frequency of somaclonal variation. HortScience 29: 1232-1237

Smith TF. and Waterman MS. 1981. Identification of common molecular subsequences. Journal of Molecular Biology 25; 147: 195-197

Smyth DR., Kalitsis P., Joseph JL. and Sentry JW. 1989. Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of *Drosophila*. Proceedings of the National Academy of Science USA 86: 5015- 5019.

Soller M. and Beckmann JS. 1986. Restriction fragment length polymorphisims in poultry breeding. Poultry Science 65: 1474- 1488

Strathdee G. and Brown R. 2002. Aberrand DNA methylation in cancer: potential clinical interventions. Expert Reviews in Molecular Medicine. 4 March pp. 1- 17

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

Suoniemi A., Anamthawat-Jonsson K., Arna T. and Schulman AH. 1996. Retrotransposon BARE-1 is a major, dispersed component of the barley (Hordeum vulgare L.) genome. Plant Molecular Biology 30: 1321- 29

Takeda S., Sugimoto K., Otsuki H. and Hirochika H. 1998. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Molecular Biology 36: 365–376.

Takeda S., Sugimoto K., Otsuki H. and Hirochika H. 1999. A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant Journal 18: 1- 11.

Thomas CA. 1971. The genetic organization of chromosomes. Annual Review of Genetics 5: 237-256

Toder R., Grutzner F., Haaf T. and Bausch E. 2001. Species-specific evolution of repeated DNA sequences in great apes. Chromosome Research 9: 431-435

Twyman RM. 1998. Advances in molecular biology: a concise reference. Bios Scientific Publishers, Springer- Verlag, New York.

Ushijima T., Morimura K., Hosoya Y., Okonogi H., Tatematsu M., Sugimura T. and Nagao M. 1997. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. Proceedings of the National Academy of Science USA. 94: 2284-2289

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

Vongs A., Kakutani T., Martienssen RA. and Richards EJ. 1993. *Arabidopsis thaliana* DNA methylation mutants. Science 260: 1926- 1928.

Vorster BJ., Kunert KJ. and Cullis CA. 2002. Use of representational difference analysis for the characterization of sequence differences between date palm varieties. Plant Cell Reports 21:271–275

Voytas DF., Cummings MP., Konieczny A., Ausubel FM. and Rodermel SR. 1992. *Copia*-like retrotransposons are ubiquitous among plants. Proceedings of the National Academy of Science USA 89: 7124-7128

Wang DG., Fan JB., Siao CJ., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris MS., Shen N., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson TJ., Lander ES., et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082

Welsh J. and McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213- 7218

White SE., Habera LF. and Wessler SR. 1994. Retrotransposons in the flanking regions of normal plant genes – a role for *copia*-like elements in the evolution of gene structure and expression. Proceedings of the National Academy of Science USA 91: 11792- 96

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

Wright WE., Brasiskyte D., Piatyszek MA. and Shay JW. 1996. Experimental elongation of telomeres extends the lifespan of immortal x normal cell hybrids. EMBO Journal 15: 1734-1741

Yoder JA., Walsh CP. and Bestor TH. 1997. Cytosine methylation and the ecology of intagenomic parasites. Trends in Genetics 13: 335-40

Zakian VA. 1996. Structure, function and replication of Saccharomyces cerevisiae telomeres. Annual Review of Genetics 30: 141-172

Zheng KL., Castiglione S., Biasini MG., 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

Zoldos V., Siljak-Yakovlev S., Papes D., Sarr A. and Panaud O. 2001. Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. Molecular Genetics and Genomics 265: 234-241

Annex

I   :   Molecular methods

II   :   Tissue culture

III   :   General solutions and buffers

IV   :   Plasmid maps

V   :   Primer sequences

I) Molecular Methods

I.I Plant material and DNA extraction

'Medjool' and 'Barhee' plant material originated from 14 plants grown in California that have been propagated solely via off-shoots from mother plant material imported in the beginning of the last century from Northern Africa (Nixon 1950). The *in vitro* plants used were 'Medjool' derived from explant material collected in California, and 'Barhee' derived from explant material collected in the United Arabic Emirates.

Total cellular DNA was isolated from leaves (1 g) using the Nucleon Phytopure Plant DNA extraction kit (Amersham Life Science, UK) according to the manufacturer's instructions. To test the quality and amount of isolated DNA, samples of isolated DNA (1 µl) were run on a 1% agarose gel in TAE buffer (0.04 M Tris-acetate; 1 mM EDTA, pH 8) as outlined by Sambrook et al. (1989). After staining of the gel with ethidium bromide for 15 min, DNA quality was determined on a white/UV-transilluminator, photographed using a Grab-IT system (Vacutec, USA) and the DNA concentration of the samples was visually determined using 4 different λ-phage DNA amounts as standards (25 ng, 50 ng, 100 ng and 250 ng) for comparison.

I.II Polymerase chain reaction (PCR)

Standard DNA amplifications by PCR were carried out in 25 µl reaction volumes containing 500 mM KCl, 25 mM $MgCl_2$, 100 mM Tris-HCl (pH 8.3), 25 mM of each dNTP and 5 units TaKaRa DNA polymerase (TaKaRa, Japan) in a GeneAmp PCR 9600 system (Perking Elmer, USA). Primers for PCR were obtained either from Operon Technologies (Operon Technologies, USA) or were designed using the online primer design tool of Molecular Biology Shortcuts (MBS), program 'Oligos and Primers' (www.mbshortcuts.com/biotools/index.htlm). The designed primers were then manufactured and purchased from MWG-Biotech AG (Germany). The standard PCR program consisted of 94°C (5 min) to denature double stranded DNA. This was followed by 35 cycles of amplification consisting of denaturing DNA at 94°C (1 min), primer annealing at 55°C or higher depending on primer pair (1 min), and extension of the DNA chain at 72°C (2 min). This was followed by another extension cycle at 72°C for 5 min.

I.III   Representational difference analysis (RDA)

I.III.I   *Preparation of RDA amplicons*

RDA was performed following the general outline described by Lisitsyn et al. (1993). Two micrograms of each of the DNAs ('Barhee' and 'Medjool') were digested with 80 units of the enzyme *Bam*HI or *Hind*III (Amersham Life Science, UK) 100 µg of yeast tRNA (Sigma-Aldrich, USA) was added to the digest to limit non-specific binding of DNA against the sides of eppendorf tubes and pipet tips. The digests were then extracted with an equal volume of phenol/chloroform. The DNA containing upper phase was removed and the DNA precipitated using 1/10 volume 3 M sodium acetate and 2.5 volumes 100% ethanol. The precipitate was washed in 70% ethanol and dried. The digested DNA was dissolved in 18 µl of TE buffer (10 mM Tris-Cl, 1 mM EDTA), pH 8, at a concentration of 100 µg/ml.

I.III.II   Ligation of adaptor sequences

One µg of the *Bam*HI and *Hind*III digested DNA were ligated to the adaptor pair RBam 12 and RBam 24 or RHind 12 and RHind 24, respectively. The DNA, primer set (0.6 µM), 10X ligase buffer and water to a final volume of 30 µl were mixed in a microcentrifuge tube. The tubes were placed in a heating block at 55°C and the block was the placed at 4°C until the temperature had fallen to 12°C (60 to 75 min). The tubes were then placed on ice for 3 min, after which 1 µl of T4 DNA ligase (1 unit) (Amersham Life Science, UK) was added and the reaction incubated at 16°C overnight. The ligation reaction was then diluted with 970 µl TE buffer, pH 8, and the amplification reaction was set up as follows:

80 µl 10X *Taq* Buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$)
64 µl dNTP mix (160 mM of each dNTP)
16 µl (1.0 µM) of one primer of the appropriate pair (J Bam 24 or R Hind 42)
80 µl of the diluted ligation mixture
560 µl water

The tube was placed in a heating block at 72°C for 5 min to melt of the 12-mer oligonucleotide. Six µl (30 units) of *Taq* polymerase (TaKaRa, Japan) was added and mixed by pipetting. The mixture was then aliquoted into eighth PCR tubes in a Perkin-Elmer GeneAmp 9600 thermocycler, the block of which was being held at 72°C. After 5 minutes, the thermocycler program run was: 20 cycles: 95°C for 30 sec; 72°C for 2 min, followed by 10 min at 72°C, followed by holding at 4°C.

The 8 tubes were combined an a 10 µl aliquot was run on a 1.5% TAE agarose gel at 100 V for 30 minutes to check the amplification and estimate the concentration by comparison to lambda phage DNA standards. The remainder was extracted with 600 µl phenol/chloroform. The upper phase was removed and the DNA precipitated with 3 M sodium acetate and an equal volume propanol. The precipitate was collected by centrifugation, washed twice with 70% ethanol and dried. The amplicons was redissolved in 80µl TE buffer, pH 8.

I.III.III Removal of the adaptors from amplicons

Of the amplicons that was to be used as driver 150 µg and 10 µg of the tester amplicons were digested with the appropriate enzyme in 800 µl and 200 µl, respectively, at an enzyme concentration of 20 units/µg DNA, at 37°C for one hour. Ten µg Yeast tRNA was added and each digest extracted with phenol/chloroform. The DNA was precipitated with an equal volume of isopropanol and 1/10 volume 3 M sodium acetate, as described above. Isopropanol is used instead of ethanol in order to eliminate very small fragments (e.g. adaptors) from the precipitate. The driver and tester DNAs were both redissolved at approximately 400 µg/ml. Both the digested driver and tester amplicons were run on a 1.5% TAE agarose gel alongside an equal aliquot of undigested amplicons to check the completeness of the digestion. On the same gel, standard lambda phage DNA was run so that the final concentrations of driver and tester DNAs could be estimated and adjusted if necessary.

I.III.IV Change of adaptors on tester amplicons

Tester DNA was prepared by adding a second adaptor pair JBam 12 and 24 for *Bam*HI digested DNA or JHind 12 and 24 for *Hind*III digested DNA to 1 µg of the first

round amplicons in the same way as described for the ligation of the first set. The ligation reaction was then diluted to a final volume of 100 µl with water. An aliquot of the ligate was amplified for 20 cycles in a reaction volume of 20 µl to check that the newly ligated adaptors would support amplification with the new primer.

I.III.V   Subtractive hybridisation and kinetic enrichment

The hybridisation reaction was set up by mixing 80 µl of the driver (~ 40 µg) with 40 µl tester (~ 0.4 µg) amplicons (a ratio of driver : tester of 100 : 1). To this was added 30 µl of 10 M ammonium acetate and 380 µl 100% ethanol. After incubation on ice for 10 min the DNA precipitate was collected by centrefugation for 10 min at full speed in a microcentrifuge. The pellet was washed twice with 70% ethanol and dried after which it was redissolved in 4 µl 3X EE buffer [30 mM N-(2-hydroxyethyl piperizine)-N'-(3-propene sulfonic acid) (EPPS), pH 8, 3 mM EDTA] followed by repeated vortexing interspersed with centrifugation. The solution was collected at the bottom of the tube and overlaid with light mineral oil so that the spherical droplet could be seen to be completely covered by oil. The DNA was denatured by placing it in a heating block at 98°C for 5 minutes. One µl of 5 M sodium chloride solution was added and the tube briefly centrifuged to mix the aqueous phases. The DNA was then incubated at 67°C overnight. Ten µl TE buffer, pH 8, was added and mixed with the aqueous phase, which was then carefully removed from under the oil by pipetting and added to clean microcentrifuge tube containing 380 µl TE buffer and 40 µg tRNA. The appropriately annealed fragments were amplified in the following reaction:

80 µl 10X *Taq* buffer
64 µl dNTP mix (160 mM of each dNTP)
80 µl of the diluted hybridisation mixture
560 µl ds H_2O
6 µl *Taq* polymerase (30 units)

The tube was placed in a heating block at 72°C for 5 min to fill in the ends. Sixteen µl of the appropriate primer (the 24mer used in the adapter ligated to the tester amplicons) was added, mixed by pipetting, and the mixture aliquoted into eighth PCR tubes. The tubes were placed in a Perkin-Elmer thermocycler, the block of which, were being held at 72°C.

The following thermocycler program was then run: 10 cycles of 95°C for 30 sec; 72°C for 2 min followed by 10 min at 72°C, followed by holding at 4°C. The tubes were combined and 20 µl of the solution subjected to a further 20 cycles of amplification under the same conditions.

Ten µg of tRNA was added to the remainder of the amplified reaction and then extracted with 600 µl phenol/chloroform. An aliqoute (750 µl) was removed and 75 µl (1/10 volume) of 3 M sodium acetate and 825 µl (one volume) propanol added. The precipitate was collected by centrifugation, after incubation on ice for 10 minutes. The pellet was washed twice with 70% ethanol, dried and dissolved in 40 µl TE buffer, pH 8. At this stage the aliquot subjected to the additional amplification cycles was run on a 1.5% TAE agarose gel to check that amplification occurred. This first round difference product (20 µl) was digested with 20 units of mung bean nuclease (Amersham Life Science, UK) at 30°C for 30 minutes to remove all single- stranded DNA. The reaction was stopped by the addition of 160 µl of TE buffer, pH 8. The digested product was amplified in a reaction mixture consisting of:

80 µl 10X *Taq* buffer

64 µl dNTP mix (160 mM of each dNTP)

16 µl of the same primer used for the 10-cycle amplification

80 µl of the diluted nuclease treated mixture

560 µl ds H_2O

6 µl *Taq* polymerase (30 units)

The mixture was aliquoted into eighth PCR tubes in a Perkin-Elmer thermocycler, the block of which was being held at 72°C. The thermocycler program run was 20 cycles: 95°C for 30 seconds; 72°C for 2 minutes followed by 10 minutes at 72°C, followed by holding at 4°C.

The 8 tubes were combined and a 10 µl aliquot was run on a 1.5% TAE agarose gel at 100 V for 30 min to check the amplification. The remainder was extracted with phenol/chloroform. The upper phase was pipetted into a clean tube and a 1/10 volume of 3 M sodium acetate and 1 volume of propanol added. The tubes were

mixed by inversion and placed on ice for 15 min. The precipitate was collected by centrifugation for 15 min at full speed in a microsentrifuge, washed twice with 70% ethanol and dried. The amplicons were redissolved in 80 μl of TE buffer, pH 8. The concentration was estimated by electrophoresis on a 1.5% TAE agarose gel with lambda phage DNA standards and adjusted to 100 μg/ml.

I.IV   Cloning of the subtraction products

Subtraction products (2 μg) were digested with the 50 units of the appropriate restriction enzyme for 30 minutes at 37°C. One hundred ng of this digest was mixed with 10 ng of pBluescriptII vector (Strategene, USA) digested with the same enzyme (a ratio of insert to vector of 10 : 1). These were extracted with phenol/chloroform and precipitated with sodium acetate and ethanol. The precipitate was collected by microcentrifugation and the pellet washed twice with 70% ethanol before being dried. The dried pellet was redissolved in 8 μl water. To this was added 1 μl 10X ligase buffer and 1 μl T4 ligase, and the mixture incubated overnight at 16°C. Two μl of the ligate was used to transform competent XL1Blue cells and 50 plasmid-containing colonies carrying an insert were selected and probed with either the 'Barhee' or 'Medjool' labeled driver amplicons using the *Gene Images* random prime-labeling module (Amersham Life Sciences UK). Ten colonies that showed a much stronger signal after hybridization with the 'Barhee' amplicons than with the 'Medjool' amplicons were selected for plasmid isolation and determination of the insert sequence and size.

I.V   Southern-blot hybridization

Total genomic DNA of 'Barhee' and 'Medjool' (1 μg) were digested with two units of each of the restriction enzymes *Bam*HI and *Eco*RI at 37°C for two hours. The restricted DNA was then separated on a 1.5% TAE agarose gel using electrophoresis, and transferred to a positively charged nylon membrane (Roche, Switzerland), as described by Sambrook et al. (1989). Cloned fragments of Dp41 and Dp36 were used as probes, which were labeled using the *Gene Images* random prime, labeling module. Membranes were pre-hybridized and hybridized at 65°C in a hybridization buffer containing 5% SSC, 0.1% SDS and 20-fold dilution of the liquid

block provided. Hybridization was carried out overnight and the membranes washed at 60°C using a 1% SSC and 0.1% SDS solution. Detection was preformed using the *Gene Images CDP-Star* detection module (Amersham Life Sciences UK), according to the manufacturer's instructions. Membranes were exposed to Hyperfilm ECL (Amersham Life Sciences UK) and the films developed.

I.VI   Sequence analysis

DNA sequence analysis was carried out with the dideoxy chain terminator method developed by Sanger et al. (1997). Recombinant plasmids were sequenced using fluorescent dye terminators and AmliTaq from the ABI PRISM™ Big DYE Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems, Inc., Foster City, Calif.), in a cycle sequence protocol according to the recommendations of the manufacturer. Sequence reactions were run on PAGE (polyacrylimide gel electrophoresis) using an ABI PRISM™ 377 Autosequencer (Perkin-Elmer Applied Biosystems). Sequence electropherograms were analysed using Sequence Navigator version 1.0.1 (Perkin-Elmer).

II)   Tissue culture and somatic embryogenesis

II.I   Tissue culture initiation

The most responsive embryo developmental stages for somatic embryogenesis initiation range from the cleavage polyembryony, to the early (precotyledonary) dominance stage (Percy *et al.*, 2000). For culture initiation, seeds were extracted from developing cones and placed in Petri dishes containing moistened filter paper. Within 1 hour of extraction, seeds were surface sterilized in preparation for dissection. Seeds were surface sterilised by placing them in metal baskets and stirring vigorously for 6 min in 200 ml of 3% (v/v) hydrogen peroxide with one drop of Tween 20. Seeds were then rinsed tree times for 2 min each with sterile water and transferred to a petridish containing a moist, sterile filter paper. The seed coat, nucellus, and membrane surrounding the metagametophyte were removed under a stereomicroscope, and the megagametophyte was placed on initiation medium. Five to eight megagametophytes were cultured in each Petri dish containing approximately 25 ml of medium, and dishes were sealed with parafilm. Explants were cultured in darkness at 24°C for 16 weeks and examined every 4 weeks for embryogenic tissue initiation.

II.II   Embryogenic tissue proliferation and culture maintenance

Explants displaying proliferation from the zygotic embryo cells were microscopically selected, then embryogenic tissue was removed, dispersed by vigorous shaking in liquid MLV maintenance medium, poured onto a sterile filter paper disk (Whatman No. 2, 5.5 cm) in a Büchner funnel and a short low-pressure pulse (55 kPa) was applied to drain the medium prior to transferring the disc to fresh medium. This approach was also used for maintenance of embryogenic tissue. Maintenance and proliferation cultures were incubated in darkness at 24°C.

II.III   Maturation of somatic embryos

The somatic embryo experiments were performed by combining embryogenic tissue of one line (from several plates), one week after subculture, in a 50 ml test tube,

adding liquid media without growth regulators and vigorously shaking the tube to break up the clumps of tissue into a fine suspension. Subsequently 3 ml containing approximately 0.2 g of suspended embryonal masses were withdrawn with a wide-mouth pipette and placed on a moist filter paper disk (Whatman No. 2, 5.5 cm in diameter) in a Büchner funnel attached to a vacuum pump. A short low-pressure pulse (5 sec, 55 kPa) was applied to drain all liquid medium and anchor the embryonal masses to the filter paper as a thin layer. Each disk of filter paper with the embryonal masses was subsequently placed on maturation medium in a 10 X 20 mm Petri dish and cultured for up to 10 weeks. The cultures were kept under dim light condition at 1.6 μmol m^{-2}s^{-1} from cool white florescence lamps (Philips F72T12/CW, 65 W) under a 16 h photoperiod at 24°C.

II.IV   Tissue culture media

II.IV.I   MLM 10X stock

	2 Liters	Final concentration in medium [g/L]
NH$_4$NO$_3$	16.5 g	0.825
KNO$_3$	19.0 g	0.95
MgSO$_4$.7H$_2$O	18.5 g	0.925
KH$_2$PO$_4$ (monobasic)	3.4 g	0.17
CaCl$_2$.2H$_2$O	0.22 g	0.011
LM Minor stock (100X)	200 ml	
LM Vitamin stock (100X)	200 ml	
Myo-Inisitol	2.0 g	0.1
Iron Solution	100 ml	
dH$_2$O to	2 L	
Stored at -20°C or 2 weeks at +4°C		

II.IV.II   MLM Minor Stock 100X

	1 Liter	Final concentration in medium [g/L]
$Na_2MoO_4.2H_2O$	0.125 g	0.00125
KI	0.415 g	0.00415
H_3BO_3	3.1 g	0.031
$MnSO_4.H_2O$ (or $MnSO_4.4H_2O$)	2.1 g (2.7 g)	0.021 (0.027)
$ZnSO_4.7H_2O$	4.3 g	0.043
$CuSO_4.5H_2O$	0.05 g	0.0005
$CoCl_2.6H_2O$	0.013	0.00013
DH_2O to	1 L	
Stored at -20°C		

II.IV.III   MLM Vitamin stock 100X

	1 Liter	Final concentration in medium [g/L]
Nicotinic acid	0.05 g	0.0005
Pyridoxine.HCL	0.01 g	0.0001
Thiamine.HCL	0.01 g	0.0001
dH_2O to	1 L	
Stored at -20°C		

II.IV.IV   MLM Iron solution

	100 ml	Final concentration in medium [g/L]
Na_2EDTA	0.746 g	0.0373
$FeSO_4.7H_2O$	0.556 g	0.0287
dH_2O to	100 ml	
Made fresh		

II.V **Growth regulators**

II.V.I   2,4-Dichlorophenoxyacetic Acid (2,4-D) stock (1 mg/ml)

2,4-D	100 mg
95% Ethanol	50 ml
DH$_2$O to	100 ml
Stored at +4°C	

II.V.II   6-Benzyl-Aminopurine (BA) stock (2.5 mg/ml)

BA	50. mg
Warmed 0.5 N HCL (or NaOH)	5 ml to dissolve
Warmed dH$_2$O to	100 ml
Stored at +4°C	

II.V.III   (±) cis-trans Abscic (ABA) stock (10mM)

ABA	0.02643 g
1 N NaOH	drops to dissolve
dH$_2$O to	10 ml
Filter sterilized and wrapped in foil. Made fresh.	
1 ml stock/ L = 10 μM	

II.V.IV   Glutamine stock (25 mg/ml) (Amino acids)

Glutamine	25.0 g
dH$_2$O to	1 L
pH 5.8. Filter sterilized. Stored frozen	

II.VI   MLV Media

II.VI.I   MLV Ultra low medium (MLV UL)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	2 %	20.0 g
2,4-D (1 mg/ml stock)	1.1 μM	0.25 ml
BA (0.5 mg/ml stock)	1.1 μM	0.5 ml
dH$_2$O to		1 L
Phytagel	0.4 %	4 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml

II.VI.II   MLV Standard medium (MLV Std)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	2 %	20.0 g
2,4-D (1 mg/ml stock)	9.5 μM	2 ml
BA (0.5 mg/ml stock)	4.4 μM	2 ml
dH$_2$O to		1 L
Phytagel	0.4 %	4 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml

II.VI.III   Maturation medium (MM)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	6 %	60.0 g
dH$_2$O to		1 L
Phytagel	1 %	10 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml
ABA (10 mM sterile stock)	120 µM	12 ml

III   General solutions and buffers

III.I   Antibiotics

III.I.I   Ampicillin stock solution (50 mg/ml)

Dissolve 2 g Ampicillin (D(-)-a-Aminobenzylpenicillin sodium salt) (Sigma-Aldrich, Germany) powder in 40 ml of sterile distilled water. Aliquote in 1.5 ml tubes and store at -20°C.

III.I.II   Tetracycline stock solution (10 mg/ml)

Dissolve 0.1 g of tetracycline powder (Sigma-Aldrich, Germany) in 5 ml 100% ethanol, adjust volume to 10 ml with sterile distilled water. Cover tube with foil and store at -20°C.

III.II   Southern blotting solutions

III.II.I   Neutralization buffer

Combine 121.1 g of Tris base (Sigma-Aldrich, Germany) and 87.7 g NaCl. Add 900 ml distilled water. Adjust pH to 8.0 with concentrated HCl. Adjust volume with distilled water to 1 L and autoclave.

III.II.II   Denaturing buffer

Combine 87.7 g NaCl and 20 g NaOH. Add distilled water to 1 L. Stir to dissolve and autoclave.

III.II.III 20X Standard Saline-Citrate (SSC)

Combine 174.4 g NaCl, 88.3 g $C_6H_5Na_3O_7.2H_2O$ (Sodium-citrate) in 800 ml distilled water. Stir to dissolve and adjust volume to 1 L. Autoclave.

III.III   Buffers and salt solutions

III.III.I   Agarose gel buffer (TAE) (50X stock)

Combine 968 g Tris base, 228.4 ml glacial acetic acid and 400 ml 0.5 M EDTA, pH 8.0. Add 3 L distilled water and stir until solids dissolve. Adjust volume to 4 L with distilled water. Dilute to 1X before use.

III.III.II   Agarose gel loading buffer (6X)

Combine 0.063 g bromophenol blue (BPB), 0.063 g xylene cyanol FF (XC) and 2.5 ml glycerol. Add distilled water to 25 ml. Aliquot into 1.5 ml tubes and store at -20°C. Add 1 µl to sample before loading onto agarose gel.

III.III.III   Ethidium Bromide solution

Dissolve 0.2 g ethidium bromide powder in 20 ml distilled water. Vortex to dissolve and cover container with foil.

III.III.IV   5 M Ammonium acetate (NH₄OAc)

Dissolve 385.4 g ammonium acetate in 500 ml distilled water by slowly adding the ammonium acetate powder into the stirring water. Adjust volume to 1 L. Autoclave and store at RT.

III.III.VI   3 M Sodium acetate (NaOAc) pH 4.8

Dissolve 40.82 g NaOAc.3H$_2$O in 80 ml distilled water and stir on a hot plate until dissolved. Adjust pH to 4.8 with glacial acetic acid. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.VII   5 M Sodium chloride

Dissolve 29.22 g NaCl in 80 ml distilled water. Adjust volume to 100 ml, autoclave and store at RT.

III.III.VIII   1 M Tris-HCl pH 8.0

Dissolve 12.11 g Tris base in 80 ml of distilled water. Adjust pH to 8.0 using concentrated HCl. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.IX   0.5 M EDTA pH 8.0

Dissolve 8 g NaOH pelles in 400 ml distilled water. Add 93.05 g Na$_2$EDTA.2H$_2$O and allow to dissolve. Adjust pH to 8.0 using more NaOH pellets. Adjust volume to 500 ml with distilled water. Autoclave and store at RT.

III.III.X   TE buffer (10 mM Tris-HCL, 1 mM EDTA)

Combine 1 ml 1M Tris-HCl pH 8.0 and 0.2 ml 0.5 M EDTA pH 8.0. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.XI   Low TE buffer (10 mM Tris-HCL, 0.1 mM EDTA)

Combine 1 ml 1M Tris-HCl pH 8.0 and 20 µl 0.5 M EDTA pH 8.0. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.XII   10 N Sodium hydroxide (NaOH)

Dissolve 40 g NaOH pellets in 70 ml distilled water. Adjust volume to 100 ml. Store at RT.

IV. Plasmid maps

IV.I   pBluescript II (Stratagene USA)

V.   Primer sequences

RAPD and SCAR primers	
OPE-01	5'-CCCAAGGTCC-3'
OPE-06	5'-AAGACCCCTC-3'
DpSL	5'- GTGTTAGGGGCAAAATGGAA–3'
DpSR	5'- TTGTCCGTCTGAGACTCCCT–3'
Date palm subtraction product specific primers	
DP36L	5'-CTATCGACGACAGGCTGACA -3'
DP36R	5'-GACCCGGACTTGTTGGAGTA-3'
DP41L	5'-CCTTCTCCCCGTAGTAACCG-3'
DP41R	5'-AGGAAAGGCAACCTACCGAG-3'
DP50L	5'-TACACGATGTCCCTCAACCA-3'
DP50R	5'-GGAACATTTCCTCGGTATCC-3'
PLM 1	5'-TTACAGAGGGGAAAGGAGGA-3'
PLM 4	5'-GGAAGGAGGTGGCTCCG-3'
PLB11	5'-CGCAATCTTGCAAGTATCAGT-3'
RDA adapters	
R Bam 24	5'-AGCACTCTCCAGCCTCTCACCGAG-3'
R Bam 12	5'-GATCCTCGGTGA-3'
J Bam 24	5'-ACCGACGTCGACTATCCATGAACG-3'
J Bam 12	5'-GATCCGTTCATG-3'
N Bam 24	5'-AGGCAACTGTGCTATCCGAGGGAG-3'
N Bam 12	5'-GATCCTCCCTCG-3'
R Hind 24	5'-AGCACTCTCCAGCCTCTCACCGCA-3'
R Hind 12	5'-AGCTTGCGGTGA-3'
J Hind 24	5'-ACCGACGTCGACTATCCATGAACA-3'
J Hind 12	5'-AGC TTGTTCATG-3'
N Hind 24	5'-AGGCAGCTGTGGTATCGAGGGAGA-3'
N Hind 24	5'-AGCTTCTCCCTC-3'

R Hpa 24	5'-AGCACTCTCCAGCCTCTCACCGAC-3'
R Hpa 11	5'-CGGTCGGTGAG-3'
J Hpa 24	5'-ACCGACGTCGACTATCCATGAAAC-3'
J Hpa 11	5'-CGGTTTCATGG-3'
N Hpa 24	5'-AGGCAACTGTGCTATCCGAGGGAC-3'
N Hpa 11	5'-CGGTCCCTCGG-3'
S Hpa 24	5'-ACTTCTACGGCTGAATTCCGACAC-3'
S Hpa 12	5'-CGGTGTCGGAAT-3'
Sequencing primers	
T7	5'-TAATACGACTCACTATAGGG-3'
Sp6	5'-CATACGATTTAGGTGACACTATAG-3'

GENETICS AND GENOMICS

B.J. Vorster · K.J. Kunert · C.A. Cullis

Use of representational difference analysis for the characterization of sequence differences between date palm varieties

Abstract Representational difference analysis was applied to subtract the genomes of the two date palm varieties, Barhee and Medjool, for identification and characterization of unique genome differences suitable for discriminating between individual plants and the two varieties. Three different DNA difference products were isolated from Barhee representing families of dispersed, repeated variable sequences present in the genome of both varieties. Several variant members of repeated DNA were detected by sequence analysis, containing base changes from C to T and G to A and short deletions. Mutated DNA sequences could be amplified in a polymerase chain reaction-based test from a much smaller number of Barhee plants than from Medjool plants allowing the differentiation between individual plants and partial discrimination between varieties.

Keywords Date palm · Representational difference analysis · Genome analysis · Hyper-variable DNA · Plant variety identification

Introduction

Genomes of closely related plants or varieties can differ by only a few coding genes or in minor genome re-organizations and a range of different approaches is available for detection of such genetic differences. Among these are the analyses of r-DNA intergenic regions (Scribner

Communicated by J.C. Register III

C.A. Cullis (✉)
Biology Department, Case Western Reserve University,
10900 Euclid Avenue, Cleveland, OH 4404, USA
e-mail: cac5@po.cwru.edu
Tel.: +1-216-3685362, Fax: +1-216-3684672

B.J. Vorster · K.J. Kunert
Botany Department,
Forestry and Agricultural Biotechnology Institute,
74 Lunnon Road, University of Pretoria, Hillcrest, Pretoria 0002,
South Africa

and Pearce 2000) and of simple sequence repeats (SSRs), which are also known as microsatellites. SSRs have been used for example for identification of varieties of *Chrysanthemum* and *Citrus* (Wolff et al. 1995; Fang and Roose 1997) as well as individual plants of oilseed rape cultivars (Charters et al. 1996). However, the two most widely used genetic techniques to detect plant variation are random amplified polymorphic DNA (RAPD) analysis, which detects DNA polymorphisms amplified by arbitrary primers (Williams et al. 1990; Welsh and McClelland 1990) and amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995; O'Hanlon et al. 2000). Recently, the RAPD technique has also been applied to identify date palm on the variety level (Corniquel and Mercier 1994; Sedra et al. 1998). All these techniques are useful for the differentiation of plants by identifying random polymorphisms. However, the comparisons are usually made on the basis of the presence or absence of a band rather than directly on any DNA sequence variation.

In contrast, representational difference analysis (RDA) allows the cloning and sequencing of fine genome differences between two highly similar genomes and further provides exact sequence information about these differences. RDA detects any kind of labile DNA region in two genomes to be compared and can be used to derive probes for genomic losses, rearrangements, amplifications, point mutations and pathogenic organisms found within any of the genomes compared (Lisitsyn et al. 1993, 1994; Michiels et al. 1998). RDA has been applied in a variety of contexts including the isolation of repetitive sequences present in only one of the compared genomes (Navin 1996). Nekrutenko et al. (2000) used RDA to create a species-specific marker for voles and Toder et al. (2001) have applied RDA in evolutionary genomics to search for overall genome differences between humans and the great apes. RDA has also been used to determine differences between two distantly related oak species where similarities of isolated RDA fragments with known retrotransposons were found (Zoldos et al. 2001). In addition, Donnison et al. (1996)

272

applied RDA to identify male-specific restriction fragments in the dioecious plant *Silene latifolia*. RDA has also been used to identify polymorphisms in banana lines that are a result of genomic rearrangements during in vitro propagation resulting in markers useful for the detection of early variation in the initiation of tissue culture plants (Cullis and Kunert 2000). One of the specific advantages of RDA is that subtractions between pooled DNA samples can be performed in order to identify specific polymorphisms that are only present in either a particular individual or a particular variety rather than relying on identification based on a pattern of polymorphic bands.

The aim of this study was to apply the RDA technology to date palm and to investigate if it is a useful technique for generation of markers suitable for identification and characterization of variable regions in the date palm genome. Such variable regions might ultimately be applicable for either individual date palm plant or variety identification/discrimination. In general, there is a need for DNA-based markers in date palm that will facilitate variety identification as well as confirmation of the genetic fidelity of individual propagated plants. The current identification based on morphological characteristics is often difficult. For example, in the Kingdom of Saudi Arabia alone almost 400 date palm cultivars have been classified (Hussain and El-Zeid 1978), based mainly on their fruit characteristics that are expressed in the mature stage of the plant. In this study, plants of the two date palm varieties Barhee and Medjool, which are commonly used in commercial tissue culture for the asexual production of date palm via somatic embryogenesis, have been investigated. A repetitive labile DNA region has been identified allowing the differentiation between individual plants and partial discrimination between varieties.

Materials and methods

Plant material and DNA extraction

The date palm varieties Medjool and Barhee were used as the source of DNA. The in vitro plants used were Medjool derived from explant material collected in California, and Barhee derived from explant material collected in the United Arabic Emirates. Non-tissue culture-derived Medjool and Barhee plant material originated from 14 plants grown in California that have been propagated solely via off-shoots from mother plant material imported in the beginning of the last century from Northern Africa (Nixon 1950). Total cellular DNA was isolated from leaves (1 g) using either the technique outlined by Aitchitt et al. (1993) or the Nucleon Phytopure plant DNA extraction kit (Amersham Life Sciences, UK) according to the manufacturer's instructions. Both techniques gave similar results.

Preparation of RDA amplicons

RDA was performed following the general outline described by Lisitsyn et al. (1993). Two micrograms of each of the DNAs (Barhee and Medjool) were digested with 80 units of the enzyme *Bam*HI or *Hin*dIII. The digests were then extracted with phenol/chloroform, precipitated and resuspended at a concentration of 100 µg/ml. The *Bam*HI and *Hin*dIII digests were ligated to the adaptor pair JBam 12 (5′-GATCCGTTCATG-3′) and JBam 24 (5′-ACCGACGTCGACTATCCATGAACG-3′) or RHind 12 (5′-AGCTTGCGGTGA-3′) and RHind 24 (5′-AGCACTCTCCA-GCCTCTCACCGCA-3′), respectively. The ligation products were amplified by polymerase chain reaction (PCR) by using the primer JBam 24 or RHind 24 to generate the first-round amplicons, followed by digestion with *Bam*HI or *Hin*dIII to remove the adaptors. Tester DNA was prepared by adding a second adaptor pair NBam 12 and 24 (5′-GATCCTCCCTCG-3′ and 5′-AGGCA-ACTGTGCTATCCGAGGGAG-3′) for *Bam*HI-digested DNA or JHind 12 and 24 (5′-AGCTTGTTCATG-3′ and 5′-ACCGAC-GTCGACTATCCATGAACA-3′) for *Hin*dIII-digested DNA to the ends of the first-round amplicons.

Subtractive RDA hybridization, kinetic enrichment and cloning of the difference products

The hybridization reaction was set up using 40 µg driver DNA (Medjool) and 0.4 µg tester DNA (Barhee) (100:1 driver/tester ratio) in a final volume of 4 µl hybridization buffer consisting of 30 m*M* EPPS [(2-hydroxyethyl piperizine)-*N*′-(3-propene sulfonic acid)], pH 8, and 3 m*M* EDTA. The DNA was denatured at 100°C for 10 min, 1 µl of sodium chloride (5 *M*) was added to a final concentration of 1 *M* and the reaction incubated at 67°C for 16 h. The hybridization mix was then diluted and an aliquot amplified using NBam 24 or JHind 24. The first round of amplification was for ten cycles, followed by digestion of the products by mung bean nuclease. The nuclease-treated product was then amplified for an additional 20 cycles. The resulting amplicons, which are called the first difference product, were used in this study. These subtraction products were digested with the appropriate restriction enzyme and ligated into the appropriately digested pBluescript II (Strategene, USA). The ligation products were transformed into XL1Blue-competent cells and 50 plasmid-containing colonies carrying an insert were selected and probed with either the Barhee- or Medjool-labeled driver amplicons using the Gene Images random prime-labeling module (Amersham Life Sciences). Colonies that showed a much stronger signal after hybridization with the Barhee amplicons than with the Medjool amplicons were selected for plasmid isolation and determination of the insert sequence and size.

Primer design and testing

Pairs of primers were designed using a standard design program (Expasy, Switzerland). The primer pairs were used in a PCR reaction using Barhee and Medjool DNA as template at various annealing temperatures to optimize the PCR reaction. The PCR reactions were carried out in 25-µl volumes containing 25 ng total genomic DNA, 15 ng primer, 100 m*M* of each dNTP, 10 m*M* TRIS-HCl, pH 8.3, 2 m*M* MgCl$_2$ and 0.5 units Taq polymerase (Takara, Japan). Amplification was performed using a Perkin Elmer Gene-Amp PCR system 9600 with the following program: (1) 94°C for 5 min×1 cycle; (2) 94°C for 1 min, 65°C or 60°C for 1 min depending on the primer pair, 72°C for 1 min×35 cycles; (3) 72°C for 5 min×1 cycle, and optional soak period at 4°C. The products were separated on a 1.5% agarose gel, stained with ethidium bromide and visualized under UV light.

Southern blot hybridization

Genomic DNA was digested with different restriction enzymes, separated on a 1.5% TAE agarose gel using electrophoresis, and transferred to a positively charged nylon membrane (Roche, Switzerland), as described by Sambrook et al. (1989). Cloned bands were used as probes and were labeled using the Gene Images random prime, labeling module (Amersham Life Sciences). Mem-

Fig. 1 Sequence analysis of isolated RDA clones DP2, DP36 and DP41 and position of the most variable DNA region of the DP41 clone

```
DP-2   CCTATCGAAC CCATTCATAC AGAGCCAGTT TTCAATGTCC CTCAACCATC GCGCGGATCT AGTAGGGTCT CCCATCCTCC  80

DP-36  CCTATGGCGA CCGGCTGACC TGGCACTGGT GTCCGCAGCA ACTCTGCTCG GATGGAAAGA AGTCGACCTC GACGAAAGCG  80

DP-41  CCTTCTCCCC GTAGTAACCG GCCTCCCCGC AAATCTTGCA AGTATCACTG AGGGGGAAGA AGGAGGAGGG GCCTCCGCAC  80


       CGATAGATAC TTAGGTATAC TAGAAGAGGA TACCGAGAAA ATGTTCCTAG TGGGAGATAG AGATCACATAC AGGAT  156

       GCTCGGTAAA GCCCCGGTAG TACTCCAACA AGTCCGGGTC AATCCGACGG TATCTCCTCG CGCTGGAT  147

       GTGCCTCACG TCGTGGGGAC ACCGTAGATG GCTCGGTAGG TTGCCTTTCC TCCGTTGGAT  141
```

branes were pre-hybridized and hybridized at 65°C in a buffer containing 5× SSC, 0.1% SDS and a 20-fold dilution of the liquid block provided overnight and washed at 60°C using a 1× SSC and 0.1% SDS solution. Detection was performed using the Gene Images CDP-Star detection module, according to the manufacturer's instructions. Membranes were exposed to Hyperfilm ECL (Amersham Life Sciences) and the films developed.

Sequence analysis

For plasmid sequencing, PCR products were cloned into the plasmid pMOS*Blue* (Amersham Pharmacia Biotech, UK) and plasmids were recovered from transformed MOS*Blue* cells selected on an appropriate antibiotic. Sequencing was performed using Sequenase (Perkin Elmer, USA) according to the manufacturer's instructions on an automated DNA sequencer (Applied Biosystems, USA).

Table 1 Sequences of primers used to amplify different parts of clone DP41 with DP41R 5'-GCAACCTACCGAGCCATCT-ACGGT-3' as the right-hand primer and the difference product DP36 with DP36R 5'-GGAACATTTTCTCGGTATCCTC-3' as the right-hand primer, and a total number of seven non-tissue culture-derived plants amplifying a DNA fragment with the predicted size. *Numbers in parentheses* indicate the total number of plants tested

Primer	Sequence (5'–3')	Amplification product	
		Barhee	Medjool
DP41L	CCTTCTCCCCGTAGTAACCG	5 (7)	7 (7)
PLM1	TTACAGAGGGGAAAGGAGGA	5 (7)	7 (7)
PLM4	GGAAGGAGGTGGCTCCG	1 (7)	7 (7)
PLB11	CGCAATCTTGCAAGTATCAGT	2 (7)	6 (7)
DP36L	TCGAACCCATTCATACAGAGC	4 (7)	7 (7)

Results

The subtractions between Barhee and Medjool were performed with either Barhee as tester and Medjool as driver or with Medjool as tester and Barhee as driver for amplicons derived from *Bam*HI- or *Hin*dIII-digested DNAs. Following a single round of subtraction using a tester to driver ratio of 1:100 only one of the four subtractions, that with Barhee *Bam*HI-digested DNA as tester and Medjool as driver, produced a DNA difference product, which is approximately 150 bp in length. The difference product was cloned and 50 *Escherichia coli* colonies containing the cloned difference product (data not shown) hybridized separately with labeled Barhee tester and Medjool driver amplicons. Both sets of amplicons hybridized to the colonies indicating that the isolated difference product or a closely related sequence was present in both varieties.

Ten *E. coli* colonies with the greatest differential signal between tester and driver hybridization were selected, plasmid DNA isolated and the insert sequence determined. This analysis revealed that the cloned difference product consisted of at least three sequences with lengths of 141 bp (DP41), 147 bp (DP36), and 156 bp (DP2) (Fig. 1) indicating a complex mixture of fragments in the difference product. Seven of the sequenced inserts were identical to DP2 and two identical to DP36. The DP41 sequence contained an AGG motif repeated in tandem 3 times. A search of these three sequences using different databases and DNA sequence analysis tools, such as Blast, FastA and the Smith-Waterman algorithm, result-

ed in a 75% homology to *Oryza sativa* genomic DNA, chromosome 1 (accession no. AP002902) when a local alignment was done with DP41 but no homology was found for DP2 or DP36.

Since most of the sequenced clones were identical to DP2, we selected only the clones DP36 and DP41, the latter of which was unique, for further characterization. Hybridization of labeled DP41 to *Eco*RI- or *Bam*HI-digested DNAs from tissue culture-derived Medjool and Barhee gave a smear pattern, which is characteristic of a dispersed, repetitive sequence (data not shown). Similar hybridization profiles were obtained when either DP36 or DP2 were used as probes.

Four primers, DP41L and DP41R and DP36L and DP36 R (Table 1), were designed from the DP41 and DP36 sequences. These primer pairs amplified the predicted 120-bp or 110-bp PCR product, respectively, from genomic DNA of six different tissue culture-derived Medjool and Barhee plants but failed to amplify a fragment in two non-tissue culture-derived Medjool and three non-tissue culture-derived Barhee plants (Fig. 2; Table 1). The DP41 amplification product was then characterized in more detail. From each of the six tissue-culture-derived Barhee and Medjool plants, three independent clones of the DP41 amplification product were sequenced (Fig. 3). These 36 clones from different Medjool and Barhee plants revealed a high degree of homology with DP41, the differences being minor base pair changes or single base deletions occurring mainly in a
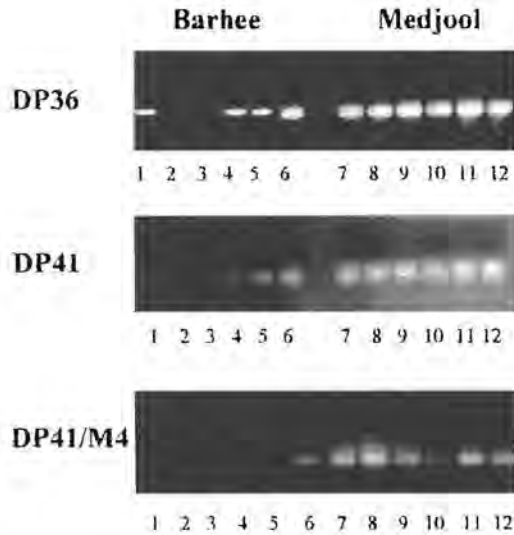
Fig. 2 PCR products of genomic DNA from six out of seven Barhee (B) and Medjool (M) plants amplified with primers for DP36, DP41 and DP41/M4 with the following template DNAs: *lanes 1–6* non-tissue culture-derived B plants; *lanes 7–12* non-tissue culture-derived M plants

Fig. 3 Sequence alignment of DP41 with amplified DNA products from genome of M and B plants showing differences to DP41. *Empty space* indicates a base pair deletion. For abbreviations, see Fig. 2

variable 45-bp region of the fragment (Fig. 1). However, in general more changes were observed in the Barhee sequences (specifically single base deletions) than in the Medjool sequences. Only those sequences showing differences in comparison to DP41 are shown in Fig. 3. From the 18 sequences analyzed for each variety, six of the Medjool sequences and eight of the Barhee sequences were identical to DP41. This indicates that the region of the DP41 sequence used to design the primers is common to both genomes. Among the variants found within the different genomes, two, M6 and B9, were identical. Two variants, M1 and M4, the latter with a six base deletion in its sequence, were unique to Medjool and one variant (B11) was unique to Barhee.

Primers PLM1, PLM4 and PLB11 (Table 1) were designed from M1, M4 and B11, to cover the variable portion of these sequences (Fig. 3) when used in conjunction with DP41R. All six tissue culture-derived Medjool plants and also all tissue culture-derived Barhee plants, which originated from a single mother plant, amplified a PCR product with the expected size with all three primers (data not shown). However, primer PLB11 only amplified a PCR product with the expected size of about 110 bp from two of seven non-tissue culture-derived Barhee and six of seven Medjool plants at an optimal annealing temperature of 65°C (Table 1). An identical result was observed with primer PLM1 at an optimal annealing temperature of 60°C (Table 1). Primer PLM4 (at 65°C annealing temperature), which covered a unique 6-bp deletion, amplified a PCR product from DNA of all seven non-tissue culture-derived Medjool plants (Fig. 3) but only from one Barhee plant (Table 1).

Discussion

RDA has been successfully applied in this study for date palm and has resulted in the isolation of three repetitive DNA sequences from the Barhee genome. A subset of each of these families of sequences appears either to be restricted to Barhee, or present at different multiplicity in Barhee, since the reverse subtraction with Medjool as tester yielded no products. The presence of these sequences in the DNA from both varieties could also be explained by methylation differences between the two varieties, with Medjool DNA being more methylated in this region, since *Bam*HI, a methylation-sensitive restriction enzyme, was used to generate the initial amplicons. However, the lack of difference products when *Bam*HI-derived amplicons of genomic Barhee DNA were used as a driver, or when the subtractions were performed with *Hin*dIII-derived amplicons, is consistent with a general high degree of genomic similarity between plants of the two varieties.

One sequence, which is highly variable, has been further characterized with respect to its presence in individual date palm plants of two date palm varieties. Sequence analysis of the amplified difference products from both date palm varieties identified the existence of several variant members of the repeated sequence, consistent with this family representing a variable "genetic hotspot" in the genome (Linacero et al. 2000). Variation included changes from C→T and G→A, deletion of single base pairs (which occurred at a higher frequency in Barhee) and deletion of several base pairs. Base pair changes represent the most commonly observed point mutations in plants, and can also be a consequence of plant tissue culture (Phillips et al. 1994). The extent of variation observed between plants indicates that this might be a rapidly evolv-

ing/changing sequence. We currently hypothesize that Barhee plants are generally more susceptible to these variations and that these variations have resulted in a heterogeneous Barhee population among non-tissue culture-derived plants including variations in copy number. Since all tissue culture-derived Barhee plants originated from a single Barhee mother plant, conclusions about the behavior of this family of sequences through tissue culture cannot be drawn. However, Pluhar et al. (2001) found an unequal copy number of repeated DNA among callus samples of alfalfa, and speculated that genomic stress induced by tissue culture may have caused that unequal copy number.

In the results reported here, RDA has proved to be useful in identifying a particular repetitive class of sequences that is highly variable in date palms, which is consistent with earlier observations that RDA can be used to isolate families of repetitive sequences (Cullis and Kunert 2000; Nekrutenko et al. 2000; Zoldos et al. 2001). These types of sequences are more difficult to identify with either AFLPs or RAPDs since they either result in many related polymorphisms or generate no size polymorphisms. An additional advantage of RDA is that it can also be performed using bulked amplicons, and can thus be used to identify polymorphisms that are restricted to a particular group of individuals. Therefore, bulking a series of Barhee and Medjool samples and then doing the subtraction will identify variety-specific polymorphisms, rather than individual specific polymorphisms. The types of sequences identified in these experiments would likely be identified as polymorphisms using other techniques, but due to the hyper-variability, each individual would have a unique pattern, or there would be a number of different patterns, none of which would be variety specific. The ideal probes for identifying genomes are those which have a unique location in a given genotype that can be identified directly. Application of RDA therefore offers the opportunity to generate such useful probes.
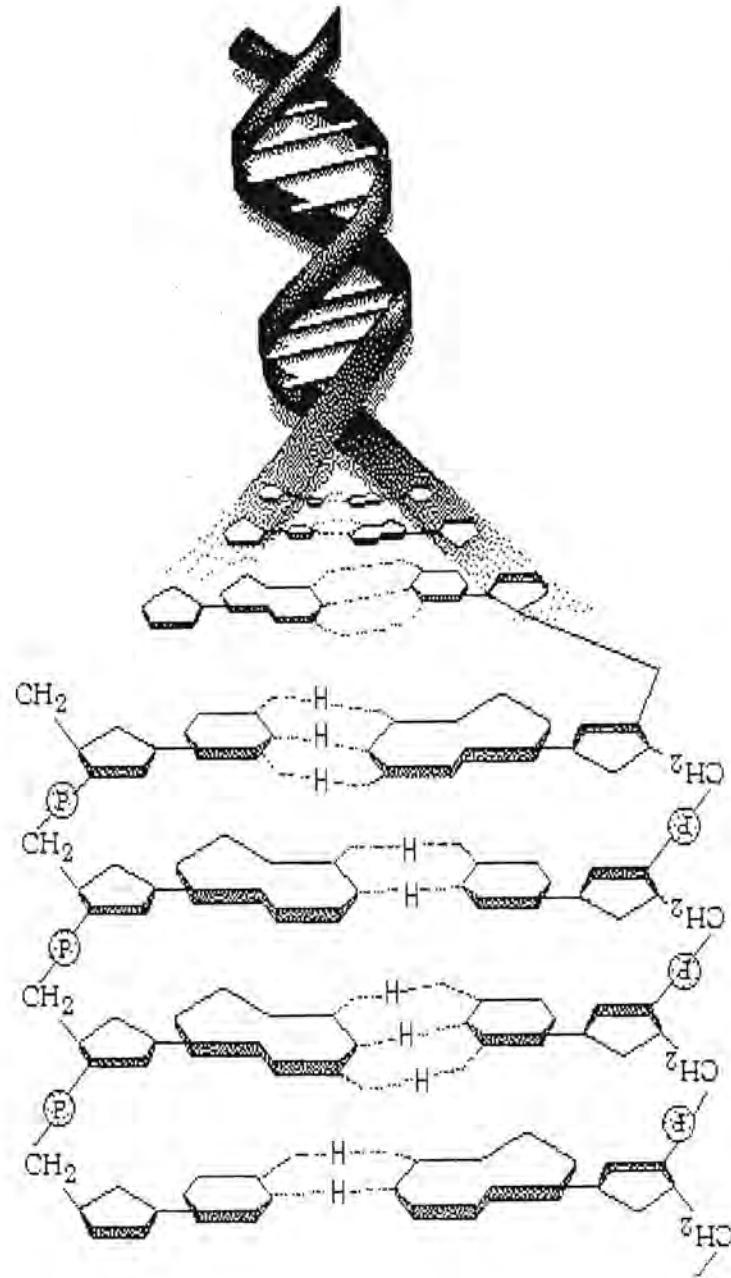
References

Aitchitt M, Ainsworth CC, Thangavelu M (1993) A rapid and efficient method for the extraction of total DNA from mature leaves of the date palm (*Phoenix dactylifera* L.). Plant Mol Biol Rep 11:317–319

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5′-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442–447

Corniquel B, Mercier L (1994) Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. Plant Sci 101:163–172

Cullis C, Kunert KJ (2000) Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. Acta Hortic 530:421–428

Donnison IS, Siroky J, Vyskot B, Saedler H, Grant SR (1996) Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. Genetics 144:1893–1901

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theor Appl Genet 95:408–417

Hussain F, El-Zeid A (1978) Studies on physical and chemical characteristics of date varieties of Saudi Arabia. Ministry of Agriculture and Water, Saudi Arabia

Linacero R, Alves EF, Vazquez AM (2000) Hot spots of DNA instability revealed through the study of somaclonal variation in rye. Theor Appl Genet 100:506–511

Lisitsyn N, Lsistsyn N, Wigler M (1993) Cloning the differences between two complex genomes. Science 259:946–950

Lisitsyn NA, Segre JA, Kusumi K, Lisitsyn NM, Nadeau JH, Frankel WN, Wigler MH, Lander ES (1994) Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nat Genet 6:57–63

Michiels L, Van Leuven F, Van den Oord JJ, De Wolf-Peeters C, Delabie J (1998) Representational difference analysis using minute quantities of DNA. Nucleic Acids Res 26:3608–3610

Navin A, Prekeris R, Lisitsyn NA, Sonti MM, Grieco DA, Narayanswami S, Lander ES, Simpson EM (1996) Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics 36:349–353

Nekrutenko A, Makova KD, Baker RJ (2000) Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene 249:47–51

Nixon RW (1950) Imported varieties of dates in the United States. US Department of Agriculture circular no. 834. US Department of Agriculture, Washington, D.C.

O'Hanlon PC, Peakall R, Briese DT (2000) A review of new PCR-based genetic markers and their utility to weed ecology. Weed Res 40:239–254

Phillips RL, Kaeppler SM, Olhoff P (1994) Genetic instability of plant tissue cultures: breakdown of normal controls. Proc Natl Acad Sci USA 91:5222–5226

Pluhar SA, Erickson L, Pauls KP (2001) Effects of tissue culture on a highly repetitive DNA sequence (E180 satellite) in *Medicago sativa*. Plant Cell Tissue Organ Cult 67:195–199

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Scribner KT, Pearce JM (2000) Microsatellites: evolutionary and methodological background and empirical applications at individual, population, and phylogenetic levels. In: Baker A (ed) Molecular methods in ecology. Blackwell, London, pp 235–271

Sedra MH, Lashermes P, Trouslot P, Combes MC, Hamon S (1998) Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. Euphytica 103:75–82

Toder R, Grutzner F, Haaf T, Bausch E (2001) Species-specific evolution of repeated DNA sequences in great apes. Chromosome Res 9:431–435

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

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisims amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wolff K, Zietkiewicz E, Hofstra H (1995) Identification of *Chrysanthemum* cultivars and stability of DNA fingerprint patterns. Theor Appl Genet 91:439–447

Zoldos V, Siljak-Yakovlev S, Papes D, Sarr A, Panaud O (2001) Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. Mol Genet Genom 265:234–41

Copy number	the number of copies of a given gene in a set of chromosomes, see multigene family.
Cross-hybridization	the binding of a probe to a DNA sequence other than the intended target sequence. This occurs because of homology between the probe and the sequence and because low stringency hybridization wash coinditions are followed.
Cultivar	a variety of plant developed through selective breeding programs.
Cytosine (C)	pyrimidine base that pairs with guanosine in DNA.
Deletion	removal of gene region or base pair from chromosome.
Denaturation	for DNA or RNA, describes separation of double-stranded molecule to a single-stranded state, usually by heating; for protein, describes change in physical shape, which is usually renders it inactive.
Digested DNA	DNA cleaved by the action of restriction enzymes.
Diploid	the normal number of chromosomes (two copies of each – 2n) in virtually all eukaryotes.
Direct repeats	multiple identical (or closely related) nucleotide sequences in the same orientation in a DNA molecule.
DNA	(deoxyribonucleic acid) the molecular basis of heredity. DNA consists of a polysugar-phosphate backbone from which the bases (nucleotides) project. DNA forms a double helix that is held together by hydrogen bonds between specific pairs of bases (thymine to adenine, guanine to cytosine). Each strand in the double helix is complementary to its partner strand in terms of its base sequence.
DNA ligase	enzyme that joins two double-stranded DNAs together, end to end, by catalyzing 3'OH and 5'P termini bond formation.
DNA polymerase	an enzyme that catalyzes synthesis of DNA under direction of a single-stranded DNA template.

References

Ait-Chitt M., Ainsworth CC. and Thangavelu M. 1993. A rapid and efficient method for extraction of total DNA from mature leaves of date palm (*Phoenix dactylifera* l.) Plant Molecular Biology Reports 11: 317-319

Altschul SF., Gish W., Miller W., Myers EW. and Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-410

Arnault C. and Dufournel I. 1994. Genome and stresses: reactions against aggressions, behaviour of transposable elements. Genetica 93: 149-160

Arnold-Schmitt B., Herterich S. and Neumann K-H. 1995. Physiological aspects of genome variability in tissue culture. I. Growth phase-dependant differential DNA methylation of the carrot genome (Daucus carota L.) during primary culture. Theoretical and Applied Genetics 91: 809–815

Ayala F. 1970. Teleological explanations in evolutionary biology. Philosophy of Science (March) 1-15

Barakat A., Han DT., Banslimane A-A., Rode A. and Bernardi G. 1999. The gene distribution in the genomes of pea, tomato and date palm. FEBS Letters 463: 139-142

Barral V., This P., Imbert-Establet D., Combes C. and Delseny M. 1993. Genetic variability and evolution of the Schistosoma genome analised by using random amplified polymorphic DNA markers. Molecular Biochemistry and Parasitology 59: 211- 222

Bender CM., Pao MM. and Jones PA. 1998. Inhibition of DNA methylation by 5-aza-2'-deoxycytidine suppresses the growth of human tumor cell lines. Cancer Research 58: 95- 101

Bennett MD. and Smith JB. (1991). Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London* B 334: 309-345

Bennetzen JL. and Kelogg EA. 1997. Do plants have a one-way ticket to genome obesity? Plant Cell 9: 1509-1514

Bennetzen JL., Schrick K., Springer PS., Brown WE. And SanMiguel P. 1994. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 37: 565– 576.

Bloom KS. and Carbon J. 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes. Cell 29: 305-317

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

Bodens C., Joandet S., Laigret F. and Kremer A. 1997. Detection of genomic regions differentiating two closely related oak species *Quercus petreae* (Matt.) and *Q. robur* (L.). Heredity 78: 433-444

Bouman H. and de Klerk GJ. 1997. Somaclonal variation. In: Biotechnology of Ornamental Plants p. 165-183. Oxford University Press, NY.

Brown PTH. 1989. DNA methylation in plants and its role in tissue culture. Genome 31: 717- 729

Brown TA. 1992. Genetics a molecular approach. Second edition. Chapman and Hall London.

Buchanan BB., Gruissem W. and Jones RL. 2000. Biochemistry and molecular biology of plants. American Society of Plant Physiologists, Rochville, Maryland.

Burns TD., White TJ. and Taylor JW. 1991. Fungal molecular systematics. Annual Review of Ecological Systematics 22: 524- 564

Buscot F., Wipf D., Di Battista C., Munch J-C, Botten B. and Martin F. 1996. DNA polymorphisim in morels: PCR/ RFLP analysis of the ribosomal DNA spacer and microsatellite primed PCR. Mycology Research 100: 63- 71

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

Casacuberta JM., Vernhettes S. and Grandbastien M-A. 1995. Sequence variability within the tobacco retrotransposon *Tnt*1 population. EMBO Journal 14: 2670-2678

Chandler V., Rivin C. and Walbot W. 1986. Stable non-mutator stocks of maize have sequences homologous to the Mu1 transposable element. Genetics 114: 1007-1021

Chang Y., Cesarman E., Pessin M.S., Lee F., Culpepper J., Knowles DM. and Moore PS. 1994. Identification of herpes virus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266: 1865-1869

Chen M., SanMiguel P. and Bennetzen JL. 1998. Sequence organizatin and conservation in sh2/ a1- homologous regions of sorgum and rice. Genetics 148: 435-443.

Colot V. and Rossignol. 1999. Eukaryotic DNA methylation as a evolutionary device. BioEssays 21: 402- 411

Copenhaver GP. and Preuss D. 1999. Centromeres in the genomic era: unravelling paradoxes. Current Opinions in Plant Biology 2: 104- 108

Corniquel B. and Mercier L. 1994. Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. Plant Science 101: 163-172

Crow JF. 1997. The high spontaneous mutation rate: Is it a health risk? Proceedings of the National Academy of Science USA 94: 8380-8386

Cullis CA. 1990. DNA rearrangements in response to environmental stress. Advances in Genetics 28: 73-97

Cullis CA. 1999. Environmental Stress - a generator of adaptive variation? In: H.R. Lerner (ed.). Plant Adaptations to Stress Environments. Marcel Dekker, New York, in press.

Cullis CA. and Kunert KJ. 2000. Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. Acta Horticulturae 530: 421-428

Cullis CA., Rademan S. and Kunert KJ. 1999. Method for finding genetic markers of somaclonal variation. International publication number WO 99/53100.

Curtis CA. and Lukaszewski AJ. 1991. Genetic linkage between C- bands and storage protein genes in chromosome 1B of tetraploid wheat. Theoretical and Applied Genetics 81: 245- 252

Deng Z., Huang S., Xiao S. and Gmitter JR. FG. 1997. Development and characterization of SCAR markers linked to the *Citrus tristeza*. Genome 40: 697-740

Donnison IS., Siroky J., Vyskot B., Saedler H. and Grant SR. 1996. Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. Genetics 144: 1893-1901

Eggleston WB., Alleman M. and Kermicle JL. 1995. Molecular organization and germinal instability of *R- stippled* maize. Genetics 141: 347- 360

Evans DA. 1988. Applications of somaclonal variation. Biotechnology in Agriculture p. 203-223

Fedoroff N. 2000. Transposons and genome evolution in plants. Proceedings of the National Academy Of Science USA 97: 7002-7007

Fedoroff N., Furtek D. and Nelson O. 1984. Cloning of the Bronze locus in maize by a simple and generalizable procedure using the transposable controlling element activator (AC). Proceedings of the National Academy Science USA 81: 3825-3829

Fedoroff N., Wessler S. and Shure M. 1983. Isolation of the transposable maize controlling elements Ac and Ds. Cell 35: 243- 251

Feldman M., Liu B., Segal G., Abbo S., Levy AA. and Vega JM. 1997. Rapid elimination of low- copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of Homoeologous chromosomes. Genetics 147: 1381- 1387.

Fitzgerald-Hayes M., Clarke L. and Carbon J. 1982. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell 29: 235-244

Flavell AJ., Dunbar E., Anderson R., Pearce SR., Hartly R. and Kumar A. 1992b. Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Research 20: 3639-44

Flavell AJ., Smith DB. and Kumar A. 1992a. Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. Molecular and General Genetics. 231:233-42

Flavell RB., Bennett MD., Smith JB. and Smith D. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochemical Genetics 12: 257-269

Finer JJ., Krebel HB. and Becwar MR. 1989. Initiation of embryogenic callus and suspension cultures of eastern white pine (*Pinus strobus* L.) Plant Cell Reports 8: 203- 206

Foster LM., Kozac KR., Loftus MG., Stevens JJ. and Ross IK. 1993. The polymerase chain reaction and its application to the filamentous fungi. Mycological Research 97: 769-781

Gill BS., Fiebe B. and Endo TR. 1991. Standard karyotype and nomenclature system for description of chromasome bands and structural aberrations in wheat (*Triticum aestivum*). Genome 34 : 830- 839

Grandbastien M-A. 1998. Activation of plant retrotransposons under stress conditions. Trends Plant Science 3: 181-187

Graur D., Shuali Y. and Li WH. 1989. Deletions in processed pseudogenes accumulate faster in rodents than in humans. Journal of Molecular Evolution 28: 279-285

Gu WK., Weeden NF., Yu J. and Wallace DH. 1995. Large-scale, cost-effective screening of PCR products in marker assisted selection application. Theoretical and Applied Genetics 91: 465- 470

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

Hadrys H., Balick M. and Schierwater B. 1992. Applications of RAPD fingerprinting in molecular ecology. Molecular Ecology

Heldt HW. 1997. Plant Biochemistry & Molecular Biology. Oxford University Press, New York.

Henry RJ. 1997. Practical Applications of Plant Molecular Biology. Chapman & Hall, London.

Hirochika H. 1993. Activation of tobacco retrotransposons during tissue-culture. EMBO Journal 12: 2521-2528

Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H. and Kanda M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. Proceedings of the National Academy of Science USA 93: 7783-7788

Hu W., Das OP. and Messing J. 1995. *Zeon*-1, a member of a new maize retrotransposon family. Molecular and General Genetics 248: 471-480

Jiang JM., Nasuda S., Dong FG., Scherrer CW. and Woo SS. *et al.*, 1996 A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proceedings of the National Academy of Science USA 93:14210-14213

Jin Y-K. and Bennetzen JL. 1989. Structure and coding properties of *Bs*1, a maize retrovirus-like transposon. Proceedings of the National Academy of Science USA 86:6235-6239

Johnston JS., Jensen A., Czeschin DG. and Price HJ. 1996. Environmentally induced nuclear DNA content instability in *Helianthus annus* (Asteraceae). American Journal of Botany 83: 1113-1120

Jones N., Ougham H. and Thomas H. 1997. Markers and mapping: we are all geneticists now. New Phytologist 137: 165-177

Judson HF. 1996. The Eighth Day of Creation. Cold Spring Harbor Press, New York

Karp A. 1993. Are your plants normal? - Genetic instability in regenerated and transgenic plants. Agro-Food-Industry Hi-Tech: May/June 7-12

Kunert KJ., Vorster J., Bey E. and Cullis C. 2000. Introduction of DNA-based quality assurance procedure for commercially micropropagated date palms in South Africa. Proceedings of the International Date Palm Symposium,Windhoek/Namibia, pp 74-80.

Larkin PJ. and Scowcroft WR. 1981. Somaclonal variation - a novel source of variability from cell cultures for plant improvement. Theoretical and Applied Genetics 60: 197-214

Laszlo E. 1996. The systems view of the world: A holistic vision for our time. (Advances in systems theory, complexity, and the human sciences). Hampton Press, Australia.

Lee M. and Phillips RL. 1986. Genomic rearrangements in maize induced by tissue culture. Genome 29: 122-128

Leroy XJ., Hily JM. and Chaumeil P. 2001. Detection of in vitro culture-induced instability through inter-simple sequence repeat analysis. Theoretical and Applied Genetics 102: 885-891

Leutwiler LS., Hough- Evans BR. and Meyerowitz EM. 1984. The DNA of *Arabidopsis thaliana*. Genetics 194: 15- 23

Levin I., Crittenden LB. and Dogson JB. 1993. Genetic map of the Z chromosome using random amplified polymorphic DNA (RAPD) markers. Genomics 16: 224- 230

Levitt J. 1972. Responses of plants to environmental stresses. Academic Press, New York and London

Lieb M. and Rehmat S. 1997. 5-Methylcytosine is not a mutation hot spot in nondividing *Escherichia coli*. Proceedings of the National Academy of Science USA 94: 940- 945

Linacero R., Alves EF. and Vazquez AM. 2000. Hot spots of DNA instability revealed through the study of somaclonal variation in rye. Theoretical and Applied Genetics 100: 506-511

Lisitsyn N., Lsistsyn N. and Wigler M. 1993. Cloning the differences between two complex genomes. Science 259: 946-950

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

MacLean DJ., Braithwaite KS., Manners JM. and Irwin JAG. 1993. How do we identify and classify fungal pathogens in the era of DNA analysis? In Advances in Plant Pathology Volume 10. Andrews JH. and Tommerup IC. (Ed.) Academic Press London

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

Martiensenn RA. and Colot V. 2001. DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science 293: 1070- 1074

Masson P., Surosky R., Kingsbury JA. and Fedoroff NV. 1987. Genetics 177: 177- 137

McClintock B. 1984. The significance of responses of the genome to challenge. Science 226: 792-801

McCouch SR. 2001. Genomics and synteny. Plant Physiology 125: 152-155

Mhiri C., De Wit PJGM. and Grandbastien M-A. 1999. Activation of the promoter of the *Tnt*1 retrotransposon in tomato after inoculation with the fungal pathogen *Cladosporium fulvum*. Molecular Plant-Microbe Interactions 12: 592-603

Mhiri C., Morel JB., Vernhettes S., Casacuberta JM., Lucas H. and Grandbastien MA. 1997. The promoter of the tobacco *Tnt*1 retrotransposon is induced by wounding and by abiotic stress. Plant Molecular Biology 33: 257-266

Miller JT., Dong F., Jackson SA., Song J. and Jiang J. 1998. Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150: 1615- 1623.

Moor G., Lucas H., Batty N. and Flavell R. 1991. A family of retrotransposons and associated genomic variation in wheat. Genomics 10: 461- 468.

Moreau-Mhiri C., Morel JB., Audeon C., Ferault M., Grandbastien MA. and Lucas H. 1996. Regulation of expression of the tobacco *Tnt*1 retrotransposon in heterologous species following pathogen related stresses. Plant Journal 9: 409-419

Murray B. (1998). Nuclear DNA amounts in gymnosperms. *Annals of Botany* 82 (Supplement A): 3-15.

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

Navin A, Prekeris R, Lisitsyn NA, Sonti MM, Grieco DA, Narayanswami S, Lander ES, Simpson EM. 1996. Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics 36: 349-353

Nekrutenko A, Makova KD, Baker RJ. 2000. Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene 249: 47-51

Nelke M., Nowak J., Wright J.M. and Mclean N.L. 1993. DNA fingerprinting of red clover (*Trifolium pratens* L.) with Jeffrey's probes: detection of somaclonal variation and other applications. Plant Cell Report 13: 72-78

Nierras CR., Liebman SW. and Warner JR. 1997. Does Saccharomyces need and organized nucleolus? Chromosoma 105: 444-451

Nixon RW. (1950) Imported varieties of dates in the United States. US Department of Agriculture Circular No. 834

O'Brien IEW., Smith DR., Gardner RC. and Murray BG. 1996. Flow cytometric determination of genome size in Pinus. Plant Science 115: 91-99

O'Hanlon PC., Peakall R. and Briese DT. 2000. A review of new PCR- based genetic markers and their utility to weed ecology. Weed Research 40: 239- 254

O'Niell RJW., O'Niell, MJ., and Graves, MA. 1998. Undermethylation associated with retroelements activation and chromosomes remodelling in an interspecific mammalian hybrid. Nature 393: 68-72

Olszewska MJ. and Osiecka R. 1982. The relationship between 2C DNA content, life-cycle type, systematic position, and the level of DNA endoreplication in nuclei of parenchyma cells during growth and differentiation of roots in some monocotyledonous species. Biochemie Und Physiologie Der Pflanzen 177: 319-336

Paran I. and Michelmore, RW. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics 85: 985-993

Pardue ML., Danilevskaya ON., Lowenhaupt K., Slot F. and Traverse KL. 1996. Drosophila telomeres: new views on chromosome evolution. Trends in Genetics 12: 48-52

Pearce SR., Harrison G., Li D., Heslop-Harrison JS., Kumar A. and Flavell AJ. 1996. The Ty1-copia group retrotransposons in Vica species: copy number, sequence heterogeneity and chromosomal location. Molecular and General Genetics 250: 305-315

Pearson WR. and Lipman DJ. 1988. Improved tools for biological sequence comparison. Proceedings of the National Academy of Science USA 85: 2444-2448

Percy RE.,m Klimaszewska K. and Cyr DR. 2000. Evaluation of somatic embryogenesis for clonal propagation of eastern white pine. Canadian Journal of Forest Reasearch 30: 1867- 1876.

Peschke VM., Phillips RL. and Gengenbach BG. 1987. Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science 238: 804-807

Petrov DA., Lozovskaya ER. and Hartl DL. 1996. High intrinsic rate of DNA loss in Drosophila. Nature 384: 346-349

Phillips RL., Kaeppler SM. and Olhoft P. 1994. Genetic instability of plant tissue cultures: Breakdown of normal controls. Proceedings of the National Academy of Science USA 91: 5222-5226

Poulsen GB., Kahl G. and Weising K. 1993. Abundance and polymorphism of simple repetitive DNA sequences in *Brassic napus* L. Theoretical and Applied Genetics 85: 994-1000

Pouteau S., Gandbasien MA., Boccara M. 1994. Microbial elicitors of plant defense response activate transcription of a retrotransposon. Plant Journal 5: 535-542

Pruitt RE., and Meyerowitz EM. 1986. Characterization of the genome of *Arabidopsis thaliana*. Journal of Molecular Biology 187: 169-183

Rabinowicz PD., Schutz K., Dedhia N., Yordan C., Parnell LD., Stein L., McCombie WR. and Martienssen RA. 1999. Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nature Genetics 23: 305- 308

Rothnie HM, McCurrach KJ., Clover LA. and Hardman N. 1991. Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum*. Nucleic Acids Research 19: 279- 286

Russell JR., Fuller JD., Macaulay M., Hatz BG., Jahoor A., Powell W. and Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics 95: 714–722

Sabir A., Newbury HJ., Todd G., Catty J. and Fors-Lloud BV. 1992. Detection of genetic stability using isozymes and RFLPs in beet plants regenerated in vitro. Theoretical and Applied Genetics 84: 113-114

Sambrook J., Fritsch EF. and Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Sandhu D. and Gill KS. 2002. Gene- containing regions of wheat and the other grass genomes. Plant Physiology 128: 803- 811

Sandhu D., Champoux JA., Bondareva SN. and Gill KS. 2001. Identification and physical localization of useful genes and markers to a major gene- rich region on wheat group *1S* chromosomes. Genetics 157: 1735- 1747

SanMiguel P., Gaut, BS., Tikhonov, A., Nakajima, Y. and Bennetzen, JL. 1998. The palaeontology of intergene retrotransposons of maize. Nature Genetics 20: 43-45

SanMiguel P., Tikhonov A., Jin YK., Motchoulskaia N. and Zakharov D, et al., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765-68

Schilling AG. 1996. Polymerase chain reaction-based assays for species-specific detection of *Fusarium culmorum*, *F.graminearum*, and *F.avenaceum*. Phytopathology 86: 515 522

Schwarz-Sommer Z., Gieri A., Cuypers H., Peterson PA. and Saedler H. 1985. Plant transposable elements generate the DNA sequence diversity needed for evolution. EMBO Journal 4: 591-597

Scribner KT. and Pearce JM. 2000. Microsatellites: evolutionary and methodological background and empirical applications at an individual, population and phylogenetic levels. In Molecular methods in ecology, Baker A (ed), Blackwell London pp. 235- 271

Sedra MH., Lashermes P., Trouslot P., Combes MC. and Hamons S. 1998. Identification and genetic diversity analysis of date palm (*Phoenix dactilyfera* L.) varieties from Morocco using RAPD markers. Euphytica 103: 75- 82.

Selker EU. and Garrett PW. 1988. DNA sequence duplications trigger gene inactivation in *Neurospora crassa*. Proceedings of the National Academy of Science USA 85: 6870- 6874.

Selker EU., Cambareri EB., Jenesen BC. and Haack KR. 1987. Rearrangement of duplicated DNA in specialized Neurospora.Cell 51: 741- 752.

Skirvin RM. and Janick J. 1976. Tissue culture-induced variation in scented *Pelargonium* spp. Journal of the American Society for Horticultural Science 101: 281-290

Skirvin RM., McPheeters KD. and Norton M. 1994. Sources and frequency of somaclonal variation. HortScience 29: 1232-1237

Smith TF. and Waterman MS. 1981. Identification of common molecular subsequences. Journal of Molecular Biology 25; 147: 195-197

Smyth DR., Kalitsis P., Joseph JL. and Sentry JW. 1989. Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the *gypsy* group of *Drosophila*. Proceedings of the National Academy of Science USA 86: 5015- 5019.

Soller M. and Beckmann JS. 1986. Restriction fragment length polymorphisims in poultry breeding. Poultry Science 65: 1474- 1488

Strathdee G. and Brown R. 2002. Aberrand DNA methylation in cancer: potential clinical interventions. Expert Reviews in Molecular Medicine. 4 March pp. 1- 17

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

Suoniemi A., Anamthawat-Jonsson K., Arna T. and Schulman AH. 1996. Retrotransposon BARE-1 is a major, dispersed component of the barley (Hordeum vulgare L.) genome. Plant Molecular Biology 30: 1321- 29

Takeda S., Sugimoto K., Otsuki H. and Hirochika H. 1998. Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Molecular Biology 36: 365–376.

Takeda S., Sugimoto K., Otsuki H. and Hirochika H. 1999. A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant Journal 18: 1- 11.

Thomas CA. 1971. The genetic organization of chromosomes. Annual Review of Genetics 5: 237-256

Toder R., Grutzner F., Haaf T. and Bausch E. 2001. Species-specific evolution of repeated DNA sequences in great apes. Chromosome Research 9: 431-435

Twyman RM. 1998. Advances in molecular biology: a concise reference. Bios Scientific Publishers, Springer- Verlag, New York.

Ushijima T., Morimura K., Hosoya Y., Okonogi H., Tatematsu M., Sugimura T. and Nagao M. 1997. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. Proceedings of the National Academy of Science USA. 94: 2284-2289

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

Vongs A., Kakutani T., Martienssen RA. and Richards EJ. 1993. *Arabidopsis thaliana* DNA methylation mutants. Science 260: 1926- 1928.

Vorster BJ., Kunert KJ. and Cullis CA. 2002. Use of representational difference analysis for the characterization of sequence differences between date palm varieties. Plant Cell Reports 21:271–275

Voytas DF., Cummings MP., Konieczny A., Ausubel FM. and Rodermel SR. 1992. *Copia*-like retrotransposons are ubiquitous among plants. Proceedings of the National Academy of Science USA 89: 7124-7128

Wang DG., Fan JB., Siao CJ., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris MS., Shen N., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson TJ., Lander ES., et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280: 1077-1082

Welsh J. and McClelland M. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213-7218

White SE., Habera LF. and Wessler SR. 1994. Retrotransposons in the flanking regions of normal plant genes – a role for *copia*-like elements in the evolution of gene structure and expression. Proceedings of the National Academy of Science USA 91: 11792-96

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

Wright WE., Brasiskyte D., Piatyszek MA. and Shay JW. 1996. Experimental elongation of telomeres extends the lifespan of immortal x normal cell hybrids. EMBO Journal 15: 1734-1741

Yoder JA., Walsh CP. and Bestor TH. 1997. Cytosine methylation and the ecology of intagenomic parasites. Trends in Genetics 13: 335-40

Zakian VA. 1996. Structure, function and replication of Saccharomyces cerevisiae telomeres. Annual Review of Genetics 30: 141-172

Zheng KL., Castiglione S., Biasini MG., 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

Zoldos V., Siljak-Yakovlev S., Papes D., Sarr A. and Panaud O. 2001. Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. Molecular Genetics and Genomics 265: 234-241

Annex

I) Molecular Methods

I.I Plant material and DNA extraction

'Medjool' and 'Barhee' plant material originated from 14 plants grown in California that have been propagated solely via off-shoots from mother plant material imported in the beginning of the last century from Northern Africa (Nixon 1950). The *in vitro* plants used were 'Medjool' derived from explant material collected in California, and 'Barhee' derived from explant material collected in the United Arabic Emirates.

Total cellular DNA was isolated from leaves (1 g) using the Nucleon Phytopure Plant DNA extraction kit (Amersham Life Science, UK) according to the manufacturer's instructions. To test the quality and amount of isolated DNA, samples of isolated DNA (1 μl) were run on a 1% agarose gel in TAE buffer (0.04 M Tris-acetate; 1 mM EDTA, pH 8) as outlined by Sambrook et al. (1989). After staining of the gel with ethidium bromide for 15 min, DNA quality was determined on a white/UV-transilluminator, photographed using a Grab-IT system (Vacutec, USA) and the DNA concentration of the samples was visually determined using 4 different λ-phage DNA amounts as standards (25 ng, 50 ng, 100 ng and 250 ng) for comparison.

I.II Polymerase chain reaction (PCR)

Standard DNA amplifications by PCR were carried out in 25 μl reaction volumes containing 500 mM KCl, 25 mM $MgCl_2$, 100 mM Tris-HCl (pH 8.3), 25 mM of each dNTP and 5 units TaKaRa DNA polymerase (TaKaRa, Japan) in a GeneAmp PCR 9600 system (Perking Elmer, USA). Primers for PCR were obtained either from Operon Technologies (Operon Technologies, USA) or were designed using the online primer design tool of Molecular Biology Shortcuts (MBS), program 'Oligos and Primers' (www.mbshortcuts.com/biotools/index.htlm). The designed primers were then manufactured and purchased from MWG-Biotech AG (Germany). The standard PCR program consisted of 94°C (5 min) to denature double stranded DNA. This was followed by 35 cycles of amplification consisting of denaturing DNA at 94°C (1 min), primer annealing at 55°C or higher depending on primer pair (1 min), and extension of the DNA chain at 72°C (2 min). This was followed by another extension cycle at 72°C for 5 min.

I.III   Representational difference analysis (RDA)

I.III.I   *Preparation of RDA amplicons*

RDA was performed following the general outline described by Lisitsyn et al. (1993). Two micrograms of each of the DNAs ('Barhee' and 'Medjool') were digested with 80 units of the enzyme *Bam*HI or *Hind*III (Amersham Life Science, UK) 100 µg of yeast tRNA (Sigma-Aldrich, USA) was added to the digest to limit non-specific binding of DNA against the sides of eppendorf tubes and pipet tips. The digests were then extracted with an equal volume of phenol/chloroform. The DNA containing upper phase was removed and the DNA precipitated using 1/10 volume 3 M sodium acetate and 2.5 volumes 100% ethanol. The precipitate was washed in 70% ethanol and dried. The digested DNA was dissolved in 18 µl of TE buffer (10 mM Tris-Cl, 1 mM EDTA), pH 8, at a concentration of 100 µg/ml.

I.III.II   Ligation of adaptor sequences

One µg of the *Bam*HI and *Hind*III digested DNA were ligated to the adaptor pair RBam 12 and RBam 24 or RHind 12 and RHind 24, respectively. The DNA, primer set (0.6 µM), 10X ligase buffer and water to a final volume of 30 µl were mixed in a microcentrifuge tube. The tubes were placed in a heating block at 55°C and the block was the placed at 4°C until the temperature had fallen to 12°C (60 to 75 min). The tubes were then placed on ice for 3 min, after which 1 µl of T4 DNA ligase (1 unit) (Amersham Life Science, UK) was added and the reaction incubated at 16°C overnight. The ligation reaction was then diluted with 970 µl TE buffer, pH 8, and the amplification reaction was set up as follows:

     80 µl 10X *Taq* Buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM MgCl$_2$)
     64 µl dNTP mix   (160 mM of each dNTP)
     16 µl (1.0 µM) of one primer of the appropriate pair (J Bam 24 or R Hind 42)
     80 µl of the diluted ligation mixture
     560 µl water

The tube was placed in a heating block at 72°C for 5 min to melt of the 12-mer oligonucleotide. Six µl (30 units) of *Taq* polymerase (TaKaRa, Japan) was added and mixed by pipetting. The mixture was then aliquoted into eighth PCR tubes in a Perkin-Elmer GeneAmp 9600 thermocycler, the block of which was being held at 72°C. After 5 minutes, the thermocycler program run was: 20 cycles: 95°C for 30 sec; 72°C for 2 min, followed by 10 min at 72°C, followed by holding at 4°C.

The 8 tubes were combined an a 10 µl aliquot was run on a 1.5% TAE agarose gel at 100 V for 30 minutes to check the amplification and estimate the concentration by comparison to lambda phage DNA standards. The remainder was extracted with 600 µl phenol/chloroform. The upper phase was removed and the DNA precipitated with 3 M sodium acetate and an equal volume propanol. The precipitate was collected by centrifugation, washed twice with 70% ethanol and dried. The amplicons was redissolved in 80µl TE buffer, pH 8.

I.III.III   Removal of the adaptors from amplicons

Of the amplicons that was to be used as driver 150 µg and 10 µg of the tester amplicons were digested with the appropriate enzyme in 800 µl and 200 µl, respectively, at an enzyme concentration of 20 units/µg DNA, at 37°C for one hour. Ten µg Yeast tRNA was added and each digest extracted with phenol/chloroform. The DNA was precipitated with an equal volume of isopropanol and 1/10 volume 3 M sodium acetate, as described above. Isopropanol is used instead of ethanol in order to eliminate very small fragments (e.g. adaptors) from the precipitate. The driver and tester DNAs were both redissolved at approximately 400 µg/ml. Both the digested driver and tester amplicons were run on a 1.5% TAE agarose gel alongside an equal aliquot of undigested amplicons to check the completeness of the digestion. On the same gel, standard lambda phage DNA was run so that the final concentrations of driver and tester DNAs could be estimated and adjusted if necessary.

I.III.IV   Change of adaptors on tester amplicons

Tester DNA was prepared by adding a second adaptor pair JBam 12 and 24 for *Bam*HI digested DNA or JHind 12 and 24 for *Hind*III digested DNA to 1 µg of the first

round amplicons in the same way as described for the ligation of the first set. The ligation reaction was then diluted to a final volume of 100 µl with water. An aliquot of the ligate was amplified for 20 cycles in a reaction volume of 20 µl to check that the newly ligated adaptors would support amplification with the new primer.

I.III.V   Subtractive hybridisation and kinetic enrichment

The hybridisation reaction was set up by mixing 80 µl of the driver (~ 40 µg) with 40 µl tester (~ 0.4 µg) amplicons (a ratio of driver : tester of 100 : 1). To this was added 30 µl of 10 M ammonium acetate and 380 µl 100% ethanol. After incubation on ice for 10 min the DNA precipitate was collected by centrefugation for 10 min at full speed in a microcentrifuge. The pellet was washed twice with 70% ethanol and dried after which it was redissolved in 4 µl 3X EE buffer [30 mM N-(2-hydroxyethyl piperizine)-N'-(3-propene sulfonic acid) (EPPS), pH 8, 3 mM EDTA] followed by repeated vortexing interspersed with centrifugation. The solution was collected at the bottom of the tube and overlaid with light mineral oil so that the spherical droplet could be seen to be completely covered by oil. The DNA was denatured by placing it in a heating block at 98°C for 5 minutes. One µl of 5 M sodium chloride solution was added and the tube briefly centrifuged to mix the aqueous phases. The DNA was then incubated at 67°C overnight. Ten µl TE buffer, pH 8, was added and mixed with the aqueous phase, which was then carefully removed from under the oil by pipetting and added to clean microcentrifuge tube containing 380 µl TE buffer and 40 µg tRNA. The appropriately annealed fragments were amplified in the following reaction:

80 µl 10X *Taq* buffer

64 µl dNTP mix (160 mM of each dNTP)

80 µl of the diluted hybridisation mixture

560 µl ds H$_2$O

6 µl *Taq* polymerase (30 units)

The tube was placed in a heating block at 72°C for 5 min to fill in the ends. Sixteen µl of the appropriate primer (the 24mer used in the adapter ligated to the tester amplicons) was added, mixed by pipetting, and the mixture aliquoted into eighth PCR tubes. The tubes were placed in a Perkin-Elmer thermocycler, the block of which, were being held at 72°C.

The following thermocycler program was then run: 10 cycles of 95°C for 30 sec; 72°C for 2 min followed by 10 min at 72°C, followed by holding at 4°C. The tubes were combined and 20 µl of the solution subjected to a further 20 cycles of amplification under the same conditions.

Ten µg of tRNA was added to the remainder of the amplified reaction and then extracted with 600 µl phenol/chloroform. An aliqoute (750 µl) was removed and 75 µl (1/10 volume) of 3 M sodium acetate and 825 µl (one volume) propanol added. The precipitate was collected by centrifugation, after incubation on ice for 10 minutes. The pellet was washed twice with 70% ethanol, dried and dissolved in 40 µl TE buffer, pH 8. At this stage the aliquot subjected to the additional amplification cycles was run on a 1.5% TAE agarose gel to check that amplification occurred. This first round difference product (20 µl) was digested with 20 units of mung bean nuclease (Amersham Life Science, UK) at 30°C for 30 minutes to remove all single- stranded DNA. The reaction was stopped by the addition of 160 µl of TE buffer, pH 8. The digested product was amplified in a reaction mixture consisting of:

80 µl 10X *Taq* buffer

64 µl dNTP mix (160 mM of each dNTP)

16 µl of the same primer used for the 10-cycle amplification

80 µl of the diluted nuclease treated mixture

560 µl ds H_2O

6 µl *Taq* polymerase (30 units)

The mixture was aliquoted into eighth PCR tubes in a Perkin-Elmer thermocycler, the block of which was being held at 72°C. The thermocycler program run was 20 cycles: 95°C for 30 seconds; 72°C for 2 minutes followed by 10 minutes at 72°C, followed by holding at 4°C.

The 8 tubes were combined and a 10 µl aliquot was run on a 1.5% TAE agarose gel at 100 V for 30 min to check the amplification. The remainder was extracted with phenol/chloroform. The upper phase was pipetted into a clean tube and a 1/10 volume of 3 M sodium acetate and 1 volume of propanol added. The tubes were

mixed by inversion and placed on ice for 15 min. The precipitate was collected by centrifugation for 15 min at full speed in a microsentrifuge, washed twice with 70% ethanol and dried. The amplicons were redissolved in 80 μl of TE buffer, pH 8. The concentration was estimated by electrophoresis on a 1.5% TAE agarose gel with lambda phage DNA standards and adjusted to 100 μg/ml.

I.IV   Cloning of the subtraction products

Subtraction products (2 μg) were digested with the 50 units of the appropriate restriction enzyme for 30 minutes at 37°C. One hundred ng of this digest was mixed with 10 ng of pBluescriptII vector (Strategene, USA) digested with the same enzyme (a ratio of insert to vector of 10 : 1). These were extracted with phenol/chloroform and precipitated with sodium acetate and ethanol. The precipitate was collected by microcentrifugation and the pellet washed twice with 70% ethanol before being dried. The dried pellet was redissolved in 8 μl water. To this was added 1 μl 10X ligase buffer and 1 μl T4 ligase, and the mixture incubated overnight at 16°C. Two μl of the ligate was used to transform competent XL1Blue cells and 50 plasmid-containing colonies carrying an insert were selected and probed with either the 'Barhee' or 'Medjool' labeled driver amplicons using the *Gene Images* random prime-labeling module (Amersham Life Sciences UK). Ten colonies that showed a much stronger signal after hybridization with the 'Barhee' amplicons than with the 'Medjool' amplicons were selected for plasmid isolation and determination of the insert sequence and size.

I.V   Southern-blot hybridization

Total genomic DNA of 'Barhee' and 'Medjool' (1 μg) were digested with two units of each of the restriction enzymes *Bam*HI and *Eco*RI at 37°C for two hours. The restricted DNA was then separated on a 1.5% TAE agarose gel using electrophoresis, and transferred to a positively charged nylon membrane (Roche, Switzerland), as described by Sambrook et al. (1989). Cloned fragments of Dp41 and Dp36 were used as probes, which were labeled using the *Gene Images* random prime, labeling module. Membranes were pre-hybridized and hybridized at 65°C in a hybridization buffer containing 5% SSC, 0.1% SDS and 20-fold dilution of the liquid

block provided. Hybridization was carried out overnight and the membranes washed at 60°C using a 1% SSC and 0.1% SDS solution. Detection was preformed using the *Gene Images CDP-Star* detection module (Amersham Life Sciences UK), according to the manufacturer's instructions. Membranes were exposed to Hyperfilm ECL (Amersham Life Sciences UK) and the films developed.

I.VI   Sequence analysis

DNA sequence analysis was carried out with the dideoxy chain terminator method developed by Sanger et al. (1997). Recombinant plasmids were sequenced using fluorescent dye terminators and AmliTaq from the ABI PRISM™ Big DYE Terminator Cycle Sequencing Ready Reaction Kit (Perkin-Elmer Applied Biosystems, Inc., Foster City, Calif.), in a cycle sequence protocol according to the recommendations of the manufacturer. Sequence reactions were run on PAGE (polyacrylimide gel electrophoresis) using an ABI PRISM™ 377 Autosequencer (Perkin-Elmer Applied Biosystems). Sequence electropherograms were analysed using Sequence Navigator version 1.0.1 (Perkin-Elmer).

II)   Tissue culture and somatic embryogenesis

II.I   Tissue culture initiation

The most responsive embryo developmental stages for somatic embryogenesis initiation range from the cleavage polyembryony, to the early (precotyledonary) dominance stage (Percy *et al.*, 2000). For culture initiation, seeds were extracted from developing cones and placed in Petri dishes containing moistened filter paper. Within 1 hour of extraction, seeds were surface sterilized in preparation for dissection. Seeds were surface sterilised by placing them in metal baskets and stirring vigorously for 6 min in 200 ml of 3% (v/v) hydrogen peroxide with one drop of Tween 20. Seeds were then rinsed tree times for 2 min each with sterile water and transferred to a petridish containing a moist, sterile filter paper. The seed coat, nucellus, and membrane surrounding the metagametophyte were removed under a stereomicroscope, and the megagametophyte was placed on initiation medium. Five to eight megagametophytes were cultured in each Petri dish containing approximately 25 ml of medium, and dishes were sealed with parafilm. Explants were cultured in darkness at 24°C for 16 weeks and examined every 4 weeks for embryogenic tissue initiation.

II.II   Embryogenic tissue proliferation and culture maintenance

Explants displaying proliferation from the zygotic embryo cells were microscopically selected, then embryogenic tissue was removed, dispersed by vigorous shaking in liquid MLV maintenance medium, poured onto a sterile filter paper disk (Whatman No. 2, 5.5 cm) in a Büchner funnel and a short low-pressure pulse (55 kPa) was applied to drain the medium prior to transferring the disc to fresh medium. This approach was also used for maintenance of embryogenic tissue. Maintenance and proliferation cultures were incubated in darkness at 24°C.

II.III   Maturation of somatic embryos

The somatic embryo experiments were performed by combining embryogenic tissue of one line (from several plates), one week after subculture, in a 50 ml test tube,

adding liquid media without growth regulators and vigorously shaking the tube to break up the clumps of tissue into a fine suspension. Subsequently 3 ml containing approximately 0.2 g of suspended embryonal masses were withdrawn with a wide-mouth pipette and placed on a moist filter paper disk (Whatman No. 2, 5.5 cm in diameter) in a Büchner funnel attached to a vacuum pump. A short low-pressure pulse (5 sec, 55 kPa) was applied to drain all liquid medium and anchor the embryonal masses to the filter paper as a thin layer. Each disk of filter paper with the embryonal masses was subsequently placed on maturation medium in a 10 X 20 mm Petri dish and cultured for up to 10 weeks. The cultures were kept under dim light condition at 1.6 μmol m^{-2}s^{-1} from cool white florescence lamps (Philips F72T12/CW, 65 W) under a 16 h photoperiod at 24°C.

II.IV Tissue culture media

II.IV.I MLM 10X stock

	2 Liters	Final concentration in medium [g/L]
NH$_4$NO$_3$	16.5 g	0.825
KNO$_3$	19.0 g	0.95
MgSO$_4$.7H$_2$O	18.5 g	0.925
KH$_2$PO$_4$ (monobasic)	3.4 g	0.17
CaCl$_2$.2H$_2$O	0.22 g	0.011
LM Minor stock (100X)	200 ml	
LM Vitamin stock (100X)	200 ml	
Myo-Inisitol	2.0 g	0.1
Iron Solution	100 ml	
dH$_2$O to	2 L	
Stored at -20°C or 2 weeks at +4°C		

II.IV.II MLM Minor Stock 100X

	1 Liter	Final concentration in medium [g/L]
$Na_2MoO_4.2H_2O$	0.125 g	0.00125
KI	0.415 g	0.00415
H_3BO_3	3.1 g	0.031
$MnSO_4.H_2O$ (or $MnSO_4.4H_2O$)	2.1 g (2.7 g)	0.021 (0.027)
$ZnSO_4.7H_2O$	4.3 g	0.043
$CuSO_4.5H_2O$	0.05 g	0.0005
$CoCl_2.6H_2O$	0.013	0.00013
DH_2O to	1 L	
Stored at -20°C		

II.IV.III MLM Vitamin stock 100X

	1 Liter	Final concentration in medium [g/L]
Nicotinic acid	0.05 g	0.0005
Pyridoxine.HCL	0.01 g	0.0001
Thiamine.HCL	0.01 g	0.0001
dH_2O to	1 L	
Stored at -20°C		

II.IV.IV MLM Iron solution

	100 ml	Final concentration in medium [g/L]
Na_2EDTA	0.746 g	0.0373
$FeSO_4.7H_2O$	0.556 g	0.0287
dH_2O to	100 ml	
Made fresh		

II.V   Growth regulators

II.V.I   2,4-Dichlorophenoxyacetic Acid (2,4-D) stock (1 mg/ml)

2,4-D	100 mg
95% Ethanol	50 ml
DH$_2$O to	100 ml
Stored at +4°C	

II.V.II   6-Benzyl-Aminopurine (BA) stock (2.5 mg/ml)

BA	50. mg
Warmed 0.5 N HCL (or NaOH)	5 ml to dissolve
Warmed dH$_2$O to	100 ml
Stored at +4°C	

II.V.III   (±) cis-trans Abscic (ABA) stock (10mM)

ABA	0.02643 g
1 N NaOH	drops to dissolve
dH$_2$O to	10 ml
Filter sterilized and wrapped in foil. Made fresh.	
1 ml stock/ L = 10 µM	

II.V.IV   Glutamine stock (25 mg/ml) (Amino acids)

Glutamine	25.0 g
dH$_2$O to	1 L
pH 5.8. Filter sterilized. Stored frozen	

II.VI   MLV Media

II.VI.I   MLV Ultra low medium (MLV UL)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	2 %	20.0 g
2,4-D (1 mg/ml stock)	1.1 μM	0.25 ml
BA (0.5 mg/ml stock)	1.1 μM	0.5 ml
dH$_2$O to		1 L
Phytagel	0.4 %	4 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml

II.VI.II   MLV Standard medium (MLV Std)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	2 %	20.0 g
2,4-D (1 mg/ml stock)	9.5 μM	2 ml
BA (0.5 mg/ml stock)	4.4 μM	2 ml
dH$_2$O to		1 L
Phytagel	0.4 %	4 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml

II.VI.III   Maturation medium (MM)

		1 Liter
MLM 10X stock		100 ml
Casein Hydrolosate (Casamino Acids)	1 g/L	1.0 g
Sucrose	6 %	60.0 g
dH$_2$O to		1 L
Phytagel	1 %	10 g
Set pH to 5.7. Autoclaved		
Glutamine (25 mg/ml sterile stock)	0.5 g/L	20 ml
ABA (10 mM sterile stock)	120 µM	12 ml

III General solutions and buffers

III.I Antibiotics

III.I.I Ampicillin stock solution (50 mg/ml)

Dissolve 2 g Ampicillin (D(-)-a-Aminobenzylpenicillin sodium salt) (Sigma-Aldrich, Germany) powder in 40 ml of sterile distilled water. Aliquote in 1.5 ml tubes and store at -20°C.

III.I.II Tetracycline stock solution (10 mg/ml)

Dissolve 0.1 g of tetracycline powder (Sigma-Aldrich, Germany) in 5 ml 100% ethanol, adjust volume to 10 ml with sterile distilled water. Cover tube with foil and store at -20°C.

III.II Southern blotting solutions

III.II.I Neutralization buffer

Combine 121.1 g of Tris base (Sigma-Aldrich, Germany) and 87.7 g NaCl. Add 900 ml distilled water. Adjust pH to 8.0 with concentrated HCl. Adjust volume with distilled water to 1 L and autoclave.

III.II.II Denaturing buffer

Combine 87.7 g NaCl and 20 g NaOH. Add distilled water to 1 L. Stir to dissolve and autoclave.

III.II.III 20X Standard Saline-Citrate (SSC)

Combine 174.4 g NaCl, 88.3 g $C_6H_5Na_3O_7.2H_2O$ (Sodium-citrate) in 800 ml distilled water. Stir to dissolve and adjust volume to 1 L. Autoclave.

III.III   Buffers and salt solutions

III.III.I   Agarose gel buffer (TAE) (50X stock)

Combine 968 g Tris base, 228.4 ml glacial acetic acid and 400 ml 0.5 M EDTA, pH 8.0. Add 3 L distilled water and stir until solids dissolve. Adjust volume to 4 L with distilled water. Dilute to 1X before use.

III.III.II   Agarose gel loading buffer (6X)

Combine 0.063 g bromophenol blue (BPB), 0.063 g xylene cyanol FF (XC) and 2.5 ml glycerol. Add distilled water to 25 ml. Aliquot into 1.5 ml tubes and store at -20°C. Add 1 μl to sample before loading onto agarose gel.

III.III.III   Ethidium Bromide solution

Dissolve 0.2 g ethidium bromide powder in 20 ml distilled water. Vortex to dissolve and cover container with foil.

III.III.IV   5 M Ammonium acetate (NH_4OAc)

Dissolve 385.4 g ammonium acetate in 500 ml distilled water by slowly adding the ammonium acetate powder into the stirring water. Adjust volume to 1 L. Autoclave and store at RT.

III.III.VI   <u>3 M Sodium acetate (NaOAc) pH 4.8</u>

Dissolve 40.82 g NaOAc.3H$_2$O in 80 ml distilled water and stir on a hot plate until dissolved. Adjust pH to 4.8 with glacial acetic acid. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.VII   <u>5 M Sodium chloride</u>

Dissolve 29.22 g NaCl in 80 ml distilled water. Adjust volume to 100 ml, autoclave and store at RT.

III.III.VIII   <u>1 M Tris-HCl pH 8.0</u>

Dissolve 12.11 g Tris base in 80 ml of distilled water. Adjust pH to 8.0 using concentrated HCl. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.IX   <u>0.5 M EDTA pH 8.0</u>

Dissolve 8 g NaOH pelles in 400 ml distilled water. Add 93.05 g Na$_2$EDTA.2H$_2$O and allow to dissolve. Adjust pH to 8.0 using more NaOH pellets. Adjust volume to 500 ml with distilled water. Autoclave and store at RT.

III.III.X   <u>TE buffer (10 mM Tris-HCL, 1 mM EDTA)</u>

Combine 1 ml 1M Tris-HCl pH 8.0 and 0.2 ml 0.5 M EDTA pH 8.0. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.XI   <u>Low TE buffer (10 mM Tris-HCL, 0.1 mM EDTA)</u>

Combine 1 ml 1M Tris-HCl pH 8.0 and 20 µl 0.5 M EDTA pH 8.0. Adjust volume to 100 ml with distilled water. Autoclave and store at RT.

III.III.XII   10 N Sodium hydroxide (NaOH)

Dissolve 40 g NaOH pellets in 70 ml distilled water. Adjust volume to 100 ml. Store at RT.

IV.   Plasmid maps

IV.I   pBluescript II (Stratagene USA)

pBluescript II SK (+/–) Multiple Cloning Site Region
(sequence shown 598–826)

IV.II   pMosBlue (Amersham UK)

pMOS*Blue*
(2887bp)

Hind III (45)
BspM (50)
Sph I (58)
Sse8387 I (61)
Pst I (61)
Sal I (63)
Acc I (64)
Hinc II (65)
Xba I (69)
Spe I (81)
Nde I (83)
EcoR V (95)
BamH I (99)
Ava I (104)
Sma I (106)
Kpn I (112)
Sac I (118)
EcoR I (120)

Dra III (503)
BsaA I (599)
Drd II (513)
Bcc I (519)
Nae I (611)

T7 start (41)
lacZ
f1 origin
Sap I (2563)
Afl III (2546)
ori
MlwN I (2137)
HgiE II (1964)
Ap
Xmn I (1086)
BsaH I (1119)
Sca I (1177)
Earn1105 I (1656)
Pfl1108 I (1639)
Bsa I (1502)

V.   Primer sequences

RAPD and SCAR primers	
OPE-01	5'-CCCAAGGTCC-3'
OPE-06	5'-AAGACCCCTC-3'
DpSL	5'- GTGTTAGGGGCAAAATGGAA–3'
DpSR	5'- TTGTCCGTCTGAGACTCCCT–3'
Date palm subtraction product specific primers	
DP36L	5'-CTATCGACGACAGGCTGACA -3'
DP36R	5'-GACCCGGACTTGTTGGAGTA-3'
DP41L	5'-CCTTCTCCCCGTAGTAACCG-3'
DP41R	5'-AGGAAAGGCAACCTACCGAG-3'
DP50L	5'-TACACGATGTCCCTCAACCA-3'
DP50R	5'-GGAACATTTCCTCGGTATCC-3'
PLM 1	5'-TTACAGAGGGGAAAGGAGGA-3'
PLM 4	5'-GGAAGGAGGTGGCTCCG-3'
PLB11	5'-CGCAATCTTGCAAGTATCAGT-3'
RDA adapters	
R Bam 24	5'-AGCACTCTCCAGCCTCTCACCGAG-3'
R Bam 12	5'-GATCCTCGGTGA-3'
J Bam 24	5'-ACCGACGTCGACTATCCATGAACG-3'
J Bam 12	5'-GATCCGTTCATG-3'
N Bam 24	5'-AGGCAACTGTGCTATCCGAGGGAG-3'
N Bam 12	5'-GATCCTCCCTCG-3'
R Hind 24	5'-AGCACTCTCCAGCCTCTCACCGCA-3'
R Hind 12	5'-AGCTTGCGGTGA-3'
J Hind 24	5'-ACCGACGTCGACTATCCATGAACA-3'
J Hind 12	5'-AGC TTGTTCATG-3'
N Hind 24	5'-AGGCAGCTGTGGTATCGAGGGAGA-3'
N Hind 24	5'-AGCTTCTCCCTC-3'

R Hpa 24	5'-AGCACTCTCCAGCCTCTCACCGAC-3'
R Hpa 11	5'-CGGTCGGTGAG-3'
J Hpa 24	5'-ACCGACGTCGACTATCCATGAAAC-3'
J Hpa 11	5'-CGGTTTCATGG-3'
N Hpa 24	5'-AGGCAACTGTGCTATCCGAGGGAC-3'
N Hpa 11	5'-CGGTCCCTCGG-3'
S Hpa 24	5'-ACTTCTACGGCTGAATTCCGACAC-3'
S Hpa 12	5'-CGGTGTCGGAAT-3'
Sequencing primers	
T7	5'-TAATACGACTCACTATAGGG-3'
Sp6	5'-CATACGATTTAGGTGACACTATAG-3'

GENETICS AND GENOMICS

B.J. Vorster · K.J. Kunert · C.A. Cullis

Use of representational difference analysis for the characterization of sequence differences between date palm varieties

Abstract Representational difference analysis was applied to subtract the genomes of the two date palm varieties, Barhee and Medjool, for identification and characterization of unique genome differences suitable for discriminating between individual plants and the two varieties. Three different DNA difference products were isolated from Barhee representing families of dispersed, repeated variable sequences present in the genome of both varieties. Several variant members of repeated DNA were detected by sequence analysis, containing base changes from C to T and G to A and short deletions. Mutated DNA sequences could be amplified in a polymerase chain reaction-based test from a much smaller number of Barhee plants than from Medjool plants allowing the differentiation between individual plants and partial discrimination between varieties.

Keywords Date palm · Representational difference analysis · Genome analysis · Hyper-variable DNA · Plant variety identification

Introduction

Genomes of closely related plants or varieties can differ by only a few coding genes or in minor genome re-organizations and a range of different approaches is available for detection of such genetic differences. Among these are the analyses of r-DNA intergenic regions (Scribner

Communicated by J.C. Register III

C.A. Cullis (✉)
Biology Department, Case Western Reserve University,
10900 Euclid Avenue, Cleveland, OH 4404, USA
e-mail: cac5@po.cwru.edu
Tel.: +1-216-3685362, Fax: +1-216-3684672

B.J. Vorster · K.J. Kunert
Botany Department,
Forestry and Agricultural Biotechnology Institute,
74 Lunnon Road, University of Pretoria, Hillcrest, Pretoria 0002,
South Africa

and Pearce 2000) and of simple sequence repeats (SSRs), which are also known as microsatellites. SSRs have been used for example for identification of varieties of *Chrysanthemum* and *Citrus* (Wolff et al. 1995; Fang and Roose 1997) as well as individual plants of oilseed rape cultivars (Charters et al. 1996). However, the two most widely used genetic techniques to detect plant variation are random amplified polymorphic DNA (RAPD) analysis, which detects DNA polymorphisms amplified by arbitrary primers (Williams et al. 1990; Welsh and McClelland 1990) and amplified fragment length polymorphisms (AFLPs) (Vos et al. 1995; O'Hanlon et al. 2000). Recently, the RAPD technique has also been applied to identify date palm on the variety level (Corniquel and Mercier 1994; Sedra et al. 1998). All these techniques are useful for the differentiation of plants by identifying random polymorphisms. However, the comparisons are usually made on the basis of the presence or absence of a band rather than directly on any DNA sequence variation.

In contrast, representational difference analysis (RDA) allows the cloning and sequencing of fine genome differences between two highly similar genomes and further provides exact sequence information about these differences. RDA detects any kind of labile DNA region in two genomes to be compared and can be used to derive probes for genomic losses, rearrangements, amplifications, point mutations and pathogenic organisms found within any of the genomes compared (Lisitsyn et al. 1993, 1994; Michiels et al. 1998). RDA has been applied in a variety of contexts including the isolation of repetitive sequences present in only one of the compared genomes (Navin 1996). Nekrutenko et al. (2000) used RDA to create a species-specific marker for voles and Toder et al. (2001) have applied RDA in evolutionary genomics to search for overall genome differences between humans and the great apes. RDA has also been used to determine differences between two distantly related oak species where similarities of isolated RDA fragments with known retrotransposons were found (Zoldos et al. 2001). In addition, Donnison et al. (1996)

applied RDA to identify male-specific restriction fragments in the dioecious plant *Silene latifolia*. RDA has also been used to identify polymorphisms in banana lines that are a result of genomic rearrangements during in vitro propagation resulting in markers useful for the detection of early variation in the initiation of tissue culture plants (Cullis and Kunert 2000). One of the specific advantages of RDA is that subtractions between pooled DNA samples can be performed in order to identify specific polymorphisms that are only present in either a particular individual or a particular variety rather than relying on identification based on a pattern of polymorphic bands.

The aim of this study was to apply the RDA technology to date palm and to investigate if it is a useful technique for generation of markers suitable for identification and characterization of variable regions in the date palm genome. Such variable regions might ultimately be applicable for either individual date palm plant or variety identification/discrimination. In general, there is a need for DNA-based markers in date palm that will facilitate variety identification as well as confirmation of the genetic fidelity of individual propagated plants. The current identification based on morphological characteristics is often difficult. For example, in the Kingdom of Saudi Arabia alone almost 400 date palm cultivars have been classified (Hussain and El-Zeid 1978), based mainly on their fruit characteristics that are expressed in the mature stage of the plant. In this study, plants of the two date palm varieties Barhee and Medjool, which are commonly used in commercial tissue culture for the asexual production of date palm via somatic embryogenesis, have been investigated. A repetitive labile DNA region has been identified allowing the differentiation between individual plants and partial discrimination between varieties.

Materials and methods

Plant material and DNA extraction

The date palm varieties Medjool and Barhee were used as the source of DNA. The in vitro plants used were Medjool derived from explant material collected in California, and Barhee derived from explant material collected in the United Arabic Emirates. Non-tissue culture-derived Medjool and Barhee plant material originated from 14 plants grown in California that have been propagated solely via off-shoots from mother plant material imported in the beginning of the last century from Northern Africa (Nixon 1950). Total cellular DNA was isolated from leaves (1 g) using either the technique outlined by Aitchitt et al. (1993) or the Nucleon Phytopure plant DNA extraction kit (Amersham Life Sciences, UK) according to the manufacturer's instructions. Both techniques gave similar results.

Preparation of RDA amplicons

RDA was performed following the general outline described by Lisitsyn et al. (1993). Two micrograms of each of the DNAs (Barhee and Medjool) were digested with 80 units of the enzyme *Bam*HI or *Hin*dIII. The digests were then extracted with phe-

nol/chloroform, precipitated and resuspended at a concentration of 100 μg/ml. The *Bam*HI and *Hin*dIII digests were ligated to the adaptor pair JBam 12 (5′-GATCCGTTCATG-3′) and JBam 24 (5′-ACCGACGTCGACTATCCATGAACG-3′) or RHind 12 (5′-AGCTTGCGGTGA-3′) and RHind 24 (5′-AGCACTCTCCA-GCCTCTCACCGCA-3′), respectively. The ligation products were amplified by polymerase chain reaction (PCR) by using the primer JBam 24 or RHind 24 to generate the first-round amplicons, followed by digestion with *Bam*HI or *Hin*dIII to remove the adaptors. Tester DNA was prepared by adding a second adaptor pair NBam 12 and 24 (5′-GATCCTCCCTCG-3′ and 5′-AGGCA-ACTGTGCTATCCGAGGGAG-3′) for *Bam*HI-digested DNA or JHind 12 and 24 (5′-AGCTTGTTCATG-3′ and 5′-ACCGAC-GTCGACTATCCATGAACA-3′) for *Hin*dIII-digested DNA to the ends of the first-round amplicons.

Subtractive RDA hybridization, kinetic enrichment and cloning of the difference products

The hybridization reaction was set up using 40 μg driver DNA (Medjool) and 0.4 μg tester DNA (Barhee) (100:1 driver/tester ratio) in a final volume of 4 μl hybridization buffer consisting of 30 m*M* EPPS [(2-hydroxyethyl piperizine)-*N*′-(3-propene sulfonic acid)], pH 8, and 3 m*M* EDTA. The DNA was denatured at 100°C for 10 min, 1 μl of sodium chloride (5 *M*) was added to a final concentration of 1 *M* and the reaction incubated at 67°C for 16 h. The hybridization mix was then diluted and an aliquot amplified using NBam 24 or JHind 24. The first round of amplification was for ten cycles, followed by digestion of the products by mung bean nuclease. The nuclease-treated product was then amplified for an additional 20 cycles. The resulting amplicons, which are called the first difference product, were used in this study. These subtraction products were digested with the appropriate restriction enzyme and ligated into the appropriately digested pBluescript II (Strategene, USA). The ligation products were transformed into XL1Blue-competent cells and 50 plasmid-containing colonies carrying an insert were selected and probed with either the Barhee- or Medjool-labeled driver amplicons using the Gene Images random prime-labeling module (Amersham Life Sciences). Colonies that showed a much stronger signal after hybridization with the Barhee amplicons than with the Medjool amplicons were selected for plasmid isolation and determination of the insert sequence and size.

Primer design and testing

Pairs of primers were designed using a standard design program (Expasy, Switzerland). The primer pairs were used in a PCR reaction using Barhee and Medjool DNA as template at various annealing temperatures to optimize the PCR reaction. The PCR reactions were carried out in 25-μl volumes containing 25 ng total genomic DNA, 15 ng primer, 100 m*M* of each dNTP, 10 m*M* TRIS-HCl, pH 8.3, 2 m*M* MgCl$_2$ and 0.5 units Taq polymerase (Takara, Japan). Amplification was performed using a Perkin Elmer Gene-Amp PCR system 9600 with the following program: (1) 94°C for 5 min×1 cycle; (2) 94°C for 1 min, 65°C or 60°C for 1 min depending on the primer pair, 72°C for 1 min×35 cycles; (3) 72°C for 5 min×1 cycle, and optional soak period at 4°C. The products were separated on a 1.5% agarose gel, stained with ethidium bromide and visualized under UV light.

Southern blot hybridization

Genomic DNA was digested with different restriction enzymes, separated on a 1.5% TAE agarose gel using electrophoresis, and transferred to a positively charged nylon membrane (Roche, Switzerland), as described by Sambrook et al. (1989). Cloned bands were used as probes and were labeled using the Gene Images random prime, labeling module (Amersham Life Sciences). Mem-

Fig. 1 Sequence analysis of isolated RDA clones DP2, DP36 and DP41 and position of the most variable DNA region of the DP41 clone

```
DP-2   CCTATCGAAC CCATTCATAC AGAGCCAGTT TTCAATGTCC CTCAACCATC GCGCGGATCT AGTAGGGTCT CCCATCCTCC 80

DP-36  CCTATGGCGA CCGGCTGACC TGGCACTGGT GTCCGCAGCA ACTCTGCTCG GATGGAAAGA AGTCGACCTC GACGAAAGCG 80

DP-41  CCTTCTCCCC GTAGTAACCG GCCTCCCCGC AAATCTTGCA AGTATCACTG AGGGGGAAGA AGGAGGAGGG GCCTCCGCAC 80


       CGATAGATAC TTAGGTATAC TAGAAGAGGA TACCGAGAAA ATGTTCCTAG TGGGAGATAG AGATCACATAC AGGAT 156

       GCTCGGTAAA GCCCCGGTAG TACTCCAACA AGTCCGGGTC AATCCGACGG TATCTCCTCG CGCTGGAT 147

       GTGCCTCACG TCGTGGGGAC ACCGTAGATG GCTCGGTAGG TTGCCTTTCC TCCGTTGGAT 141
```

branes were pre-hybridized and hybridized at 65°C in a buffer containing 5× SSC, 0.1% SDS and a 20-fold dilution of the liquid block provided overnight and washed at 60°C using a 1× SSC and 0.1% SDS solution. Detection was performed using the Gene Images CDP-Star detection module, according to the manufacturer's instructions. Membranes were exposed to Hyperfilm ECL (Amersham Life Sciences) and the films developed.

Sequence analysis

For plasmid sequencing, PCR products were cloned into the plasmid pMOS*Blue* (Amersham Pharmacia Biotech, UK) and plasmids were recovered from transformed MOS*Blue* cells selected on an appropriate antibiotic. Sequencing was performed using Sequenase (Perkin Elmer, USA) according to the manufacturer's instructions on an automated DNA sequencer (Applied Biosystems, USA).

Table 1 Sequences of primers used to amplify different parts of clone DP41 with DP41R 5′-GCAACCTACCGAGCCATCT-ACGGT-3′ as the right-hand primer and the difference product DP36 with DP36R 5′-GGAACATTTTCTCGGTATCCTC-3′ as the right-hand primer, and a total number of seven non-tissue culture-derived plants amplifying a DNA fragment with the predicted size. *Numbers in parentheses* indicate the total number of plants tested

Primer	Sequence (5′–3′)	Amplification product	
		Barhee	Medjool
DP41L	CCTTCTCCCCGTAGTAACCG	5 (7)	7 (7)
PLM1	TTACAGAGGGGAAAGGAGGA	5 (7)	7 (7)
PLM4	GGAAGGAGGTGGCTCCG	1 (7)	7 (7)
PLB11	CGCAATCTTGCAAGTATCAGT	2 (7)	6 (7)
DP36L	TCGAACCCATTCATACAGAGC	4 (7)	7 (7)

Results

The subtractions between Barhee and Medjool were performed with either Barhee as tester and Medjool as driver or with Medjool as tester and Barhee as driver for amplicons derived from *Bam*HI- or *Hin*dIII-digested DNAs. Following a single round of subtraction using a tester to driver ratio of 1:100 only one of the four subtractions, that with Barhee *Bam*HI-digested DNA as tester and Medjool as driver, produced a DNA difference product, which is approximately 150 bp in length. The difference product was cloned and 50 *Escherichia coli* colonies containing the cloned difference product (data not shown) hybridized separately with labeled Barhee tester and Medjool driver amplicons. Both sets of amplicons hybridized to the colonies indicating that the isolated difference product or a closely related sequence was present in both varieties.

Ten *E. coli* colonies with the greatest differential signal between tester and driver hybridization were selected, plasmid DNA isolated and the insert sequence determined. This analysis revealed that the cloned difference product consisted of at least three sequences with lengths of 141 bp (DP41), 147 bp (DP36), and 156 bp (DP2) (Fig. 1) indicating a complex mixture of fragments in the difference product. Seven of the sequenced inserts were identical to DP2 and two identical to DP36. The DP41 sequence contained an AGG motif repeated in tandem 3 times. A search of these three sequences using different databases and DNA sequence analysis tools, such as Blast, FastA and the Smith-Waterman algorithm, result-

ed in a 75% homology to *Oryza sativa* genomic DNA, chromosome 1 (accession no. AP002902) when a local alignment was done with DP41 but no homology was found for DP2 or DP36.

Since most of the sequenced clones were identical to DP2, we selected only the clones DP36 and DP41, the latter of which was unique, for further characterization. Hybridization of labeled DP41 to *Eco*RI- or *Bam*HI-digested DNAs from tissue culture-derived Medjool and Barhee gave a smear pattern, which is characteristic of a dispersed, repetitive sequence (data not shown). Similar hybridization profiles were obtained when either DP36 or DP2 were used as probes.

Four primers, DP41L and DP41R and DP36L and DP36 R (Table 1), were designed from the DP41 and DP36 sequences. These primer pairs amplified the predicted 120-bp or 110-bp PCR product, respectively, from genomic DNA of six different tissue culture-derived Medjool and Barhee plants but failed to amplify a fragment in two non-tissue culture-derived Medjool and three non-tissue culture-derived Barhee plants (Fig. 2; Table 1). The DP41 amplification product was then characterized in more detail. From each of the six tissue-culture-derived Barhee and Medjool plants, three independent clones of the DP41 amplification product were sequenced (Fig. 3). These 36 clones from different Medjool and Barhee plants revealed a high degree of homology with DP41, the differences being minor base pair changes or single base deletions occurring mainly in a
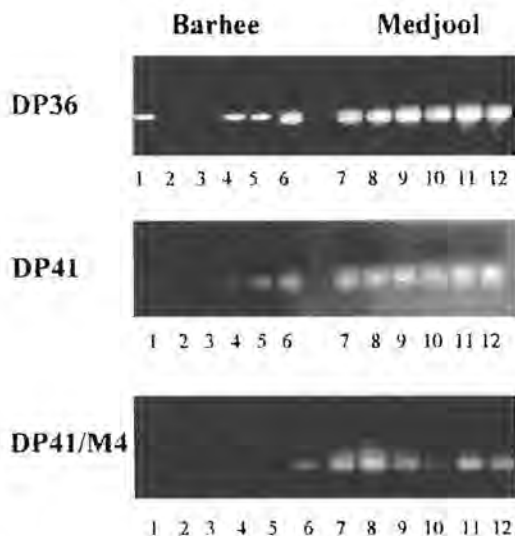
274

Fig. 2 PCR products of genomic DNA from six out of seven Barhee (B) and Medjool (M) plants amplified with primers for DP36, DP41 and DP41/M4 with the following template DNAs: *lanes 1–6* non-tissue culture-derived B plants; *lanes 7–12* non-tissue culture-derived M plants

Fig. 3 Sequence alignment of DP41 with amplified DNA products from genome of M and B plants showing differences to DP41. *Empty space* indicates a base pair deletion. For abbreviations, see Fig. 2

variable 45-bp region of the fragment (Fig. 1). However, in general more changes were observed in the Barhee sequences (specifically single base deletions) than in the Medjool sequences. Only those sequences showing differences in comparison to DP41 are shown in Fig. 3. From the 18 sequences analyzed for each variety, six of the Medjool sequences and eight of the Barhee sequences were identical to DP41. This indicates that the region of the DP41 sequence used to design the primers is common to both genomes. Among the variants found within the different genomes, two, M6 and B9, were identical. Two variants, M1 and M4, the latter with a six base deletion in its sequence, were unique to Medjool and one variant (B11) was unique to Barhee.

Primers PLM1, PLM4 and PLB11 (Table 1) were designed from M1, M4 and B11, to cover the variable portion of these sequences (Fig. 3) when used in conjunction with DP41R. All six tissue culture-derived Medjool plants and also all tissue culture-derived Barhee plants, which originated from a single mother plant, amplified a PCR product with the expected size with all three primers (data not shown). However, primer PLB11 only amplified a PCR product with the expected size of about 110 bp from two of seven non-tissue culture-derived Barhee and six of seven Medjool plants at an optimal annealing temperature of 65°C (Table 1). An identical result was observed with primer PLM1 at an optimal annealing temperature of 60°C (Table 1). Primer PLM4 (at 65°C annealing temperature), which covered a unique 6-bp deletion, amplified a PCR product from DNA of all seven non-tissue culture-derived Medjool plants (Fig. 3) but only from one Barhee plant (Table 1).
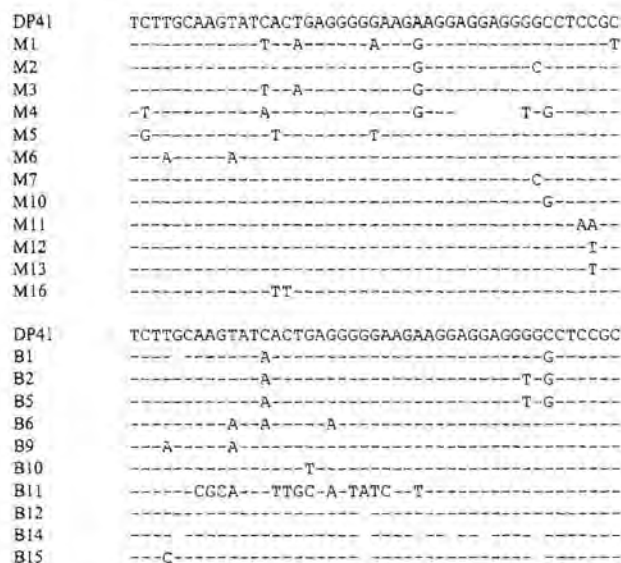
Discussion

RDA has been successfully applied in this study for date palm and has resulted in the isolation of three repetitive DNA sequences from the Barhee genome. A subset of each of these families of sequences appears either to be restricted to Barhee, or present at different multiplicity in Barhee, since the reverse subtraction with Medjool as tester yielded no products. The presence of these sequences in the DNA from both varieties could also be explained by methylation differences between the two varieties, with Medjool DNA being more methylated in this region, since *Bam*HI, a methylation-sensitive restriction enzyme, was used to generate the initial amplicons. However, the lack of difference products when *Bam*HI-derived amplicons of genomic Barhee DNA were used as a driver, or when the subtractions were performed with *Hin*dIII-derived amplicons, is consistent with a general high degree of genomic similarity between plants of the two varieties.

One sequence, which is highly variable, has been further characterized with respect to its presence in individual date palm plants of two date palm varieties. Sequence analysis of the amplified difference products from both date palm varieties identified the existence of several variant members of the repeated sequence, consistent with this family representing a variable "genetic hotspot" in the genome (Linacero et al. 2000). Variation included changes from C→T and G→A, deletion of single base pairs (which occurred at a higher frequency in Barhee) and deletion of several base pairs. Base pair changes represent the most commonly observed point mutations in plants, and can also be a consequence of plant tissue culture (Phillips et al. 1994). The extent of variation observed between plants indicates that this might be a rapidly evolv-

ing/changing sequence. We currently hypothesize that Barhee plants are generally more susceptible to these variations and that these variations have resulted in a heterogeneous Barhee population among non-tissue culture-derived plants including variations in copy number. Since all tissue culture-derived Barhee plants originated from a single Barhee mother plant, conclusions about the behavior of this family of sequences through tissue culture cannot be drawn. However, Pluhar et al. (2001) found an unequal copy number of repeated DNA among callus samples of alfalfa, and speculated that genomic stress induced by tissue culture may have caused that unequal copy number.

In the results reported here, RDA has proved to be useful in identifying a particular repetitive class of sequences that is highly variable in date palms, which is consistent with earlier observations that RDA can be used to isolate families of repetitive sequences (Cullis and Kunert 2000; Nekrutenko et al. 2000; Zoldos et al. 2001). These types of sequences are more difficult to identify with either AFLPs or RAPDs since they either result in many related polymorphisms or generate no size polymorphisms. An additional advantage of RDA is that it can also be performed using bulked amplicons, and can thus be used to identify polymorphisms that are restricted to a particular group of individuals. Therefore, bulking a series of Barhee and Medjool samples and then doing the subtraction will identify variety-specific polymorphisms, rather than individual specific polymorphisms. The types of sequences identified in these experiments would likely be identified as polymorphisms using other techniques, but due to the hyper-variability, each individual would have a unique pattern, or there would be a number of different patterns, none of which would be variety specific. The ideal probes for identifying genomes are those which have a unique location in a given genotype that can be identified directly. Application of RDA therefore offers the opportunity to generate such useful probes.

References

Aitchitt M, Ainsworth CC, Thangavelu M (1993) A rapid and efficient method for the extraction of total DNA from mature leaves of the date palm (*Phoenix dactylifera* L.). Plant Mol Biol Rep 11:317–319

Charters YM, Robertson A, Wilkinson MJ, Ramsay G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5′-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442–447

Corniquel B, Mercier L (1994) Date palm (*Phoenix dactylifera* L.) cultivar identification by RFLP and RAPD. Plant Sci 101:163–172

Cullis C, Kunert KJ (2000) Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. Acta Hortic 530:421–428

Donnison IS, Siroky J, Vyskot B, Saedler H, Grant SR (1996) Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. Genetics 144:1893–1901

Fang DQ, Roose ML (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. Theor Appl Genet 95:408–417

Hussain F, El-Zeid A (1978) Studies on physical and chemical characteristics of date varieties of Saudi Arabia. Ministry of Agriculture and Water, Saudi Arabia

Linacero R, Alves EF, Vazquez AM (2000) Hot spots of DNA instability revealed through the study of somaclonal variation in rye. Theor Appl Genet 100:506–511

Lisitsyn N, Lsistsyn N, Wigler M (1993) Cloning the differences between two complex genomes. Science 259:946–950

Lisitsyn NA, Segre JA, Kusumi K, Lisitsyn NM, Nadeau JH, Frankel WN, Wigler MH, Lander ES (1994) Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nat Genet 6:57–63

Michiels L, Van Leuven F, Van den Oord JJ, De Wolf-Peeters C, Delabie J (1998) Representational difference analysis using minute quantities of DNA. Nucleic Acids Res 26:3608–3610

Navin A, Prekeris R, Lisitsyn NA, Sonti MM, Grieco DA, Narayanswami S, Lander ES, Simpson EM (1996) Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics 36:349–353

Nekrutenko A, Makova KD, Baker RJ (2000) Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. Gene 249:47–51

Nixon RW (1950) Imported varieties of dates in the United States. US Department of Agriculture circular no. 834. US Department of Agriculture, Washington, D.C.

O'Hanlon PC, Peakall R, Briese DT (2000) A review of new PCR-based genetic markers and their utility to weed ecology. Weed Res 40:239–254

Phillips RL, Kaeppler SM, Olhoff P (1994) Genetic instability of plant tissue cultures: breakdown of normal controls. Proc Natl Acad Sci USA 91:5222–5226

Pluhar SA, Erickson L, Pauls KP (2001) Effects of tissue culture on a highly repetitive DNA sequence (E180 satellite) in *Medicago sativa*. Plant Cell Tissue Organ Cult 67:195–199

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Scribner KT, Pearce JM (2000) Microsatellites: evolutionary and methodological background and empirical applications at individual, population, and phylogenetic levels. In: Baker A (ed) Molecular methods in ecology. Blackwell, London, pp 235–271

Sedra MH, Lashermes P, Trouslot P, Combes MC, Hamon S (1998) Identification and genetic diversity analysis of date palm (*Phoenix dactylifera* L.) varieties from Morocco using RAPD markers. Euphytica 103:75–82

Toder R, Grutzner F, Haaf T, Bausch E (2001) Species-specific evolution of repeated DNA sequences in great apes. Chromosome Res 9:431–435

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

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Wolff K, Zietkiewicz E, Hofstra H (1995) Identification of *Chrysanthemum* cultivars and stability of DNA fingerprint patterns. Theor Appl Genet 91:439–447

Zoldos V, Siljak-Yakovlev S, Papes D, Sarr A, Panaud O (2001) Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. Mol Genet Genom 265:234–41

Copy number	the number of copies of a given gene in a set of chromosomes, see multigene family.
Cross-hybridization	the binding of a probe to a DNA sequence other than the intended target sequence. This occurs because of homology between the probe and the sequence and because low stringency hybridization wash coinditions are followed.
Cultivar	a variety of plant developed through selective breeding programs.
Cytosine (C)	pyrimidine base that pairs with guanosine in DNA.
Deletion	removal of gene region or base pair from chromosome.
Denaturation	for DNA or RNA, describes separation of double-stranded molecule to a single-stranded state, usually by heating; for protein, describes change in physical shape, which is usually renders it inactive.
Digested DNA	DNA cleaved by the action of restriction enzymes.
Diploid	the normal number of chromosomes (two copies of each – 2n) in virtually all eukaryotes.
Direct repeats	multiple identical (or closely related) nucleotide sequences in the same orientation in a DNA molecule.
DNA	(deoxyribonucleic acid) the molecular basis of heredity. DNA consists of a polysugar-phosphate backbone from which the bases (nucleotides) project. DNA forms a double helix that is held together by hydrogen bonds between specific pairs of bases (thymine to adenine, guanine to cytosine). Each strand in the double helix is complementary to its partner strand in terms of its base sequence.
DNA ligase	enzyme that joins two double-stranded DNAs together, end to end, by catalyzing 3'OH and 5'P termini bond formation.
DNA polymerase	an enzyme that catalyzes synthesis of DNA under direction of a single-stranded DNA template.