Stress-induced genome alterations in plants

by

CHRISTELL VAN DER VYVER

Thesis submitted in partial fulfillment of the requirements for the degree

PHILOSOPHIAE DOCTOR

Forestry and Agricultural Biotechnology Institute (FABI)
Department of Botany

in the

Faculty of Natural and Agricultural Sciences

University of Pretoria

SUPERVISORS: PROF. K.J. KUNERT
PROF. C. CULLIS

December 2002

TABLE OF CONTENTS

Stress-induced genome changes in plants.

Christell van der Vyver
Department of Botany, Forestry and Agricultural Biotechnology Institute, University of Pretoria, 74 Lunnon Road, Hillcrest, 0002, Pretoria, South Africa.

Supervisor: Karl J. Kunert
Department of Botany, Forestry and Agricultural Biotechnology Institute, University of Pretoria, 74 Lunnon Road, Hillcrest, 0002, Pretoria, South Africa.

Co-supervisor: Chris Cullis
Department of Biology, Case Western Reserve University, 10900 Eucid Avenue, Cleveland, Ohio, 44106, USA.

Dissertation submitted in partial fulfillment of the requirements for the degree Philosophiae Doctor.

ABSTRACT

Stress can alter the genomic composition of a plant. Among the stresses that can change the genome is the introduction of foreign DNA molecules via a plant tissue culture process. Representational difference analysis (RDA) was used as a novel technique to isolate and characterize genomic changes that may be associated with transgene insertion. Three different DNA sequences were isolated by RDA as subtraction products from tobacco expressing an exogenous cysteine proteinase inhibitor (cystatin) transgene and showing an conditional phenotype, namely reduced stem elongation, under low light. Subtraction products represent possible DNA sequence differences between wild-type plants and plants derived from a plant tissue culture/gene insertion process. Two methylation-sensitive subtraction products, Hp12 and Hp14, were similar to part

of the tobacco chloroplast genome and the tobacco 18S rRNA gene, respectively. A third non-methylation sensitive DNA subtraction product, Hi30, had no significant homology to any reported DNA sequences. Screening a genomic library derived from a cystatin expressing plant, the Hi30 sequence could be localized in regions with homology to known repetitive DNA sequence families. Flanking regions of the Hi30 subtraction further revealed homology to DNA sequences of ribosomal RNA genes and to general cloning vectors. A variety of both wild-type and plant tissue culture/gene insertion derived tobacco plants were further screened by PCR for the presence/absence of the RDA subtraction products. Amplification products showed both changes in the copy number and DNA sequence variability. However, these changes could not clearly differentiate between the two types of plants. Due to the homology of one subtraction product to rDNA, plants were also screened for possible changes in the rDNA repeat unit. A similar pattern for the rDNA transcribed regions using Southern blot analysis were found regardless if wild-type or plants derived from a plant tissue culture/gene insertion process were used. However, generally less rDNA was present in plants derived from a plant tissue culture/gene insertion process. Overall, only minor genome changes could be identified with RDA in plants derived from a stressful tissue culture/gene insertion process, which could however not be clearly associated with to the genetic modification process.

Research Objectives

There is evidence that stress can alter the genomic composition of the plant. The introduction of foreign DNA molecules into the plant genome to produce genetically modified plants might be among such stresses. Plant transformation involves a plant tissue culture process requiring plant growth regulators, antibiotics to regenerate and select from an explant and the transfer of an exogenous transgene. In this PhD project, the potential of the technique of Representational Difference Analysis (RDA) was therefore evaluated to possibly isolate and characterize genome changes that might have occurred as a direct result of the plant tissue culture/gene insertion process in tobacco plants. In particular (1) the evaluation of the RDA technique to identify genome changes in *in vitro* propagated genetically modified tobacco, (2) the identification and characterization of such possible genome alterations associated with plant transformation, (3) the location of such genome differences in the plant genome and (4) the general occurrence of such genome changes in a variety of both genetically modified and non-modified wild-type plants were studied. The experiments carried out in this PhD project were therefore focused on three primary objectives. These were (1) the evaluation of the usefulness to isolate genome changes in tobacco, which were possibly induced by a stressful event such as plant tissue culture/gene insertion, (2) the isolation and characterization of possible genomic variations associated with plant tissue culture/gene transfer in genetically modified tobacco, (3) genome localization of possibly altered genome sequences and (4) the possible detection of such changes in a variety of genetically modified tobacco plants. An additional avenue, the relationship of an unusual plant phenotype, which was observed in the transgenic lines, to transformation and exogenous gene expression, was also pursued in a preceding study.

Thesis Composition

Chapter 1 of this thesis presents an introduction into stress in plants and what is currently known about stress-induced changes of the plant genome with a primary emphasis on changes in the DNA sequence. This first chapter summarizes research results so far obtained and includes the techniques that have been applied for characterization of genome changes and the advantages and disadvantages of these techniques. **Chapter 2** is an outline of a preceding physiological/biochemical study with genetically modified plant material carrying and expressing a cysteine proteinase inhibitor gene. This study formed the basis and objectives for a detailed analysis of the tobacco genome analysis *by Representational Difference Analysis* (RDA) to characterize genetically modified plants derived from an *in vitro* propagation/gene insertion process. **Chapter 3** focuses on the application of (RDA) on genetically modified tobacco. This chapter outlines the results obtained from application of this subtractive technique executed on genomic DNA digested with methylation-sensitive and non-sensitive restriction enzymes including the isolation and cloning of subtraction products. This chapter further outlines the results of their analysis using DNA sequencing and bio-informatics tools. **Chapter 4** focuses on experiments carried out identifying the location of subtraction products in genetically modified tobacco plants by screening a constructed genomic library with cloned subtraction products. This chapter also outlines the results of the application of a two-step PCR method with biotinylated primers to identify and characterize tail-end flanking regions of one of the RDA subtraction products. **Chapter 5** describes the results obtained from the screening of different types of tobacco lines with DNA primers designed from analyzed subtraction products with the emphasis to identify a possible genetic marker for transformation in genetically modified plants. **Chapter 6** the **Summary and Perspective** outline 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. Finally, in the **Annex** the methods and sequence data used in this study are described.

Acknowledgements

First and foremost I would like to thank my heavenly father without whom none of this would have been possible. You have guided me thus far, being a companion through all lives joys and sorrows. I look forward to traveling this path that you have set for me in your constant presence.

Thank you to my family. My mother and farther for years of support and guidance. Never ones falling to encourage me to reach the goals that I've set for myself.

This thesis would not have been completed if not for the help of friends and lab colleagues. Thanks for all your advice and support throughout my project.

I would like to thank the University of Pretoria, the National Research Foundation and the Mellon Foundation for financial assistance that allowed me to conduct and complete my research. A special thanks also to Prof. Karl Kunert for years of patience and support in teaching and sharing his endless amount of knowledge with me. My sincere gratitude and appreciation also goes to Prof Cullis who allowed me to work in his laboratory, sharing his knowledge and spending his time teaching and guiding me. A big thanks also to you and your wife, Margaret, for housing me when working and visiting the USA. My appreciation also goes to Prof. C. Foyer at Rothamsted Research, UK for allowing the use of their facilities and the close collaboration in completing and publishing the results of the physiological and biochemical analysis of the genetically modified tobacco plants.

Lastly but definitely not the least, I want to thank my husband Riaan, for always having my best interest at hart, your constant support and love.

ABBREVIATIONS AND SYMBOLS

BAP	-	6-Benzylaminopurine
bp	-	Base pair
°C	-	Degrees Celsius
dH_2O	-	Distilled water
DNA	-	Deoxyribonucleic acid
E.coli	-	*Escherichia coli*
f.w	-	Fresh weight
g	-	Grams
gus	-	ß-glucoronidase reporter gene
h	-	Hour
H_2O	-	Water
kDa	-	Kilo dalton
kPa	-	Kilo pascal
L	-	Litres
LB	-	Luria Bertani
M	-	Molarity
mg	-	Milligrams
ml	-	Millilitres
mM	-	Millimolar
MS	-	Murashige and Skoog
NaOAc	-	Sodium acetate
NaCl	-	Sodium chloride
NaOH	-	Sodium hydroxide
ng	-	Nanograms
nm	-	Nanometre
NPT II	-	Neomycin phosphotransferase II
OC-I	-	Oryza cyctatin I
PAGE	-	Polyacrylamide gel electrophoresis

pfu	-	Plaque forming units
pH	-	Log hydrogen ion concentration
RNA	-	Ribodeoxynucleic acid
sdH_2O	-	Sterile distilled water
SDS	-	Dodecylsulphate sodium salt
sec	-	Seconds
SSC	-	Sodium chloride; Sodium citrate
T-DNA	-	Transfer DNA
U	-	Unit
µg	-	Micrograms
µl	-	Microlitres
µM	-	Micromolar
UV	-	Ultra violet
%	-	Percentage

INDEX OF FIGURES

of the Hi30 subtraction product aligned at different locations with the original flanking sequence represented in A. (T-C1 and T-C2) Tobacco plants ('*Samsun/UK*'), which went through a transformation process but not carrying a transgene. (NT2 and NT3) Genomic DNA-derived from non-modified wild-type individual tobacco plants of cultivar '*Samsun/UK*'. (T1 and T2) Amplification of genomic target DNA derived from genetically modified tobacco plants ('*Samsun/UK*') carrying the OC-I coding sequence. (San1 and San2) DNA from non-modified tobacco cultivar *'Samsun/San1191'*. (NT1) Genomic DNA-derived from the non-modified wild-type individual tobacco plant used in the RDA procedure. Primer 3 used in the amplification of the original flanking sequence (A) are underlined as well as the Hi30R and H30H30 primer sequences, Gap=(-), Identical base pairs aligned with A=(.).

Figure 5.9. **136**

Amplification of genomic target DNA using PCR with primers designed for Hp12 with a single base pair change as identified in subtraction product Hp12. (A) DNA amplification with genomic DNA-derived from either non-modified wild-type individual tobacco plants of cultivar '*Samsun/UK*' (NT1 to NT4) (NT1 represents the original wild-type DNA used in the RDA procedure) and genetically modified tobacco plants ('*Samsun/UK*') containing a *gus* transgene (G1 to G3) as template. Lane M represents a 100 bp marker. (B) Amplification of genomic target DNA derived from genetically modified tobacco plants ('*Samsun/UK*') carrying the *OC-I* coding sequence (T1 to T4); tobacco plants ('*Samsun/UK*'), which went through a transformation process but not carrying a transgene (T-C1 to T-C4) and DNA from tobacco cultivar *'Samsun/San1191'* (San1-4). Lane M represents a 100 bp marker.

INDEX OF TABLES

Genome variations in transgenic tobacco explored by representational difference analysis

Chapter 1: Plant genome variation under stress

Any change in an environmental condition reducing or adversely modifying growth and development of an organism can be considered as stressful, and potentially affecting the genome of an organism. In comparison to the great number of reports focusing on variable gene expression under stress, investigations about genomic variation at the DNA sequence level as a response to stress are rather limited. This introductory review outlines the current knowledge about stress-induced genome variations mainly investigated in plant tissue culture and their detection on the DNA sequence level. Variation includes both direct changes in the DNA sequence and epigenetic variation due to DNA methylation. Detection of such genome variations and possible related phenotypic changes have been investigated with a variety of test methods at the morphological, cytological, cytochemical, biochemical, and molecular levels.

I) The plant genome and stressful environments

Conditions for growth are seldom optimal and any change in an environmental condition that results in a response of an organism might be considered as stressful with the potential for modifying genome composition, growth and development of the organism (Levitt, 1972; Koehn and Bayne, 1989).

Many research groups have investigated the stressful factors that vary gene expression. Typical environmental stress factors *in vivo* include pathogenic micro-organisms, a wide variety of abiotic stresses, such as drought, extreme temperatures, toxic minerals and pollutants, and also *in vitro* plant processes such as tissue culture and genetic engineering of plants (Chapin, 1991;

Fowden et al., 1993; Ditt et al., 2001; Cassells and Curry, 2001). Figure 1.1 outlines a variety of stresses affecting the plant genome.

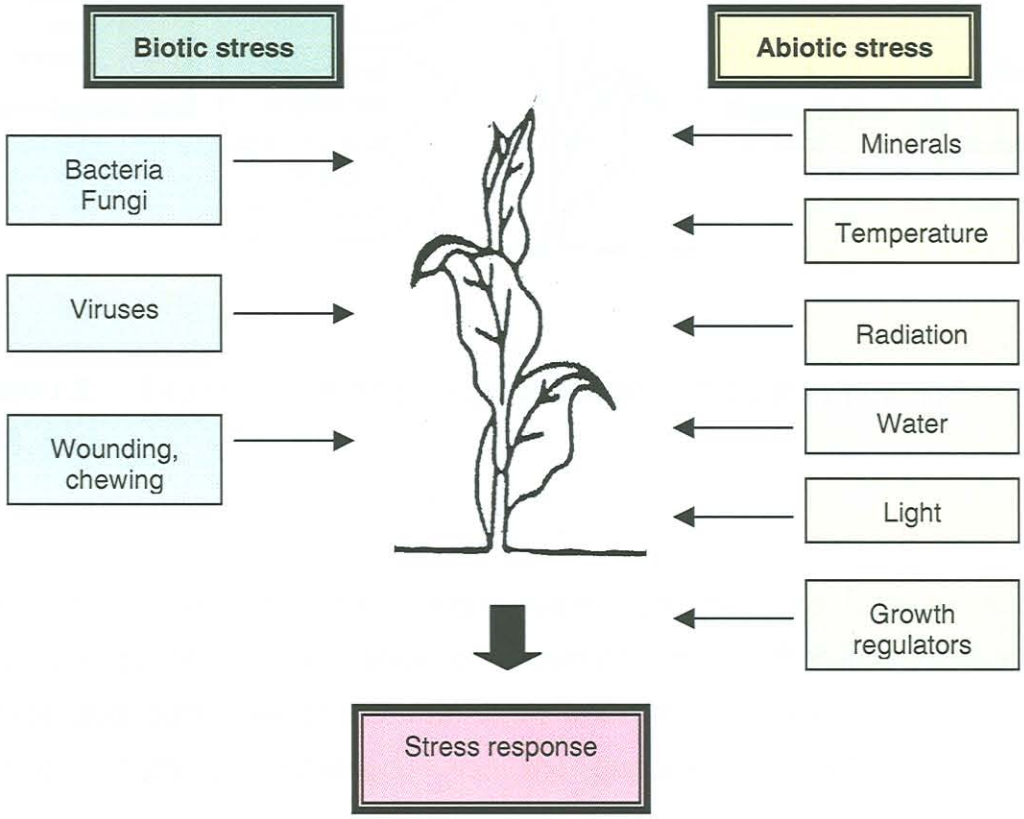

Figure 1.1: Abiotic and biotic stresses affecting the plant genome (Salisbury and Ross, 1992).

II) Stress and the response of the plant genome

A plant needs to adapt to stress in order to survive and many factors determine how the plant genome responds to stress. The genetic make-up of a plant, its developmental circumstances, the duration and severity of the stress, the number of times the plant is subjected to stress and any synergistic effects of multiple stress influences this genome response (Figure 1.2). If adaptation and repair mechanisms are not sufficient and the effect of the stress factors is ultimately not alleviated, the outcome of stress will be death of the organism.

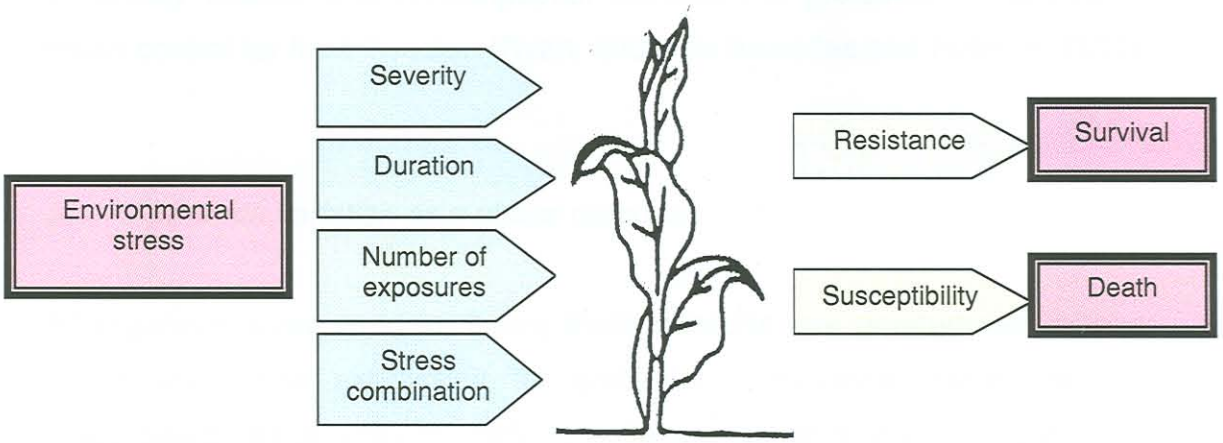

Figure 1.2: Factors determining the stress response of a plant (Buchanan et al., 2000).

Plants can survive by either avoiding or tolerating stress. Simple avoidance of stress can be through expression of certain phenotypic characteristics, such as extended roots down to the water table to avoid drought stress. Tolerance of stress might be achieved when the genome expresses temporarily protective proteins to modulate metabolism such as the heat shock proteins. A large number of research groups have investigated variable gene expression of protective systems under stress, including the function of a single protective gene in transgenic plants, or production of a complete set of new proteins, such as heat shock proteins (Viswanathan and Khannachopra, 1996; Sabehat et al., 1998; Smirnoff, 1998; Bartels and Nelson, 1994; Savenstrand et al., 2002; Savenstrand et al., 2000). Among the specific targets for investigation of up-regulation of protective systems by the genome are pathogenesis-related proteins and components of the cellular antioxidative system, which protect against oxidative stress (Kitajima and Sato, 1999; Foyer and Noctor, 2000; Foyer et al., 1994; Perl-Treves and Perl, 2002). Protective systems are sometimes transmitted to and expressed in distant tissues and even in neighboring plants. For example, the chewing of insects or wounding activates proteinase inhibitors that initially release a signal molecule, such as systemin, leading to a signaling cascade that

eventually results in a transcriptional activation of proteinase inhibitors for insect control far from invasion (Ryan, 2000; de Bruxelles and Roberts, 2001).

A) Genome variation as a stress response

An organism is prepared for facing many stresses in a programmed manner by variable gene expression for protection. However, there are also responses of the genome to challenges that are not so precisely programmed. This might result in chromosome breakage, DNA mutations and ultimately changed gene expression (McClintock, 1984). In comparison to the great number of reports focusing on variable gene expression under stress, fewer reports about the direct effects of stress on the DNA sequence are available. 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 on the DNA sequence level for these challenges (Figure 1.3; Capy, 1998; Cullis, 1990; Johnston et al., 1996).

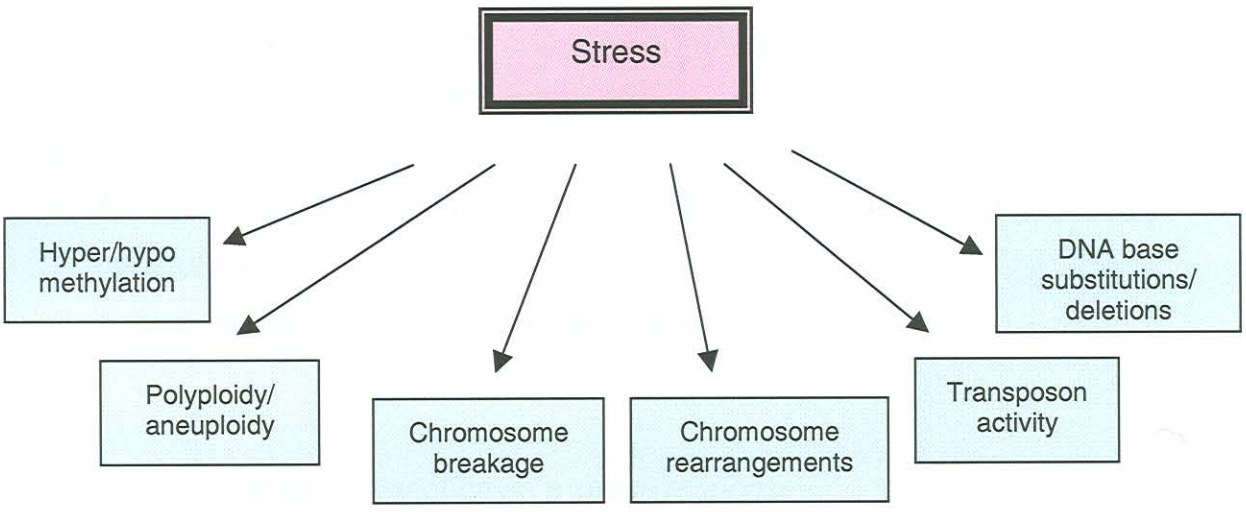

Figure 1.3: Stress and genome modifications (Cassels and Curry, 2001).

The genome has been considered for long as stable to stress despite occasional changes in chromosome structure or inversions. However, recent research suggests that the genome is rather flexible and can undergo changes, which are often referred to as plasticity (Capy, 1998). Such changes might occur naturally over long time periods during evolution. But plant tissue culture, where plants are rapidly propagated in a stressful, artificial environment, seemingly induces such changes in a relatively short time period. Chromosome rearrangements were detected in regenerated tissue culture plants of *Zea mays* (Lee and Phillips, 1986) and changes in the amount of rDNA and peroxidase isozyme band patterns were found in flax exposed to stress (Cullis, 1981). Since plants are unable to move and search for favourable growth conditions, they have to adapt their genome to the changing environment. Walbot and Cullis (1983 and 1985) proposed that once the ordinary physiological responses to an environmental stress are exhausted, the plant genome has to adapt to the new environment by rearranging its DNA in limited genomic regions, which might be related to phenotypic effects. According to Cullis (personal communication), the genomic response in plants needs to have the following properties to have an adaptive advantage towards stress:

- A sensing system that activates the adaptive mechanism when normal physiological responses are exhausted.
- A phenotypic variation as a result of genome variation.
- An advantage from the change in its DNA sequence for current and future generations.
- A genome alteration, which is reversible and limited.

B) DNA sequence variation as a stress response

Genome variation in plants as a response to stress can be either genetic or epigenetic. Figure 1.4 summarizes the genetic and epigenetic genome variations, which have been reported (Cassells and Curry, 2001; Kaeppler et al., 2000; Abe et al., 2002). Genetic changes include both chromosomal gross

rearrangements and changes in the DNA sequence, whereas epigenetic changes are primarily alterations in DNA methylation. All parts of the genome may not be equally susceptible so that variation in the genome is dependent on a particular stress being experienced. However, some regions in the DNA sequence of the genome may be more susceptible during stress than others and therefore might alter irrespective of the inducing stress. Plant tissue culture has been investigated as a stressful process in greater detail for these variable regions, due to its commercial implications.

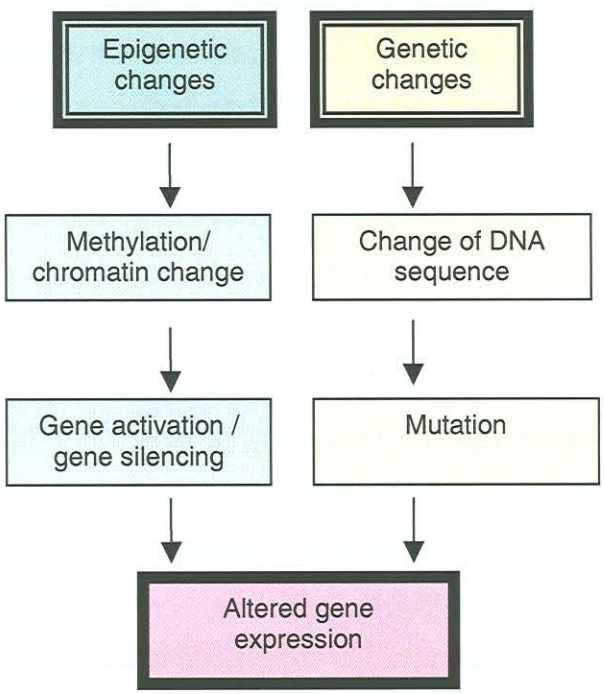

Figure 1.4: Genetic and epigenetic changes in plants due to stress (Capy et al., 2000; Nakao, 2001)).

Genomic variation in tissue culture can result in aneuploidy, chromosomal rearrangements, activation of transposable elements, point mutations, genome rearrangements, changes in ploidy level, methylation changes and even altered copy number of sequences (Cullis, 1990; Peschke et al., 1987; Hirochika, 1993; Phillips et al., 1994). Changes in the ploidy level and

chromosome rearrangements are quite common in tissue culture plants (Kaeppler et al., 2000). In barley ploidy changes were the most prevalent cytological changes followed by chromosome breakage (Hang and Bregitzer, 1993). There is evidence that late replication of hetero-chromatin in tissue culture plants might cause chromosome breakage (Johnson et al., 1987). McCoy et al. (1982) performed a meiotic analysis on regenerated oat plants and found loss of part or the entire chromosome arms, and chromosomal instabilities were also found among *in vitro* grown maize plants (Lee and Phillips, 1986). Translocations were a commonly observed chromosome abnormality with inversions, insertions and deletions occurring in the DNA sequences. Sequence families, which are subject to change can include both repetitive families and low and / or single copy sequences. Blundy et al. (1987) found an almost three-fold reduction in the ribosomal RNA genes in callus cultures of flax and the extent of chromosome abnormalities in maize culture was dependent on the time isolated cells were cultured (Chandler et al., 1986).

Most plant and also animal genomes consist largely of repetitive DNA. Stretches of nucleotide sequence that occur one or only a few times in the genome of a plant can represent as little as 5% of the DNA, while repetitive sequences, typically one to 10 000 nucleotides long, are present in hundred or thousands of copies in the genome (Schmidt and Heslop-Harrison, 1998). Repetitive DNA sequences are especially sensitive to stress-related DNA changes and account for a large portion of variation in sequence copy numbers. Plant tissue culture for example influences such repetitive DNA sequences. *Cymbidium* protocorms exposed to a chemical stress using an auxin-type plant growth regulator, amplified AT-rich satellite DNA, whereas exposure to the plant hormone gibberellic acid increased a GC-rich fraction (Nagl and Rucker, 1976). Highly repeated sequences were amplified up to 75-fold in rice suspension cultures (Zheng et al., 1987) and reduction in copy number of a highly repetitive DNA sequence in plant tissue culture of *Medicago* was also recently reported (Pluhar et al., 2001).

Ribosomal RNA sequences are another highly repetitive sequence family, which can be affected by stress (Blundy et al., 1987). Ribosomal RNA (rRNA) is transcribed from DNA as a large RNA precursor that is subsequently processed. Two types of ribosomes are known in higher plants termed the 70S and 80S (Ting, 1982). The 80S ribosomes are located in the cytoplasm and the 70S are located in the chloroplast and mitochondria. These ribosomes contain smaller subunits and are repeated and arranged in one or more tandem arrays (Nierras et al., 1997). With the exception of some legumes, almost all plant chloroplast genomes, including tobacco, contain two copies of a large inverted repeat, with a size of between 20 and 25 kb. The inverted repeat regions contain the 16S, 23S and 5S rRNA genes as well as some tRNA and ribosomal protein genes, and separate the large single-copy (LSC) and small single-copy regions (SSC) (Lu et al., 1996) (Figure 1.5). In contrast, the rRNA unit in the cytosol consists of the 18S, 5.8S and 25S rRNA coding regions with non-coding spacers with the 5S rRNA genes being present as tandem arrays elsewhere in the genome (Haberer and Fischer, 1996) (Figure 1.6). In the mitochondrion, rRNA is made up by the 18S, 5S and 26S coding units and non-coding spacers (Heldt, 1997). Copy numbers of rRNA genes are highly variable between plants species ranging from a few hundred to thousands of copies per haploid genome, for example *Linum usitatissimum* (flax) contains about a 1000 copies per haploid genome, while *Arabidopsis thaliana* contains about 570 repeats per haploid genome (Cullis, 1979; Pruitt and Meyerowitz, 1986). Some of these rRNA genes, such as the 5S rRNA gene, are highly conserved in the coding region and are useful tools to study evolutionary relationships in organisms. Although the ribosomal RNA sequences are highly variable, stress-induced DNA changes in these regions have not been investigated in great detail. So far, only a decrease in ribosomal RNA genes in callus culture of flax and changes in the amount of rDNA and peroxidase isozyme band patterns in flax exposed to stress have been reported (Cullis, 1981; Blundy et al., 1987). A study conducted by Bettini et al. (1998) also investigated tissue culture-induced variability of the rDNA in the presence or absence of stress but found no qualitative differences with either RFLPs or RAPDs.

Figure 1.5: Ribosomal RNA genes in the chloroplast: arrangement of the 16S-23S-5S RNA gene complex. Transcription of the 16S and 23S rRNA genes is from right to left. SSC = Small single copy region (Nierras et al., 1997; Haberer et al., 1996).

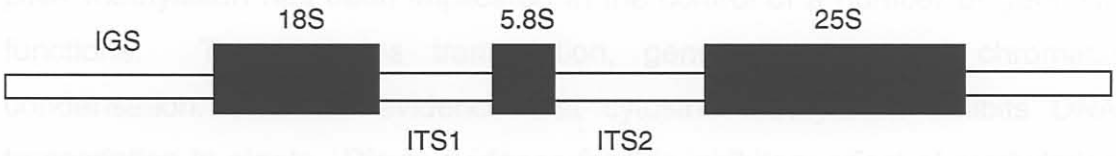

Figure 1.6: Ribosomal RNA genes in the cytosol: arrangement of the 18S-5,8S-25S RNA gene complex. IGS = intergenic spacer; ITS = internal transcribed spacer (Henry, 1997).

C) DNA methylation and stress

Epigenetic variation refers to any somatic or meiotic heritable alteration in gene expression, which is potentially reversible and not due to a DNA sequence change. Epigenetic variation involves mechanisms, such as gene silencing / activation, and can occur due to stress (Waddington, 1953; Kaeppler et al., 2000).

The most common epigenetic variation of DNA reported in plant cells is DNA methylation. Methylation occurs at carbon 5 of cytosines and up to 30% of the cytosine in plant DNA can be modified (Ohki et al., 2001). However,

levels of cytosine methylation can vary considerable between plant species. Rye (*Secale cereale*) has 33% methylation (Thomas and Sherratt, 1956) whereas *Arabidopsis thaliana,* with a small genome, has only 4.6% methylation (Leutwiler et al., 1984). In plants, 70-80% of the CG di-nucleotides are methylated and additional methylation is found for the tri-nucleotide CNG, where N can be any base (Finnegan et al., 1993). The CG dinucleotide has symmetrical cytosine residues in the two DNA strands and, when modified, both cytosines are methylated (Cedar et al., 1979). This symmetry allows the pattern of methylation to be maintained through DNA replication and 5-mC might allow interference with normal protein-DNA interactions and influence of gene expression.

DNA methylation has been implicated in the control of a number of genomic functions. This includes transcription, gene silencing and chromatin condensation. There is evidence that cytosine methylation inhibits DNA transcription in plants. Direct evidence for this inhibitory effect of methylation on transcription in plants comes from experiments in which *in vitro* methylated DNA was introduced into protoplasts of either tobacco or petunia (Weber and Graessmann, 1989). These experiments indicated that to inhibit gene expression the level of methylation must exceed a certain threshold and/or cover specific sites. Introduction of *in vitro* methylated DNA into the protoplasts of tobacco suppressed the expression of the *gus* gene under the control of the 35S promoter sequence from cauliflower mosaic virus (Hershkovitz et al., 1989). These losses of transcription in methylated regions can be due to either the prevention of transcription initiation or by impending transcript elongation. It is thought that this phenomenon serves as a genome-defence mechanism that guards against the deleterious effects of multicopy transposable elements and aberrant gene duplications.

Many studies also investigated the relation between methylation and inactivation of genes (Jost and Saluz, 1993). By linking methylation with ribosomal RNA genes in tobacco, Fulnecek et al. (1998) found that the 5S rRNA gene is highly methylated, exceeding the average methylation density of the tobacco genome. Methylation might also be involved in the control of

inactivation of different 25S rRNA domains in *Brassica* species (Chen and Pikaard, 1997) and by investigating the 5S RNA genes in soybeans, Quemada et al. (1987) demonstrated a decrease in methylation in newly initiated callus and suspension cultures. However, methylation patterns are not always altered. rDNA was quantitatively unchanged in methylation in crown gall callus of flax despite a dramatic reduction in the number of these sequences (Blundy et al., 1987).

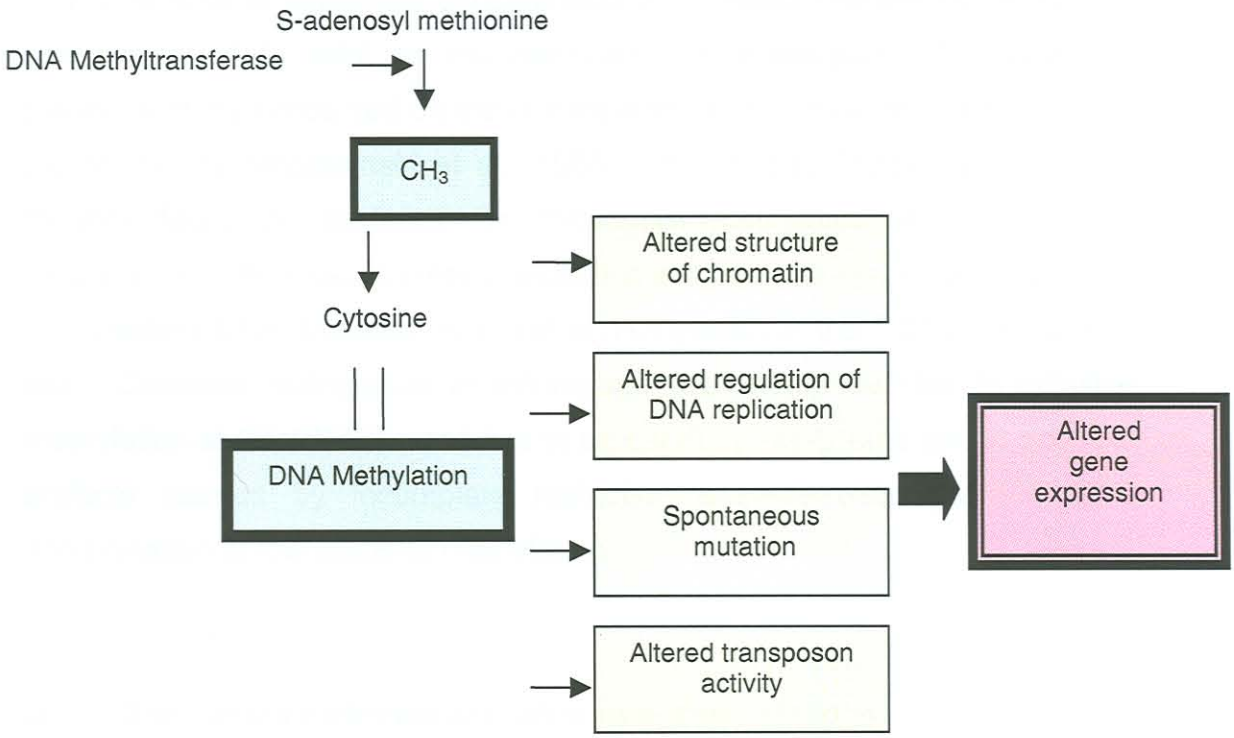

***Figure 1.7*:** DNA methylation in plants showing the enzyme DNA methyltransferase transferring a methyl group to cytosine residues and the resulting consequences for the plant (Nakao, 2001; Jost and Saluz, 1993).

Methylation is carried out by DNA methyltransferases. These enzymes catalyse the transfer of a methyl-group from S-adenosylmethionine to the pyramidine ring of cytosine residues (Figure 1.7). Plants have at least three classes of cytosine methyl-transferases, which differ in their protein structure and function (Nakao, 2001). The MET1 family most likely functions as

maintenance methyl-transferases, but they may also play a role in *de novo* methylation. The chromomethylases, which are unique to plants, seem to preferentially methylate DNA in heterochromatin. Putative *de novo* methyl-transferases finally represent the remaining class and the various classes of methyl-transferases show differential activity on cytosines (Nakano et al., 2000; Bender, 1998).

In plants, methylation is mainly restricted to the nuclear genome, where methylcytosine is especially concentrated in repeated sequences. However, controversial data exist on the methylation of chloroplast DNA in higher plants. A study conducted on the chloroplasts of peas detected methylation in the plastid (Ngernprasirtsiri et al., 1988). In contrast, Fojtova et al. (2001) recently found no evidence for chloroplast methylation in tobacco and *Arabidopsis*. The isoschizomeric restriction enzyme pair *Msp*I and *Hpa*II used for digesting DNA detected no possible methylation at the CCGG recognition site. Detected methylation in chloroplast DNA might be due to cytosine methylation at *Eco*RII sites and has to be carefully interpreted due to possible artifacts caused by incomplete restriction enzyme cleavage and false interpretation of low levels of methylation.

D) Transposable element activation as a stress response

Transposable elements can be activated in the plant genome by stress including plant tissue culture (McClintock, 1984; Peschke et al., 1987). Transposable elements are discrete sequences of DNA that are distinguished by their ability to move from one chromosomal site to another. Two families of transposable elements are known, the transposons and the retrotransposons. The retrotransposons, unlike the DNA transposons, such as the Ac transposon from maize (Brettell and Dennis, 1991), propagate not by "cutting and pasting", but by a mechanism of reverse transcription followed by integration of the new cDNA copy back into the genome (Boeke and Corces, 1989). Due to the replicative nature of retrotransposon mobilization combined with the large size of the elements, which is between 5 and 10 kb, these

elements have the potential to be major contributors to genome variation in stressful environments (Vicient et al., 2001). Larkin and Scowcroft (1981) proposed that the activation of transposable elements might be responsible for tissue culture-induced mutations and Hirochika et al. (1996) found that three of five reported rice retrotransposons were activated under tissue culture conditions and that their copy number increased with a prolonged culture period. Similar retrotransposon activation was also reported for tobacco (Hirochika, 1993) and transposition of the tobacco retrotransposon Tnt1 was observed in transformed *Arabidopsis* plants regenerated *in vitro* (Lucas et al., 1995).

There is further evidence that auxin-like compounds, such as 2,4-D, play a role in activating the promoter of the tobacco retrotransposon Tnt1 (Pauls et al., 1994) and that activity of transposons is related to DNA methylation, which is itself influenced by cellular stress (Banks et al., 1988). Brettell and Dennis (1991) reported that when plants containing a quiescent, unstable Ac element were cultured, the regenerated plants had a high frequency of element reactivation and Ac activation was related to the expected change in DNA methylation. Further, transposable elements are frequently present as dispersed repeats with up to 50 – 100 copies per cell (Sutton et al., 1984). Such a high copy number might also be one of the reasons why transposable elements are targeted for methylation.

Methylation of transposable element sequences can silence the expression of transposon-encoded genes and prevent transposon-mediated DNA re-arrangements. Due to cytosine methylation, a loss of RNA-polymerase-II-dependent transcription in the methylated regions is caused either by preventing transcription initiation or by impeding transcript elongation (Barry et al., 1993). Further, methylation of transposable elements can also silence read-through transcription from transposon promoters into neighbouring genes and thereby prevent inappropriate expression of those genes (Bender, 1998). Passage through tissue culture frequently results in reactivation of an inactive transposable element (Peschke and Phillips, 1991). This observation was first detected in the reactivation of transposable elements when Ac

activity was detected in maize plants regenerated from cultures derived from explants that contained no active Ac elements (Peschke et al., 1987). This study confirmed a link between activation of previously silent transposable elements following tissue culture and demethylation of the transposable element sequences.

III) Plant tissue culture and stress

Cultivation of plants *in vitro* is stressful and always carries the risk of genome variation, which might cause phenotypic variation. Recently, Cassells and Curry (2001) hypothesised that much of the variability expressed in tissue culture plants might be the consequence of, or related to, oxidative stress damage. This might be caused to the plant tissues during explant preparation and due to media and environmental factors in the propagation process. Typical stresses in plant tissue culture include high salt concentrations, water imbalance, mineral deficiency, excess in metal ions, overexposure to plant growth regulators, such as auxins and cytokinins, and the introduction of foreign genetic material during plant transformation (Figure 1.8) (Phillips et al. 1994; Skirvin et al., 1994; Cullis, 1999; Arnault and Dufournel, 1994).

A) Stress-induced variation of somaclonal cells

Genome variation in somatic tissue culture cells as a response to stress is a widespread phenomenon. This variation is also called somaclonal variation (Larkin and Scowcroft, 1981) and is not limited to any particular propagation technique or group of plants. In all organisms, spontaneous mutations occur from one generation to the next. However, somaclonal variation specifically describes the additional mutations in plants produced through stressful tissue culture (Bouman and de Klerk, 1997). Theoretically, all cells, organs or regenerated plants should be genetically identical in asexual plant tissue culture to the original explant from which plants have been generated. 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).

A variety of morphological phenotypic variations, possibly due to somaclonal variation, occur in plants regenerated from tissue culture. It has been reported in ornamentals, plantation crops, vegetable and food crops, forest species and fruit trees (Rani and Raina, 2000). The economic consequences of somaclonal variation can be enormous especially in forest trees with long life cycles. Therefore, an analysis of micropropagated plants using a multidisciplinary approach, especially at genome level, is essential. Linacero et al. (2000) found hot spots of DNA instability in rye plants regenerated from immature embryos. At least 40% of the studied rye plants showed at least one variation and the number of mutations per plant was high ranging from 1 to 12. In a study using shoot-tip culture for banana micropropagation, six families of the cultivar 'Williams' showed no variation towards dwarfism but five families did produce dwarf variants (Israeli et al., 1996). Leroy et al. (2001) also found with callus of cauliflower using the microsatellite technology that only 6 out of 224 calli had stable original DNA patterns. Somaclonal variation was also found in *in vitro*-cultured beet plants (Sabir et al., 1992), red clover (Nelke et al., 1993) and *Brassica napus* (Poulsen et al., 1993). However, somoclonal variation has not always been confirmed in cultured plants. Using random amplified polymorphic DNA, Goto et al. (1998) found no genetically instability in micropropagated shoots of the Japanese black pine and interior spruce also showed no genetic instability in the embryogenic cultures when morphological characteristics and isozyme patterns were analyzed (Eastman et al., 1991).

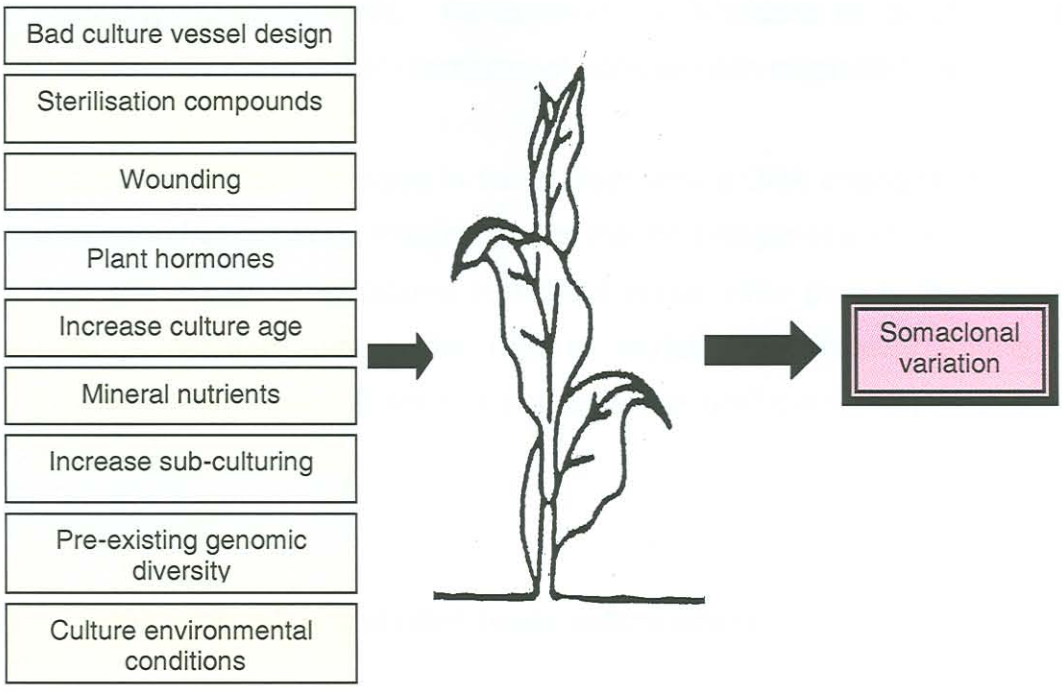

Figure 1.8: Stressful factors inducing genome modifications and plant phenotypes found as a consequence of plant tissue culture (McCoy et al., 1982; Brar and Jain, 1998; Ziv, 1991; Hagege, 1995).

Variation of somatic cells in plant tissue culture and its causes are still not well understood. Tissue culture plants are normally regenerated from a series of cell divisions in meristematic tissues and the apical meristem is formed during the early stages of plant embryo development from which the axillary meristem develops. An apical meristem may, however, originate also from non-zygotic cells, in particular from somatic or callus cells. Plants regenerated from adventitious meristems are often genetically different from the mother plant (Bouman and de Klerk, 1997) and somaclonal variation is therefore often associated with callus formation (Skirvin and Janick, 1976). The use of growth regulators and specifically the auxin-type compound 2,4-D, which induces callus formation, and also the cytokinin-type compound BA (6-benzylaminopurine) have been found to play an important role in the induction

of variability (Evans, 1988). As outlined by Smulders et al. (1994) and Bouman and de Klerk (1997) somaclonal variation can originate from:

- Programmed DNA changes in the explant where DNA changes are not an exception but rather an integral part of the morphogenesis of plants.
- Increase in rate of mutations in explant tissue other than in the apical or axillary meristem where the rate of mutations increases during cell divisions after the initial start of tissue culture until the formation of a new meristem.

B) DNA methylation and plant tissue culture stress

The stressful tissue culture environment also frequently alters DNA methylation patterns. Studies of regenerated plants with methylation-sensitive restriction enzymes have revealed both hyper- and hypo-methylation (Larkin and Scowcroft, 1981). Specifically, the hormone composition of the tissue culture medium can affect the level of DNA methylation in cultured cells. The antibiotics kanamycin and cefotaxime that are commonly used as selective agents in the production of genetically modified plants cause DNA hyper-methylation (Schmitt et al., 1997). Schmitt et al. (1997) also observed increased methylation in repetitive DNA, but the reversal of methylation was not obtained on removal of the antibiotics. Only plants grown from seeds in the progeny and in the absence of these antibiotics showed reduced methylation levels. Besides antibiotics, growth regulators added to the tissue culture medium also affect DNA methylation. Carrot cultures had reduced levels of methylation when grown on a cytokinin-containing medium (Arnholdt-Schmitt et al., 1991). Auxin-type plant growth regulators, such as 2,4-D and NAA, have also been implicated in alterations of DNA methylation. Methylation in carrot cell cultures increased during somatic embryogenesis when cultures were exposed to a high auxin concentration (Lo-Schiavo et al., 1989). However, alterations in methylation patterns in cultured cells have rather a tendency towards demethylation of cytosine residues (Jost and

Saluz, 1993). This might result from either imbalances between the activity of the enzymes involved in maintaining methylation, DNA replication and cell division or chromosome breakage and repair (Peschke et al., 1991).

C) Genetically modified plants and stress

Genome variations occur in genetically modified plants that have been derived from stressful plant tissue culture. In general, genetically modified plants are expected to integrate and express one or more foreign gene(s) in an otherwise unmodified genomic environment. Variation originating from the transformation process might be in addition to any changes arising from direct integration of the transgene. T-DNA integration after *Agrobacterium* infection can cause complex rearrangements and modifications around T-DNA/plant DNA junctions including base substitutions, duplication of border and genomic sequences and small nucleotide deletions (Ohba et al., 1995; Windels et al., 2001; Zheng et al., 2001; Stahl et al., 2002). In addition, increased chromosomal variation in transgenic barley plants, perhaps due to the additional stresses that occur during the transformation process, have also been reported (Choi et al., 2000a and b; Choi et al., 2000).

Known transformation procedures, such as *Agrobacterium tumefaciens* infection, which itself has been recently shown to be a stressful event (Ditt et al., 2001), particle bombardment and DNA transfer to protoplasts, include the use of cultured cells as an intermediate process in the regeneration of genetically modified plants. Labra et al. (2001) showed in rice the occurrence of genomic changes in genetically modified plants produced by infecting calli with *Agrobacterium tumefaciens*. Frequency of genomic variation from the original rice genotype was the highest in genetically modified rice recovered from protoplasts with the longest *in vitro* treatment, intermediate using *Agrobacterium* transformation and a callus phase, but the lowest with particle bombardment or intact cell electroporation.

Disagreement, however, exists about the most dominant reaction leading to variation. According to Sala et al. (2000), *in vitro* cell culture plays a more dominant role in inducing DNA changes than the insertion of the foreign gene itself. Bregitzer et al. (1998) also observed somaclonal variation in the progeny of genetically modified barley plants, but they found that the transformation procedure induced greater variation than the tissue culture process in the absence of transformation. However, regeneration of plants in the absence of a transformation process generally does not encounter the restraints imposed by the conditions used to select the genetically modified cells and also not the pre-existing genetic variability frequently present in the cultivar used.

D) Genetically modified plants and altered phenotypes

Phenotypic changes in genetically modified plants are well documented and include mainly alterations such as chlorophyll deficiencies, altered flowering time and reduced stem elongation. For example, phenotypic alterations found in transgenic rice that may not be due to the transgene integration or expression include longer flowering time, smaller plants and reduced fertility (Arencibia et al., 1998; Bao et al., 1996; Lynch et al., 1995). Altered growth characteristics were also found in genetically modified tobacco plants expressing the *gus* gene selfed over several generations (Caligari et al., 1993). The tobacco variety '*Samsun*' and genetically modified '*Samsun*' in which a rice cystatin gene (*OC-I*) had been introduced by *Agrobacterium tumefaciens*-mediated transform also showed a conditional phenotype. In the first year of this study, a conditional phenotype namely reduced elongation with significantly inhibited stem elongation, when grown under low light intensities were found. Physiological and biochemical analysis of these tobacco plants were done in a study carried out in collaboration with the research group of Prof. C. Foyer at Rothamsted Research (UK) (Van der Vyver et al., 2002, in press).

IV) Detection of stress-induced plant and genome variations

A range of different approaches is available for detecting genome variations and altered phenotypes. The available test methods differ, however, in their sensitivity, technical complexity, ease of use, and stage at which they can be applied. Screening at the morphological, cytological (chromosome number and structure), cytochemical (genome size), biochemical (proteins and isozymes), and molecular (nuclear and organelle genomes) levels provides a useful and often easy tool for detection of variation. Such detection of variation is specifically demanded in plant tissue culture, where severe environmental and chemical stresses might be imposed on plants (Rani and Raina, 2000; Cloutier and Landry, 1994).

A) Morphological and cytological screening techniques

Morphological screening and using chromosome structure for determination of genome and plant variation generally has the advantage of simplicity and the avoidance of any sophisticated analytical procedure. For chromosome structure, chromosome abnormalities can be screened before plant maturation in explants, such as roots, shoot apices, inflorescences, and also in protoplast and callus. Variation in chromosome structure can result in translocations, inversions, duplications and deletions of the plant genome (Karp, 1993). Tetraploid and hexaploid hybrids of potatoes have shown numerical and structural chromosome mutations (Waara et al., 1992).

In morphological screening, plant phenotypes are simply examined, often subjectively, via the description of easily detectable plant characteristics, such as plant form and structure. Change in appearance has been discovered in several tissue culture-derived graminaceous species, such as barley (*Hordeum vulgare*) and rye (*Secale cereale*) (Linacero and Vazquez, 1992), where 1% in barley and 50% in rye showed a morphological change, with chlorophyll deficiency as the most frequently observed change. Other morphological changes were the occurrence of an increased number of

flowers in each spikelet and production of poly-embryos. The frequency of morphological variation might differ, however, between plant cultivars and also appears to be dependent on the genotype.

B) Protein and DNA-based screening techniques

Highly discriminatory profiling methods using very sensitive molecular techniques ease the finding of minor genetic variations. Altered gene expression and changed genomic DNA patterns allow the evaluation of variation at the genome level. In contrast to morphological assessment, the assessment of variation at the genome level describes the internal make-up of a plant and identifies the variance in either the production of a plant protein expressed from certain regions of the DNA or in the DNA sequence. Figure 1.9 outlines the general process for plant identification by either protein or DNA analysis.

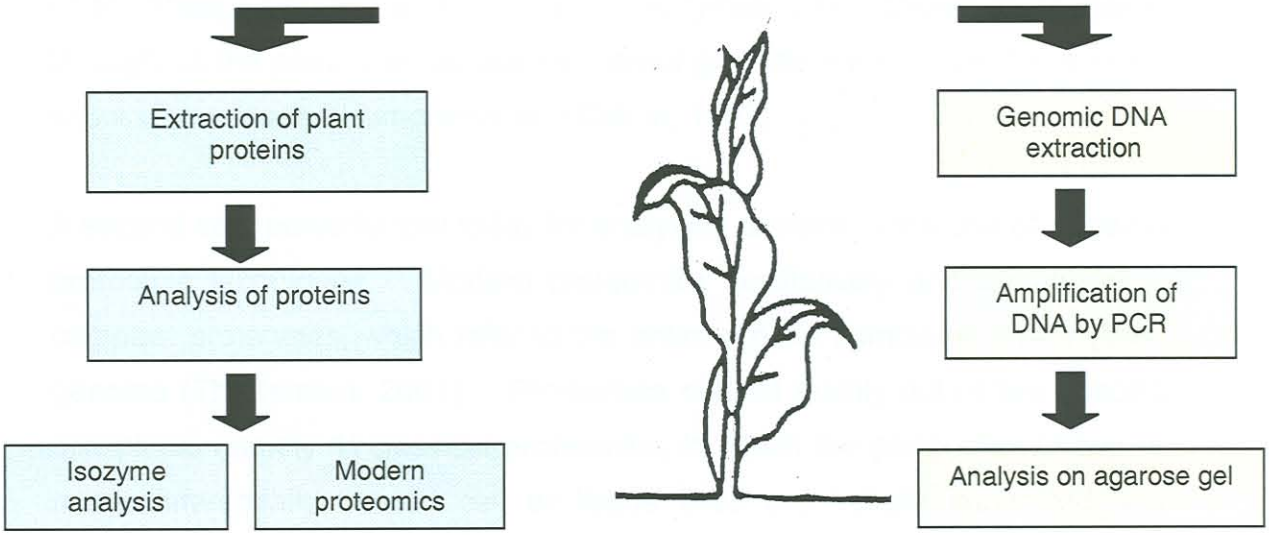

Figure 1.9: A general outline of the steps involved in plant identification by either protein or DNA analysis. As outlined below in detail, proteins are mainly analyzed as isozymes whereas DNA is analyzed by different amplification and subtraction techniques (O'Hanlon et al., 1999).

C) Protein-based screening techniques

Variation in protein patterns can be analysed as isozymes, which are different molecular forms of a protein actively controlling identical biochemical processes of a living cell. Technically, isozymes are separated in an electrical field supported in a polyacrylamide or starch matrix. A characteristic pattern of different isozyme bands called and "isozyme fingerprint" is visualised after staining. A difference between isozyme fingerprints derived from two types of plants is further called a "polymorphism". Enzymes most commonly measured include esterase (EST), glutamate oxaloacetate transaminase (GOT), endopeptidase (ENP), alcohol dehydrogenase (ADH), peroxidase (PER) polyphenol oxidase (POD) matrix (Bebeli et al., 1990). In the past, different isozyme patterns were found in different plant tissue, within and between plant species and between plant populations (Kawarasaki et al., 1996; Diaz et al., 1998; Elisiario et al., 1999). Variation in isozymes patterns were even found between individual *in vitro* grown shoots of Brassica species as well as in tomato, which has been exposed to abiotic stress (Samantaray et al., 1999; Lawrence et al., 1996). Isozymes, when shown to be stable throughout the plant, can be used to detect genome variation as it has been shown for grasses (Humphreys and Dalton, 1992).

A second very powerful tool today for analysing proteins is the use of modern proteomic techniques. Modern proteomics qualitatively and quantitatively compare proteomes, which refer to the entire protein complement of a given genome (Theillement, 2001). Proteomics consist mainly out of two specific disciplines namely 1) classical proteomics, in which the proteomes of two or more differentially treated cell or tissue lines are initially separated and visualized by 2D gel electrophoresis upon which proteins that differ in abundance between the gels are identified by mass spectrometry or 2) functional proteomics, where usually a subset of proteins has been isolated from a given starting material. Each protein in the subset has a common feature, which was used in the isolation procedure. Following some bioinformatics validation work, the common feature can give evidence of the function of each characterized protein (Klose, 1975; O'Farrell, 1975).

Drawbacks, however, for both classical and functional proteomics exist (Fey et al., 1997). Classical proteomics requires little or no prior knowledge about the proteome to be examined but cannot be use for example with all membrane proteins. Also, only proteins with sizes between 10 kDa and 100 kDa can migrate well in 2D gels, while proteins present in low copy numbers are not well detectable on 2D gels (Fey and Larsen, 2001). Functional proteomics in contrast require a good prior knowledge about the system to be studied. Regardless of any disadvantages presently existing in techniques available to study proteomes, it still remain a necessary approach to answer a great number of basic or applied questions. The recent success in discovering great amount of genes through modern genomics makes proteomics a necessary and complementary research field for deciphering the role and function of these newly revealed genes.

D) DNA-based screening techniques

In contrast to protein-based systems, DNA-based systems have the advantage that the DNA content of a cell is independent of environmental conditions, organ specificity or growth stage. Each cell of a living individual contains DNA as genetic material, and the DNA determines the individual characteristics via the control of protein synthesis in the cell. However, except for genome analysis in plant breeding and cultivar identification, DNA-based techniques to detect specifically stress-induced genome variations have been carried out in limited number to determine true-to-typeness of tissue culture-derived plants (Linacero et al., 2000; Skirvin et al., 1994).

Genome variations on the DNA level can be detected by techniques such as RFLP, AFLP, RAPDs and simple sequence repeats (SSRs), also known as microsatellites, (Figure 1.10) (Brown et al., 1990; Muller et al., 1990; Damasco et al., 1996; O'Hanlon et al., 2000). Among the recently introduced methods to detect genome variations, which still have to demonstrate their potential to reliably detect variations, are inter-simple sequence repeats (ISSR) (Albani and Wilkinson, 1998; Leroy et al., 2001) and representational difference

analysis (Powell et al., 1996; Cullis et al., 1999). A review article published by Cloutier and Landry (1994) outlines a variety of DNA-based techniques, which are applicable especially in plant tissue culture.

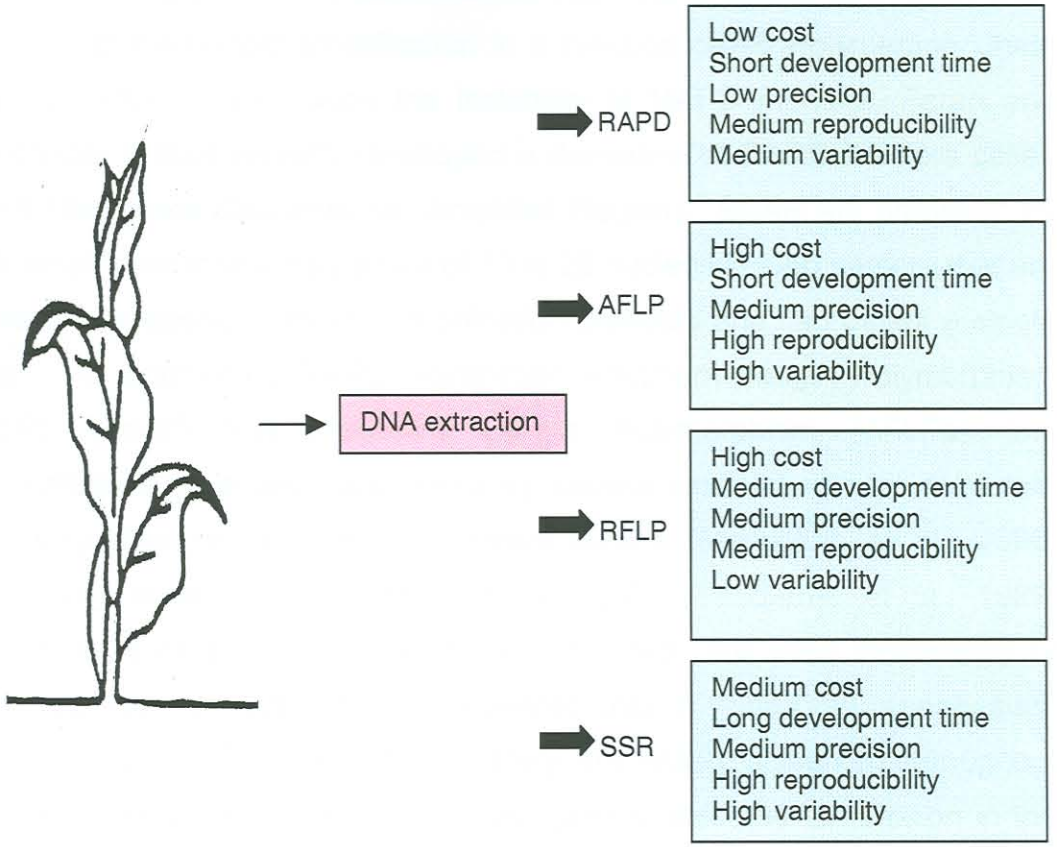

Figure 1.10: A summary of the qualitative characteristics of DNA based techniques to detect genome variations in plants. Cost = financial requirements to prepare a laboratory and obtain results; Development time = time required to develop genetic assays, depend on availability of primers; Precision = diversity present within a sample; Reproducibility = the ability to obtain the same genetic result for the same sample in repeated assays; Variability = inherent capacity of a DNA marker to reveal variation (adapted from O'Hanlon et al., 2000; Powell et al., 1996; Rafalski and Tingey, 1993).

D.1) RAPD, RFLP, AFLP and microsatellites

Characterisation of genome variation in plant tissue using Random Amplified Polymorphic DNA or RAPD analysis (Welsh and McClelland, 1990), which is

often referred to as DNA "fingerprinting", has been widely used to determine plant variation on the DNA sequence level (Heinze et al., 1996; Wolff, 1996). RAPD is possibly the simplest test of all recently applied DNA-based techniques consisting of the production of duplicate of segments of plant DNA and several million-fold amplification in a reaction called Polymerase Chain Reaction (PCR). To reduce the instability of RAPD markers, Paran and Michelmore (1993) recently developed a derivative of RAPD markers called SCAR (Sequence Characterized Amplified Region). SCAR are produced by PCR amplification driven by a pair of 20 to 25 nucleotide long primers that are derived by sequencing RAPD amplification products and can detect a single locus. In addition to RAPD, Restriction Fragment Length Polymorphism (RFLP), Amplification Fragment Length Polymorphism (AFLPs) and microsatellites have been also used by several research groups to detect DNA sequence variation in plant tissue culture (Mandolino et al., 1996; Chowdhury et al., 1994; Matthes et al., 2001; Vendrame et al., 1999; Ruskortekaas et al., 1994). Microsatellites, which are often moderately or highly repeated, consist of tandem repeated units of <6bp DNA repeats such as $(GA)_n$, $(GT)_n$, $(TG)_n$ or $(AAT)_n$. They are widely dispersed throughout eukaryotic genomes and are often highly polymorphic due to variation in the number of repeats. They are greatly informative and locus-specific in many plant species and they detect genome instability of a different type when compared for example with RAPDs. Microsatellites are isolated by screening genomic libraries, by genome walking with the labelled repeat of interest or from the analysis of expressed sequence tags (ESTs).

Technically, basic RFLP markers are detected by hybridisation of a probe, such as a cDNA clone, to restriction digested genomic DNA. In comparison, RAPDs use random ten base primers to generate large numbers of polymorphisms. AFLP markers are generated by PCR amplification of RFLPs. The AFLP procedure further uses adaptors being added to the ends of restriction fragments, and these adaptors are then used as primers in a PCR reaction. Amplified bands in AFLP are separated on a sequencing gel to identify differences in DNA band patterns. In addition, a second technique termed STS (Sequence Tagged Site) also derives from RFLP's. STSs are

short, unique sequences also amplified by PCR, but the primers are derived from sequence information of the ends of a genomic or cDNA clone that had produced an RFLP. An important advantage of STS is the elimination of the need for storage and distribution of clones (Cloutier and Landry, 1994). In order to accelerate gene tagging with molecular DNA markers a number of strategies have been in use for some time, which are all based on high resolution mapping of a target sites. These accelerated techniques all involve the pooling of DNA from individuals identical for a particular trait or phenotype and subsequently extracting the DNA as a pool to perform the comparative analysis (Churchill et al., 1993; Michelmore et al., 1991).

D.2) Representational difference analysis

Representational Difference Analysis (RDA) has attracted much attention especially in our group as a technique to identify genome variations in plants (Cullis and Kunert, 1999; Cullis and Kunert, 2000; Kunert et al., 2002; Vorster et al., 2002). RDA is a powerful DNA-based molecular subtractive technique to isolate labile hyper-variable DNA regions of the plant genome, which might have changed due to adaptation to the environment. Essentially, the method consists of a subtraction of all sequences that are held in common between two individuals, which might be morphologically identical, but differ for example significantly in their tolerance to environmental conditions or in the expression of certain morphological or biochemical characteristics. Technically, the RDA technology combines representation, subtractive hybridization, and kinetic enrichment. Representation means a production of the sub-population of DNA fragments derived from a given DNA population, such that the sequence complexity of the sub-population is lower than the sequence complexity of the initial DNA. Representations, which reduce complexity at least ten-fold over the complexity of the genome of higher organisms with genomes as complex as for grasses, are generally required for the success of the subsequent steps. Subtractive hybridization can be explained as eliminating similar sequences by hybridization between two representations and obtaining unique sequences present in only one of the

representations. Kinetic enrichment is based on the second order kinetics of DNA re-annealing. The rate of formation of double stranded DNA is higher for DNA species of higher concentration. The first round of RDA is mainly dependent on subtractive enrichment, but subsequent rounds do heavily rely on kinetic enrichment. In RDA, kinetic enrichment and subtractive enrichment are combined in a single step called hybridization/amplification.

The RDA technique has the potential, in comparison to the other DNA-based detection techniques, to give in a relatively short time period direct information about genomic losses, rearrangements and amplifications and also insertion of transposable elements into the genome (Lisitsyn et al., 1993). One of the most important advantages of RDA performed on genomes of different plants is its ability to scan in comparison to RFLPs, RAPDS and AFLPs up to 15% of the genome of most plants in each subtraction. The use of 300 random primers in a RAPD analysis for example would scan less than 1% of the same genome. Furthermore, RFLPs, RAPDS, AFLPs and microsatellites, generate random polymorphisms that are useful, for example, to define a population structure, but do not give an indication of the underlying causes of the population differentiation. The RDA technique is generally less suitable for such population studies, but is able to isolate and characterize any DNA sequences that might have changed in response to, or as an adaptation to, an environmental cue.

Another important advantage of RDA is its ability to preferentially isolate families of repetitive sequences that are unique to one of the compared genomes. Such families of repetitive DNA are homoplasy-free characters that can be converted into genetic markers for plant identification in a high throughput PCR-based assay (Nekrutenko et al., 2000). Identified labile regions in the plant genome by the RDA technique have also the potential to be applied as a genetic marker for a variable quantitative trait. A trait-linked genetic marker can be easily converted to a PCR-based marker applicable in studies to determine plant relationships based on traits or in a plant-breeding program.

E) Limitations of detection techniques

The applicability of each technique to detect stress-induced plant and genome variation is still limited. Morphological characteristics may vary widely with the environment and the growth stage of the plant. Some characteristics, however, which change through somaclonal variation, are obvious once plants mature. This includes a significantly changed plant structure or leaf discoloration and deformation. According to Karp (1993) several dis-advantages occur when using morphological screening or evaluation of chromosome structure in studying genome variation. Morphological variation may be epigenetic and not transmitted to the progeny and individual plants have to be grown to full maturity before any assessment can be completed. Cytology is further a time consuming staining technique and cannot be used to screen large numbers of plants or cultures. The absence of morphological variation or a normal chromosome complement does not mean that the plants are normal. Plants might carry small mutations, which are only manifested in the progeny of regenerated plants and single regenerated plants may also carry more than one small mutation, which may only segregate in the progeny.

Isozyme analysis, although easy to apply, has several general drawbacks (Karp, 1993). These include the dependence of isozyme expression on environmental conditions, the organ-specific presence of an isozyme and the often-limited amount of detectable polymorphism between individual plants. Further, isozyme analysis lacks a direct assessment of genomic variation at the genome level, which comprises the bulk of somaclonal variants in tissue culture plants.

RAPDs are normally found to be easy to perform but have the major disadvantage that reproducibility is difficult to achieve between different laboratories and often even between different people in the same laboratory (Jones et al., 1997). Constant detection of identical DNA amplification products has to be confirmed by several-fold repeated experiments preferably by different people. There are also several reports on the importance of *Taq*

polymerase and the thermocycler used in the PCR reaction for RAPDs and variability in RAPD profiles due to the use of different brands of both (Khandka et al., 1997; This et al., 1997). For example, different brands of DNA *Taq* polymerase amplify differently, which results in varying profiles of DNA amplification products. Skroch and Nienhaus (1995) examined the impact of this irreproducibility on the scoring of RAPDs. When expressed as the percentage of RAPD bands scored that were also scored in replicate data, only 75% reproducibility was obtained for 50 RAPD primers. Both RAPDs and AFLPs simply compare the DNA from any number of different samples and can be used to detect the level of difference between them. In both cases only those differences specific to a particular primer set are detected in any reaction. Thus, if the material is only different at a few sites within the genome (closely related) then a large number of primers have to be used in order to detect variation. For example, in experiments with flax, the use of 300 different RAPD primers only covered about 1% of the genome (Oh and Cullis; personal communication).

Microsatellites have clearly demonstrated their merits in population studies determining the gene flow between plant populations but have limitations to identify any kind of unknown variable region in the plant genome. They suffer from a similar drawback as RFLPs, namely, they need to be isolated and then to be characterized. RFLP also requires both large amounts of DNA and the isolation of informative probes that yield differences between the sources of the DNA. Identified genome changes detected with these two techniques are also rarely directly linked to a trait. Both ISSR and RDA, although known for some time, have not been widely evaluated as a technique. For RDA this is partially due to its complexity and therefore one the objectives of the following study was to evaluate the potential of the RDA technique for identification and characterization of genome differences possibly induced in genetically modified plants.

Conclusions

It is been well established that an organism needs to adapt to a stressful environment in order to survive. The majority of these adaptive responses for survival are at the gene expression level aimed to overcome a short period of stress. They are not designed to generate a permanent genetic change in the DNA sequence of the plant genome. A relative small number of reports indicate, however, that such DNA sequence changes can occur in response to exposure of a plant to a stressful environment. Therefore, there is a need to study such possible stress-induced DNA sequence changes in greater detail to which this study has contributed by comparing the genomes of a unstressed plant and a plant following exposure to a stressful event such as plant transformation, which includes a tissue culture process. Several molecular techniques have been further applied to detect such changes in the genome. But, so far these techniques have yielded little or no specific sequence information of the genome parts that might change under stress. Consequently, there is also a need to evaluate new techniques for identification, isolation and characterization of such susceptible genome regions that vary following stress exposure. The results reported here have further contributed to this end by evaluating the RDA technique. Finally, there is a need to develop markers for detection of these possible sequence changes, which might be ultimately related to either beneficial or detrimental phenotypical plant alterations under stress. This study therefore focuses on experiments to possibly relate any detected genome variation to a stressful condition.

References

Abe T., Li N., Togashi A. and Sasahara T. 2002. Large deletions in chloroplast DNA of rice calli after long-term culture. Journal of Plant Physiology 159: 917-923.

Albani M.C. and Wilkinson M.J. 1998. Inter simple sequence repeat polymerase chain reaction for the detection of somaclonal variation. Plant Breeding 117: 573-575.

Arencibia A., Gentinetta E., Cuzzoni E., Castiglione S., Kohli A., Vain P., Leech M., Christou P. and Sala F. 1998. Molecular analysis of the genome of transgenic rice (*Oryza sativa* L.) plants produced via particle bombardment or intact cell electroporation. Molecular Breeding 4: 99-109.

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

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

Banks J.A., Masson P. and Federoff N. 1988. Molecular mechanisms in the developmental regulation of the maize suppressor mutator transposable element. Genes and Development 2: 1364-1380.

Bao P.H., Granata S., Castiglione S., Wang G.J., Giordani C., Cuzzoni E., Damiani G., Bandi C., Datta S.K., Potrykus I., Callegarin A. and Sala F. 1996. Evidence for genomic changes in transgenic rice (*Oryza sativa* L) recovered from protoplasts. Transgenic Research 5: 97-103.

Bartels D. and Nelson D. 1994. Approaches to improve stress tolerance using molecular genetics. Plant, Cell and Environment 17: 659-667.

Barry C., Faugeron G. and Rossignol J-L. 1993. Methylation induced premeiotically in ascobolus: Coextension with DNA repeat lengths and effect on transcript elongation. Proceedings of the National Academy of Science, U.S.A. 90: 4557-4561.

Bebeli P.J., Karp A. and Kaltsikes P.J. 1990. Somalclonal variation from cultured immature embryos of sister lines of rye differing in heterochromatic content. Genome 33: 177-183.

Bender J. 1998. Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing. Reviews TIBS 23: 252-256.

Bettini P., Chiarugi P. and Buiatti M. 1998. An *in vitro* molecular study of the *Nicotiana tabacum* L. genome in the presence or absence of the herbicide atrazine. Theoretical and Applied Genetics 96: 242-250.

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

Boeke J.D. and Corces V.G. 1989. Transcription and reverse transcription of retrotransposons. Annual Review of Microbiology 43: 403-434.

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

Brar D.S. and Jain S.M. 1998. Somaclonal variation: mechanism and applications in crop improvement. In: Somaclonal variation and induced mutations in crop improvement, p.15-37. Kluwer Academic Publishers, Dordrecht.

Bregitzer P., Halbert S.E. and Lemaux P.G. 1998. Somalclonal variation in the progeny of transgenic barley. Theoretical and Applied Genetics 96: 421-425.

Brettel R.I.S. and Dennis E.S. 1991. Reactivation of a silent Ac following tissue culture is associated with heritable alterations in methylation pattern. Molecular and General Genetics 229: 365-372.

Brown P.T.H., Kyozla J., Suleklyo V., Kimura Y., Shimamoto K. and Lorz H. 1990. Molecular changes in protoplast-derived rice plants. Molecular and General Genetics 223: 324- 328.

Buchanan B.B, Gruissem W. and Jones R.L. 2000. Biochemistry and molecular biology of plants. American Society of Plant Physiologists, Rochville, Maryland.

Caligari P.D.S., Yapabandara Y.M.H.B., Paul E.M., Perret J., Roger P. and Dunwell J.M. 1993. Field performance of derived generations of transgenic tobacco. Theoretical and Applied Genetics 86: 875-879.

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

Capy P., Gasper G., Biemont C. and Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity 85: 1001-106.

Cassells A.C. and Curry R.F. 2001. Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micropropagators and genetic engineers. Plant Cell, Tissue and Organeses 64: 145-157.

Cedar H., Solage A., Glaser G. and Razin A. 1979. Direct detection of methylated cytosine in DNA by use of the restriction enzyme *Msp*I. Nucleic Acids Research 6: 2125-2132.

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.

Chapin F.S. 1991. Integrated responses of plants to stress. BioScience 41: 29-36.

Chen Z.J. and Pikaard C.S. 1997. Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance. Genes and Development 11: 2124-2136.

Choi H.W., Lemaux P.G. and Cho M.J. 2000. High frequency of cytogenetic aberration in transgenic oat (*Avena sativa* L.) plants. Plant Science 156: 85-94.

Choi H.W., Lemaux P.G. and Cho M.J. 2000. Increased chromosomal variation in transgenic versus nontransgenic barley (*Hordeum vulgare* L.) plants. Crop Science 40: 524-533.

Chowdhury M.K.U., Vasil V. and Vasil I.K. 1994. Molecular analysis of plants regenerated from embryogenic cultures of wheat (*Triticum aestivum* L). Theoretical and Applied Genetics 87: 821-828.

Churchill G.A., Giovannoni J.J. and Tanksley S.D. 1993. Pooled-sampling makes high-resolution mapping practical with DNA markers. Proceedings of the National Academy of Science, U.S.A. 90: 16-20.

Cloutier S. and Landry B.S. 1994. Molecular markers applied to plant tissue culture. In Vitro Cellular and Developmental Biology-Plant 30: 32-39.

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

Cullis C.A. 1981. Environmental induction of heritable changes in flax: defined environments inducing changes in rDNA and peroxidase isozyme band pattern. Heredity 47: 87-94.

Cullis C.A. 1990. DNA rearrangements in reponse to environmental stress. Advances in Genetics 28: 73-97.

Cullis C.A. 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 C.A. and Kunert K.J. 1999. Isolation of tissue culture-induced polymorphisms in bananas by representational difference analysis. In: A. Altman, M. Ziv and S. Izhar (eds.). Plant Biotechnology and In Vitro Biology in the 21st Century, Kluwer Academic Press, The Netherlands, p. 245-248.

Cullis C.A., Rademan S. and Kunert K.J. 1999. Method for finding genetic markers of somaclonal variation. International publication number WO 99/53100.

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

Damasco O.P., Graham G.C., Henry R.J., Adkins S.W., Smith M.K. and Godwin I.D. 1996. Random amplified polymorphic DNA (RAPD) detection of dwarf off-types in micropropagated Cavendish (*Musa* spp. AAA) bananas. Plant Cell Reports 16: 118-123.

de Bruxelles G.L. and Roberts M.R. 2001. Signals regulating multiple responses to wounding and herbivores. Critical Reviews in Plant Sciences 20: 487-521.

Diaz O., Salomon B. and Bothmer von R. 1998. Description of isozyme polymorphisms in *Elymus* species by using starch gel electrophoresis. In: Jaradat A.A. (ed.). Third Triticeae Symposium, ICARDA, Aleppo, pp. 199-208.

Ditt R.F., Nester E.W. and Comai L. 2001. Plant gene expression response to *Agrobacterium tumefaciens*. Proceedings of the National Academy of Science, U.S.A. 98: 10954-10959.

Eastman P.A.K., Webster F.B., Pitel J.A. and Roberts D.R. 1991. Evaluation of somaclonal variation during somatic embryogenesis of interior spruce (*Picea glauca engelmannii* complex) using culture morphology and isozyme analysis. Plant Cell Report 10: 425-430.

Elisiario P.E., Justo E.M. and Leitao J.M. 1999. Identification of mandarin hybrids by isozyme and RAPD analysis. Scientia Horticulterae 81: 287-299.

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

Fey S.J. and Larsen P.M. 2D or not 2D. Two-dimensional gel electrophoresis. Current Opinion in Chemical Biology 5: 26-33.

Fey S.J., Nawrocki A., Larsen M.R., Gorg A., Roepstorff P., Skews G.N., Williams R. and Larsen M.P. 1997. Proteome analysis of *Saccharomyces cerevisiae*: a methodological outline. Electrophoresis 18: 1361-1372.

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

Fojtova M., Kovarik A. and Matyasek R. 2001. Cytosine methylation of plastid genome in higher plants. Fact or artefact? Plant Science 160: 585-593.

Fowden C.H., Mansfieeld T. and Stoddard J. 1993. Plant adaptation to environmental stress. Chapman and Hall, London.

Foyer C.H., Lelandais M. and Kunert K.J. 1994. Photooxidative stress in plants. Physiologia Plantarum 92: 696-717.

Foyer C.H. and Noctor G. 2000. Oxygen processing in photosynthesis: regulation and signalling. New Phytologist 146: 359-388.

Fulnecek J., Matyasek R., Kovarik A. and Bezdek M. 1998. Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. Molecular and General Genetics 259: 133-141.

Goto S., Thakur R.C. and Ishii K. 1998. Determination of genitic stability in long-term micropropagated shoots of *Pinus thunbergii* Parl. using RAPD markers. Plant Cell Reports 18: 193-197.

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

Hagege D. 1995. Habituation in plant cell cultures: adaptation to free radicals attacks. Comptes Rendus- Society of Biology 189: 1183-1190.

Hang A. and Bregitzer P. 1993. Chromosomal variations in immature embryo-derived calli from six barley cultivars. Journal of Heredity 84: 105-108.

Heinze B., Westcott R., Schmidt J. and Glossl J. 1996. Application of random amplified polymorphic DNA (RAPD) to detect genetic variation in Norway spruce. New Forests 11: 173-184.

Heldt H.W. 1997. Plant Biochemistry & Molecular Biology. Oxford University press, New York.

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

Hershkovitz M., Gruenbaum Y., Zakai N. and Loyter A. 1989. Gene transfer in plant protoplasts. Inhibition of gene activity by cytosine methylation and expression of single stranded DNA constructs. FEBS Letters 253: 167-172.

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.

Humphreys M.W. and Dalton S.J. 1992. Stability at the phosphogluco-isomerase (PGI/2) locus in *Lolium multiflorum* (2n=4x=28) x *Festuca arundinacea* (2n=4x=42) plants regenerated from cell suspension. Genome 35: 461-467.

Israeli Y., Ben-Bassat D. and Reuveni O. 1996. Selection of stable banana clones which do not produce dwarf somaclonal variants during in vitro culture. Scientia Horticulturae 67: 197-205.

Johnson S.S., Phillips R.L. and Rines H.W. 1987. Possible role of heterochromatin in chromosome breakage induced by tissue culture in oats (*Avena sativa* L.) Genome 29: 439-446.

Johnston J.S., Jensen A., Czeschin D.G. and Price H.J. 1996. Environmentally induced nuclear 2C DNA content instability in *Helianthus annus* (Asteraceae). American Journal of Botany 83: 1113-1120.

Jones C.J., Edwards K.J., Castaglione S., Winfield M.O., Sala F., van de Wiel G., Bredemeijer B., Vosman B., Matthes M., Daly A., Brettscheider R., Bettini P., Buiatti M., Maestri E., Malcevschi A., Marmiroli N., Aert R., Volckaert G., Ruada J., Linacero R., Vasquez A. and Karp A. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European Laboratories. Molecular Breeding 3: 381-390.

Jost J.P. and Saluz H.P. 1993. DNA Methylation: Molecular biology and biological significance. Birkhauser Verlag, Berlin, p. 245-249.

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

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

Kawarasaki Y., Nakano H. and Yamane T. 1996. Purification and some properties of wheat germ acid phosphatase. Plant Science 119:67-77

Khandka D.K., Tuna M., Tal M., Nejidat A. and Golan-Goldhirsh A. 1997. Variability in the pattern of random amplified polymorphic DNA. Electrophoresis 18:2852-2856.

Kitajima S. and Sato F. 1999. Plant pathogenesis-related proteins: Molecular mechanisms of gene expression and protein function. Journal of Biochemistry 125: 1-8.

Koehn R.K. and Bayne B.L. 1989. Towards a physiological and genetical understanding of the energitics of the stress response. Biological Journal of the Linnean Society 37: 157-171.

Klose J. 1974. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. Humangenetik 26: 231-243.

Kunert K.J., Voster J., Bester C. and Cullis C.A. 2002. DNA microchip technology in the plant tissue culture industry. In K Rajasekaran, TJ Jacks and JW Finley (eds.). *Crop Biotechnology, Symposium Series No 829.* American Chemical Society, Washington DC (in press).

Labra M., Savini C., Bracale M., Pelucchi N., Colombo L., Bardini M. and Sala F. 2001. Genomic changes in transgenic rice (*Oryza sativa* L.) plants produced by infecting calli with *Agrobacterium tumefaciens*. Plant Cell Reports 20: 325-330.

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

Lawrence C.B., Joosten M.H.A.J. and Tuzun S. 1996. Differential induction of pathogenesis-related proteins in tomato by *Altenaria solani* and the association of a basic chitinase isozyme with resistance. Physiological and Molecular Plant Pathology 48: 361-377.

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

Leroy X. J., Hily J.M. 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 L.S., Hough-Evans B.R. and Meyerowitz E.M. 1984. The DNA of *Arabidopsis thaliana*. Molecular and General Genetics 194:15-23.

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

Linacero R. and Vazquez A.M. 1992. Cytogenetic variation in rye regenerated plants and their progeny. Genome 35: 428-430.

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

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

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

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

Lucas H., Feuerbach F., Kunert K., Grandbastien M.A. and Caboche M. 1995. RNA-mediated transposition of the tobacco retrotransposon Tnt1 in *Arabidopsis thaliana*. The EMBO Journal 14: 2364-2373.

Lynch P.T., Jones J., Blackhall N.W., Davey M.R., Bower J.B., Cocking E.C., Nelson M.R., Bigelow D.M., Orum T.V., Orth C.E. and Schuh W. 1995. The phenotypic characterisation of R_2 generation transgenic rice plants under field and glasshouse conditions. Euphytica 85: 395-401.

Mandolino G., Demarco S., Faeti V., Bagatta M., Carboni A. and Ranalli P. 1996. Stability of fingerprints of *Solanum tuberosum* plants derived from conventional tubers and vitrotubers. Plant Breeding 115: 439-444.

Matthes M., Singh R., Cheah S.C. and Karp A. 2001. Variation in oil palm (*Elaeis guineensis Jacq.*) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. Theoretical and Applied Genetics 102: 971-979.

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

McCoy T.J., Phillips R.L. and Rines H.W. 1982. Cytogenetic analysis of plants regenerated from oat (*Avena sativa*) tissue cultures: high frequency of partial chromosome loss. Canadian Journal of Genetics and Cytology. 24: 37-50.

Michelmore R.W., Paran I. and Kesseli R.V. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. Proceedings of the National Academy of Science, U.S.A. 88: 9828-9832.

Muller E., Brown P.T.H., Martks S. and Lorz H. 1990. DNA variation in tissue-culture-derived rice plants. Theoretical and Applied Genetics 80: 673-679.

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.

Nakano Y., Steward N., Sekine M., Kusano T. and Sano H. 2000. A tobacco NtMET1 cDNA encoding a cDNA methyltransferase: Molecular characterization and abnormal phenotypes of transgenic tobacco plants. Plant Cell Physiology 41(4): 448-457.

Nakao M. 2001. Epigenetics: interaction of DNA methylation and chromatin. Gene 278: 25-31.

Nekrutenko A., Makova K.D. and Baker R.J. 2000. Isolation of binary species-specifiec 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.

Ngernprairtsiri J., Kobyashi T., Sugiyama T. and Akazawa T. 1988. DNA methylation as a mechanism of transcription regulation in non-photosynthetic plastids in plant cells. Proceedings of the National Academy of Science, USA 85: 4750-4754.

Nierras C.R., Liebman S.W. and Warner J.R. 1997. Does Saccharomyces need and organized nucleolus? Chromosoma 105: 444-451.

O'Farrel P.H. 1975. High resolution two-dimensional electrophoresis of poteins. Journal of Biological Chemistry 250: 4007-4021.

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

Ohba T., Yoshioka Y., Machida C. and Machida Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco - an example for multiple duplications of DNA around the integration target. The Plant Journal 7: 157-164.

Ohki I., Shimotake N., Fujita N., Jee J., Ikegami T., Nakao M. and Shirakawa M. 2001. Solution structure of the methyl-cpg binding domain of human MBDI in complex with methlated DNA. Cell 105: 487-497.

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

Pauls P.K., Kunert K., Huttner E. and Grandbastien M.A. 1994. Expression of the tobacco Tnt1 retrotransposon promoter in heterologous species. Plant Molecular Biology 26: 393-402.

Perl-Treves R. and Perl A. 2002. Oxidative stress: An introduction. In: Oxidative Stress in Plants. D. Inzé and M.V. Van Montagu (eds.), Taylor and Francis, London and New York, p. 1-33.

Peschke V.M. and Phillips R.L. 1991. Activation of the maize transposable element *Suppressor-mutator (Spm)* in tissue culture. Theoretical and Applied Genetics 81: 90-97.

Peschke V.M., Phillips R.L. and Gengenbach B.G. 1987. Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science 238: 804-807.

Peschke V.M., Phillips R.L. and Gengenbach B.G. 1991. Genetic and molecular analysis of tissue-culture-derived Ac elements. Theoretical and Applied Genetics 82:121-129.

Phillips R.L., Kaeppler S.M. and Olhoft P. 1994. Genetic instability of plant tissue cultures: breakdown of normal controls. Proceedings of the National Academy of Sciences, U.S.A. 91: 5222-5226.

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

Poulsen G.B., 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.

Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S. and Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP, and SSR microsatellite markers for germplasm analysis. Molecular Breeding 2: 225-238.

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

Quemada H., Roth E.J. and Lark K.G. 1987. Changes in methylation of tissue cultured soybean cells detected by digestion with the restriction enzymes *Hpa*II and *Msp*I. Plant Cell Report 6: 63-66.

Rafalski J.A. and Tingey S.V. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends in Genetics 9: 275-280.

Rani V. Raina S.N. 2000. Genetic fidelity of organized meristem-derived micropropagated plants: A critical reappraisal. In Vitro Cellular and Developmental Biology - Plant 36: 319-330.

Ruskortekaas W., Smulders M.J.M., Arens P. and Vosman B. 1994. Direct comparison of levels of genetic variation in tomato detected by a GACA-containing microsatellite probe and by random amplified polymorphic DNA. Genome 37: 375-381.

Ryan C.A. 2000. The systemin signaling pathway: differential activation of plant defensive genes. Biochimica et Biophysica Acta - Protein Structure & Molecular Enzymology 1477: 112-121.

Sabehat A., Weiss D. and Lurie S. 1998. Heat-shock proteins and cross-tolerance in plants. Physiologia Plantarum 103: 437-441.

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

Sala F., Arencibia A., Castiglione S., Yifan H., Labra M., Savini C., Bracale M. and Pelucchi N. 2000. Somaclonal variation in transgenic plants. Proceedings of the International Symposium on Methods and Markers for Quality Assurance in Micropropagation (Eds.). A.C. Cassells, B.M. Doyle, R.F. Curry Acta Hort. 530, ISHS.

Salisbury F.B. and Ross C.W. 1992. Plant Physiology. Wadsworth, Belmont, California.

Samantaray S., Rout G.R. and Das P. 1999. Changes in isozyme patterns during *in vitro* regeneration from cotyledon explants of *Brassica* species. Biologia Plantarum 42: 169-176

Savenstrand H., Brosche M., Angehagen M. and Strid A. 2000. Molecular markers for ozone stress isolated by suppression subtractive hybridization: specificity of gene expression and identification of a novel stress-regulated gene. Plant, Cell and Environment 23: 689-700.

Savenstrand H., Brosche M. and Strid A. 2002. Regulation of gene expression by low levels of ultraviolet-B radiation in Pisum sativum: Isolation of novel genes by suppression subtractive hybridisation. Plant and Cell Physiology 43: 402-410.

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

Schmitt F., Oakeley E.J. and Jost J.P. 1997. Antibiotics induce genome-wide hypermethylation in cultured *Nicotiana tabacum* plants. Journal of Biological Chemistry 272 (3): 1534-1540.

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

Skirvin R.M., McPheeters K.D. and Norton M. 1994. Sources and frequency of somaclonal variation. HortScience 29: 1232-1237.

Skroch P. and Nienhaus J. 1995. Impact of scoring error and reproducibility of RAPD data on RAPD-based estimates of genetic distance. Theoretical and Applied Genetics 91: 1086-1091.

Smirnoff N. 1998. Plant resistance to environmental stress. Current Opinion in Biotechnology 9: 214-219.

Smulders M.J.M., Rus-Kortekaas W. and Gilissen L.J.W. 1994. Development of polysomaty during differentiation in diploid and tetraploid tomato (*Lycopersican esculentum*) plants. Plant Science 97: 53-60.

Stahl R., Horvath H., Van Fleet J., Voetz M., von Wettstein D. and Wolf N. 2002. T-DNA integration into the barley genome from single and double cassette vectors. Proceedings of the National Academy of Science, U.S.A. 99: 2146-2151

Sutton W.D., Gerlach W.L., Schwartz D. and Peacock W.J. 1984. Molecular analysis of *Ds* controlling element mutations at the *Adh* locus of maize. Science 223: 1265-1268.

Theillement H. 2001. Potential of proteomics for plant improvement. AgBiotechNet 3, ABN 076.

This P., Cuisset C. and Boursiquot J.M. 1997. Development of stable RAPD markers for the identification of grapevine rootstocks and the analysis of genetic relationships. American Journal of Enology and Viticulture 48: 492-501.

Thomas A.J. and Sherratt H.S.A. 1956. The isolation of nucleic acid fractions from plant leaves and their purine and pyrimidine composition. Biochemical Journal 62:1-4.

Ting I.P. 1982. Plant Physiology. Addison-Wesley publishing company, California.

Van der Vyver C., Schneidereit J., Driscoll S., Turner J., Kunert K.J. and Foyer C.H. 2002. Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances cold tolerance. Plant Biotechnology, in press.

Vendrame W.A., Kochert G. and Wetzstein H.Y. 1999. AFLP analysis of variation in pecan somatic embryos. Plant Cell Reports 18: 853-857.

Vicient C.M., Jaaskelainen M.J., Kalendar R. and Schulman A.H. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiology 125: 1283-1292.

Viswanathan C. and Khannachopra R. 1996. Heat shock proteins – role in thermotolerance of crop plants. Current Science 71: 275-284.

Vorster J.B., Kunert K.J. and Cullis C.A. 2002. Use of representational difference analysis for the characterization of sequence differences between date palm varieties. Plant Cell Reports 21: 271-275.

Waara S., Pljnacker L.P., Ferwerda M.A., Wallin A. and Eriksson T. 1992. A cytogenetic and phenotypic characterization of somatic hybrid plants obtained after fusion of two diffeent dihaploid clones of potato (*Solanum tuberosum* L.). Theoretical and Applied Genetics 85: 470-479.

Waddinton C.H. 1953. Principles of Embryology. Symposia of the Society for Experimental Biology 7: 186.

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

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

Weber H. and Graessmann A. 1989. Biological activity of hemimethylated and single-stranded DNA after direct gene transfer into tobacco protoplasts. FEBS Letters 253: 163-166.

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

Windels P., Taverniers I., Depicker A., Van Bockstaele E. and De Loose M. 2001. Characterisation of the Roundup Ready soybean insert. European Food Research Technology 213: 107-112.

Wolff K. 1996. RAPD analysis of sporting and chimerism in chrysanthemum. Euphytica 89: 159-164.

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

Zheng S.J., Henken B., Sofiari E., Jacobsen E., Krens F.A. and Kik C. 2001. Molecular characterization of transgenic shallots (*Allium cepa* L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders. Transgenic Research 10: 237-245.

Ziv M. 1991. Vitrification: morphological and physiological disorders of *in vitro* plants. In: Micropropagation: Technology and Applications, pp 45-70. Kluwer Academic Publishers, Dordrecht.

Chapter 2: Preliminary study to physiologically and biochemically analyse oryzacystatin I expressing tobacco.

In collaboration with the group of Prof. C. Foyer at Rothamsted Research, (UK), possible improved tolerance to abiotic stress has been evaluated in genetically modified tobacco expressing the cysteine proteinase inhibitor oryzacystatin I (*OC-I*) in the cytosol. This study carried out in benefit to the PhD thesis, proceeded the study on the analysis of the tobacco genome and was mainly conducted in the first year of the PhD thesis.

Abstract

The primary goal of this study was to assess whether expression of exogenous *OC-I* in genetically modified tobacco plants provides improved tolerance to chilling stress. The plants had a conditional phenotype, possible not directly connected to the expression of the transgene, affecting plant morphology and general plant growth characteristics. A marked effect on stem elongation was observed in plants grown under low light intensities. After 7 weeks of growth at low light, the plants expressing *OC-I* were smaller with fewer expanded leaves and a slightly lower total biomass than wild-type plants. Maximal rates of photosynthesis (A_{max}) were also decreased, the inhibitory effect being greatest in the plants wth highest *OC-I* expression. After 12 weeks of growth at low light, however, the plants expressing *OC-I* performed better in terms of shoot biomass production, which was nearly double that of wild-type controls. All plants showed similar responses to drought, however photosynthesis was better protected against chilling injury in plants constitutively expressing *OC-I*. Photosynthesis CO_2 assimilation was decreased in all plants following exposure to 5°C, but the inhibition was significantly less in the *OC-I* expressing plants than in controls. The results of this study are reported below and are important to bring in line with the genetic analysis of the plants used in the study for my PhD thesis.

Introduction

Cystatins bind tightly and reversibly to the papain-like group of cysteine proteinases. Cystatins have been used in attempts to engineer better pest control in plants by targeting the digestive system of Coleopteran insects and nematodes (Leplé et al., 1995; Urwin et al., 2001). The best characterized of these is the oryzacystatin-I (*OC-I*) from rice (Abe et al., 1987) which has been successfully expressed in tobacco without any deleterious effect on the plant (Masoud et al., 1993).

Several cystatins have been isolated from plants and two observations have led to the conclusion that cystatins are involved in the regulation of protein turnover during seed development (Abe et al., 1987; Abe et al., 1992). Other recent studies have also shown that endogenous cystatins are specifically induced during cold or salt stress (Pernas et al., 2000), wounding and/or following treatment with methyl jasmonate (Botella et al., 1996), or by prosystemin overexpression (Jacinto et al., 1998). All of these observations support the hypothesis that cystatins play a crucial role in general plant defence mechanisms. However, additional benefits (or disadvantages) arising from constitutive overexpression of cystatins controlling the action of cysteine proteinases in transgenic plants have been largely ignored. Such cysteine proteinases have acidic pH optima *in vitro*, suggesting that they are localized to the vacuole *in vivo* (Callis, 1995). They are expressed mainly in young and senescent leaves and flowers (Buchanan-Wollastan and Ainsworth, 1997; Guerrero et al., 1998; Xu and Chye, 1999) and accumulate in response to oxidative stress (Schaffer and Fischer, 1988). Recently, a role of cysteine proteinases has also been proposed in programmed cell death (Solomon et al., 1999; Xu and Chye, 1999) and an involvement in developmentally regulated programmed suicide pathways has been found (Hadfield and Bennett 1997; Penell and Lamb, 1997). Linthorst et al. (1993) further reported that expression of the tobacco cysteine proteinase, CYP-8, is regulated by a circadian rhythm and that the proteinase is involved in the wound response in tobacco. However, the interaction between endogenous

cysteine proteinases and endogenous or exogenous cystatins, as their natural inhibitors, during abiotic stress remains to be elucidated.

Results

A) OC-I expression

Plants of 3 selected genetically modified lines T4/5, T4/3-1 and T4/3-2 expressed the *OC-I* coding sequence (Van der Vyver et al., 2002; in press). After immunoblotting, a band at the predicted size of about 12 kDa equivalent to the 102 amino acid *OC-I* protein was observed following SDS-PAGE and using an antiserum raised against *OC-I* (Leplé et al., 1995) (Figure 2.1).

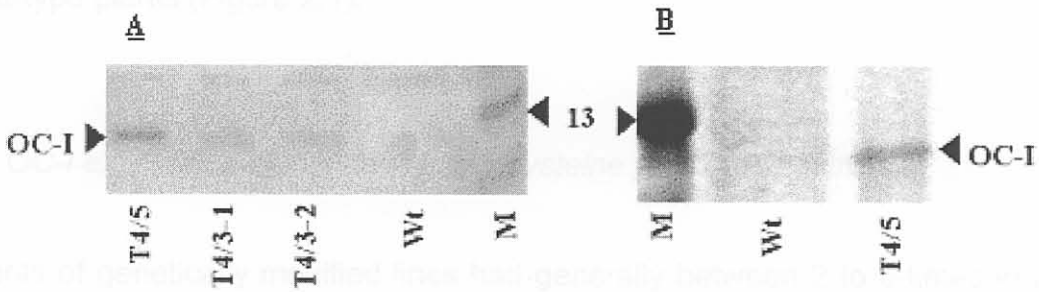

Figure 2.1: Immunoblotting of leaf extracts from plants of genetically modified lines T4/5, T4/3-1 and T4/3-2 and of the wild-type line (Wt). For blotting a polyclonal antibody raised against *OC-I* (A) was used. Separation of plant extracts on SDS PAGE and Coomassie Blue staining of separated plant extracts after heat treatment and concentration of extracts (B). For each line, 10 µg of protein from an extract of a fully expanded leaf (A) or from a concentrated extract (B) was loaded onto the gel. Size marker (M) is shown in kDa.

An accurate direct measurement of the inhibitor level in plant extracts is difficult due to possible binding of the inhibitor to endogenous cysteine proteinases. In the future this problem might be solved under strong

denaturing conditions, breaking the binding between *OC-I* and putative cysteine proteinases. The strongest expression after immunoblotting was detected in modified plants of line T4/5. Recently, it was suggested that *OC-I* is poorly expressed in transgenic plants (Womack et al., 2000). This conclusion perhaps arose because the detection of *OC-I* transgene expression (based solely on immunoblotting) can be problematic. For this reason, we have included two selectable marker genes (*nptII* and *gus*), in addition to *OC-I*, in the transgenic lines. This allowed us to avoid selection of false positives for kanamycin resistance and to increase selection efficiency by using *gus* expression together with *OC-I*, because it is unlikely that the *gus* gene would segregate away from *OC-I* in the progeny. Expression of *OC-I* was also detected after separating a plant extract on a SDS PAGE following heat treatment of the plant extract to remove the bulk of heat-labile proteins and concentrating the expressed heat-stabile *OC-I* by freeze drying (Figure 2.1). No reaction with the antiserum or a protein band for *OC-I* was found in wild-type plants (Figure 2.1).

B) OC-I expressing plants have lower cysteine proteinase activity.

Plants of genetically modified lines had generally between 2 to 5-times lower endogenous cysteine proteinase activity than wild-type control plants when grown under *in vitro* conditions at 25°C on a half-strength MS medium (Figure 2.2). However, a great variability was found in endogenous cysteine proteinase activity between individual plants and batches of plants. One possible explanation is that different levels or forms of cysteine proteinases are present at different stages of plant development, which might be differentially inhibited by exogenous *OC-I* in the different lines (Figure 2.2).

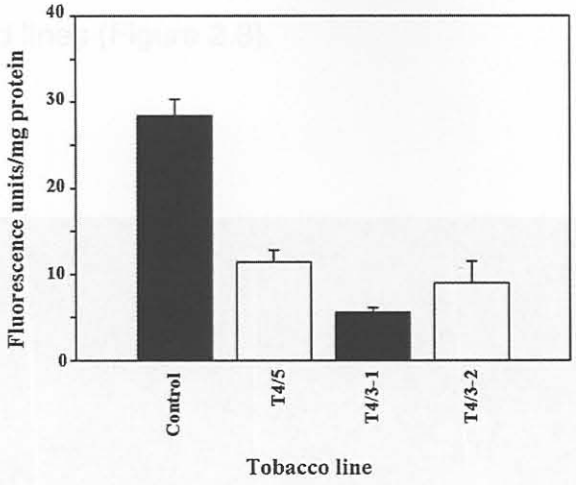

Figure 2.2: Cysteine proteinase activity present in leaves of the lines T4/5, T4/3-1 and T4/3-2 and controls (control). Plants were grown in culture at and irradiance of 50 µmol m^{-2} s^{-1} light intensity. After 4 weeks at 10°C (A) the plants were allowed to recover 2 weeks at 25°C (B). In each case values represent the mean ± standard error of 5 different plants.

C) Conditional phenotype and growth characteristics – Rothamsted Research.

When *OC-I* expressing plants were grown for 7 weeks at a relatively low light intensity (300-350 µmol m^{-2} s^{-1}) in a greenhouse at 20°C or during fall/winter in South Africa without extra light supplementation, stem elongation was substantially decreased compared to controls (Figure 2.3). This phenotype was observed in all the *OC*-I expressing (*OC-I/nptII/gus*) lines but not in the Tgus (*nptII/gus*) control line, which showed a similar phenotype to the non-transformed wild-type controls. This indicates that decreased stem elongation is linked to *OC-I* expression and was not due to somaclonal variation consecutive to the transformation or tissue culture processes. Under low light conditions, flower development was further delayed in plants of the genetically modified lines when compared to wild-type plants and flowering was initiated

after 7 weeks in wild-type plants compared to 10 weeks in the genetically modified lines (Figure 2.3).

Figure 2.3: Growth of plants of genetically engineered lines T4/5, T4/3-2, T4/3-1 and Tgus compared to that of a plant from a wild-type tobacco line (Wt) grown for 7 weeks under a low light intensity (300-350 μmol m^{-2} s^{-1}).

These phenotypic differences between 7 weeks old wild-type and *OC-I* expressing plants were, however, much less apparent at higher growth light intensities, for example 900 μmol m^{-2} s^{-1} when higher light intensity suppresses rapid stem elongation (Figure 2.4).

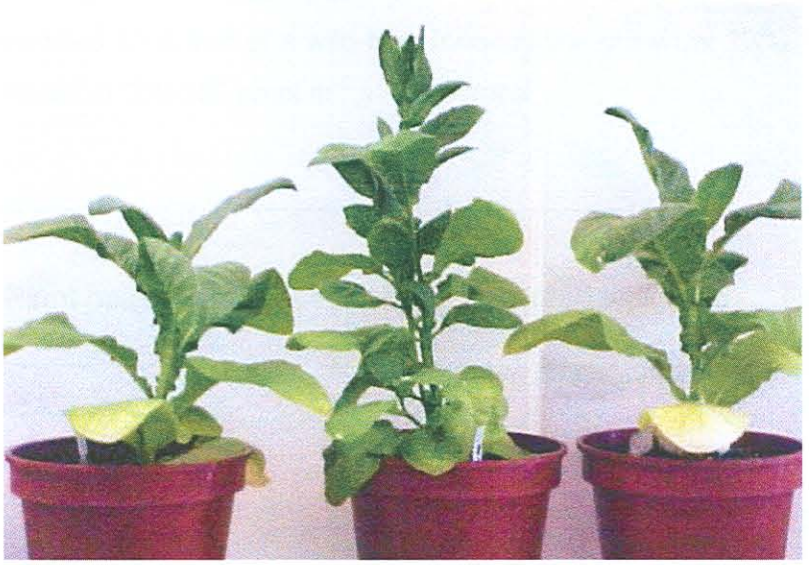

Figure 2.4: Growth of plants of genetically modified lines T4/3-1 and T4/5 and of a wild-type plant (Wt) grown for 7 weeks under a high light intensity (900 µmol m^{-2} s^{-1}) (Van der Vyver et al., 2000).

The leaves of the wild-type expanded more rapidly under low light conditions than those of the *OC-I* expressing plants after 7 weeks (Table 2.1). Total leaf area was thus slightly less in *OC-I* expressing lines than in the wild-type at 7 weeks. At 7 weeks the *OC-I* expressing plants had also lower total biomass (fresh and dry weight) than the wild-type (Table 2.1). After 12 weeks growth at a lower light intensity (300-350 µmol m^{-2} s^{-1}) the stems of the *OC-I* expressing plants were still much shorter (55% - 70%) and thicker than the stems of the wild-type plants although all lines had similar leaf numbers.

Table 2.1: Height, leaf area, fresh and dry weights of plants of individual genetically modified lines and of a wild-type tobacco line grown at 20°C for 7 weeks in the greenhouse at 300-350 μmol m^{-2} s^{-1} irradiance.

Line	Plant height (cm)	Leaf area (cm^2)	Fresh weight (g)	Dry weight (g)
Wild-type	51.2±4.8	2400±166	110±3	16.8±1.2
T4/5	12.0±0.8	2100±83	96±7	13.0±1.1
T4/3-1	11.6±1.2	2166±167	95±5	12.1±1.1
T4/3-2	15.6±0.4	2116±103	103±4	12.4±1.2

In each case values represent the mean ± standard error of 3 different plants.

D) Photosynthesis and respiration – Rothamsted Research

Photosynthesis was lower ($p \geq 0.05$) in *OC-I*-expressing tobacco plants grown at 20°C with an irradiance of 300-350 μmol m^{-2} s^{-1} for 7 weeks, than in the wild-type (Table 2.2). Maximal rates of photosynthesis (A_{max}; 18.1±0.2μmol CO_2 m^{-2} s^{-1}) were decreased as a result of *OC-I* expression (Table 2.2). All *OC-I*-expressing plants had significantly lower ($p \leq 0.05$) rates of CO_2 assimilation than wild-type plants. The inhibitory effect was greatest (11.0±0.6 μmol CO_2 $m^{-2}s^{-1}$) in transformed plants of line T4/5, which had also the highest level of *OC-I* protein. However, the apparent quantum efficiencies of photosynthesis (AQE) were similar in all lines. This observation indicates that while the absolute amount of photosynthetic machinery is decreased in the *OC-I*-expressing plants there is no photoinhibition (Table 2.2).

Dark respiration rates varied between the plants of the different lines (Table 2.2). Leaves from line T4/3-1 had similar rates of dark respiration to those of wild-type plants. Respiration rates in T4/5 plants with highest *OC-I* expression were half ($p \leq 0.05$) of those of the wild-type. Moreover, respiration was significantly higher in transformed plants of line T4/3-2 ($p \leq 0.05$; Table 2.2).

Table 2.2: Apparent quantum efficiencies (AQE), photosynthetic CO_2 assimilation rates (A_{max}) and dark respiration rates (R) in leaves from genetically modified and wild-type plants grown at 20°C in the greenhouse at 300-350 μmol m^{-2} s^{-1} irradiance.

Line	AQE (mol $[CO_2]$ mol^{-1}[light]) x 10^{-2}	A_{max} (μmol $[CO_2]$ $m^{-2}s^{-1}$)	R (μmol $[CO_2]$ $m^{-2}s^{-1}$)
Wild-type	3.56±0.36	18.1±0.2	0.82±0.10
T4/5	2.92±0.24	11.0±0.6	0.44±0.21
T4/3-1	2.88±0.16	15.6±1.4	0.89±0.19
T4/3-2	3.12±0.24	15.8±1.8	1.34±0.09

In each case values represent the mean ± standard error of samples of 6 different plants. The statistical significance of the difference between the mean values was determined by the Student's two-tailed *t* test and P values ≤ 0.05 were considered significant.

E) OC-I expression protects against chilling stress – Rothamsted Research

Photosynthetic CO_2 assimilation was decreased following exposure to 5°C for two days, with all plants showing lower A_{max} (65-80% lower; Figure 2.6) and AQE (20-75% lower; Figure 2.5) values. However, the decline in AQE in two *OC-I*-expressing lines (T4/3-1 and T4/3-2) was significantly less (20% and 31% respectively; $p \leq 0.05$) than that measured in wild-type plants in which AQE declined by 75% (Figure 2.5A). While the low-temperature-induced decline in AQE (65%) was also less in T4/5 plants (with highest *OC-I* expression), this was not significantly different ($p \geq 0.05$) to that observed in the wild-type. In all cases, low-temperature-induced changes in measured values for photochemical and non-photochemical quenching of chlorophyll *a* fluorescence were in agreement with the changes in AQE (data not shown). Following the 2 days exposure to low growth temperatures, plants were allowed to recover at 20°C for 2 days (Figure 2.5B). After 2 days recovery, AQE had returned to values measured in non-chilled plants in all lines (Figure 2.5B). There were no significant differences in AQE between plants of the different lines ($p \geq 0.05$) in the recovery phase (Figure 2.5B). A_{max} was reduced by 84% in wild-type plants after two days at 5°C (Figure 2.6A). The chilling-induced decrease in A_{max} was, however, less in the *OC-I*-expressing lines. Compared to values measured at 20°C, A_{max} was decreased by 72% in plants of lines T4/3-1 and T4/5 and by 65% in line T4/3-2 (Figure 2.6A). Two days after return to 20°C, A_{max} had recovered to 70% of the original values in line T4/3-1 and in wild-type plants. A trend to higher recoveries was observed in plants of lines T4/3-2 (78%) and T4/5 (85%), but A_{max} being significantly higher in line T4/5 with the highest *OC-I*-expression ($p \leq 0.05$; Figure 2.6B).

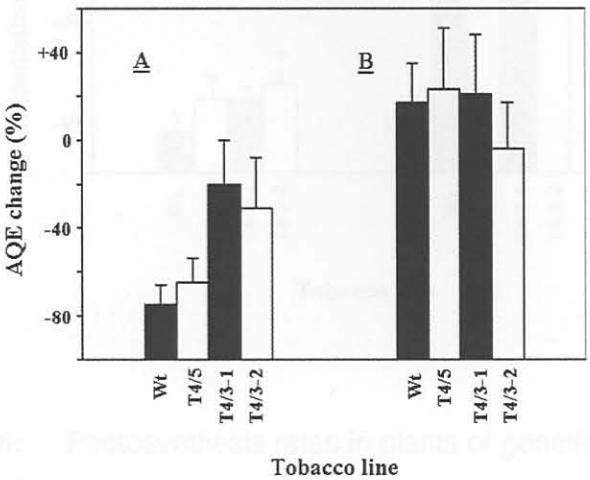

Figure 2.5: Effect of chilling on photosynthesis AQE in plants of genetically modified lines T4/5, T4/3-1 and T4/3-2 and a non-modified wild-type line (Wt). All lines were grown at low light (300-350 μmol m^{-2} s^{-1}). Measurements were made after 2 days at 5°C (A) and after a subsequent 2 days recovery at 20°C (B). Values are expressed as percentages of those measured before cold treatment. These were 3.56±0.36 mol [CO_2] mol^{-1}[light] x 10^{-2} for control plants, 2.92±0.24 mol [CO_2] mol^{-1}[light] x 10^{-2} for T4/5, 2.88±0.16 mol [CO_2] mol^{-1}[light] x 10^{-2} for T4/3-1 and 3.12±0.24 mol [CO_2] mol^{-1}[light] x 10^{-2} for T4/3-2. In each case values represent the mean ± standard error of leaves of 6 different plants.

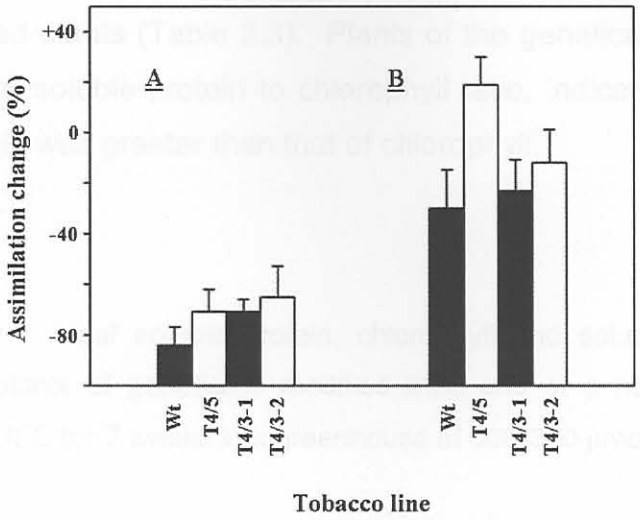

Figure 2.6: Photosynthesis rates in plants of genetically modified lines T4/5, T4/3-1 and T4/3-2 and in a non-modified wild-type line (Wt). In all cases plants were grown under low light (300-350 µmol m^{-2} s^{-1}). Measurements were made after 2 days at 5°C (A) and after a subsequent 2 days recovery at 20°C (B). Values are expressed as percentages of those measured before cold treatment. These were 18.1±0.2 µmol [CO_2] $m^{-2}s^{-1}$ for controls, 11.0±0.6 µmol [CO_2] $m^{-2}s^{-1}$ for T4/5, 15.6±1.4 µmol [CO_2] $m^{-2}s^{-1}$ for T4/3-1 and 15.8±1.8 µmol [CO_2] $m^{-2}s^{-1}$ for T4/3-2. In each case values represent the mean ± standard error of leaves of 6 different plants.

F) Protein and chlorophyll content – Rothamsted Research

The study carried out at Rothamsted also showed that leaves of plants of all genetically modified lines grown under low light intensity had a significantly higher soluble protein content ($p \leq 0.05$) than leaves of wild-type plants under the same growth conditions. Plants from all genetically modified lines T4/5, T4/3-1 and T4/3-2 had between 1.37 and 1.48 times as much soluble protein in their leaves as non-engineered plants and there was no significant difference between the engineered lines ($p \geq 0.05$) (Table 2.3). Plants of the engineered lines also contained significantly higher leaf chlorophyll contents ($p \leq 0.05$) than wild-type plants. Leaves from lines T4/5, T4/3-1 and T4/3-2 had

between 1.17 and 1.32-times as much chlorophyll as the leaves of non-engineered plants (Table 2.3). Plants of the genetically modified lines had a higher leaf soluble protein to chlorophyll ratio, indicating that the increase in leaf protein was greater than that of chlorophyll.

Table 2.3: Leaf soluble protein, chlorophyll and soluble protein to chlorophyll ratios of plants of genetically modified lines and of a non-modified wild-type line grown at 20°C for 7 weeks in a greenhouse at 300-350 μmol m^{-2} s^{-1} irradiance.

Line	*Protein* (μg cm^{-2})	*Chlorophyll* (μg cm^{-2})	*Ratio*
Wild-type	348±24	46±4	7.6
T4/5	*480±56*	*57±5*	*8.4*
T4/3-1	492±60	61±3	8.1
T4/3-2	516±20	54±2	9.6

In each case values for protein and chlorophyll represent the mean ± standard error of samples from 6 different plants. The statistical significance of the difference between the mean values was determined by the Student's two-tailed *t* test and P values ≤ 0.05 were considered significant.

G) Growth of OC-I expressing and wild-type tobacco under low temperature.

When wild-type control plants and plants of genetically modified lines were grown for extended time periods *in vitro* on MS medium and exposed for 4 weeks to 10°C at a light intensity of 50 μmol m^{-2} s^{-1}, the fresh weight doubled.

But no significant differences were observed between the lines ($p>0.05$; Figure 2.7A). After 4 weeks at low temperature, however, the wild-type plants developed necrosis on the expanding leaves (Figure 2.8). This was not observed in plants of the genetically modified lines expressing *OC-I*. In these lines the leaves did not expand but stayed green and compact.

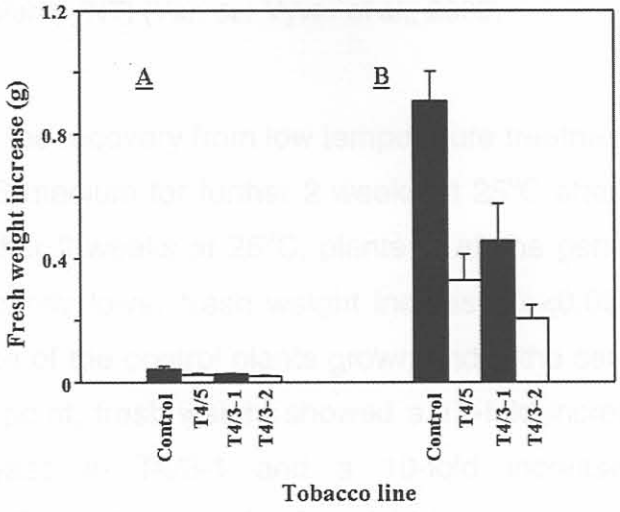

Figure 2.7: Influence of constitutive *OC*-I expression on plant growth. The increase in growth (g fresh weight) in plants of genetically modified lines T4/5, T4/3-1 and T4/3-2 and in wild-type control plants was measured in plants grown in culture at 50 µmol m^{-2} s^{-1} light intensity. They were grown for 4 weeks at 10°C (A) and then for 2 weeks at 25°C (B). At the beginning of the experiment the average fresh weight of individual plants from each line was about 12 mg. At this stage each plant was 2 cm in height. In each case values represent the mean ± standard error of 5 different plants.

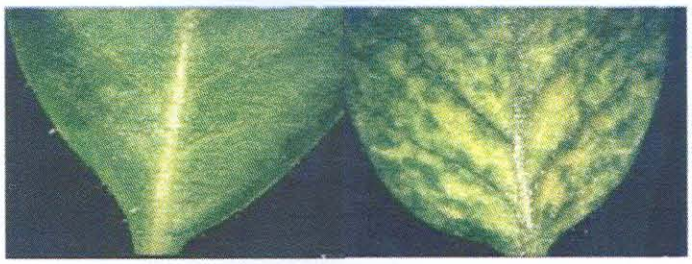

Figure 2.8: Chilling-induced chlorosis in leaves of genetically modified plants (T) and wild-type plants (NT) (Van der Vyver et al., 2000).

To determine the recovery from low temperature treatment, plants were grown *in vitro* on MS medium for further 2 weeks at 25°C after the low temperature treatment. After 2 weeks at 25°C, plants of all the genetically modified lines had a significantly lower fresh weight increase ($p \leq 0.05$) and remained only half of the size of the control plants grown under the same conditions (Figure 2.9). At this point, fresh weight showed a 12-fold increase in T4/5 plants, a 17-fold increase in T4/3-1 and a 10-fold increase in T4/3-2 plants, respectively. By contrast the fresh weight had increased 22-fold in wild-type plants (Figure 2.7B).

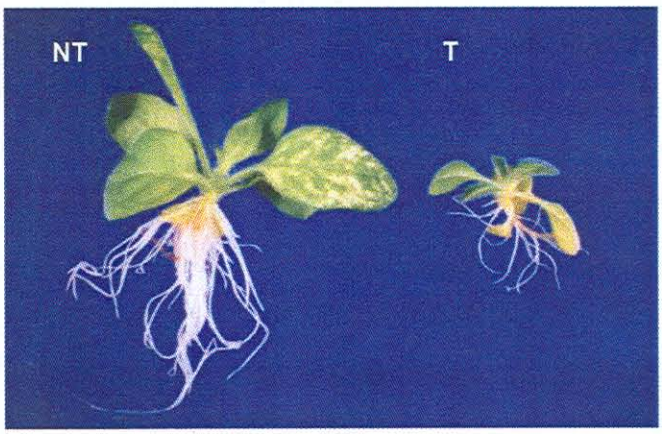

Figure 2.9: Growth of wild-type plants (NT) and genetically modified tobacco plants (T) at 10°C for 4 weeks followed by a further 2 weeks at 25°C (Van der Vyver et al., 2000).

Discussion

Several genetically modified plants expressing the proteinase inhibitor *OC-I* in the cytosol had phenotypic changes when compared with wild-type tobacco plants. Results provided evidence that under sub-optimal growth conditions modified plants could overcome chilling through improved stress acclimation. Photosynthesis was better protected against chilling-induced photoinhibition (determined by AQE) in these plants when compared to wild-type plants.

Constitutive *OC-I* expression also facilitated better post-chilling recovery of CO_2 assimilation and rates of photosynthesis were higher in the recovery phase than before cold treatment. In addition, modified plants had also less chlorophyll degradation after long-term chilling conditions than wild-type plants. This suggests that cystatin expression has the potential to increase abiotic stress tolerance by better recovery of the photosynthetic apparatus and limiting cell degradation after chilling in genetically modified plants.

A fine line seems, however, to exist between possible disadvantages and benefits attained from constitutive *OC-I* expression. Although highest recovery from chilling was observed in modified plants of line T4/5 with the highest *OC-I* expression, this line had both the lowest CO_2 assimilation rate and the lowest dark respiration rate under greenhouse conditions. Under these conditions constitutive *OC*-I expression tend to have negative rather than positive effects on photosynthesis. One possible major source of negative interference by *OC-I* could be the modification of the protein turnover rates in the cytosol. This could have major consequences for proteins whose rapid turnover is essential for the maintenance of the photosynthetic apparatus. These effects could be an advantage at sub-optimal low temperatures but not at an optimal growth temperature and such features would, therefore, have severe implications if a high cystatin expression were, for example, to be exploited for optimal pest control in crop species.

Unclear is why *OC-I* expression in the cytosol has such a profound effect on photosynthesis in the chloroplast. However, most chloroplast proteins are synthesised in the cytosol and then transported across the chloroplast envelope. This includes, for example, the small subunit of Rubisco and also chaperons required for Rubisco assembly. Also there is evidence that cysteine proteinases are expressed during chilling (Schaffer and Fischer, 1988). While the chloroplast itself has its own complement of proteinases that degrade proteins such as D1, degradation of other chloroplast proteins may be shared between different compartments of the photosynthetic cell. To date, cysteine proteinases have been found in the vacuole and recently evidence was provided that ER bodies appear to be a novel proteinase-storing system that assists in cell death under stressed conditions (Hayashi et al., 2001). Pompe-Novak et al. (2002) also found the potato leaf cysteine proteinase, PLCP-2, in protein bodies in the vacuoles, cytoplasm and in cell walls of shoots tips, leaves, stems and root tips presumably involved in organogenesis. But their location in other parts of the cell, such as the chloroplast, remains to be elucidated.

Growth characteristics were also significantly changed in genetically modified plants carrying *OC-I* when compared to wild-type plants. When plants were grown in the greenhouse with supplemented lighting of relative low intensity, which promotes rapid stem elongation in plants, a significant inhibition of plant elongation was observed in the modified plants irrespectively of the expression level of *OC-I*. However, while elongation was significantly affected, plants of the modified lines and the wild-type line had no significant differences in total leaf areas and biomass (dry weight), leaf number, and fresh weight. Leaf chlorophyll contents increased in modified plants but this did not modify photosynthesis, expressed on a surface area basis, grown under artificial low light in the greenhouse at 20°C, which is similar to growth in shaded environments or within crowded plant communities (Ballaré et al., 1994).

The "dwarf" phenotype, which has not been reported by other investigators studying cystatin expression in plants (Masoud et al., 1993; Leple et al., 1995)

was, however, dependent on the light environment in which the modified plants were grown. Differences in stem elongation were much less obvious when genetically modified plants were grown at a much higher light intensity or under natural light conditions during the summer at 25°C, which does not induce a rapid stem elongation. From this study the possibility could not be excluded that the transformation process itself contributed to altered elongation characteristics and not the transgene expression as changes in growth characteristics were inherent in all genetically modified plants tested irrespectively of the expression level of the transgene. For example, reduced growth and delayed flowering has also been reported in the tobacco cultivar *Samsun* expressing *gus* (Caligari et al., 1993). The occurrence of such genomic changes in transgenic plants produced by *Agrobacterium tumefaciens* infection has also been recently verified in rice by molecular tools (Labra et al., 2001). To identify possible genome differences between modified and wild-type plants, a detailed genetic analysis on genetically modified plants was therefore conducted after this initial physiological and biochemical characterization. This characterization study is outlined in the following chapters.

References

Abe K., Kondo H. and Arai S. 1987. Purification and characterization of a rice cysteine proteinase inhibitor. Agricultural and Biological Chemistry 51: 2763-2768.

Abe M., Abe K., Kuroda M. and Arai S. 1992. Corn kernel cysteine proteinase inhibitor as a novel cystatin superfamily member of plant origin: molecular cloning and expression studies. European Journal of Biochemistry 209: 933-937.

Ballaré C.L., Scopel A.L., Jordan E.T. and Viersta R.D. 1994. Signalling among neighboring plants and the development of size inequalities in plant populations. Proceedings of the National Academy of Science, U.S.A. 91: 10094-10098.

Botella M.A., Xu Y., Prabha T.N., Zhao Y., Narasimhan M.L., Wilson K.A., Nielsen S.S., Bressan R.A. and Hasegawa P.M. 1996. Differential expression of soybean cysteine proteinase inhibitor genes during development and in response to wounding and methyl jasmonate. Plant Physiology 112: 1201-1210.

Buchanan-Wollastan V. and Ainsworth C. 1997. Leaf senescence in Brassica napus: cloning of senescence related genes by subtractive hybridization. Plant Molecular Biology 33: 821-834.

Caligari P.D.S., Yapabandara Y.M.H.B., Paul E.M., Perret J., Roger P. and Dunwell J.M. 1993. Field performance of derived generations of transgenic tobacco. Theoretical and Applied Genetics 86: 875-879.

Callis J. 1995. Regulation of protein degradation. Plant Cell 7: 845-857.

Guerrero C., de la Calle M., Reid M.S. and Valpuesta V. 1998. Analysis of the expression of two thioprotease genes from daylily (Hermerocallis spp.) during flower senescence. Plant Molecular Biology 36: 565-571.

Hadfield K.A. and Bennett A.B. 1997. Programmed senescence of plant organs. Cell Death and Differentiation 4: 662-670.

Hayashi Y., Yamada K., Shimada T., Matsushima R., Nishizawa N.K., Nishimura M. and Hara-Nishimura I. 2001. A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of Arabidopsis. Plant and Cell Physiology 42: 894-899.

Jacinto T., Fernandes K.V.S., Machado O.L.T. and Siqueirajunior C.L. 1998. Leaves of transgenic tomato plants overexpressing prosystemin accumulate high levels of cystatin. Plant Science 138: 35-42.

Labra M., Savini C., Bracale M., Pelucchi N., Colombo L., Bardini M. and Sala F. 2001. Genomic changes in transgenic rice (*Oryza sativa* L.) plants produced by infecting calli with *Agrobacterium tumefaciens*. Plant Cell Reports 20: 325-330.

Leple J.C., Bonadebottino M., Augustin S., Pilate G., Letan V.D., Delplanque A., Cornu D. and Jouanin L. 1995. Toxicity to *Chrysomela tremulae* (Coleoptera, Chrysomelidae) of transgenic poplars expressing a cysteine proteinase inhibitor. Molecular Breeding 1: 319-328.

Linthorst H.J.M., Van der Does C., Brederode F.T. and Bol J.F. 1993. Circadian expression and induction by wounding of tobacco genes for cysteine proteinase. Plant Molecular Biology 21: 685-694.

Masoud S.A., Johnson L.B., White F.F. and Reeck G.R. 1993. Expression of a cysteine proteinase inhibitor (oryzacystatin-I) in transgenic tobacco plants. Plant Molecular Biology 21: 655-663.

Penell R.I. and Lamb C. 1997. Programmed cell death in plants. Plant Cell 9: 1157-1168.

Pernas M., Sanchez-Mong R. and Salcedo G. 2000. Biotic and abiotic stress can induce cystatin expression in chestnut. FEBS Letters 467: 206-210.

Pompe-Novak M., Poljsak-Prijatelj M., Popovic T., Strukelj B. and Ravnikar M. 2002. The impact of potato cysteine proteinases in plant growth and development. Physiological and Molecular Plant Pathology 60: 71-78.

Schaffer M.A. and Fischer R.L. 1988. Analysis of mRNAs that accumulate in response to low temperature identifies a thiol protease gene in tomato. Plant Physiology 87: 431-436.

Solomon M., Bellenghi B., Delledonne M., Menachem E. and Levine A. 1999. The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. Plant Cell 11: 431-443.

Urwin P.E., Troth K.M., Zubko E.I. and Atkinson H.J. 2001. Effective transgenic resistance to *Globodera pallida* in potato field trials. Molecular Breeding 8: 95-101.

Van der Vyver C., Kunert K.J. and Whitehead C. 2000. Cystatin over expression and cold tolerance in plants. South African Association of Botanists, Potchefstroom.

Van der Vyver C., Schneidereit J., Driscoll S., Turner J., Kunert K.J. and Foyer C.H. 2002. Oryzacystatin I expression in transformed tobacco produces a conditional growth phenotype and enhances cold tolerance. Plant Biotechnology, in press.

Womack J.S., Randall J. and Kemp J.D. 2000. Identification of a signal peptide for oryzacystatin-I. Planta 210: 844-847.

Xu F.X. and Chye M.L. 1999. Expression of cysteine proteinase during developmental events associated with programmed cell death in brinjal. Plant Journal 17: 321-327.

Chapter 3: Detection of genome modifications in genetically modified tobacco plants using representational difference analysis.

Abstract

Representational Difference Analysis was applied to identify and characterize genome differences in genetically modified tobacco plants. The plant material used was wild-type tobacco of the variety '*Samsun/UK*' and genetically modified '*Samsun/UK*' in which the rice cystatin gene had been introduced by *Agrobacterium tumefaciens*-mediated transformation. RDA was performed using the combined DNA from two independent transformants as the tester and wild-type, non-regenerated 'Samsun' as the driver. Three different DNA sequences were isolated from the two genetically modified plants representing possible DNA sequence differences between the genomes of genetically modified and a non-modified wild-type tobacco plant. Two subtraction products from a subtraction using the methylation-sensitive restriction enzyme *Hpa*II were characterized. One had 96% homology to part of the tobacco chloroplast 23S and 4.5S rRNA genes and the other a 99% homology to the 18S rRNA gene of tobacco. A third DNA subtraction sequence derived from *Hind*III-digested genomic DNA had no significant homology to any already reported DNA sequences. Results indicated that the genetic modification process both alters genome methylation and also causes minor sequence alterations.

Introduction

Introducing foreign DNA molecules into the plant genome using plant tissue culture is a stressful event and carries the risk of genomic variation (Lee and Phillips, 1987; Brown et al., 1990; Phillips et al., 1994; Cullis, 1999). Phenotypic changes in genetically modified plants are well documented

(Lynch et al., 1995; Bao et al., 1996) including altered growth characteristics as found for genetically modified tobacco expressing the *gus* gene selfed over several generations (Caligari et al., 1993). In addition, a high frequency of cytogenetic aberrations in transgenic oat, has also been reported (Choi et al., 2000). Molecular tools have further recently verified the occurrence of transformation-specific genomic changes in several genetically modified plants (Sala et al., 2000; Labra et al., 2001). AFLP (amplified fragment length polymorphisms) and RAPD (random amplified polymorphic DNAs) analysis of genetically modified and non-modified rice plants using the *Agrobacterium tumefaciens* system for plant transformation has demonstrated genomic homogeneity among non-modified rice plants and verified genomic changes within the genetically modified plants (Labra et al., 2001). Genome variation is in addition to any changes arising from T-DNA integration and has been shown to cause complex rearrangements during T-DNA/transgene integration. Modifications found around T-DNA/plant DNA junctions include base substitutions, duplication of border and genomic sequences and small nucleotide deletions (Ohba et al., 1995; Windels et al., 2001; Zheng et al., 2001; Stahl et al., 2002). Strategies to eliminate such variation include the selection of the most "normal" plants, which will be used in subsequent crossings to obtain a plant, which is morphologically, but not necessarily genetically, identical to the mother plant (Labra et al., 2001).

In order to assess possible genetic changes in the genome of genetically modified tobacco plants other than around junction sites, the technique of representational difference analysis (RDA) was used in this study. RDA is a DNA subtractive technology and can be applied to derive probes for genomic losses, rearrangements, amplifications, point mutations and pathogenic organisms found within any of two genomes to be compared (Lisitsyn et al., 1993; Ushijima et al., 1997; Michiels et al., 1998). Although the RDA technique was first reported in 1993 and has been widely applied with cDNAs (Hubank and Schatz, 2000), the technique has been applied only on a limited scale with genomic DNAs. In RDA, target DNA fragments are sequentially enriched by favorable hybridization kinetics and subsequently amplified by PCR. The technique is based on altering the ends of the DNA sequences and

amplifying certain combinations, which results in a steady depletion of common sequences. Depletion is carried out by hybridization with increasingly large amounts of driver DNA (the source of the sequences to be eliminated) and with reducing amounts of the tester DNA (the source of the sequences of interest to be isolated). After each round of hybridization and amplification, only those sequences, which have formed a double stranded DNA fragment, with both strands arising from the tester DNA, will be amplified by PCR.

Results

A) Characterization of genetically modified tobacco with RDA.

Non-modified wild-type and genetically modified tobacco plants (*Nicotiana tabacum* L., var 'Samsun') expressing a rice cystatin as described in Annex (A) under "Materials and method" for Section I: "Plant transformation and selection" were used for execution of RDA. Genome subtractions between genetically modified and wild-type tobacco plants were performed with pooled DNA of two genetically modified plants mixed in a 1:1 ratio used as tester and wild-type plant genomic DNA as driver (experiment 1 and 3) and *vice versa* (experiment 2 and 4). Figures 3.1A and 3.2A shows DNA profiles after electrophoretic separation on an agarose gel of genomic DNA from modified and wild-type tobacco DNA digested with the restriction enzymes *Hpa*II and *Hind*III, respectively. Several subtraction products were produced for experiments 1 and 3 (modified DNA as tester and wild-type DNA as driver) after applying three rounds of subtractive hybridization and amplification using driver to tester ratios for first, second and third round subtractions of 75:1, 300:1 and 15000:1 for *Hpa*II-digested DNA (Figure 3.1B, C and D) and 3:1, 100:1 and 4500:1 for *Hind*III-digested DNA (Figure 3.2B, C and D). In experiments 2 and 4 (non-modified DNA tester/modified DNA driver) subtraction products were only found for *Hpa*II amplicons (experiment 2).

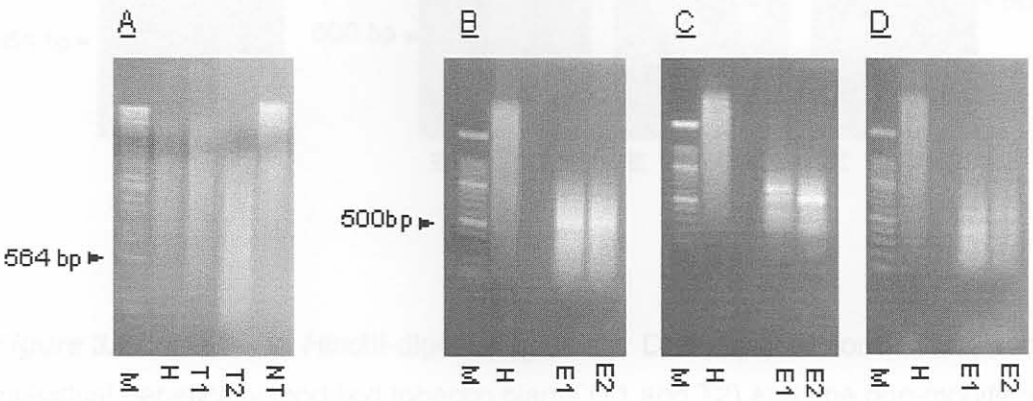

Figure 3.1: RDA of genetically modified and non-modified wild-type tobacco plant DNA. (A) Genomic *Hpa*II-digested DNA after separation on an agarose gel and stained with ethidium bromide from two individual genetically modified tobacco plants (T1 and T2) and from one wild-type tobacco plant (NT). Lane M represents molecular size marker (*Eco*RI/*Hind*III-digested λ DNA) and lane H represents 500ng herring sperm DNA. (B), (C) and (D) PCR amplified subtraction products from experiment 1 and 2 (E1 and E2) generated from *Hpa*II-digested DNA after the first (B), second (C) and third round (D) of subtractions. Lane M represents a 100 bp size marker and lane H represents 1µg of herring sperm DNA.

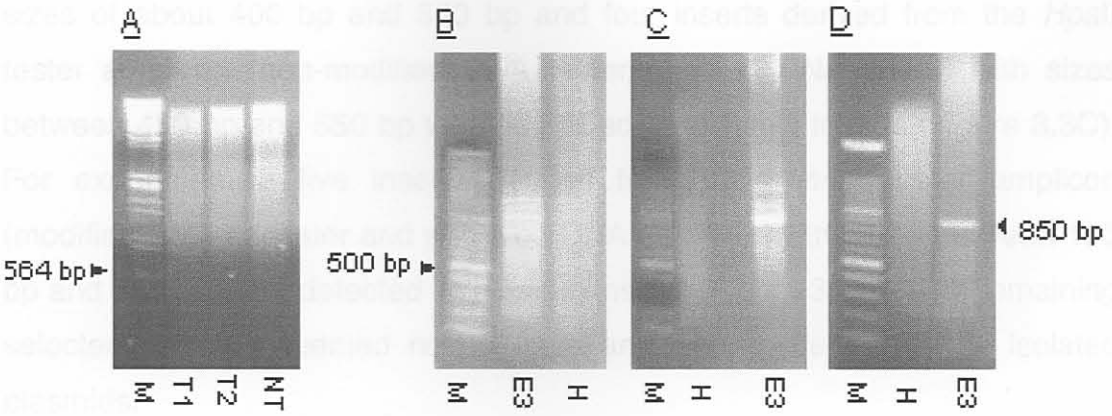

Figure 3.2: RDA on *Hind*III-digested genomic DNA. (A) Genomic DNA from two individual genetically modified tobacco plants (T1 and T2) and one non-modified wild-type plant (NT), digested with the restriction enzyme *Hind*III after separation on an agarose gel, stained with ethidium bromide. Lane M represents molecular size marker (*Eco*RI/*Hind*III-digested λ DNA). (B), (C) and (D) PCR amplified subtraction products from experiment 3 (E3) generated from *Hind*III-digested genomic DNA after the first (B), second (C) and third round (D) of subtractions. Lane M represents a 100 bp DNA marker and lane H represents 1µg of herring sperm DNA.

All pooled amplified third-round *Hpa*II and *Hind*III subtractions were cloned into the cloning vector p*MOSBlue* (Amersham Pharmacia Biotech, UK) and *E. coli* competent cells (Amersham Pharmacia Biotech, UK) were transformed with the plasmid allowing blue/white selection on an ampicillin-containing medium. Sixty white *E. coli* colonies derived from the three transformations were hybridized separately with either labeled *Hpa*II or *Hind*III driver amplicons. From each hybridization, ten colonies with the lowest hybridization signal were selected using the *Gene Images* random prime labeling module (Amersham Pharmacia Biotech, UK) (Figures 3.3 and 3.4). From these colonies plasmids were isolated, which possibly contained putative specific tester DNA subtraction products. Low level hybridization could also be due to hybridization to vector sequences or sections of probes hybridizing to parts of clone inserts. For experiments 1 and 2, two inserts derived from the *Hpa*II tester amplicon (modified DNA as tester and wild-type DNA as driver) with

sizes of about 400 bp and 550 bp and four inserts derived from the *Hpa*II tester amplicon (non-modified DNA tester/modified DNA driver) with sizes between 450 bp and 550 bp were identified as plasmid inserts (Figure 3.3C). For experiment 3 five inserts derived from the *Hind*III tester amplicon (modified DNA as tester and wild-type DNA as driver) with sizes between 150 bp and 400 bp were detected as plasmid inserts (Figure 3.4B). The remaining selected colonies seemed not to have any inserts present in the isolated plasmids.

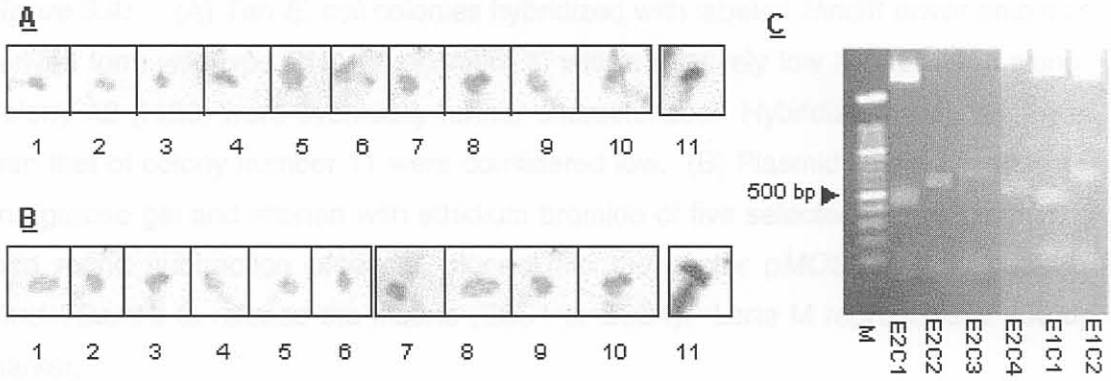

Figure 3.3: (A and B) Ten *E. coli* colonies each for the two transformations hybridized separately with labeled *Hpa*II driver amplicon derived either from wild-type DNA (A; experiment 1) or genetically modified DNA (B; experiment 2) with a relatively low hybridization signal. Hybridization signals lower than that of colony number 11 were considered low. Colonies A7 and B5 (Hp12 and Hp14) were the final clones, which were eventually further characterized. (C) Third round subtraction products cloned into the cloning vector P*MOSBlue* and cut with *Hind*III/*Bam*HI to release the cloned inserts E1C1 and E1C2 derived from experiment 1 and E2C1, E2C2, E2C3 and E2C4 derived from experiment 2. Cloned inserts were separated on an agarose gel stained with ethidium bromide. Lane M represents a 100 bp DNA marker.

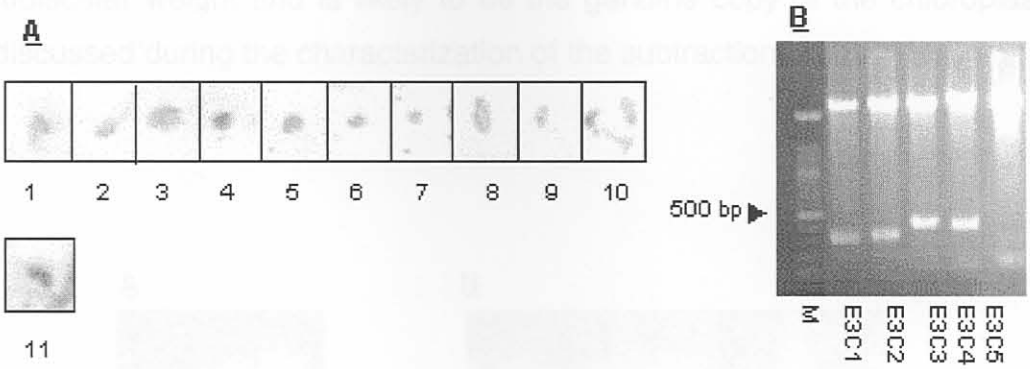

Figure 3.4: (A) Ten *E. coli* colonies hybridized with labeled *Hind*III driver amplicon derived form wild-type DNA (experiment 3) with a relatively low hybridization signal. Colony A2 (Hi30) were eventually further characterized. Hybridization signals lower than that of colony number 11 were considered low. (B) Plasmid DNA separated on an agarose gel and stained with ethidium bromide of five selected clones containing third round subtraction products, cloned into the vector p*MOSBlue* and cut with *Hind*III/*Bam*HI to release the inserts (E3C1 to E3C4). Lane M represents a 100 bp marker.

DNAs of the different rounds of subtraction were probed against all cloned and labeled final amplified subtraction products where pooled genetically modified DNA acted as tester DNA and wild-type tobacco DNA as driver. Three final subtraction products, Hp12, Hp14 and Hi30, did not hybridize with the original driver DNA amplicons derived from amplification of *Hpa*II or *Hind*III-digested genomic DNAs (Figures 3.5 and 3.6). Southern blot hybridization with other labeled isolated clones showed that these clones were not unique to the tester amplicon but were also present in the driver amplicon, indicating possible incomplete subtraction allowing background sequences through or possible size differences resulted in the enrichment of certain DNA fragments. All three final subtraction products hybridized with amplified DNAs of first, second and third round subtractions and very weakly with tester DNA amplicons (Figures 3.5 and 3.6). However, subtraction product Hp12 also

seemed to be related to the driver amplicon. The driver band is at a different molecular weight and is likely to be the genuine copy of the chloroplast (as discussed during the characterization of the subtraction product).

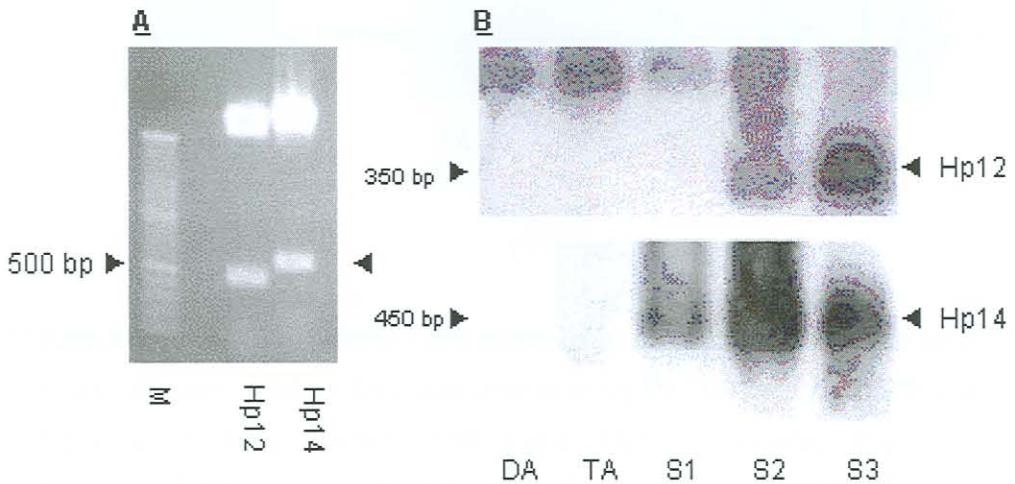

Figure 3.5: (A) Third round amplified Hp12 and Hp14 subtraction products separated on an agarose gel and stained with ethidium bromide after cloning into plasmid p*MOSBlue* and release of cloned product by a *BamH*I/*EcoR*I digest of isolated plasmid. Digestion of plasmid with *BamH*I/*EcoR*I added an additional 50 bp of cloning vector to the insert size. Arrow indicates position of Hp12 and Hp14 on the gel. Lane M represents a 100 bp marker. (B) Hybridization of subtraction products Hp12 (upper part) and Hp14 (lower part) to a filter containing *Hpa*II-digested and amplified DNAs derived from a wild-type tobacco plant (DA/driver amplicon), two pooled DNAs of genetically modified tobacco plants (TA/tester amplicon) and amplified subtraction products after first (S1), second (S2), and third round (S3) of subtraction.

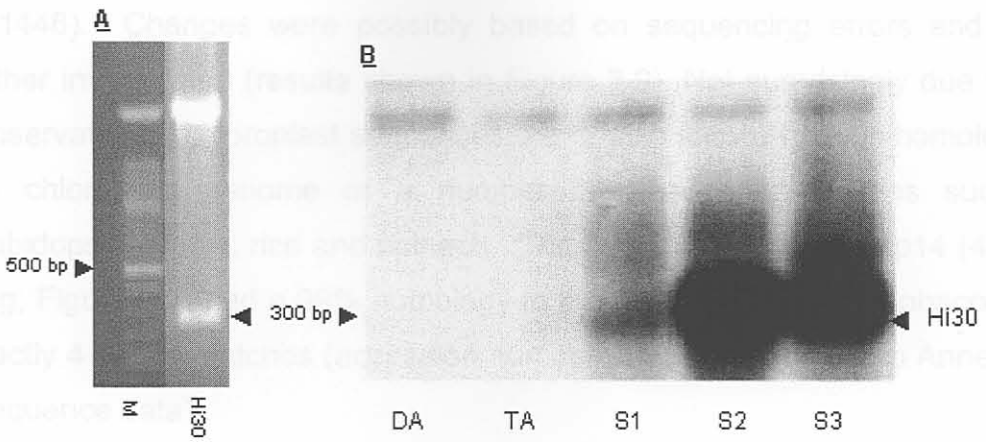

Figure 3.6: (A) Third round Hi30 subtraction product separated on an agarose gel and stained with ethidium bromide after cloning into plasmid p*MOSBlue* and release of cloned product by a *BamH*I/*EcoR*I digest of isolated plasmid. Digestion of plasmid with *BamH*I/*EcoR*I added an additional 50 bp of cloning vector to the insert size. Arrow indicates position of Hi30 on the gel. Lane M represents a 100 bp marker. (B) Hybridization of Hi30 to a filter containing *Hind*III-digested and amplified DNAs derived from a wild-type tobacco plant (DA/driver amplicon), two pooled DNAs of genetically modified tobacco plants (TA/tester amplicon) and amplified subtraction products after first (S1), second (S2), and third round (S3) of subtraction.

B) Characterization of methylation sensitive and non-sensitive RDA subtraction products.

Hp12, Hp14 and Hi30 were sequenced and obtained sequence was analyzed using the DNA analysis tools Blast, FastA and the Smith-Waterman algorithm. Remaining undigested adaptor sequences present on the subtraction products were not included in the sequence analysis. The subtraction product Hi30 was 214 bp long (Figure 3.6) and had no significant homology to any already reported DNA sequence. In contrast, product Hp12 (331 bp long; Figure 3.5) had a 97% homology to part of the tobacco chloroplast genome DNA, specifically to part of the tobacco chloroplast 23S and 4.5S rRNA genes

with 9 bp changes at different locations (accession numbers Z00044 and J01446). Changes were possibly based on sequencing errors and were further investigated (results shown in Figure 3.9). Not surprisingly due to the conservation of chloroplast sequences, Hp12 also showed a high homology to the chloroplast genome of a number of other plant species such as Arabidopsis, maize, rice and spinach. The subtraction product Hp14 (444 bp long; Figure 3.5) had a 99% homology to the 18S rRNA gene of tobacco with exactly 4 bp mismatches (accession number AJ236016) (see also Annex (B): "Sequence data").

The subtraction products Hp12 and 14 were used to determine the methylation status of these regions in the genetically modified and wild-type genomic DNAs. The two labeled subtraction products were hybridized to *HpaII*-digested genomic DNA derived from either the two original genetically modified tobacco plants or the one wild-type plant used in the experiments. Target detection was optimized regarding signal background ration, to two hours of filter exposure to a hyperfilm. After exposure, two additional larger size hybridization products were identified in wild-type DNA when the subtraction product Hp12 was the probe (Figure 3.7; Hp12/NT). The lower molecular weight band most likely represented the non-methylated plastidic DNA. The hybridization pattern observed with Hp14 gave a relatively larger amount at the higher molecular weight regions of the DNAs from the genetically modified plants (T1 and T2), and a missing band in these same DNAs at the lowest molecular weight site (Figure 3.7).

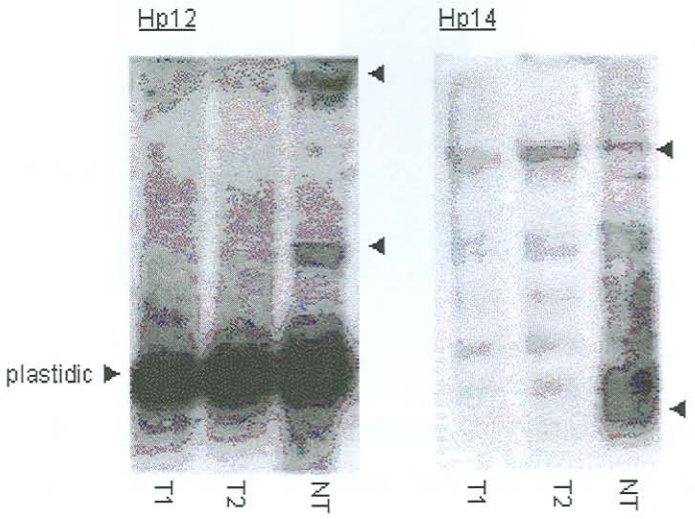

Figure 3.7: Hybridization of final amplified Hp12 and Hp14 subtraction products to a filter containing *Hpa*II-digested genomic DNA derived from two individual genetically modified tobacco plants used in the experiments for pooling DNAs (T1 and T2) and from the non-modified wild-type tobacco plant (NT). Detected band of plastidic DNA with Hp12 is indicated.

In a second experiment, wild-type genomic tobacco plant DNA was digested with *Msp*I to determine possible loss of methylation between modified and non-modified tobacco plants. After exposing the filter for three hours to hyperfilm, a single hybridization product was observed in wild-type plant DNA possibly representing the non-methylated plastidic DNA indicating an decrease in methylation in genetically modified plants when compared to wild-type plants (Figure 3.8).

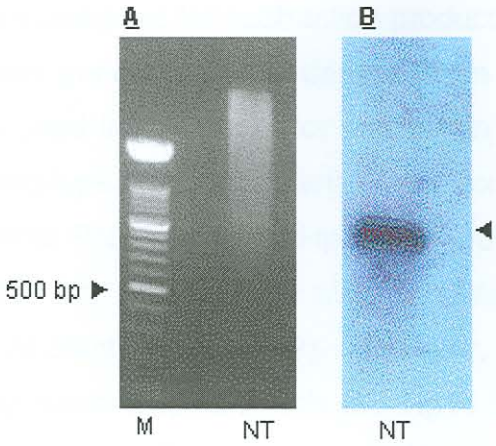

Figure 3.8: (A) Genomic *Msp*I-digested DNA after separation on an agarose gel and stained with ethidium bromide from one non-modified wild-type tobacco plant (NT). Lane M represents a 100 bp marker. (B) Hybridization of final amplified Hp12 subtraction products to a filter containing *Msp*I-digested genomic DNA derived from a non-modified wild-type tobacco plant (NT). Detected band of plastidic DNA with Hp12 is indicated.

The sequence information obtained for Hi30 and Hp12 and 14 was used to design pairs of primers to test for genomic variation in a PCR-based assay. Pairs of primers were designed using a standard design program (Expasy, Switzerland). Primers used to amplify subtraction product Hp12 were Hp12R and Hp12L, Hp14R and Hp14L for product Hp14 and Hi30R and Hi30L for subtraction product Hi30 (Table 3.1).

Table 3.1: Sequences of primers used for PCR amplification.

Name	Sequence
Hp12L	5' CGA GCC AAT GTT CGA ATA CC 3'
Hp12R	5' CCG AAG TTA CGG GGC TAT TT 3'
Hp14L	5' TGT CGG CCA AGG CTA TAA AC 3'
Hp14R	5' TTC CGT TAA CGA ACG AGA CC 3'
Hi30L	5' GGA ATG ATT TCC CAA AAC TCC 3'
Hi30R	5' CGA CGT CGA CTA TCC ATG AAC 3'

Primers designed for subtraction products Hp12 and Hp14 did not differentiate between genomic DNAs derived from either genetically modified tobacco plants used in this study for production of amplicons and DNA derived from the wild-type tobacco plant. An expected band from chloroplast and ribosomal RNA genes for Hp12 and Hp14, with the predicted size of 190 bp and 196 bp, respectively, were amplified out of the genomic DNA of both types of plants (Figure 3.9). However, a second bigger band with a size of 390 bp were amplified with the Hp14 primers from both the wild-type and genetically modified genomic DNAs. Subsequent cloning of amplification products into the vector p*MOSBlue* and sequencing of the amplified products showed that amplification products were identical for genetically modified and wild-type DNA with no base pair changes. The bigger amplification product amplified with the Hp14 primers were a direct repeat of the smaller amplification product found with the same primers. When the sequences of the amplification products were aligned with the two DNA sequences of the subtraction products Hp12 and Hp14, 1 bp and 4 bp mismatches were observed, respectively. However, these sequences aligned 100% with the known sequences of the tobacco chloroplast genome and the tobacco 18S rRNA gene. Although primers designed for subtraction product Hi30 amplified a predicted 199 bp product from genomic DNA derived from the two genetically modified plants, these primers failed to amplify, after 42 cycles of PCR, a fragment from genomic DNA from the wild-type plant (Figure 3.9). This result was confirmed when labeled Hi30 subtraction product was hybridized to a filter containing *Hind*III-digested genomic DNA from the non-modified wild-type tobacco plant. No hybridization products were detected on this filter, indicating the absence of the Hi30 subtraction product in the wild-type plant DNA used for the production of the RDA amplicons (data not showed).

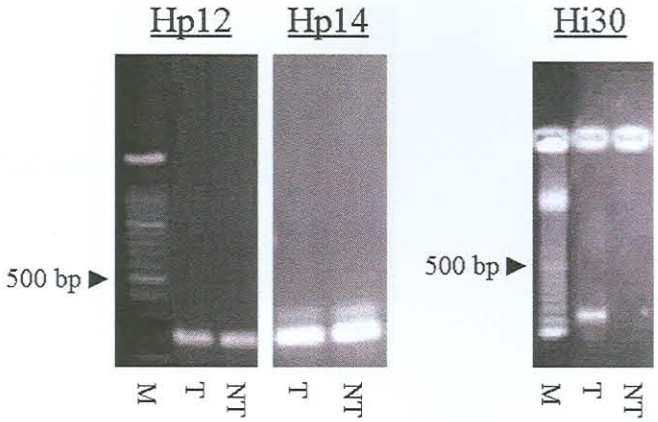

Figure 3.9: PCR amplification of genomic target DNAs with primers designed for Hi30, Hp12 and Hp14 with genomic template DNA derived from pooling DNAs of two genetically modified tobacco plants (T) and genomic DNA derived from the non-modified wild-type tobacco plant (NT). Lane M represents a 100 bp marker.

The Hp12 subtraction product was further characterized by designing the primer pair Hp12A and Hp12B (Table 3.2). Both primers had a single base pair change between the difference product and the known tobacco chloroplast sequence. DNA amplification using PCR with this primer pair resulted in an amplification product from genetically modified genomic DNA (Figure 3.10). In contrast, amplification of the identical product was much weaker in wild-type genomic DNA, which might suggest a base pair change in the primer site in the genetically modified tobacco DNA.

Table 3.2: Sequences of primers used for the characterization of subtraction product Hp12.

Name	Sequence
Hp12A	5' TTGTCTCGCGCCCCTAGGTAC '3
Hp12B	5' TACCAGGCGCTACGGCGCTGG '3
Hp12C	5' CGCAAACATTGGTGAGAATC '3

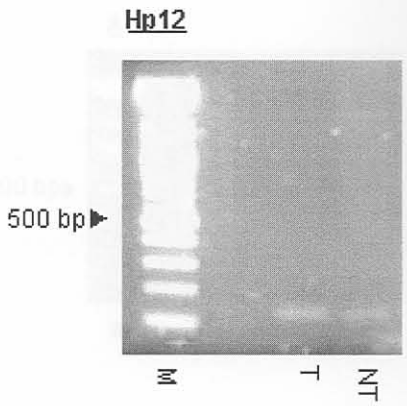

Figure 3.10: PCR amplification of genomic DNA with primer pair Hp12A and Hp12B with genomic template DNA derived from pooling DNAs of two genetically modified tobacco plants (T) and genomic DNA derived from the wild-type tobacco plant (NT). Lane M represents a 100 bp marker.

A third primer, Hp12C, was also designed outside the sequence of the Hp12 difference product. Design of Hp12C was based on available sequence information obtained from known tobacco chloroplast DNA. PCR analysis with primers Hp12C and Hp12R included the region containing the *Hpa*II cutting site present at the end of the Hp12 subtraction product, which are not present in the known tobacco chloroplast sequence. Amplification of both types of genomic DNAs by PCR, with primers Hp12C and Hp12R, followed by digestion of amplified products with the restriction enzyme *Hpa*II, showed almost identical patterns in both genetically modified and wild-type DNA (Figure 3.11A). When compared to undigested amplified DNA, a slight shift of the size of the amplified and *Hpa*II-digested DNA was found after separation on an agarose gel (Figure 3.11A). When amplifications products were transferred to a filter and probed with the Hp12 subtraction product all amplified products hybridized to Hp12 and no difference was detected between modified and wild-type tobacco DNAs (Figure 3.11B). This pattern was the same when the genomic DNAs of both genetically modified and wild-type tobacco plants were first digested with *Hpa*II or *Msp*I followed by amplification with primers Hp12C and Hp12R (Figure 3.12).

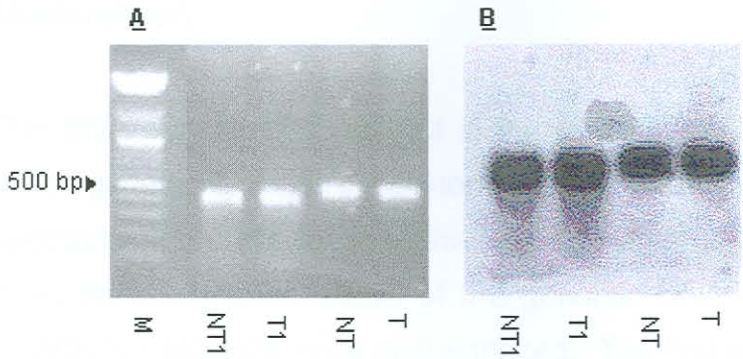

Figure 3.11: (A) PCR amplification of genomic target DNAs with primers Hp12R and Hp12C with genomic template DNA derived from pooling DNAs of two genetically modified tobacco plants (T) and genomic DNA derived from wild-type tobacco plant (NT). Amplified products from modified and wild-type plants were either undigested (NT and T) or digested with restriction enzyme *Hpa*II (NT1 and T1). Lane M represents a 100 bp marker. (B) Hybridization of Hp12 to a filter containing amplified products from A.

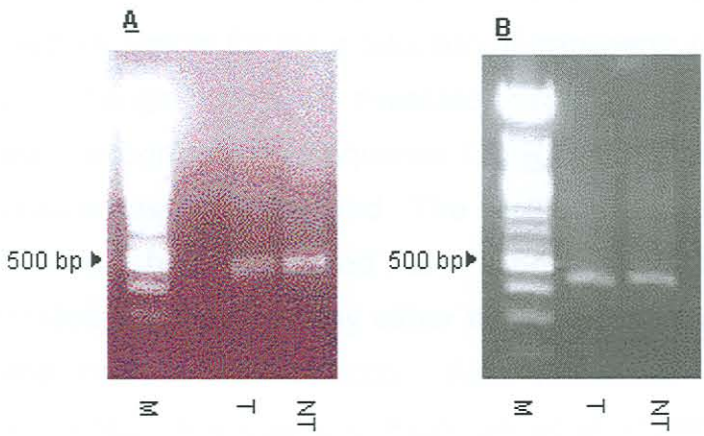

Figure 3.12: PCR amplification of genomic target DNAs with primers Hp12R and Hp12C with genomic template DNA derived from pooling (A) *Msp*I-digested (B) *Hpa*II-digested DNAs of two genetically modified tobacco plants (T) and digested genomic DNA derived from wild-type tobacco plant (NT). Lane M represents a 100 bp marker.

Discussion

The RDA experiments resulted in the isolation of DNA sequences limited to either genetically modified or non-modified wild-type tobacco plants. These subtraction products could be the result of either DNA changes in the genome or in the methylation status of the genomic DNA or both. Three of these subtraction products were characterized. The first subtraction product, Hp14, was isolated from *Hpa*II digestion and was identical to part of the 18S rRNA gene. It is therefore likely that the isolation of Hp14 was due to changes in the methylation patterns between genetically modified and wild-type tobacco plants. Changes in genome methylation due to stressful events including foreign gene insertion into a plant has been previously reported and examination of regenerated plants with methylation-sensitive restriction enzymes has revealed both hyper- and hypo-methylation (Phillips et al., 1994). Since plant tissue culture, as part of the transformation/gene insertion process, can change DNA methylation and also occurs as a consequence of treatment with an antibiotic commonly used in a tissue culture process, as selective agents for the production of transgenic plants (Schmitt et al., 1997), such changes could be expected. The methylation-sensitive endonulease *Hpa*II recognizes the sequence CCGG and does not cleave if the internal cytosine base is methylated. The section of the plant genome detected might therefore have acquired an altered methylation status during the transformation, caused by either the tissue culture process or by the foreign gene insertion into tobacco. Also, ribosomal RNA genes can be highly methylated. For example, Fulnecek et al. (1998) found that the density of methylation along the 5S rRNA genes exceeds the average methylation density in the tobacco genome. In the past is has also been proposed that methylation patterns change in correlation with gene activity. This was found to be true especially in rRNAs where methylation controlled the transcription potential of rRNA genes, which served as a mechanism to direct the number of active rRNAs (Flavell and O'Dell, 1988). When methylation-sensitive RDA was applied to different date palm varieties, some of the isolated difference products showed high homologies to the 18S, 25S and 4.5S rDNA of tobacco

indicating potential changes in methylation patterns of the rDNA in different varieties (J. Vorster, personal communication). Altered rRNA gene methylation can have a dramatic effect on protein expression, subsequently associated with a plant phenotype such as dwarfism (Neves et al., 1995). Such an association might also be present in the genetically modified tobacco characterized in this study, where the modified plants expressed a conditional dwarf phenotype.

In contrast to Hp14, subtraction product Hp12 clearly hybridized to two larger DNA fragments derived from wild-type genomic DNA. These fragments were not found in the genomic DNA derived from genetically modified plants. Since tobacco chloroplast DNA (cpDNA) is among a few well-characterized chloroplast genomes in which sequence of the complete cpDNA has been reported (Shinozaki et al., 1986), Hp12 could be identified being of plastidic origin when searching DNA sequence databases. From the hybridization experiment it was concluded that this subtraction product also seems to be present in the nuclear genome, and that the nuclear DNA copy derived from a genetically modified plant is either less methylated than in the wild-type plant nuclear DNA or is absent from the genetically modified plant.

In plants, methylation is mainly restricted to the nuclear genome, where methyl-cytosine is especially concentrated in repeated sequences (Finnegan et al., 1998). Since Hp12 has a high level of similarity to chloroplast DNA of tobacco with differences only in a few base pairs, Hp12 is possibly a nuclear localized copy (Ayliffe et al., 1998). Chloroplast and nuclear DNA are in separate cellular compartments and DNA generally does not pass in either direction through the membranes of organelles (Lewin, 1994). However, that does not exclude that such exchanges of DNA between organelles and the nucleus have occurred during evolution in endo-symbiotic events with prokaryotic organisms (Leon et al., 1998). Sequence analysis of the Hp12 difference product also showed a small number of base pair changes including changes of C to G and T to A when compared to the known tobacco chloroplast sequence. According to Phillips et al. (1994), such base pair changes represent the most commonly observed point mutations in plants,

which are also a consequence of plant tissue culture procedures using plant growth regulators. Such changes can occur either as a result of deamination of methylated cytosine or loss of precision in the DNA replication repair mechanism. Genomic DNA amplification by PCR using the primer pair Hp12A and Hp12B, which were specifically designed to detect single base pair mutations, resulted indeed in better amplification of genomic DNA of genetically modified plants than of DNA from a wild-type plant. It is therefore likely, that these single mutations have occurred in DNA of genetically modified plants, which consequently resulted in less efficient amplification from the wild-type DNA due to base pair changes in the primer site. When genomic DNA from both genetically modified and non-modified plants were first digested with *Msp*I or *Hpa*II followed by PCR amplification using the primers Hp12C and Hp12R, the same amplification pattern were observed in both types of plants. *Msp*I, is an isoschizomer of *Hpa*II and when the external C in the sequence CCGG is methylated, *Msp*I and *Hpa*II cannot cleave DNA. However, unlike *Hpa*II, *Msp*I can cleave the DNA sequence when the internal C residue is methylated. Since methylation is considered to be restricted to the plant nuclear genome, the non-methylated chloroplast DNA would have been digested with both methylation- sensitive restriction enzymes, *Hpa*II and *Msp*I, preventing amplification of a product due to digestion of DNA. No amplification would also have occurred when nuclear DNA would have been non-methylated at the *Hpa*II site allowing digestion with the enzymes. In contrast, amplification would have occurred when either the *Hpa*II site in the nuclear DNA is absent or the *Hpa*II / *Msp*I site is methylated at the internal and external C preventing digestion with the enzymes. Since amplification of a fragment was found despite digestion of DNA with *Hpa*II and *Msp*I, the *Hpa*II site is seemingly missing in the genomic plant DNA. Since the *Hpa*II site was found in the Hp12 subtraction product but not in the sequence data base of the known tobacco chloroplast genome, the *Hpa*II site present in the subtraction product very likely originates from a mutation in the DNA sequence by a A to C exchange.

Beside two methylation-sensitive products, one non-methylation sensitive subtraction product from genomic DNA of genetically modified tobacco plants

(Hi30) was identified in the study. This subtraction product was absent in genomic DNA of the wild-type tobacco plant used for RDA amplicon production. Hi30 might represent a variable DNA region, which has easily changed under plant transformation conditions. However, a primer pair designed for a subtraction product with a mutated sequence derived from DNA of a genetically modified plant seems not have allowed the amplification of a fragment with DNA of a wild-type plant. Minor DNA variations between individual wild-type plants can not be completely excluded although the original '*Samsun/UK*' cultivar used for this study is both an inbred and very likely a true-breeding cultivar. However, before any generalization about the occurrence and nature of Hi30 in the DNA of wild-type tobacco could be made a more detailed study with a greater number of genetically modified and wild-type plants was carried out, which is outlined in Chapter 5 of this thesis.

References

Ayliffe M.A., Scott N.S. and Timmis J.N. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. Molecular Biology and Evolution 15: 738-745.

Bao P.H., Granata S., Castiglione S., Wang G.J., Giordani C., Cuzzoni E., Damiani G., Bandi C., Datta S.K., Datta K., Potrykus I., Callegarin A. and Sala F. 1996. Evidence for genomic changes in transgenic rice (*Oryza sativa* L.) recovered from protoplasts. Transgenic Research 5: 97-103.

Brown P.T.H., Kyozla J., Suleklyo V., Kimura Y., Shimamoto K., and Lörz H. 1990. Molecular changes in protoplast-derived rice plants. Molecular and General Genetics 223: 324-328.

Caligari P.D.S., Yapabandara Y.M.H.B., Paul E.M., Perret J., Roger P. and Dunwell J.M. 1993. Field performance of derived generations of transgenic tobacco. Theoretical and Applied Genetics 86: 875-879.

Choi H.W., Lemaux P.G. and Cho M.J. 2000. High frequency of cytogenetic aberration in transgenic oat (*Avena sativa* L.) plants. Plant Science 156: 85-94.

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

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

Finnegan E.J., Genger R.K., Peacock W.J. and Dennis E.S. 1998. DNA methylation in plants. Annual Review of Plant Physiology and Plant Molecular Biology 49: 223-247.

Flavell R.B. and O'Dell M. 1988. Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. Journal of Molecular Biology 204: 523-534.

Fulnecek J., Matyasek R., Kovarik A. and Bezdek M. 1998. Mapping of 5-methylcytosine residues in *Nicotiana tabacum* 5S rRNA genes by genomic sequencing. Molecular and General Genetics 259: 133-141.

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

Hubank M. and Schatz D.D. 2000. Representational difference analysis of cDNA. In: Hunt S.P. and Livesey F.J. (eds.), Functional Genomics. Oxford University Press, Oxford New York, p. 45-112.

Labra M., Savini C., Bracale M., Pelucchi N., Colombo L., Bardini M. and Sala F. 2001. Genomic changes in transgenic rice (*Oryza sativa* L.) plants produced by infecting calli with *Agrobacterium tumefaciens*. Plant Cell Reports 20: 325-330.

Lee M. and Phillips R.L. 1987. Genomic rearrangements in maize are induced by tissue culture. Genome 29: 122-128.

Leon P., Arroyo A. and Mackenzie S. 1998. Nuclear control of plastid and mitochondrial development in higher plants. Annual Review of Plant Physiology and Molecular Biology 49: 453-480.

Lewin B. 1994. Genes. Oxford University Press, Oxford, UK.

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

Lisitsyn N.A., Segre J.A., Jusumi K., Lisitsyn N.M., Nadeau J.H., Frankel W.N., Wigler M.H. and Lander E.S. 1994. Direst isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nature Genetics 6: 57-63.

Lynch P.T., Jones J., Blackhall N.W., Davey M.R., Power J.B., Cocking E.C., Nelson M.R., Bigelow D.M., Orum T.V., Orth C.E. and Schuh W. 1995. The phenotypic characterisation of F(2) generation transgenic rice plants under field and glasshouse conditions. Euphytica 85: 395-401.

Michiels L., Van Leuven F., van den Oord J.J., De Wolf-Peeters C. and Delabie J. 1998. Representational difference analysis using minute quantities of DNA. Nucleic Acids Research 26: 3608-3610.

Neves N., Heslop-Harrison J.S. and Viegas W. 1995. rRNA gene activity and control of expression mediated by methylation and imprinting during embryo development in wheat x rye hybrids. Theoretically and Applied Genetics 91: 529-533.

Ohba T., Yoshioka Y., Machida C. and Machida Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco - an example for multiple duplications of DNA around the integration target. Plant Journal 7: 157-164.

Phillips R.L., Kaeppler S.M. and Olhoff P. 1994. Genetic instability of plant tissue cultures: Breakdown of normal controls. Proceedings of the National Academy of Science, U.S.A. 91: 5222-5226.

Sala F., Arencibia A., Castiglione S., Yifan H., Labra M., Savini C., Bracale M. and Pelucchi N. 2000. Somaclonal variation in transgenic plants. Proceedings of the International Symposium On Methods and Markers for Quality Assurance in Micropropagation (eds). A.C. Cassells, B.M. Doyle, R.F. Curry Acta Horticulturae 530, ISHS.

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

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

Shinozaki K., Ohme M., Tanaka M., Wakasugi T., Hayashida N., Matsubayashi T., Zaita N., Chunwongse J., Obokata J., Yamaguchi-Shinozaki K., Ohto C., Torazawa K., Meng B.Y., Sugita M., Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A., Tohdoh N., Shimada H. and Suguira M. 1986. The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. EMBO Journal 5: 2043-2049.

Stahl R., Horvath H., Van Fleet J., Voetz M., von Wettstein D. and Wolf N. 2002. T-DNA integration into the barley genome from single and double cassette vectors. Proceedings of the National Academy of Science, U.S.A. 99: 2146-2151.

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, U.S.A. 94: 2284-2289.

Windels P., Taverniers I., Depicker A., Van Bockstaele E. and De Loose M. 2001. Characterisation of the Roundup Ready soybean insert. European Food Research Technology 213: 107-112.

Zheng S.J., Henken B., Sofiari E., Jacobsen E., Krens F.A. and Kik C. 2001. Molecular characterization of transgenic shallots (*Allium cepa* L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders. Transgenic Research 10: 237-245.

This work has been submitted to Molecular Breeding entitled "Representational difference analysis as a tool to detect possible genome changes in genetically modified tobacco plants" by C Van der Vyver, KJ Kunert and CA Cullis.

Chapter 4: Isolation of flanking sequences of RDA subtraction products using a genomic library and tail PCR.

Abstract

A genomic DNA library from genetically modified tobacco was constructed to identify the flanking sequences of the isolated RDA subtraction products in the tobacco genome. Library clones that contained the isolated Hp12 and Hp14 subtraction products derived from *Hpa*II-digested genomic DNA were purified. The isolated clones were homologous to DNA sequences of the tobacco chloroplast genome and the 18S ribosomal RNA. Library clones were also isolated containing the subtraction product Hi30 derived from *Hind*III-digested genomic DNA. These showed homology to a number of known repetitive DNA sequence families. Tail PCR was used to isolate flanking regions of the Hi30 subtraction product, a DNA flanking sequence was isolated with similarities to DNA sequences of general cloning vectors and the 16S RNA gene of alga.

Introduction

Exact sequence information about genome regions can be obtained by either the construction of a genomic DNA library or by a PCR-based technique amplifying DNA flanking sequences of a known region of the plant genome. A genomic DNA library allows analysis of large regions in the plant genome. Such regions are also likely to contain highly abundant DNA sequences and/or also multi-gene families ((Nouzova et al., 2000; Khoudi et al., 1997; Belkhiri et al., 1997). In general, a genomic library consists of single enzyme-digested genomic DNA, which has been cloned into a vector, packaged and then transformed into competent *E. coli* cells. Plaques obtained after *E. coli* transformation are used for colony hybridisation with a labelled DNA probe to identify cloned fragments of genomic DNA (Sambrook et al., 1989). The DNA

can be sequenced from each selected clone and total sequence length to be analysed can be more than a hundred thousand base pairs for BAC clones.

PCR-based techniques generally use genome walking from a known to an unknown genomic DNA region and are based on (1) inverse PCR, (2) randomly primed PCR and (3) adaptor ligation PCR or a combination of the different techniques (Siebert et al., 1995; Ochman et al., 1988). In this study we used an extension of the polymerase chain reaction that permits the amplification of regions that flank RDA subtraction products of a known sequence, based on randomly primed PCR. This technique called tail PCR, which has been developed by Sorensen et al. (1993), requires no prior knowledge beyond the priming sites in the known DNA region. It includes two steps of PCR separated by a simple purification procedure. In the first PCR reaction a specific biotinylated primer from the known sequence and a partly degenerated random flanking primer (FP) are used. The FP primers are made up by a 3'-end of five fixed nucleotides to prevent primer hybridisation everywhere within a DNA sequence, followed by a stretch of seven random nucleotides. Statistically, every five nucleotides would occur every 4^5 = 1024 bp in the genome, thus resulting in an average amplification product size within the range for the PCR reaction to take place. All unspecific fragments produced by the FP primer alone are removed in an additional second PCR step in which only fragments originating from the specific primer are amplified. The amplified fragments from the second PCR are seen as distinct bands when analysed on an agarose gel. This technique has been successfully applied in studies to isolate integration sites of retroviruses, the isolation of promoter / enhancer regions, exon/intron junction analyses and in small-scale chromosome walking (Amtoft et al., 1997; Sorenson et al., 1993; Sorenson et al., 1996; Laskus et al., 1999).

In this part of the study, the flanking sequences of RDA subtraction products have been identified by using either a genomic library from genomic DNA of tobacco plants derived from a tissue culture/gene insertion process or by carrying out the tail PCR technique. Both techniques ultimately allowed locating RDA subtraction products in the genomic structure of the tobacco.

Results

A) Constructing a genomic library from a genetically modified tobacco plant.

Genomic DNA form genetically modified *OC-I* expressing tobacco plants (*Nicotiana tabacum* L., var Samsun) as outlined in Annex (A) under "Materials and methods" for Section I: "Plant transformation and selection", was used for library construction. The constructed library was screened by Southern blot analysis for presence of subtraction products. For library screening gel-purified single-stranded DNA from the three RDA subtraction products, Hp12, Hp14 and Hi30 were used as probes. After the final round of library purification positive clones hybridising with subtraction products were selected. Inserts were released from the ZAP vector by digestion of purified plasmid DNA with the restriction enzyme *Bam*HI after separation on a 1% agarose gel. Insert sizes for three individual clones ranged from 3.4 kb to 7 kb (Hp12) and from 1.1 kb to 1.4 kb (Hp14) (Figure 4.1). Four putative clones containing the Hi30 difference product were also isolated from the genomic DNA library. The inserts of the four clones had different sizes and were 1350 bp (Hi30L1), 1850 bp (Hi30L2), 460 bp and 270 bp (Hi30L3) and 750 bp and 3.4 kb (Hi30L4) long (Figure 4.2).

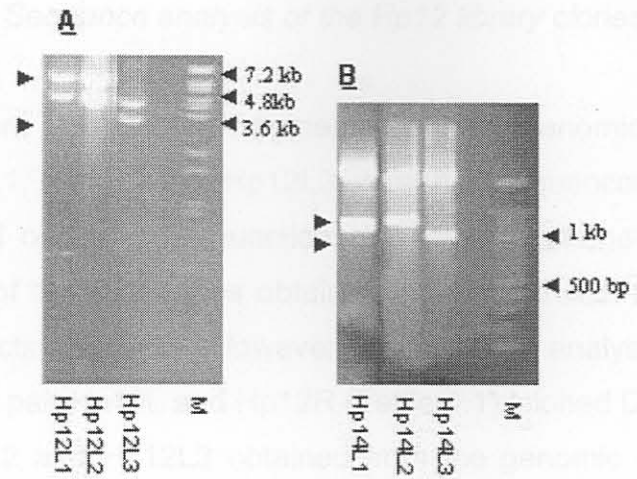

Figure 4.1: (A) DNA fragments from genomic library clones Hp12L1, Hp12L2 and Hp12L3 excised with *Bam*HI from the phagemid vector pBK-CMV separated on an agarose gel and stained with ethidium bromide. Arrows indicate position on the gel of excised inserts. Lane M represents *BstE*II-digested marker DNA with respective sizes. (B) DNA fragments isolated from library clones Hp14L1, Hp14L2 and Hp14L3 after excision with *Bam*HI from the phagemid vector pBK-CMV separated on an agarose gel and stained with ethidium bromide. Arrows indicate position of excised inserts on the gel. Lane M represents a 100 bp DNA marker.

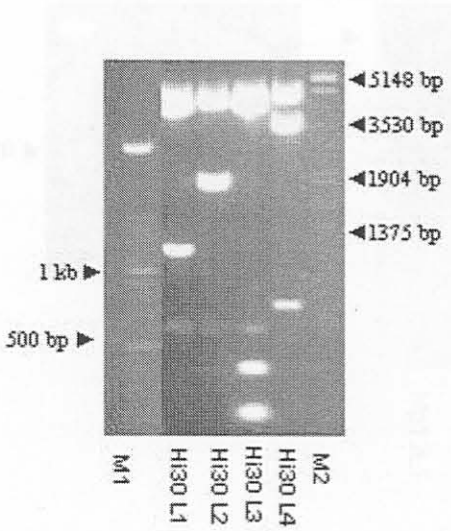

Figure 4.2: Inserts from genomic library clones Hi30L1, Hi30L2, Hi30L3 and Hi30L4 after excision with restriction enzyme *BamH*I from the phagemid vector pBK-CMV, separation on an agarose gel and stained with ethidium bromide. Lane M1 represents a 100 bp DNA marker and lane M2 represents marker DNA from *Hind*III/*Eco*RI-digested lambda DNA.

B) Sequence analysis of the Hp12 library clones.

Different cloned DNA fragments from the genomic DNA library representing Hp12L1, Hp12L2 and Hp12L3 were then sequenced. Sequence analysis was carried out from the junction site with the phagemid vector pBK-CMV and none of the sequences obtained overlapped with the sequence of the Hp12 subtraction product. However, when a PCR analysis was carried out with the primer pair Hp12L and Hp12R (Table 4.1), cloned DNA fragments for Hp12L1, Hp12L2 and Hp12L3 obtained from the genomic DNA library contained the expected 190 bp fragment representing the subtraction product Hp12 (Figure 4.3). Southern blot analysis also showed hybridization of the subtraction product Hp12 with *Bam*HI-digested plasmid DNA of cloned DNA fragments from library clones Hp12L1, Hp12L2 and Hp12L3 (Figure 4.3). Background hybridization present in Hp12L1 and Hp12L2 are most likely due to digested plasmid hybridization to the probe.

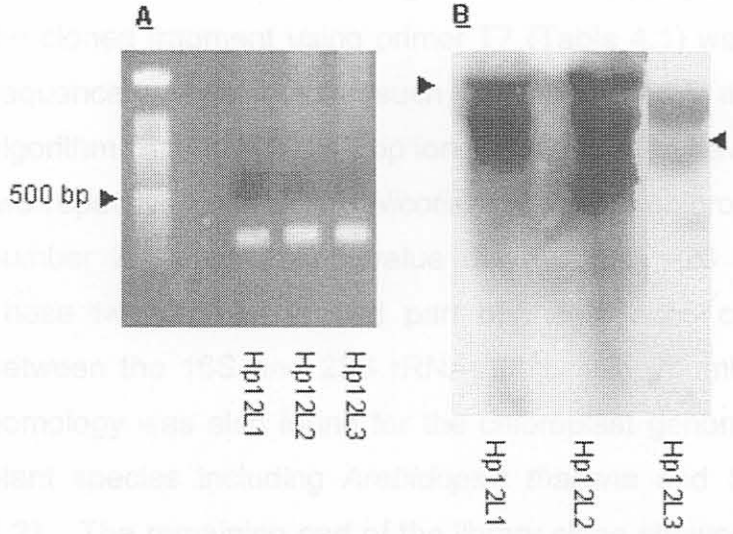

Figure 4.3: (A) PCR amplification of cloned DNA fragments with primers designed for Hp12 subtraction product with plasmid DNA derived from Hp12L1, Hp12L2 and Hp12L3 library clones as templates. Lane M represents a 100bp DNA marker. (B) Hybridization of Hp12 subtraction product to a filter (Figure 4.1) containing *Bam*HI-digested plasmid DNA from library clones Hp12L1, Hp12L2 and Hp12L3.

Table 4.1: Sequences of primers used for PCR to amplify subtraction products from vector DNA of library clones.

Name	Sequence
T3	5' ATT AAC CCT CAC TAA AGG GA 3'
T7	5' TAA TAC GAC TCA CTA TAG GG 3'
Hp12L	5' CGA GCC AAT GTT CGA ATA CC 3'
Hp12R	5' CCG AAG TTA CGG GGC TAT TT 3'
Hp12EX	5' ATTCATGCATGCTCACTTGG 3'
Hp14L	5' TGT CGG CCA AGG CTA TAA AC 3'
Hp14R	5' TTC CGT TAA CGA ACG AGA CC 3'
Hi30L	5' GGA ATG ATT TCC CAA AAC TCC 3'
Hi30R	5' CGA CGT CGA CTA TCC ATG AAC 3'

For library clone Hp12L1, a DNA sequence of 647 bp using the primer T3 (Table 4.1) and a 652 bp fragment DNA sequence from the opposite end of the cloned fragment using primer T7 (Table 4.1) were obtained. Using DNA sequence analysis tools, such as Blast, FastA and the Smith-Waterman algorithm, Hp12L1T3 (647 bp long) was found to have significant homology to two repetitive parts of the *Nicotiana tabacum* chloroplast genome (accession number Z00044). An E-value of less than –20 was deemed significant. These two regions formed part of the tobacco chloroplast spacer region between the 16S and 23S rRNAs (accession number V00166). Significant homology was also found for the chloroplast genomes of a number of other plant species including *Arabidopsis thaliana* and *Spinacia oleracea* (Table 4.2). The remaining part of the library clone showed no significant homology to known DNA sequences. The BLAST search for the total Hp12L1T7 sequence revealed a significant homology (E=0.0) to two parts of the *Nicotiana tabacum* chloroplast genome (accession number Z00044) as well as to the chloroplast genome of a number of other plant species (AJ316582, AP000423; Table 4.2). The sequence also had a 92% homology to the

Arabidopsis thaliana mitochondrial genome (accession number Y08501; Table 4.2).

From the library clone Hp12L2, a 626 bp fragment sequence with a T3 primer and a 667 bp fragment sequenced with a T7 primer were obtained (Table 4.1). The BLAST search using the program, BlastN, GenBank, EMBL, DDBJ and PDB sequences (no EST, STS, GSS or phase 0, 1 or 2 HTGS sequences) revealed that the sequence Hp12L2T3 had significant homology to four parts of the *Nicotiana tabacum* chloroplast genome. These homologous parts ranged from 80 to 522 bp (accession number Z00044) forming part of a tRNA gene with a 27 bp gap without any homology to the chloroplast genome. In order to extend the Hp12L2T3 sequence the primer Hp12EX (Table 4.1) at the tail of the DNA sequence was designed to extend the sequence analysis. Extended sequence analysis resulted again in sequencing of further parts of chloroplastic DNA. BLAST search for the sequence Hp12L2T7 showed that the total sequence had significant homology to two parts of the *Nicotiana tabacum* chloroplast genome (accession number Z00044) (Table 4.2).

For library clone Hp12L3, a 1007 bp fragment sequence was obtained with a T3 primer and a 792 bp fragment sequence with a T7 primer (Table 4.1). A BLAST search showed that sequence Hp12L3T3 had no significant homology to known DNA sequences. An E-value of less than –20 was deemed significant. The library clone however did show some homology in base pair overlap to two parts of the *Nicotiana tabacum* RENT3 repetitive sequence family (accession number AY049964; E = 0.006). Hp12L3T3 also had a high homology (over 80%) to the regions of the RENT7, 5 and 2 repetitive sequence families in *Nicotiana tabacum* (accession numbers AY049966 / 5 / 3; Table 4.2). BLAST search for sequence Hp12L3T7 revealed that the sequence had no significant homology to known DNA sequences. Some sequence homology was found to parts of the *Nicotiana tabacum* RENT2 and RENT1 repetitive sequences (accession number AY049963; Table 4.2).

***Table 4.2*:** Sequence homology of library clones Hp12L1, Hp12L2 and Hp12L3 to known DNA data banks. An E-value of less than –20 was deemed significant. Sequence data are outlined in Annex (B); Section II: "Sequence data of genomic library clones". Overlap = the number of homologous base pairs to a known DNA sequence fragment.

Library clone	Homology	E Value	Overlap (Homologous bp)	% homology of overlap
Hp12L1T3	*Nicotiana tabacum*	0.0	489 bp	94%
	chloroplast genome	0.0	489 bp	94%
	Spinacia oleracea	-128	384 bp	93%
	chloroplast genome	-128	384 bp	93%
		-21	97 bp	91%
	Arabidopsis thaliana	-123	375 bp	93%
	chloroplast genome	-123	375 bp	93%
		-34	109 bp	94%
		-34	109 bp	94%
Hp12L1T7	*Nicotiana tabacum*	0.0	654 bp	99%
	chloroplast genome	0.0	654 bp	99%
	Atropa belladonna	0.0	654 bp	98%
	chloroplast genome	0.0	654 bp	98%
	Arabidopsis thaliana	0.0	567 bp	92%
	chloroplast genome	0.0	567 bp	92%
	Arabidopsis thaliana	-115	325 bp	92%
	mitochondrial genome			
Hp12L2T3	*Nicotiana tabacum*	0.0	522 bp	96%
	chloroplast genome	0.0	522 bp	96%
		-13	80 bp	90%
		-13	80 bp	90%
Hp12L2T7	*Nicotiana tabacum*	0.0	667 bp	96%
	chloroplast genome	0.0	667 bp	96%

Hp12L3T3	*Nicotiana tabacum* RENT3 repetitive sequence family	0.006 1.4	45 bp 41 bp	91% 87%
	Nicotiana tabacum RENT7, repetitive sequence family	0.092	46 bp	91%
	Nicotiana tabacum RENT5 repetitive sequence family	1.4	41 bp	87%
	Nicotiana tabacum RENT2 repetitive sequence family	1.4	45 bp	88%
Hp12L3T7	*Nicotiana tabacum* RENT2 repetitive sequence family	0.005 1.1	41 bp 41 bp	90% 87%
	Nicotiana tabacum RENT1 repetitive sequence family	0.018	108 bp	81%

C) Sequence analysis of the Hp14 library clones

For library clones Hp14L1, Hp14L2 and Hp14L3, sequence data showed an overlap with the Hp14 subtraction product with overlap sizes of 417 bp (Hp14L1), 444 bp (Hp14L2) and 82 bp (Hp14L3). PCR analysis with the primer pair Hp14R and Hp14L (Table 4.1) designed for amplification of the Hp14 subtraction product and also Southern blot analysis with subtraction product Hp14 as a probe confirmed the presence of the Hp14 subtraction product within the plasmids of the library clones (Figure 4.4).

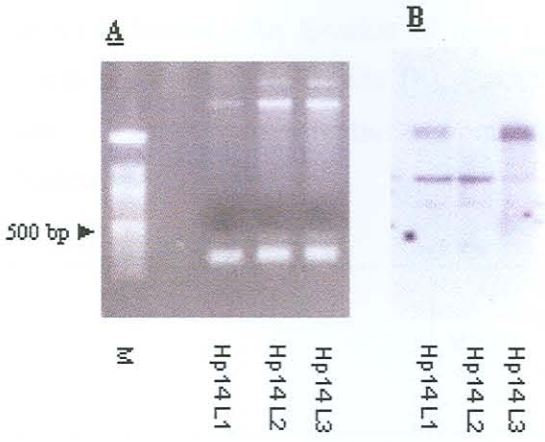

***Figure 4.4*:** (A) PCR amplification of DNA fragments from library clones with primers designed for Hp14 subtraction product using plasmid DNA of Hp14L1, Hp14L2 and Hp141L3 library clones as templates. Lane M represents a 100 bp DNA marker. (B) Hybridization of Hp14 difference product to a filter containing *Bam*HI-digested plasmid DNA from library clones Hp14L1, Hp14L2 and Hp14L3.

BLAST analysis using the program, BastN, GenBank, EMBL, DDBJ and PDB sequences (no EST, STS, GSS or phase 0, 1 or 2 HTGS sequences) of library clone Hp14L1T3, revealed significant homology to the *Nicotiana tabacum* 18S rRNA gene, while significant homology to the large subunit ribosomal RNA gene of a number of *Ceratocystis* species was found for the Hp14L1T7 sequence (accession numbers AJ236016, U47824). An E-value of less than –20 was deemed significant. When the sequence of library clones Hp14L2T3 / T7 was analyzed, BLAST analysis showed 95% homology to a 612 bp overlap from the 26S rRNA gene of *Nicotiana tabacum* as well as a significant homology to the *Nicotiana tabacum* 18S rRNA gene (AF479172, AJ236016). In contrast, the sequence of the library clones Hp14L3T3 / T7 had significant homology to the 18S rRNA gene of *Nicotiana tabacum* and the *Nicotiana tabacum* 5.8S rRNA gene (accession numbers AJ236016, AJ012365) (Table 4.3).

Table 4.3: Sequence homology of library clones Hp14L1, Hp14L2 and Hp14L3 to known DNA data banks. An E-value of less than –20 was deemed significant. Sequence data are outlined in Annex (B); Section II: "Sequence data of genomic library clones". Overlap = the number of homologous base pairs to a known DNA sequence fragment.

Library clone	Homology	E value	Overlap (Homologous bp)	% Homology of overlap
Hp14L1T3	*Nicotiana tabacum* 18S rRNA gene	0.0	623 bp	98%
Hp14L1T7	Large subunit ribosomal RNA gene of *Ceratocystis* species.	0.0	641 bp	97%
Hp14L2T3	*Nicotiana tabacum* 26S rRNA gene	0.0	612 bp	95%
Hp14L2T7	*Nicotiana tabacum* 18S rRNA gene	0.0	640 bp	98%
Hp14L3T3	*Nicotiana tabacum* 18S rRNA gene	0.0	621 bp	95%
Hp15L3T7	*Nicotiana tabacum* 5.8S rRNA gene	0.0	611 bp	98%

D) Sequence analysis of Hi30 library clones

For library clones Hi30L1, Hi30L2 and Hi30L4 only right and left tail sequences were obtained due to the large size of insert in the ZAP vector. None of the sequences obtained overlapped with the sequence for the Hi30 subtraction product. PCR analysis with primer pair Hi30L and Hi30R (Table 4.1) designed to amplify subtraction product Hi30 resulted in the expected 199 bp amplified fragment when vector DNA from library clones Hi30L1, Hi30L2, Hi30L3 and Hi30L4 was used as template. This confirmed the presence of

the subtraction product Hi30 in the vector DNA of library clones (Figure 4.5). However, when *Bam*HI-digested vector DNA was transferred to a filter and hybridized with the Hi30 subtraction product, signals for both the cloning vector and the Hi30 subtraction product was found (Figure 4.5).

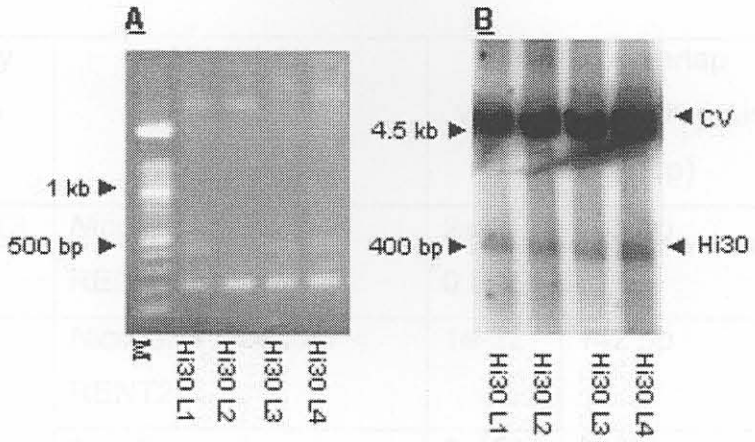

Figure 4.5: (A) PCR amplification of Hi30 subtraction product from different library clones with primers designed for Hi30 subtraction product using vector DNA from library clones Hi30L1, Hi30L2, Hi30L3 and Hi30L4 as template. Lane M represents a 100 bp DNA marker. (B) Hybridization of Hi30 subtraction product to a filter containing the amplified Hi30 subtraction product from library clones Hi30L1, Hi30L2, Hi30L3 and Hi30L4. Arrows indicate the position of the Hi30 subtraction product (Hi30) and the cloning vector (CV).

BLAST analysis of library clone, Hi30L1T3 indicated over 80% homology to the *Nicotiana tabacum* RENT1/2/3/7 repetitive sequence families with overlapping sequences of up to 403 bp (accession numbers AY049962/4/6/3). Library clone Hi30L2T3 / T7 had, however, no significant homology to any known DNA sequences. An E-value of less than –20 was deemed significant. Blast analysis of library clone Hi30L3 revealed significant homology to a *Nicotiana tabacum Bam*HI tandem repeat element DNA in up to four regions (accession numbers X15068, X12489). Library clone Hi30L4T7 had 84% homology to a 150 bp overlapping sequence of the *Nicotiana tabacum* RENT3 repetitive sequence (accession number AY049964).

Table 4.4: Sequence homology of library clones Hi30L1, Hi30L2, Hi30L3 and Hi30L4 to known DNA data banks. An E-value of less than –20 was deemed significant. Sequence data are outlined in Annex (B); Section II: "Sequence data of genomic library clones". Overlap = the number of homologous base pairs to a known DNA sequence fragment.

Library clone	Homology	E value	Overlap (Homologous bp)	% Homology of overlap
Hi30L1T3	*Nicotiana tabacum* RENT1	2e-85 0.005	403 bp 61 bp	85% 88%
	Nicotiana tabacum RENT2	1e-22	142 bp	85%
	Nicotiana tabacum RENT3	2e-54 8e-05	255 bp 36 bp	86% 94%
	Nicotiana tabacum RENT7	9e-54 8e-05	274 bp 36 bp	85% 94%
Hi30L1T7	No significant homology to known DNA			
Hi30L2T3	No significant homology to known DNA			
Hi30L2T7	No significant homology to known DNA			
Hi30L3	*Nicotiana tabacum* *Bam*HI tandem repeat element DNA.	-141 -129 -119 1e-71	361 bp 361 bp 354 bp 184 bp	93% 91% 90% 95%
Hi30L4T3	No significant homology to known DNA			
Hi30L4T7	*Nicotiana tabacum* RENT3 repetitive sequence family.	1e-20	150 bp	84%

E) Tail PCR: Hi30

Flanking regions adjacent to the Hi30 subtraction product in genomic DNA of genetically modified tobacco was isolated by applying a two-step tail PCR technique. In this procedure biotinylated primers were designed from the sequence of the Hi30 subtraction product and random forward primers in the flanking regions of the subtraction product (Table 4.5).

Table 4.5: Sequences of primers used for tail PCR.

Name	Sequence
Random forward primers	
FP1	5' CAGTTCAAGCTTGTCCAGGAATTCNNNNNNGGCCT 3'
FP2	5' CAGTTCAAGCTTGTCCAGGAATTCNNNNNNGCGCT 3'
FP3	5' CAGTTCAAGCTTGTCCAGGAATTCNNNNNNCCGGT 3'
FP4	5' CAGTTCAAGCTTGTCCAGGAATTCNNNNNNCGCGT 3'
Primer 3	5' CAGTTCAAGCTTGTCCAGGAATTC 3'
Hi30 primers	
Hi30R2B	5' CGTCGGATGTCATTTACACG 3' biotinylated
Hi30R2	5' CGGATGTCATTTACACGTTTG 3'
Hi30L2B	5' TAAAAATCGAGCCCGAAATC 3' biotinylated
Hi30L2	5' AAATCGAGCCCGAAATCC 3'

In the two-step tail PCR reaction, primers Hi30R2B and FP1-4 and genetically modified tobacco genomic DNA as template were used in a first PCR reaction for amplification. This was followed by a second PCR reaction using an aliquot from the first PCR reaction as DNA template and primers Hi30R2 and Primer 3. In the second reaction, a 390 bp DNA flanking sequence (Hi30RB) was amplified (Figure 4.6A). In contrast, no flanking sequence was amplified when the primer pair Hi30L2B and Hi30L2 was used. Amplification product

Hi30RB was cloned into the cloning vector p*MOSBlue* and then sequenced (Figure 4.6B).

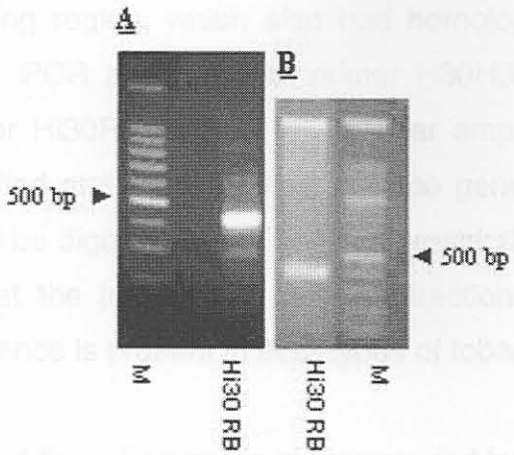

Figure 4.6: (A) PCR analysis of amplified DNA flanking sequence (Hi30RB) adjacent to subtraction product Hi30. Lane M represents a 100 bp DNA marker. (B) Cloned Hi30RB amplification product excision from the cloning vector p*MOSBlue* after digestion with restriction enzymes *Hind*III and *Bam*HI to release the cloned insert. Insert separation on an 1.5 % TAE agarose gel and staining with ethidium bromide to visualise DNA. Lane M represents a 100 bp DNA marker.

Alignment of cloned Hi30RB product with subtraction product Hi30, showed a 96% homology for the first 57 bp of the subtraction product as well as 100% homology for a 25 bp repeat, at the other end of the Hi30 subtraction product. The remaining flanking sequence had significant homology to both a 276 bp overlapping sequence of alga 16S RNA genes (E = -143; accession number Af005250) and as much as 260 bp overlapping sequence to a number of cloning vector DNAs (E = - 139; accession numbers AF327711; Y14836/35; L37382). Significant homology was also found for the *Arabidopsis thaliana* mRNA for mitochondrial ATP synthase beta subunit (E = -122; accession number AJ271468) Sequence data are outlined in Annex (B); Section III: "Sequence data of tail PCR".

In order to determine whether a *Hind*III site at the junction sites of the Hi30 subtraction product is present in genetically modified and non-modified DNA, the primer H30H30 was designed using sequence information from the Hi30 flanking region, which also had homology to known cloning vectors (Table 4.6). PCR analysis with primer H30H30 and the Hi30 subtraction product primer Hi30R, resulted in a similar amplified pattern in both the genetically modified and non-modified tobacco genomic DNA. Both amplified products could be digested with the *Hind*III restriction enzyme indicating that the *Hind*III site at the junction of Hi30 subtraction product and the adjacent flanking sequence is present in both types of tobacco (Figure 4.7).

Table 4.6: Sequences of primers used for Hi30 junction site determination.

Name	Sequence
H30H30	5' ATTAGGCACCCCAGGCTTTA 3'
Hi30R	5' CGA CGT CGA CTA TCC ATG AAC 3'

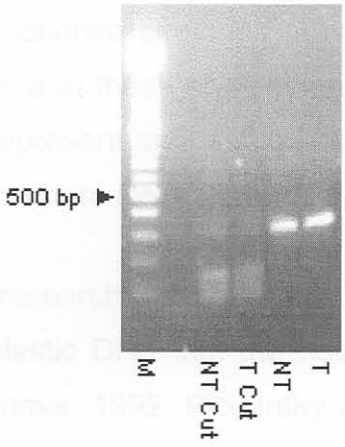

Figure 4.7: PCR amplification of genomic target DNAs with primers Hi30R and H30H30 where the genomic DNA template derived from genetically modified (T) and non-modified (NT) tobacco plant. Amplified products were digested with *Hind*III and then separated on an agarose gel and stained with ethidium bromide. NTCut = amplified non-modified DNA digested with *Hind*III and TCut = amplified genetically modified DNA digested with *Hind*III. Lane M represents a 100 bp marker.

Discussion

Construction of a genomic DNA library from tobacco was time-consuming and sequence walking has to be applied to ultimately obtain a complete sequence of a library clone. Sequence walking was required due to the rather large size of library clones after *Bam*HI digestion of genomic tobacco DNA with cutting sites located on average every 4^6 = 4096 bp in the tobacco genome. In this study, for detection of each of the three subtraction products $2.5x10^5$ library clones were ultimately screened. To screen the complete tobacco genome, which has a size of $3.7x10^9$ bp, the screening of about 10^6 clones would have been required. Therefore, the probability of finding a single copy, of a 4 kb DNA fragment in $2.5x10^5$ library clones will be not more than 23.7%, using the probability equation as described by Sambrook et al. (1989).

A number of library clones for the RDA subtraction products Hp12 and Hp14 derived from methylation-sensitive *Hpa*II-digested genomic tobacco DNA was finally isolated and analysed in more detail. Two of these clones, Hp12L1 and Hp12L2, were homologous to chloroplastic DNA. However they both had small sequence stretches of non-chloroplastic DNA adjacent to chloroplastic DNA. Due to these non-chloroplastic stretches, the Hp12 subtraction product might represent an insertion of chloroplastic DNA into the nuclear genome of tobacco where DNA changes have occurred.

Other research groups have previously also identified such insertions of chloroplastic DNA into the nuclear genome in several plant species (Ayliffe and Timmis, 1992; Pichersky and Tanksley, 1988; Du Jardin, 1990; Timmis and Scott, 1983). This study however, gave no evidence that specifically the transformation or tissue culture process has caused such insertion of chloroplastic DNA into nuclear DNA. It might be rather speculated that during evolution sections of the chloroplast genome have been transferred to the nucleus with gradual changes of the DNA sequence over time. By such transfer the nuclear genome might have acquired control over the metabolic activities of the chloroplast or reducing the total amount of cellular DNA due to

multi-copies of chloroplastic DNA (Ayliffe and Timmis, 1992; Thorsness and Weber, 1996).

The nuclear genome of most higher plants is further extensively methylated and therefore resistant to *Hpa*II digestion. In contrast, the chloroplastic genome is non-methylated or methylated only at very few sites (Jeddoloh and Richards, 1996; Scott and Possingham, 1980; Ngernprasirtsiri et al., 1989). By using in this study the methylation sensitive restriction enzyme *Hpa*II in RDA, nuclear DNA seemingly have been cut at non-methylated sites susceptible to *Hpa*II digestion. Since two subtraction products (Hp12 and Hp14) were obtained in this study after *Hpa*II digestion, genomic DNA of the genetically modified plants was possibly less methylated than the respective DNA of the non-modified wild-type plant. It has to be investigated in a future study if such possible methylation changes at specific sites are directly linked to a changed gene expression pattern affecting the phenotype of the plant as found for genetically modified plants expressing the *OC-I* gene.

Isolated library clone Hp12L3 was identified as part of a repetitive tobacco sequence family (*Nicotiana tabacum* RENT repetitive sequence families; Foster et al., 2001, submitted). The origin of this repetitive sequence family is not known. Due to the relative large size of the tobacco genome, repetitive DNA is quite abundant in the nuclear genome. Repetitive DNA accounts for most of differences in genome size and genomic sequence composition in higher plants (Schmidt and Heslop-Harrison, 1998; Favell et al., 1974). The exact amount of repetitive DNA present in tobacco is not known but certain repetitive families, which can be up to 5% of the tobacco genome have already been found and characterized (Gazdova et al., 1995; Matyasek et al., 1997; Jakowitsch et al., 1998; Suzuki et al., 1994). In general, the percentages of repetitive DNA in plant genomes vary considerable between species. For example up to 80% of grass genomes, 20% of the tomato genome, belonging like tobacco to the *Solanaceae* family, and 70% of the pea genome consist of repetitive DNA (Vicient et al., 2001; Barakat et al., 1999). The chance of an insertion event to occur in repetitive DNA is, therefore, much greater than for other parts of the genome. Ayliffe et al. (1998)

suggested that the lack of chloroplast DNA within the nuclear genome of *Arabidopsis* could be due to the low levels of non-coding or repetitive regions present in this plant species. The different locations of the Hp12 subtraction product in chloroplast DNA and repetitive DNA might further suggest that the transposition of plastid sequences to the nucleus is still an ongoing process as part of the plant's natural evolution or the continued movement of the nuclear localized copy. It still has to be shown if this process can be accelerated by a stressful event like a plant tissue culture / gene insertion process. The nuclear background of the individual plants might also affect the rate of such DNA transfer as found for yeast (Thorsness and Fox, 1990) and this might explain the observed heterogeneity of transfer of these sequences in investigated plants so far.

The adjacent putative library clones for the Hi30 subtraction product were also homologous to tobacco repetitive sequence families. Repetitive sequences are mostly dispersed repeats represented by various families of mobile elements, such as the retrotransposons or tandem repeats, with various sizes and their monomer units arranged in a head-to-tail orientation (Voytas et al., 1992; Schmidt and Heslop-Harrison, 1998; Nouzova et al., 2000). Transposable elements can represent between 3 to 50 % of the content of the genome, depending on the species and are known to be mobilized in response to stress (Capy et al., 1997; Capy et al., 2000). These repetitive elements could therefore easily have been mobilized in response to the stressful tissue culture environment / genetic transformation, incorporating itself at different location in the tobacco genome. Therefore, when the tobacco genomic DNA was digested for the RDA procedure different length DNA fragments were generated resulting in altered PCR products leading to the isolation of the subtraction product.

Southern blot analysis of the Hi30 subtraction product to a filter containing isolated Hi30 library clones showed an unexpected hybridization with phagemid vector pBK-CMV. By applying the technique of tail PCR to investigate the flanking sequences of the Hi30 subtraction product in more detail, a short piece of DNA that showed homology to a number of cloning

vectors was identified. Since this DNA sequence showed homology to cloning vectors, this short sequence might have originated from transfer of the gene construct used in the transformation process. Unfortunately, exact sequence data for the vector used in the transformation process pKYLX6 (Schardl et al., 1987) are not currently available in any search DNA library. PCR analysis indicated, however, that this sequence is also present in non-modified tobacco plants consequently eliminating the possibility that this sequence might originate from the transformation vector. In a study conducted by White et al. (1983) sequence homologies were also found between the root-inducing plasmid pRiA4b form *Agrobacterium rhizogenes* and the genomes of the untransformed host plant *Nicotiana glauca* (White et al., 1983).

In summary, this part of the study provided no evidence that any of the isolated subtraction products are directly linked to any plant tissue culture/ gene insertion process in genetically modified plants.

References

Amtoft H.W., Sorensen A.B., Bar-Hurry C., Schmidt J., Luz A. and Pedersen F. S. 1997. Stability of AML1 (core) site enhancer mutations in T lymphomas induce by attenuated SL3-3 murine leukemia virus mutants. Journal of Virology 71: 5080-5087.

Ayliffe M.A. and Timmis J.N. 1992. Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. Theoretical and Applied Genetics 85: 229-238.

Ayliffe M.A., Scott N.S. and Timmis J.N. 1998. Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. Molecular Biology and Evolution 15: 738-745.

Barakat A., Han D.T., Benslimane 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.

Belkhiri A., Integan H. and Klassen G.R. 1997. A tandem array of 5S ribosomal RNA genes in *Pythium irregulare*. Gene 186: 155-159.

Capy P., Bazin C., Higuet D. and Langin T. 1997. Dynamic and evolution of transposable elements. R.G. Landes Company, Austin, Texas, USA.

Capy P., Gasperi G., Biemont C. and Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity 85: 101-106.

Du Jardin P. 1990. Homologies to plastid DNA in the nuclear and mitichondrial genomes of potato. Theoretical and Applied Genetics 79: 807-812.

Favell R.B., Bennet M.D., Smith J.B. and Smith D.B. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochemical Genetics 12: 257-269.

Foster E., Hattori J., Zhang P., Labbe H., Martin-Heller T., Li-Pook-Than J., Ouellet T., Malik K. and Miki B. 2001. The new RENT family of repetitive elements in Nicotiana species harbor cryptic gene regulatory elements related to the tCUP promoter. Submitted, ECORC, Agriculture and Agri-Food Canada.

Gazdova B., Siroky J., Fajkus J., Brzobohaty B., Kenton A., Parokonny A., Heslop-Harrison J.S., Palme K. and Bezdek M. 1995. Characterization of a new family of tobacco highly repetitive DNA, GRS, specific for the *Nicotiana tomentosisformis* genomic component. Chromosome Research 3: 245-254.

Jakowitsch J., Papp I., Matzke M.A. and Matzke A.J. 1998. Identification of a new family of highly repetitive DNA, NTS9, that is located predominantly on the S9 chromosome of tobacco. Chromosome Reseach 6: 649-651.

Jeddoloh J.A. and Richards E.J. 1996. CCG methylation in angiosperms. Plant Journal 9: 579-586.

Khoudi H., Vezina L.P., Mercier J., Castonguay Y., Allard G. and Laberge S. 1997. An alfalfa rubisco small subunit homologue shares *cis*-acting elements with the regulatory sequences of the RbcS-3A gene from pea. Gene 197: 343-351.

Laskus T., Radkowski M., Wang L.F., Nowicki M. and Rakela J. 1999. Detection and sequence analysis of Hepatitis B virus integration in peripheral blood mono-nuclear cells. Journal of Virology 73: 1235-1238.

Matyasek R., Gazdova B., Fajkus J. and Bezdek M. 1997. NTRS, a new family of highly repetitive DNAs specific for the T1 chromosome of tobacco. Chromosoma 106: 369-379.

Ngernprasirtsiri J., Chollet R., Kobayashi H., Sugiyama T. and Akazawa T. 1989. DNA methylation and the differential expression of C4 photosynthesis genes in the mesophyll and bundle sheath cells of greening maize leaves. Journal of Biological Chemistry 264: 8241-8248.

Nouzova M., Neumann P., Navratilova A., Galbraith D. and Macas J. 2000. Microarray-based survey of repetitive genomic sequences in *Vicia* spp. Plant Molecular Biology 45: 229-244.

Ochman H., Gerber A.S. and Hartl D.L. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120: 621-623.

Pichersky E. and Tanksley S.D. 1988. Chloroplast DNA sequences integrated into an intron of a tomato gene. Molecular and General Genetics 215: 65-68.

Sambrook J., Fritsch E.F., Maniatis T. 1989. Molecular cloning: A laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schardl C.L., Byrd A.D., Bension G., Altschuler M.A., Hildebrand D.F. and Hunt A.G. 1987. Design and construction of a versatile system for the expression of foreign genes in plants. Gene 61: 1-11.

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

Scott N.S. and Possingham J.V. 1980. Chloroplast DNA in expanding spinach leaves. Journal of Experimental Botany 123: 1081-1092.

Siebert P.D., Chenchik A., Kellogg D.E., Lukyanov K.A. and Lukyanov S.A. 1995. An improved PCR method for wlking in uncloned genomic DNA. Nucleic Acids Research 23: 1087-1088.

Sorensen A.B., Duch M., Jorgensen P. and Pedersen F.S. 1993. Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. American Society of Microbiology 67: 7118-7124.

Sorensen A.B., Duch M., Amtoft H.W., Jorgensen P. and Pedersen F.S. 1996. Sequence tags of provirus integration sites in DNAs of tumors induced by the murine retrovirus SL3-s. Journal of Virology 70: 4063-4070.

Suzuki K., Yamagiwa Y., Matsui T. and Yoshida K. 1994. Restriction enzyme-resistant high molecular weight telomeric DNA fragments in tobacco. DNA Research 1: 129-138.

Thorsness P.E. and Fox T.D. 1990. Escape of DNA from the mitochondria to the nucleus in *Saccharomyces cerevisiae*. Nature 346: 376-379.

Thorsness P.E. and Weber E.R. 1996. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. International Review of Cytology 165: 207-234.

Timmis J.N. and Scott N.S. 1983. Sequence homology between spinach nuclear and chloroplast genomes. Nature 305: 65-67.

Vicient C.M., Jaaskelainen M.J., Kalendar R. and Schulman A.H. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiology 125: 1283-1292.

Voytas D.F., Cummings M.P., Koniexzny A.K., Ausubel F.M. and Rodermel S.R. 1992. *Copia*-like retrotransposons are ubiquitous among plants. Proceedings of the National Academy of Science, U.S.A. 89: 7124-7128.

White F.F., Garfinkel D.J., Huffman G.A., Gordon M.P. and Nester E.W. 1983. Sequence homolgous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. Nature 301: 348-350.

Chapter 5: PCR and sequence analysis for genome modifications in tobacco plants.

Abstract

Different types of tobacco were screened for the presence/absence of a RDA subtraction product Hi30 isolated from *Hind*III-digested genomic DNA from genetically modified tobacco. Primers designed from Hi30 were used to test the genomic DNAs from two different selections of the tobacco cultivar '*Samsun*' and a range of genetically modified *'Samsun'* plants carrying different transgenes as template DNA. Hi30 was not unique to genetically modified plants. A 38 bp DNA region in Hi30, which varies, was identified in all tobacco plants tested, seemingly representing a variable DNA region in the tobacco genome. The flanking sequence of the Hi30 subtraction product contained a 185 bp conserved sequence fragment adjacent to a variable region, which differed in size and sequence between the tested tobacco plants. Only single base pair changes were detected when plants were screened for the presence/absence of the RDA subtraction product Hp12, which has been isolated from *Hpa*II-digested genomic DNA. Plants of different tobacco lines were further screened to detect possible changes in the rDNA repeat unit present in tobacco. Although a similar pattern for the rDNA transcribed regions were found in both genetically modified and non-modified tobacco plants, generally less rDNA was present in plants of the genetically modified tobacco lines.

Introduction

Introducing foreign DNA molecules into the plant genome, regardless of the method, using plant tissue culture is considered a stressful event and carries the risk of genome variation (Ditt et al., 2001; Labra et al., 2001; Choi et al., 2000). So far, scientists have predominantly used molecular characterization

of plants together with conventional morphological plant characterization based on expressed characteristics to describe variation and genetic diversity between individual plants (Arencibia et al., 1998; Lynch et al., 1995). The range of such morphological characters has been further increased by the use of electron microscopy and biochemical or phytochemical assays. The extensive use of morphological markers is due to the simplicity of testing, the avoidance of expensive analytical laboratory setups and a traditional school of thought, which favors plant discrimination based on expressed characters. However, morphological and biochemical identification frequently requires large sets of phenotypic, enzymatic or secondary compound data. These are often difficult to assess and sometimes variable due to environmental influences and the maturation stage of the plant causing differential expression of otherwise identical genes in plants (Karp and Bright, 1985).

In contrast to a morphological or biochemical marker, a molecular genomic DNA marker describes the internal make-up of a plant and identifies the variance of total composition of DNA. An obvious advantage of molecular assays is the immense number of characters that they reveal as well as the general advantage that the DNA content of a cell is independent of environmental conditions, organ specificity or growth stage. However, except for some major crops, for example maize, rice and wheat, DNA based identification for plants has not been well developed.

Development of a genetic marker to detect regions of the genome, which vary possibly due to plant tissue culture/gene insertion has so far not been attempted. Such a marker should have the ability to detect a labile region in an *in vitro* produced plant after genetic transformation. This changes should be in addition to the normal genetic variation between individuals to indicate if a plant originates from a genetic transformation process involving plant tissue culture. A number of studies previously carried out have focused both on the integration of the transgene itself and on the characterization of the surrounding junction sites of the transgene. These studies demonstrated the general occurrence of genomic changes in genetically modified plants in various regions of the plant genome (Sala et al., 2000; Labra et al., 2001;

Ohba et al., 1995; Windels et al., 2001; Zheng et al., 2001). The objective of this part of the study was therefore, to analyze and screen a large number of genetically modified and non-modified wild-type tobacco plants to identify such possible variable DNA regions in genetically modified plants using available sequence information from two of the isolated and characterized RDA subtraction products.

Results

A) *Hi30 subtraction product*

Genomic DNAs from two different selections of the tobacco cultivar '*Samsun*' ('*Samsun/UK*' and '*Samsun/San 1191*') and a range of genetically modified '*Samsun*' tobacco plants (*Samsun/UK*) carrying different transgenes were analyzed for the presence/absence of the subtraction product Hi30 derived from genetically modified tobacco. Tobacco used for screening included the i) non-modified wild-type (*Nicotiana tabacum*) L. of two selections of the cultivar '*Samsun*' ('*Samsun/San 1191*' and '*Samsun/UK*'); ii) genetically modified plants '*Samsun/UK*' carrying a *OC-I* transgene; iii) genetically modified plants '*Samsun/UK*' carrying a *gus* transgene; iv) transformed tobacco *'Samsun/UK'* which went through an Agrobacterium transfomation process but was identified in the F1 generation after selfing to carry no transgene. In the screening experiments the non-modified tobacco plants were identified as NT ('*Samsun/UK*') and as San ('*Samsun/San1191*'), while the genetically modified tobacco plants were identified as T (*OC-I* transgene), G (*gus* transgene) or T-C (Agrobacterium process but no transgene). Table 5.1 shows the sequence of the primers Hi30R and Hi30L used for amplification of Hi30 from different tobacco genomes.

Table 5.1: Sequences of primers used for the PCR analysis of tobacco plants.

Name	Sequence
Hi30L	5' GGA ATG ATT TCC CAA AAC TCC 3'
Hi30R	5' CGA CGT CGA CTA TCC ATG AAC 3'

With the Hi30R and Hi30L primers a PCR product of around 200 bp was amplified after 30 cycles of PCR in genomic DNA from several plants of the tobacco cultivar '*Samsun/San1119',* which is a selection originating from Greece (Figure 5.1A). However, detection of this amplification product could only be achieved, in the majority of plants, after two rounds of 30 amplification cycles from the tobacco cultivar *'Samsun/UK'*. This was regardless of being genetically modified and carrying the transgene or deriving from the *Agrobacterium* transformation process without a transgene insert (Figures 5.1B and 5.1C). Plants without a transgene insert, but deriving from the transformation process, were selected after selfing of a genetically modified plant. When genomic DNA of the non-modified wild-type tobacco cultivar '*Samsun/UK*' was used as template DNA, a clear amplification product could only be detected in two of the four tested plants (NT1 and NT2). NT3 was template DNA from the original wild-type '*Samsun/UK*' plant used in the RDA procedure. No amplification product could be detected in one of the genetically modified '*Samsun/UK*' plants carrying and expressing the *OC-I* gene (T4) as well as in one of the genetically modified plants carrying and expressing the *gus* gene (G4). Two additional amplification products of around 300 bp and a weak product at around 500 bp were also amplified after 30 cycles of PCR in one of the non-modified tobacco plants of the variety '*Samsun/San1191*' (San3) (Figure 5.1A). The additional 300 bp amplification product were also present in some of the non-modified tobacco plants from the variety '*Samsun/UK*' after two rounds of 30 amplification cycles (NT1 and NT2) (Figure 5.1B). The 300 bp DNA fragment in the non-modified tobacco plants were sequenced.

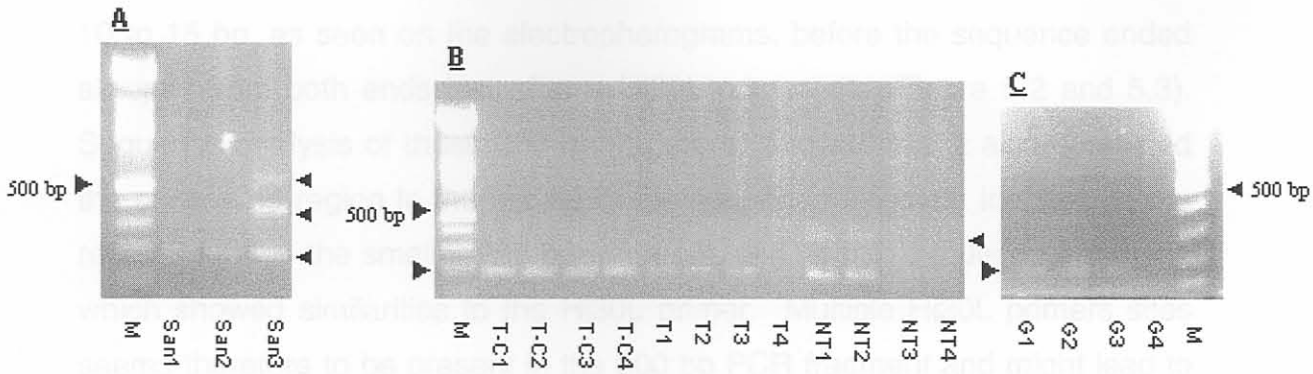

Figure 5.1: PCR amplification of genomic target DNAs with primers designed for subtraction product Hi30. (A) PCR amplification with genomic template DNA derived from tobacco cultivar '*Samsun/San1191'* (San 1-3). Lane M represents a 100 bp marker. (B) PCR products after 60 amplification cycles with genomic template DNA of tobacco plants (*'Samsun/UK'*) derived from a Agrobacterium transformation process without a transgene insert (T-C1 to T-C4), from genetically modified tobacco plants carrying the *OC-I* transgene (T1-T4) and non-modified wild-type tobacco plants (NT1 to NT4). (C) Genetically modified tobacco plants (*'Samsun/UK'*) containing the *gus* gene insert (G1 to G4). Lane M represents a 100 bp marker. Arrows indicate the amplified products.

Direct sequencing and analysis of the Hi30 subtraction product PCR fragments revealed a conserved 145 bp sequence in Hi30 adjacent to the Hi30L primer site, which was detected in all amplified DNA Hi30 products from the different plants tested (Figure 5.4). Adjacent to this conserved sequence region is a 38 bp sequence, which was variable between the different individual plants tested (Figure 5.4). Sequence data showed no homology to any known DNA sequences when a DNA database was searched. In some of the non-modified plants, a second amplification product of about 300 bp was also identified (Figure 5.1) when DNA primers for the Hi30 subtraction product were used for amplification. These 300 bp DNA fragments were sequenced directly from both sides to obtain the complete sequence. Sequencing the

300 bp PCR fragment from both ends resulted in sequence data of 100 bp and 170 bp, respectively, followed by a distinct decrease in peak heights for 10 to 15 bp, as seen on the electropherograms, before the sequence ended abruptly from both ends just after a Hi30L primer site (Figure 5.2 and 5.3). Sequence analysis of these 300 bp fragments showed that it also contained the conserved region in the middle of the sequence fragment identical to the region found in the smaller 200 bp fragment, ending in a sequence fragment, which showed similarities to the Hi30L primer. Multiple Hi30L primers sites seems therefore to be present in the 300 bp PCR fragment and might lead to more than one possible alignment of the 200 bp PCR fragment with the 300 bp PCR fragment. Premature overlapping of the two sequences from the ends might conceal a direct DNA repeat present in the unreadable middle part of the sequence possibly explaining the sudden decrease in peak heights and the short sequence data (220 bp) obtained .

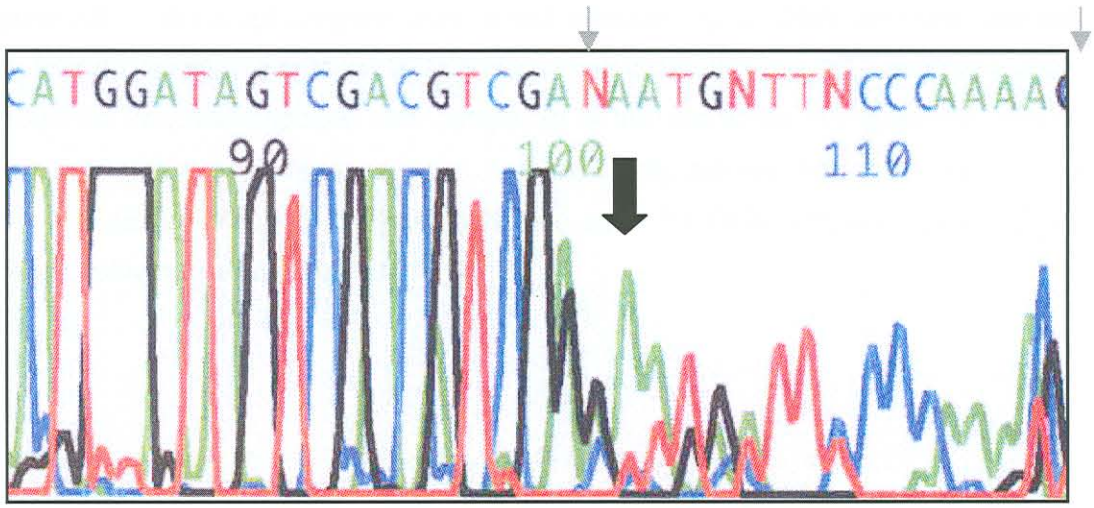

Figure 5.2: Electropherogram from direct sequencing of DNA template derived from tobacco cultivar '*Samsun/UK*' (NT1) showing the end of the DNA fragment when sequenced from the one side. Black arrow indicates the decrease in peak heights. Grey arrows indicate Hi30L primer site. A similar electropherogram was also seen for DNA template derived from tobacco cultivar '*Samsun/San1191*' (San3).

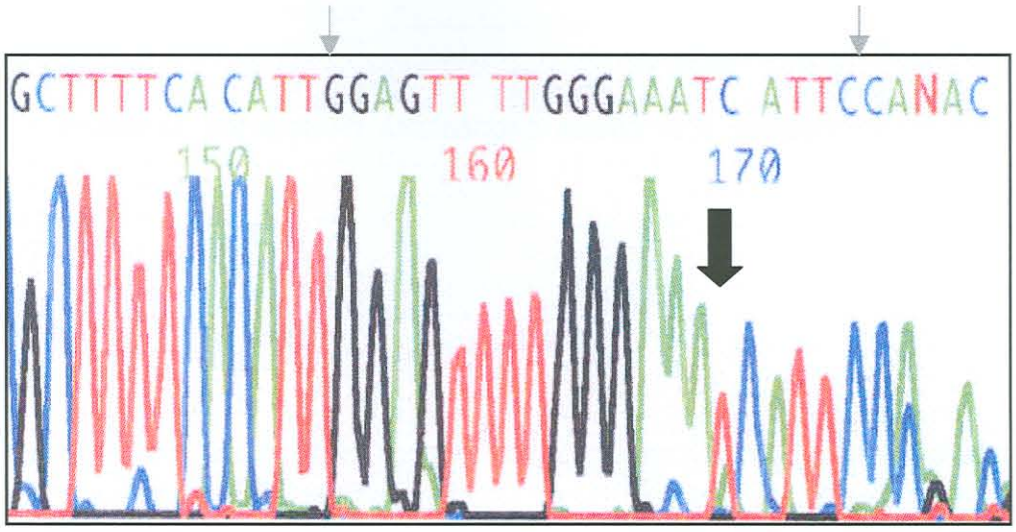

Figure 5.3: Electropherogram from direct sequencing of DNA template derived from tobacco cultivar '*Samsun/UK*' (NT1) showing the end of the DNA fragment when sequenced from the opposite side compared to the sequence in Figure 5.2. Black arrow indicates the decrease in peak heights. Grey arrows indicate Hi30L primer site. A similar electropherogram was also seen for DNA template derived from tobacco cultivar '*Samsun/San1191*' (San3).

```
                10         20         30         40         50         60
Hi30    ---------- --------AC CGACGTCGAC TATCCATGAA CAACGAAGAG ATGGCCGCGC
T-C1    ---------- ---------- ---------- ----------  ----..-... .-........
T-C2    ---------- ---------- ---------- ---------- -----..... ..........
San2    ---------- ---------- ---------- ---------- -----G.A.. .A........
San3    ---------- ---------- ---------- ---------- ---.CG.... .A........
NT1     ---------- ---------- ---------- ---------- ---...-... .A........
NT2     ---------- ---------- ---------- ---------- -........- .A........
San3B   ---------A AATCTATTCT .T........ ..........  .......... ..........
NT1B    CGTTTTGGGG AATC-ATTCT .......... ..........  .......-.A ..........

                70         80         90        100        110        120
Hi30    --TTGAACAA AGTAATT--G AAGAGACGAC ACAATTGGCC AGGGATTTCG GGCTCGATTT
T-C1    --........ -...-..--. C......... .......... .......... ..........
T-C2    --.....-.. ....-..--. .......... .......... .......... ..........
San2    C-.....-.. ....-..--. TC........ .......... .........C ..........
San3    CT.....-.. ....-..C-T .......... .......... .........C ..........
NT1     --.....-.. ....-.C--T -......... .......... .......... ..........
NT2     C-...C.... ....-..CTC .......... .......... .......... ..........
San3B   --........ ....-..T-. .......... .......... .......... ..........
NT1B    --.....-.. ....-..--. .......... .......... .......... ..........

                130        140        150        160        170        180
Hi30    TTATCC-GAT GCGTTATGAG GTTTGTCCGT CGGATGTCAT TTACACGTTT GGGGCCTACG
T-C1    ......-... .......... .......... .......... .......... ..........
T-C2    ......-... .......... .......... .......... .......... ..........
San2    ......-... .......... .......... .......... .......... ..........
San3    ......C... .......... .......... .......... .......... ..........
NT1     ......-... .......... .......... .......... .......... ..........
NT2     ......-... .......... .......... .......... .......... ..........
San3B   ......-... .......... .......... .......... .......... ..........
NT1B    ......-... .......... .......... .......... .......... ..........

                190        200        210        220
Hi30    GGAT-GCCGA CGCGCTTTTC ACATTGGAGT TTTGGGAAAT CATTCCA
T-C1    ....-..... .......... .....          ...
T-C2    ....-..... .......... .....          ...
San2    ....-..... .......... .....          ...
San3    ....A..... .......... .....          ...
NT1     ....-..... .......... .....          ...
NT2     ....-..... .......... .....          ...
San3B   ....-..... .......... .....          ...
NT1B    ....-..... .......... .....          ...
```

Figure 5.4: Alignment of DNA regions amplified by PCR with primers Hi30R and Hi30L using tobacco genomic DNA from different tobacco plants as template. Hi30 represents the sequence of the original Hi30 subtraction product isolated from genetically modified plants. Sequence data shown are from the 200 bp amplification product from transformed tobacco (*'Samsun/UK'*) without a gene insert (T-C1 and T-C2); tobacco cultivar *'Samsun/San1191'* (San2 and San3); NT1 and NT2 sequence data from non-modified wild-type tobacco DNA; 300 bp amplified product using *'Samsun/San1191'* (San3B) and wild-type tobacco *'Samsun/UK'* (NT1B) DNA as template. Underline areas indicated Hi30L and Hi30R primers, Gap=(-), Identical base pairs aligned with Hi30=(.).

B) *Hi30 flanking sequence*

Genomic DNA from the different modified and non-modified tobacco plants were also screened for the presence/absence of the flanking sequence from the Hi30 subtraction product. Two different selections of the tobacco cultivar '*Samsun/UK*' and '*Samsun/San1191*' and a range of genetically modified tobacco plants ('*Samsun/UK*') carrying different transgenes were used as DNA template. Primer Hi30R designed to bind and amplify within the subtraction product and primer H30H30 designed to bind and amplify within the flanking sequence adjacent to the subtraction product was used for 42 amplification cycles (Table 5.2).

Table 5.2: Sequences of primers used for the PCR analysis of Hi30 flanking sequences.

Name	Sequence
Hi30R	5' CGA CGT CGA CTA TCC ATG AAC 3'
H30H30	5' ATTAGGCACCCCAGGCTTTA 3'
Primer 3	5' CAGTTCAAGCTTGTCCAGGAATTC 3'
Hi30R2	5' CGGATGTCATTTACACGTTTG 3'

With the H30H30 and Hi30R primers numerous PCR products of between 250 bp and 700 bp were amplified from genomic DNA from all tested tobacco plants, regardless of being genetically modified or non-modified (Figure 5.5) (Table 5.3). The PCR bands with the highest intensity from each sample were isolated and sequenced.

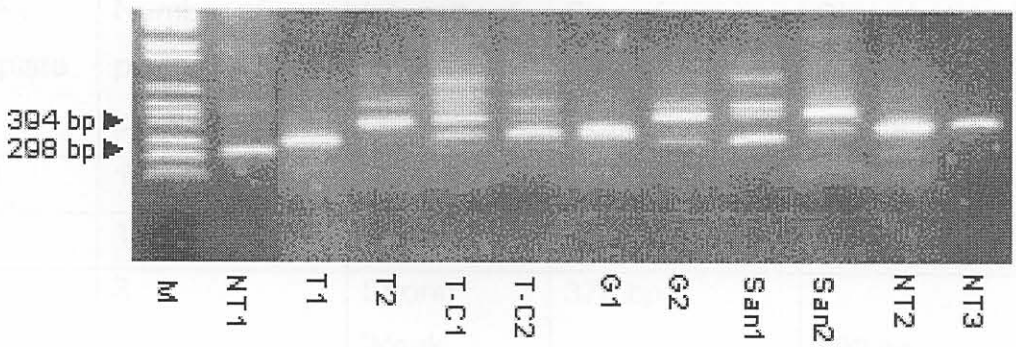

Figure 5.5: PCR amplification of genomic target DNAs with primers designed for flanking sequence of the Hi30 subtraction product. PCR amplification with genomic template DNA derived from tobacco cultivar '*Samsun/UK*' (NT1-3), with NT1 representing the original non-modified tobacco plant used in the RDA procedure, genetically modified tobacco plants carrying the *OC-I* transgene (T1 and T2), from an Agrobacterium transformation process but without a transgene insert (T-C1 and T-C2), from genetically modified tobacco plants (*'Samsun/UK'*) containing the *gus* gene insert (G1 and G2) and from tobacco cultivar '*Samsun/San1191'* (San1 and San2). Lane M represents molecular weight marker VI.

Table 5.3: Numerous PCR products amplified by using the primers Hi30R and H30H30, designed from the Hi30 subtraction product and the adjacent flanking sequence. PCR amplification with genomic template DNA derived from tobacco plants as outlined in Figure 5.5. The intensity of the amplified PCR products were considered as weak or strong when compared to the PCR band amplified in NT1.

DNA template	Number of prominent PCR products	Intensity of PCR products	Size of sequenced PCR products	Size of non-sequenced PCR products
NT1	1	Strong	234 bp	-
T1	1	Strong	278 bp	-
T2	3	Strong Weak Weak	378 bp	 500 bp 700 bp
T-C1	4	Strong Weak Weak Weak	375 bp	 270 bp 520 bp 650 bp
T-C2	3	Strong Weak Weak	278 bp	 450 bp 650 bp
G1	1	Strong	275 bp	-
G2	3	Strong Weak Weak	378 bp	 236 bp 400 bp
San1	4	Strong Weak Weak Weak	236 bp	 378 bp 400 bp 650 bp
San2	2	Strong Weak	378 bp	 275 bp
NT2	1	Strong	275 bp	-
NT3	1	Strong	278 bp	-

Sequence analysis of the Hi30 flanking sequences revealed a conserved 185 bp sequence fragment adjacent to the H30H30 primer site in the flanking sequence, detected in all amplified DNAs from the different plants tested (Figure 5.8). Adjacent to this conserved sequence region in the flanking sequence, is a variable sequence region, which differs in length and sequence between different individual plants. The shortest variable region (50 bp) was found for the flanking sequence of the original non-modified wild-type individual tobacco plant used in the RDA procedure and in as one of the non-modified '*Samsun/San1191'* plants. The second longest variable region (85 bp) was observed in two non-modified '*Samsun/UK'* tobacco plants, in one genetically modified tobacco plant carrying the *OC-I* transgene and in one tobacco plant, which originated from an Agrobacterium transformation process but without a transgene insert. The longest variable region (190 bp) was also present in one of the non-modified '*Samsun/San1191'* tobacco plants, in one genetically modified *OC-I* tobacco plant and in one tobacco plant exposed to an Agrobacterium transformation but without a transgene insert. However, none of the different regions could differentiate between the different types of plants investigated. Sequence data of the 185 bp conserved region of the flanking sequence showed significant homology to the 16S rRNA gene in *Shewanella* alga with a 184 bp overlap in identical base pairs (E = -93; Accession number: AF005250) as well as to a number of cloning vectors with up to 169 bp overlap in identical base pairs (E =-89; Accession numbers: AF 327711, PTZ19UCS etc.). However, sequence analysis of the variable regions showed no significant homology to known DNA sequences. Sequence data for the two genetically modified tobacco plants (*'Samsun/UK'*) containing the *gus* gene insert could not be obtained because the completed sequence with the exception of a few unclear base pairs at the start of the sequence showed consistently double peaks (Figure 5.6 and 5.7).

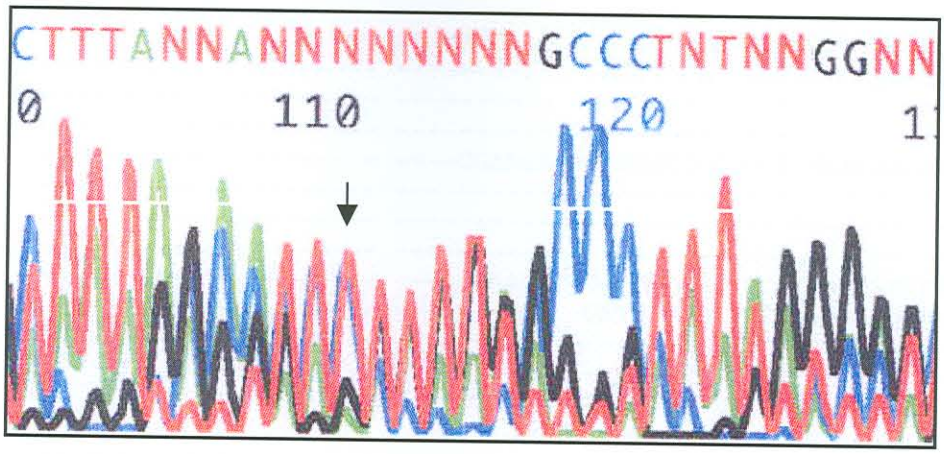

Figure 5.6: Electropherogram from sequencing of DNA template derived from a genetically modified tobacco plant (*'Samsun/UK'*) containing the *gus* gene insert (G1). Arrow indicate an example of a double peak.

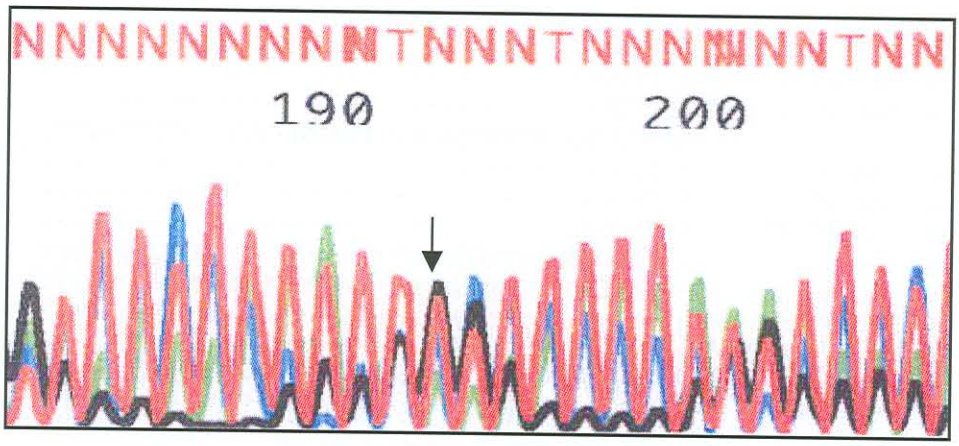

Figure 5.7: Electropherogram from sequencing of DNA template derived from a genetically modified tobacco plant (*'Samsun/UK'*) containing the *gus* gene insert (G2). Arrow indicate an example of a double peak

```
                10         20         30         40         50         60
A       ---------- ---------- ---------- ---------- ---------- ----------
Hi30R   ACCGACGTCG ACTATCCATG AACAACGAAG AGATGGCCGC GCTTGAAC-A AAGTATTG-A
Hi30A   ---------- ---------- ---------- --AGCTTCAT CCTATGGA-A TGATTTCC-C
Hi30B   ---------- ---------- ---------- ---------- ---------- ----------
T-C2    ---------- ---------- ---------- ---------- ---------- ----------
T-C1    ---------- ---------- ----CGAAGA AGATGGCCGC GCTTGAACAA AAGTATTGCA
NT3     ---------- ---------- ---------- ---------- ---------- ----------
NT2     ---------- ---------- ---------- ---------- ---------- ----------
T1      ---------- ---------- ---------- ---------- ---------- ----------
T2      ---------- ---------- -----GAAGA -GATGGCCGC GCTTGAACAA AAGTATTTCT
San2    ---------- ---------- -------AGA -GATGGCCGC GCTTGAACAA --GTATTGCA
San1    ---------- ---------- ---------- ---------- ---------- ----------
NT1     ---------- ---------- ---------- ---------- ---------- ----------

                70         80         90        100        110        120
A       ---------- ---------- ---------- ---------- ---------- ----------
Hi30R   AGAGACGACA CAATTGGCCA GGGATTTC-G GGCTCGATTT TTATCCGATG CGTTATGAGG
Hi30A   AAAACTCCAA TGTGAAAAGC GCGTCGGC-A TCCCGTAGGC CCCAAACGTG TAAATGACAT
Hi30B   ---------- ---------- ---------- ---------- ---------- ----------
T-C2    ---------- ---------- ---------- ---------- ---------- ----------
T-C1    AGAGACGACA CAATTGGCCA GGGATTTC-G GGCTCGATTT TTATCCGATG CGTTATGAGG
NT3     ---------- ---------- ---------- ---------- ---------- ----------
NT2     ---------- ---------- ---------- ---------- ---------- ----------
T1      ---------- ---------- ---------- ---------- ---------- ----------
T2      AGAGACGACA CAATTGGCCA GGGATTTCTC GGCTCGATTT TTATCCGATG CGTTATGAGG
San2    AGAGACGACA CAATTGGCCA GGGATTTC-G GGCTCGATTT TTATCCGATG CGTTATGAGG
San1    ---------- ---------- ---------- ---------- ---------- ----------
NT1     ---------- ---------- ---------- ---------- ---------- ----------

               130        140        150        160        170        180
A       ---------- ---------- ---------- -----GC-TC GGGAATGCCG ACGCGCTTTT
Hi30R   TTTGTCCGTC GGATGTCATT TACACGTTTG GGGCCTA--. ....-..... ..........
Hi30A   CCGACGGACA AACCTCATAA CGCATCGGAT AAAAAT.GAG CCCG.AAT.C CT.GC.AA..
Hi30B   ---------- ---------- ---------- -AGCTT.A.. C-T.TG.AAT GATTT.CCAA
T-C2    ---------- ---------- ---------- --AC-T.A.. --T.TG.AAT GATTT.CCAA
T-C1    TTTGTCCGTC GGATGTCATT TACACGTTTG GGGCCTA--. ....-..... ..........
NT3     ---------- ---------- ---------- --GC-T.A.. CTT.TG.AAT GATTT.CCAA
NT2     ---------- ---------- ---------- -----T.A.. C-T.TG.AAT GATTT.CCAA
T1      ---------- ---------- ---------- -AGC-T.A.. -TT.TG.AAT GATTT.CCAA
T2      TTTGTCCGTC GGATGTCATT TACACGTTTG GGGCCTA--. ....-..... ..........
San2    TTTGTCCGTC GGATGTCATT TACACGTTTG GGGCCTA--. ....-..... ..........
San1    ---------- ---------- ---------- ---------- ---------- ----------
NT1     ---------- ---------- ---------- ---------- ---------- ----------

               190        200        210        220        230        240
A       CACATTGCAG TTTTGGGAAA TCATT-CCAT AGGATGAAGC T----TGTTC ATGGATAGTC
Hi30R   .......G.. .......... .....-.... .......... .--------- ----------
Hi30A   GTGTCGT.TC ..CAATACTT .GT.CAAGCG C..CCATCT. .TCGT..... ..........
Hi30B   A.-C.CCA.- .G.GAAA.GC -GCG.CGGCA TCCCGT.G.. C----CCAAA CGT.TA.A.G
T-C2    A--C.CC-.A .G.GAAA.GC -GCG.CGGCA TCCCGT.G.. C----CCAAA CGT.TA.A.G
T-C1    .......G.. .......... .....A..GA C--------- ---------- -------...
NT3     A.-C.CCT.A .G.GAAA.GC AGCG.CGGCA TCCCGT.G.. C----CCAAA CGT.TA.A.G
NT2     A.-C.CTT.- .G.GAAA.GC -GCG.CGGCA TCCCGT.T.. C----CCAAA CGT.TA.A.G
T1      A.AC.CCT.- .G.GAAA.GC -GCG.CGGCA TCCCGT.G.. C----CCAAA CGT.TA.A.G
T2      .......G.. .......... .....A..GA C--------- ---------- -------...
San2    .......G.. .......... .....A..GA C--------- ---------- -------...
San1    -----A..TT .CC.TTATTC .TGA.CGTTC ..A.G.GC.A .----C.CGG GG--------
NT1     ------..TT .CC.-TATTC .-GA.CGTTC ..A-G.GCTA .----C.CGG GG--------
```

```
                250        260        270        280        290        300
A       GACGTCGGTA TCCATATGAC TAGTAGATCC TCTAGAGTC- GACCTGCAGG CATGCAAGCT
Hi30R   ---------- ---------- ---------- ---------- ---------- ----------
Hi30A   ........-- ---------- ---------- ---------- ---------- ----------
Hi30B   ACATC..ACG GA..A.CCT. ATAAC.CAT. GGATA.AAAT CGAGCC.GAA ATCC.TG..C
T-C2    ACATC..A-- .......... .......... .........- .......... ..........
T-C1    ...-----.. .......... .......... .........- .......... ..........
NT3     ACATC..A-- .......... .......... .........- .......... ..........
NT2     ACATC..A-- .......... .......... .........- .......... ..........
T1      ACATC..A-- .......... .......... .........- .......... ..........
T2      ...-----.. .......... .......... .........- .......... ..........
San2    ...-----.. .......... .......... .........- .......... ..........
San1    -T.ACT.AA. .......... .......... .........- .......... ..........
NT1     -TGAAA.A-. .......... .T...C.... .......AAA .......... ..........

                310        320        330        340        350        360
A       TTCCCTATAG TGAGTCGTAT TAGAGCTTGG CGTAATCATG GTCATAGCTG TTTCCTGTGT
Hi30R   ---------- ---------- ---------- ---------- ---------- ----------
Hi30A   ---------- ---------- ---------- ---------- ---------- ----------
Hi30B   AATTG.G.C. .CTC.TCA.. ACTTTG..CA A.CGCGGCCA TCTC.TCG.T G...A..GA.
T-C2    .......... .......... .......... .......... .......... ..........
T-C1    .......... .......... .......... .......... .......... ..........
NT3     .......... .......... .......... .......... .......... ..........
NT2     .......... .......... .......... .......... .......... ..........
T1      .......... .......... .......... .......... .......... ..........
T2      .......... .......... .......... .......... .......... ..........
San2    .......... .......... .......... .......... .......... ..........
San1    .......... .......... .TT.C..... .......... .......... ..........
NT1     .......... .......... .......... .......... .......... ..........

                370        380        390        400        410        420
A       GAAATCGTTA TCCGCTCACA ATTCCACACA ACATACGAGC CGGAAGCATA AAGTGTAAAG
Hi30R   ---------- ---------- ---------- ---------- ---------- ----------
Hi30A   ---------- ---------- ---------- ---------- ---------- ----------
Hi30B   AGTCGACG.C GGT------- ---------- ---------- ---------- ----------
T-C2    .....T.... .......... .......... .......... .......... .....-----
T-C1    .....T.... .......... .......... .......... .......... .....-----
NT3     .....T.... .......... .......... .......... .......... .....-----
NT2     .....T.... .......... .......... .......... .......... .....-----
T1      .....T.... .......... .......... .......... .......... .....-----
T2      .....T.... .......... .......... .......... .......... .....-----
San2    .....T.... .......... .......... .......... .......... .....-----
San1    .....T.... .......... .......... .......... .......... .....-----
NT1     .....T.... .......... .......... .......... .......... .....-----

                430        440        450        460        470        480
A       CCTGCGG-TG CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCCCT
Hi30R   ---------- ---------- ---------- ---------- ---------- ----------
Hi30A   ---------- ---------- ---------- ---------- ---------- ----------
Hi30B   ---------- ---------- ---------- ---------- ---------- ----------
T-C2    .......G.. ......A.-- ---------- ---------- ---------- ----------
T-C1    .......G.. ......A.-- ---------- ---------- ---------- ----------
NT3     .......-.. ......A.-- ---------- ---------- ---------- ----------
NT2     .......-.. ......A..A ---------- ---------- ---------- ----------
T1      .......-.. ......A.-- ---------- ---------- ---------- ----------
T2      .......-.. ......A.-- ---------- ---------- ---------- ----------
San2    .......G.. ......A.-- ---------- ---------- ---------- ----------
San1    .......-.. ......A--- ---------- ---------- ---------- ----------
NT1     .......G.. ......A.-- ---------- ---------- ---------- ----------
```

```
            490         500         510         520         530         540
A           TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAGCG CCCCACNCGA
Hi30R       ---------- ---------- ---------- ---------- ---------- ----------
Hi30A       ---------- ---------- ---------- ---------- ---------- ----------
Hi30B       ---------- ---------- ---------- ---------- ---------- ----------
T-C2        ---------- ---------- ---------- ---------- ---------- ----------
T-C1        ---------- ---------- ---------- ---------- ---------- ----------
NT3         ---------- ---------- ---------- ---------- ---------- ----------
NT2         ---------- ---------- ---------- ---------- ---------- ----------
T1          ---------- ---------- ---------- ---------- ---------- ----------
T2          ---------- ---------- ---------- ---------- ---------- ----------
San2        ---------- ---------- ---------- ---------- ---------- ----------
San1        ---------- ---------- ---------- ---------- ---------- ----------
NT1         ---------- ---------- ---------- ---------- ---------- ----------

                  550
A           AATCCTGGAC
Hi30R       ----------
Hi30A       ----------
Hi30B       ----------
T-C2        ----------
T-C1        ----------
NT3         ----------
NT2         ----------
T1          ----------
T2          ----------
San2        ----------
San1        ----------
NT1         ----------
```

Figure 5.8: Sequence alignment of original Hi30 flanking sequence with PCR products amplified with primers Hi30R and H30H30 using tobacco genomic DNA derived from different tobacco plants as template DNA. (A) Original flanking sequence DNA derived from amplification of genetically modified tobacco genomic DNA with primers H30R2 and Primer3 (Table 4.2). (Hi30R) DNA sequence of the Hi30 subtraction product in the reverse orientation. (Hi30A and Hi30B) DNA sequence of the Hi30 subtraction product aligned at different locations with the original flanking sequence represented in A. (T-C1 and T-C2) Tobacco plants ('*Samsun/UK*'), which went through a transformation process but not carrying a transgene. (NT2 and NT3) Genomic DNA-derived from non-modified wild-type individual tobacco plants of cultivar '*Samsun/UK*'. (T1 and T2) Amplification of genomic target DNA derived from genetically modified tobacco plants ('*Samsun/UK*') carrying the OC-I coding sequence. (San1 and San2) DNA from non-modified tobacco cultivar *'Samsun/San1191'*. (NT1) Genomic DNA-derived from the non-modified wild-type individual tobacco plant used in the RDA procedure. Primer 3 used in the amplification of the original flanking sequence (A) are underlined as well as the Hi30R ▒ and H30H30 ▒ primer sequences, Gap=(-), Identical base pairs aligned with A=(.).

C) *Hp12 subtraction product*

A range of different tobacco plants was further screened to identify possible DNA mutations present in the RDA subtraction product Hp12, which is part of the plastid genome. For that, PCR analysis was done using the primers Hp12A and Hp12B (Table 5.4). These primers were designed to end in a single base pair change, as found in the Hp12 subtraction product when compared to the known tobacco chloroplast sequence.

Table 5.4: Sequences of primers used for the PCR analysis of chloroplastic tobacco DNAs. Underlined base pairs indicate the base pair change present in the Hp12 subtraction product when compared to the know tobacco chloroplast sequence.

Name	Sequence
Hp12A	5' TTGTCTCGCGCCCCTAGGTAC 3'
Hp12B	5' TACCAGGCGCTACGGCGCTGG 3'

The original base pairs present in the known tobacco chloroplast genome were T and A in the sequence for Hp12A and Hp12B, respectively.

After 42 PCR cycles with the primer pair Hp12A and Hp12B, a DNA product of about 135 bp was amplified from all genetically modified and wild-type tobacco plants of the cultivar '*Samsun/UK*' with the strongest amplification in two plants carrying the *OC-I* coding sequence (T3 and T4) (Figures 5.9A and B). When genomic DNA of the non-modified wild-type tobacco cultivar '*Samsun/San1191*' was used as template DNA, a clear amplification product could only be detected in three of the four tested plants (San 1, San3 and San4), while only two of the three tested genetically modified tobacco plants expressing a *gus* gene, showed an amplification product (G1 and G2) (Figure 5.9A and B). However, when amplification products were directly sequenced only a minor variability could be detected between the DNA sequences from the different types of plants (Figure 5.10).

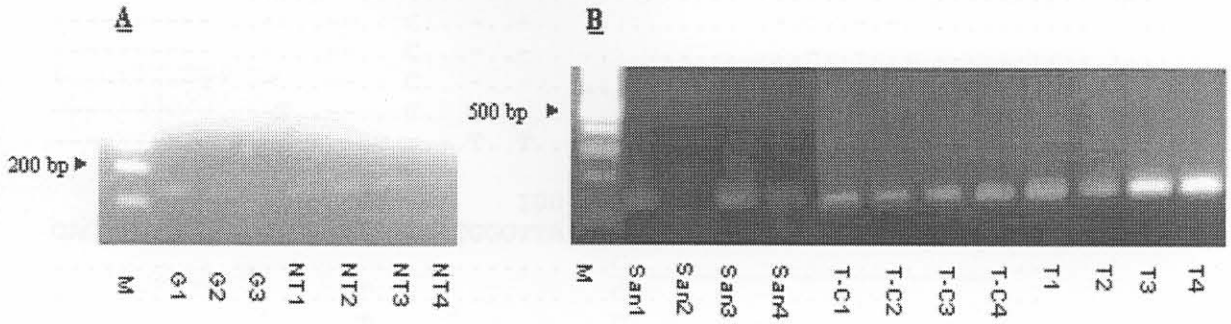

Figure 5.9: Amplification of genomic target DNA using PCR with primers designed for Hp12 with a single base pair change as identified in subtraction product Hp12. (A) DNA amplification with genomic DNA-derived from either non-modified wild-type individual tobacco plants of cultivar '*Samsun/UK*' (NT1 to NT4) (NT1 represents the original wild-type DNA used in the RDA procedure) and genetically modified tobacco plants ('*Samsun/UK*') containing a *gus* transgene (G1 to G3) as template. Lane M represents a 100 bp marker. (B) Amplification of genomic target DNA derived from genetically modified tobacco plants ('*Samsun/UK*') carrying the *OC-I* coding sequence (T1 to T4); tobacco plants ('*Samsun/UK*'), which went through a transformation process but not carrying a transgene (T-C1 to T-C4) and DNA from tobacco cultivar *'Samsun/San1191'* (San1-4). Lane M represents a 100 bp marker.

```
                10         20         30         40         50         60         70
CHL1    TAGGTATTCT CTACCTA-CC ACCT-GT-GT CGGTTTCGGG TACAGG-TAC CCTCTTGCTC AAAGTCGTTC
Hp12    ......C... .......C.. G...-..-.. .......... ......-... .......... ..........
T3      ---------- ..-....--.. C...-..-.. .......... .CAG..-..- .......... ..........
T2      ---------- ----..--N. C...-..-.. .......... ......-... .......... ..........
T1      ---------. ..-....--.. C...-..-.. .......... ......-... .......... ..........
T-C1    ---------- --T.N.--.. C...-..-.. .......... .-....-.-. .......... ..........
T-C2    ---------- ---T..--.. C...-..-.. .......... ..-...-... .-........ ..........
T-C3    ---------- ..-....--.. C...-..-.. .......... .-....-..- .......... ..........
G1      ---------- ..-....--.. C...-..-.. .......... ......-.-. .......... ..........
G2      ---------- ..-....--.. C...-..-.. .......... .-....-.-. .......... ..........
San1    ---------- ..-....--.. C...-..-.. .......... ...-..-... .......... ..........
San3    ---------. ..-....--.. C...-..-.. .......... ......G... .......... ..........
San4    ---------- ---T..--.. C...G..-.. .......... .-....-.-. .......... ..........
NT1     --------.. ..----.-.. ..-.T..T.. .......... ...-..-... .......... ..........

                80         90        100        110        120
CHL1    GAGCTTTTCC TGGGAGTACG GCATGGGTTA CTTCAGCGCC -GTAGCGCCT GGTAACC
Hp12    .......... .......... .......... ..C....... -.......... .......
T3      .......... ........T. .......... ..C....... -.......... .......
T2      .......... ........T. .......... ..C....... -.......... .......
T1      .......... ........T. .......... ..C....... G.......... .......
T-C1    .......... ........T. .......... ..C....... -.......... .......
T-C2    .......... ........T. .......... ..C....... -.......... .......
T-C3    .......... ........T. .......... ..C....... -.......... .......
G1      .......... ........T. .......... ..C....... G.......... .......
G2      .......... ........T. .......... ..C....... -.......... .......
San1    .......... ........T. .......... ..C....... -.......... .......
San3    .......... ........T. .......... ..C....... -.......... .......
San4    .......... ........T. .......... ..C....... -.......... .......
NT1     .......... ........T. .......... ..C....... -.......... .......
```

Figure 5.10: Sequence alignment of PCR amplification products with primers Hp12A and Hp12B and tobacco genomic DNA from different tobacco plants as template. Sequence data of tobacco chloroplast genome obtained from Blast search (CHL), Hp12 RDA subtraction product (Hp12), genomic DNA from genetically modified plants carrying a OC-I transgene (T1, T2 and T3); DNA from transformed plants without a transgene insert (T-C1, T-C2 and T-C3); genomic DNA from genetically modified tobacco plants carrying a *gus* transgene (G1 and G2); genomic DNA from wild-type tobacco *'Samsun/ San1191'* (San1, San3 and San4) and *'Samsun/UK'* (NT1). Hp12A and Hp12B primer sequences are underlined, Gap=(-). Identical base pairs aligned with the known tobacco chloroplast genome (CHL)=(.).

D) *rDNA repeat unit in tobacco plants*

A number of genetically modified and non-modified tobacco plants were screened to detect possible variation in the copy number of the total rDNA

repeat unit present in plants. A single-stranded probe was constructed by digesting the conserved rRNA repeat unit from flax with the restriction enzyme *Bam*HI. *Bam*HI-digested genomic DNA from different tobacco lines were transferred to a filter and hybridized with the constructed probe (Figure 5.11). The concentration of genomic *Bam*HI-digested DNA used to detect possible changes in RNA copy number were identical (1.5µg) for all plants, except for NT2, which had 3-times more genomic DNA digested. A maximum of five rDNA bands were found in the following tested tobacco plants: non-modified *'Samsun/UK'* tobacco and transgenic tobacco expressing a *gus* reporter gene (Figure 5.11B). Differences in the intensity of the hybridization products were found between the different tobacco plants with the exception of NT2 where an unequal amount of digested DNA was loaded onto the gel (Figure 5.11A). Intensity of hybridization products seemed to be higher in NT1 and San (Figure 5.11B).

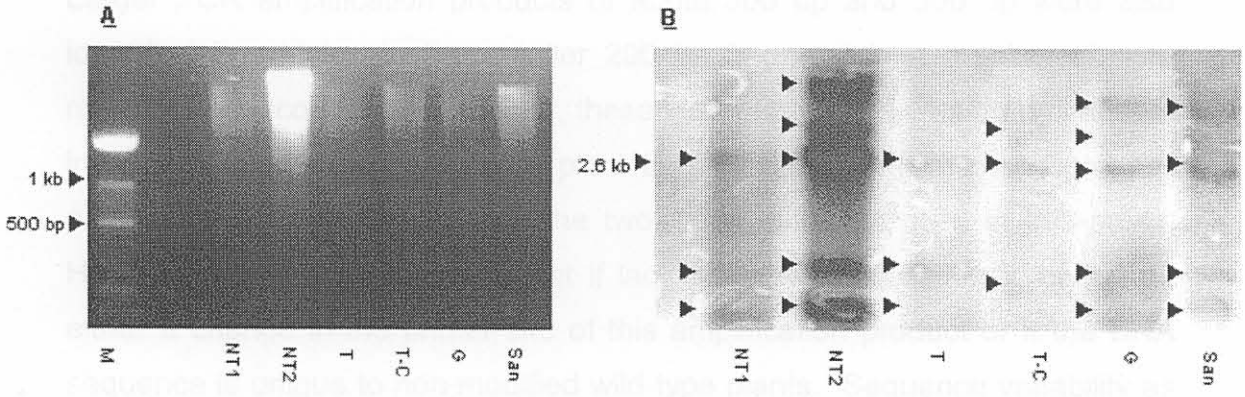

Figure 5.11: (A) Genomic *Bam*HI-digested DNA after separation on an agarose gel and stained with ethidium bromide from non-modified wild-type *'Samsun/UK'* plants (NT1 and NT2). NT1 represents the original non-modified tobacco plant DNA used in the RDA procedure; genetically modified plant DNA carrying and expressing a *OC-I* transgene (T); DNA from a transformed plant without a transgene insert but derived from a transformation process (T-C); genomic DNA from genetically modified tobacco plants carrying a *gus* transgene (G); genomic DNA from wild-type tobacco *'Samsun/ San1191'* (San). Lane M represent a 100 bp marker. (B) Hybridization of a conserved total rDNA repeat unit probe to a filter containing *Bam*HI-digested genomic DNA from genetically modified and non-modified plants as outlined under A.

Discussion

In this part of the study, the DNA regions isolated by RDA as subtraction products were characterized in greater detail by various bio-informatic tools. Further, the presence of subtraction products in a variety of genetically modified and non-modified tobacco plants was investigated. PCR analysis using primers for the Hi30 subtraction product resulted in the amplification of same size PCR products of about 200 bp in the majority of tested tobacco plants regardless if plants were genetically modified or not. The amplified 200 bp fragment contained also a 38 bp region, which was variable between plants, and changes consisted mainly of single base pair changes and deletions. The lack of detection of any amplification product in some plants further indicates a possible variability of the amplified product likely in the primer sites of the product disallowing DNA amplification.

Larger PCR amplification products of about 300 bp and 500 bp were also identified in addition to the smaller 200 bp fragment in a number of non-modified tobacco plants. However, these larger fragments was always absent in tobacco plants derived from a plant tissue culture/gene insertion process and could differentiate between the two types of plants used in this study. However, there is no evidence yet if lack of DNA amplification is based on either a change in the primer site of this amplification product or if the DNA sequence is unique to non-modified wild-type plants. Sequence variability as a consequence of a tissue culture process and specifically as a consequence of callus formation has been previously reported (Leroy et al., 2001). By sequence analysis carried out for the 300 bp amplification product, a DNA sequence was further identified in the fragment adjacent to the Hi30R primer site. This identified sequence was homologous to the sequence of the Hi30L primer. Detection of this almost homologous sequence in the fragment possibly indicates multiple primer location sites within the plant genome. In addition, the 200 bp fragment was identified as part of the larger 300 bp fragment very likely representing a repeated sequence in the plant genome. This result confirms the findings by Navin et al. (1996) that the RDA procedure is able to isolate repetitive sequences unique to only one of the compared

genomes. Also, Vorster et al. (2002) showed that a particular class of repetitive sequences could be isolated from date palms by applying the RDA technique.

PCR amplification using the primers Hi30R and H30H30 designed to bind within the Hi30 subtraction product and the adjacent flanking sequence resulted in the amplification of DNA products of various sizes in both plants derived from a tissue culture/gene insertion process and wild-type tobacco plants. A conserved region next to the H30H30 primer site was identified, which was present in all tested tobacco plants. Computer-based sequence analysis of this conserved region did not identify any sequence homology to tobacco DNA but revealed significant homology to the 16S rRNA gene of the alga *Shewanella* as well as partially to various cloning vectors. Although possible contamination of template DNA with foreign DNA cannot be completely excluded, the 16S rRNA is located within the chloroplast of both alga and higher plants, which might explain the sequence homology. Further, the complete Hi30 flanking sequence revealed homology to the 16S rRNA of alga but also to *Arabidopsis thaliana* mRNA for mitochondrial F1 ATP synthase beta subunit confirming the plant origin of the flanking sequence. It might be speculated, that these two sequences were connected during evolution and formed the sequence present in the mitochondrial Arabidopsis genome similar to the findings made by Kanazawa and Shimamoto (1999) for soybean chloroplast and mitochondrial DNA.

Adjacent to this conserved region a variable region was found, which differed in size (50, 85 and 190 bp) in individual plants. Since the conserved region was found in all fragments regardless of the size of the variable region, this conserved region seemingly represents a repeated sequence in the plant genome. This assumption is further supported by the sequence data obtained for two genetically modified tobacco plants expressing the *gus* gene. The eletropherogram of obtained sequences showed double peaks for individual base pairs indicating the sequencing of more than one identical individual sequence. However, the possibility of a double template used in the sequencing reaction can not be completely excluded although purification of

PCR products were carried out identically for all tested tobacco plants and detection of double peaks was only found in two of the tested plants.

Further, the intensity of amplified bands with identical size between the different plants and also the intensity of amplified bands with different sizes from individual plants varied. Differences in intensities might be due to an unequal number of copies of repeated elements at different locations in the genome of individual plants. A trend for production of multiple amplification fragments was evident for plants derived from a plant tissue culture/gene insertion process when compared to non-modified wild-type '*Samsun/UK*' plants used for plant transformation. Recently, 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 an unequal copy number of elements. Transformation as a stressful process might, therefore, have contributed to the appearance of such multiple fragments from repeated elements with different copy numbers in the genome. Since multiple fragments were also found in '*Samsun/San1191*' cultivar any rigorous selection process might also contribute to multiple fragment production.

By analyzing the Hp12 subtraction product, PCR analysis of amplification products amplified with changed primers to cover a single base mutation in the primer site showed minor variability in a 18 bp and 12 bp region in all tested tobacco plants. Two genetically modified tobacco plants carrying either the *OC-I* or *gus* reporter gene, showed an additional single base pair mutation in the Hp12B primer site. Application of the primers covering the base pair change resulted in the amplification of weak DNA products in the majority of plants indicating that chloroplast DNA of these plants do not have these single base pair mutations. Such DNA single base mutations have been found to be dispersed throughout the plant genome as a consequence of stress (Cassells and Curry, 2001; McClintock, 1984). These single base pair mutations might have further created new or destroyed existing restriction sites explaining creation/absence of the *Hpa*II restriction site at the junction between the Hp12 subtraction product and the flanking chloroplast DNA. Restriction site

modification might consequently have allowed the amplification of different length products in the RDA process.

Different tobacco plants were also screened in this part of the study to detect possible changes in the pattern of the coding regions for the total rRNA repeat unit in plants. Multiple ribosomal RNA genes are present in all eukaryotes with gene arrangement usually including tandem repeats of the transcription unit separated by non-transcribed spacers (Long and Dawid, 1980). The transcribed regions within the total rRNA repeat unit are conserved regions and are homologous in different plant species. Therefore, the available DNA for the rRNA repeat unit from flax (8.6 kb) can be used for detection of the rRNA coding regions in tobacco plants (Goldsbrough and Cullis, 1981). The amount of rDNA was generally lower in tobacco plants derived from a tissue culture/gene insertion process than in non-modified wild-type plants. This confirms results by Cullis (1976) showing heritable changes in the rDNA copy number induced by growth under different environmental conditions. However, comparable rDNA patterns with similar size hybridization products were found in the different tested tobacco plants but the number of hybridization products in the range varied between plants. The restriction enzyme *Bam*HI used for genomic DNA digestion recognizes the site GGATCC and is susceptible to methylation at the internal cytosine. This cytosine methylation might prevent complete DNA digestion resulting in the formation of different size hybridization products in tested plants. Smulders et al. (1995) already showed that a tissue culture process can change the methylation pattern of repetitive DNA in tomato calli. Based on available sequence data, methylation of the *Bam*HI site in tobacco would result in fragments of at least 2.5 kb and larger. Such larger size fragments were indeed found in this study (Figure 4.11). However, the possibility that a particular class of rDNA is not produced under a certain condition resulting in heterogeneity of tobacco rDNA repeats has also to be considered. Both processes might have ultimately allowed the isolation of subtraction products in RDA.

References

Arencibia A., Gentinetta E., Cuzzoni E., Castiglione S., Kohli A., Vain P., Leech M., Christou P. and Sala F. 1998. Molecular analysis of the genome of transgenic rice (*Oryza sativa* L.) plants produced via particle bombardment or intact cell electroporation. Molecular Breeding 4: 99-109.

Cassells A.C. and Curry R.F. 2001. Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micropropagators and genetic engineers. Plant Cell, Tissue and Organ Cuture 64: 145-157.

Choi H.W., Lemaux P.G. and Cho M.J. 2000. Increased chromosomal variation in transgenic versus nontransgenic barley (*Hordeum vulgare* L.) plants. Crop Science 40: 524-533.

Cullis C.A. 1976. Environmentally induced changes in ribosomal RNA cistron number in flax. Heredity 36: 73-79.

Ditt R.F., Nester E.W. and Comai L. 2001. Plant gene expression response to *Agrobacterium tumefaciens*. Proceedings of the National Academy of Science, U.S.A. 98: 10954-10959.

Goldsbrough P.B. and Cullis C.A. 1981. Characterisation of the genes for ribosomal RNA in flax. Nucleic Acids Research 9: 1301-1309.

Kanazawa A. and Shimamoto Y. 1999. Soybean recombination sites are present as dispersed segments in Arabidopsis and Liverwort mitochondrial DNA. Plant Molecular Biology Reporter 17: 19-29.

Karp A. and Bright S.W.J. 1985. On the causes and origins of somaclonal variation. Oxford Surveys of Plant Molecular and Cell Biology 2:199-234.

Labra M., Savini C., Bracale M. and Pelucchi N. 2001. Genomic changes in transgenic rice (*Oryza sativa* L.) plants produced by infecting calli with *Agrobacterium tumefaciens*. Plant Cell Reports 20: 325-330.

Leroy X.J., Leon K., Hilly J.M., Chaumeil P. and Branchard M. 2001. Detection of in vitro culture-induced instability through inter-simple sequence repeat analysis. Theoretical and Applied Genetics 102: 885-891.

Long E.O. and Dawid I.B. 1980. Repeated genes in eukaryotes. Annual Reviews in Biochemistry 49: 727-764.

Lynch P.T., Jones J., Blackhall N.W., Davey M.R., Bower J.B., Cocking E.C., Nelson M.R., Bigelow D.M., Orum T.V., Orth C.E. and Schuh W. 1995. The phenotypic characterisation of R_2 generation transgenic rice plants under field and glasshouse conditions. Euphytica 85: 395-401.

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

Navin A., Prekeris R., Lisitsyn N.A., Sonti M.M., Grieco D.A., Narayanswami S., Lander E.S. and Simpson E.M. 1996. Mouse Y-specific repeats isolated by whole chromosome representational difference analysis. Genomics 36: 349-353.

Ohba T., Yoshioka Y., Machida C. and Machida Y. 1995. DNA rearrangement associated with the integration of T-DNA in tobacco - an example for multiple duplications of DNA around the integration target. Plant Journal 7: 157-164.

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

Sala F., Arencibia A., Castiglione S., Yifan H., Labra M., Savini C., Bracale M. and Pelucchi N. 2000. Somaclonal variation in transgenic plants. Proceedings of the International Symposium On Methods and Markers for Quality Assurance in Micropropagation (eds). A.C. Cassells, B.M. Doyle, R.F. Curry Acta Horticulturae 530, ISHS.

Smulders M.J.M., Rus-Kortekaas W. and Vosman B. 1995. Tissue culture-induced DNA methylation polymorphisms in repetitive DNA of tomato calli and regenerated plants. Theoretical and Applied Genetics 91: 1257-1264.

Vorster B.J., Kunert K.J. and Cullis C.A. 2002. Use of representational difference analysis for the characterization of sequence differences between date palm varieties. Plant Cell Reports 21: 271-275.

Windels P., Taverniers I., Depicker A., Van Bockstaele E. and De Loose M. 2001. Characterisation of the Roundup Ready soybean insert. European Food Research Technology 213: 107-112.

Zheng S.J., Henken B., Sofiari E., Jacobsen E., Krens F.A. and Kik C. 2001. Molecular characterization of transgenic shallots (*Allium cepa* L.) by adaptor ligation PCR (AL-PCR) and sequencing of genomic DNA flanking T-DNA borders. Transgenic Research 10: 237-245.

Chapter 6: Summary and Perspective

One of the most challenging tasks in plant science is the understanding of the genome changes in plants that occur as a response to stress. There is accumulating evidence that introduction of foreign DNA molecules into the plant genome via an *in vitro* plant tissue culture process might be a stressful event carrying the risk of genome alterations. At the onset of this PhD study the basic assumption was therefore made, which was based on the detection of a genetically modified plant off-type, that the plant genome possesses plasticity and might reorganize in response to a stress induced by a plant tissue culture/gene insertion process. For this study, it was therefore considered important to obtain more detailed information about susceptible DNA regions that may change due to stress. Such regions might well have a hierarchy, in which they vary, and stress-induced changes might occur only in certain labile genome regions ultimately affecting plant performance and also compromise the bio-safety of plants.

As a first new result this PhD study allowed the identification of an unusual phenotype following a plant tissue culture process/gene insertion process using the Agrobacterium method and a tobacco cultivar, which has been previously widely selected for gene insertion approaches. The observed phenotype in a plant expressing a cystatin transgene, namely reduced stem elongation under growth conditions of low light intensities, has so far not been reported. However, this part of the study, which preceded the genetic analysis of genetically modified tobacco plants, did not provide sufficient evidence for the exact cause of the phenotypic change. Beside transgene interaction with metabolic processes, transgene gene insertion at a locus responsible for elongation, genetic variation present in the original inbreds, somaclonal variation due to the plant tissue culture or gene insertion process and changes in DNA methylation could not be ruled out.

This study therefore accomplished as a second new result the application of the RDA technique to identify and isolate possible putative genome changes from a complex tetraploid tobacco genome. RDA also allowed the detailed characterization of these changes with bio-informatic tools and extending the application of the technique to a further plant species. In general, execution of the RDA technique required only basic laboratory equipment and was relatively inexpensive and resulted in the isolation of three putative altered DNA sequences from tobacco plants derived from a tissue culture/gene insertion process. This study provided therefore first evidence that variability in these regions might be a direct result of a tissue culture/gene insertion process. Two of these variable regions were successfully identified as similar to part of the tobacco chloroplast genome and tobacco ribosomal RNA. In this regard, RDA has proved to be useful in identifying a particular repetitive class of sequences in tobacco. This is consistent with earlier observations that RDA can be used to isolate such families of repetitive sequences. The second chloroplast variable sequence might further confirm possible interchange of chloroplastic DNA with nuclear DNA. However, the possibility still exist that this DNA fragment originated from nuclear DNA that subsequently changed possibly due to a tissue culture/gene insertion process. Future research might therefore focus on the possibility that a plant tissue culture/gene insertion process stimulate the acceleration of chloroplast / nuclear transfer normally considered as a long-term evolutionary process. The third isolated sequence could not be matched to any previously reported genomic sequence and maybe unique to the tobacco genome.

A third new result of this study was the successful identification of adjacent sequence fragments of these variable DNA sequences in the plant genome. The plastid origin of two of the isolated DNA sequences was confirmed by matching the isolated library clones to known DNA sequences. By applying the Tail PCR method, DNA flanking sequences for the third RDA-derived sequence could also be determined. This sequence was partially homologous to sequences of

ribosomal RNA and to general cloning vectors. However, vector homology could not be linked to an Agrobacterium-derived sequence.

This study could further achieve as a fourth new result to provide first evidence that both the copy number and DNA sequence changes in certain variable regions of the tobacco genome. There is a trend to more genome variation in plants derived from a plant tissue culture/gene insertion process. Unclear is still why this trend was also found in a selection of tobacco cultivars. The question if any stressful event might result in similar genome changes found in this study has to be answered in a further future study.

None of the RDA-derived DNA sequences could be clearly linked in this study to detection of plants either derived from the plant tissue culture/gene insertion process or to plants with a phenotypic change expressing an exogenous cystatin transgene. Failure might be due to the very high degree of genomic identity between plants. Genetically modified plants might differ, if at all, only in a very small portion of the genome. Possible genome differences might be consequently too small, possibly consisting only of point mutations, to be easily detected by RDA. One should be also aware that a subtractive technology, such as RDA, is inherently subject to several sources of bias. The representation of the genome is based on digestion of the genomic DNA with single restriction enzymes. The genomic subset obtained depends, therefore, on the sequence of the restriction site and particularly its GC content. Further, tester/driver ratios used for subtractive hybridization are critical for the elimination of common regions and enrichment of specific sequences. Also, the initial representation is influenced by the size of the restriction enzyme-digested fragments from total genomic DNA, where larger fragments amplify less efficient by PCR than smaller fragments. These factors might ultimately have also accounted for the inability to isolate, for example, the inserted exogenous cystatin coding sequence by the RDA procedure although plants selected for the experiments clearly showed the expression and insertion of the transgene. In addition, a single or very low copy

number DNA difference sequence might also not have been efficiently amplified and enriched by the RDA protocol applied in this study.

Actions to overcome current failure of a clear identification, if possible at all, of plants from a stressful event, such as plant tissue culture/gene insertion process, might involve the usage of a greater range of different restriction enzymes for genome digestion. This will allow limiting the genome bias and the selectivity in the genome digestion step. By using different restriction enzymes several representations of the same genome can be scanned in each subtraction. So far the RDA technology has been developed only for four different restriction enzymes namely *Hind*III, *Bgl*II, *Bam*HI and *Hpa*II. However, many more restriction enzymes could possibly be used for genomic digestion, if they generate the same staggered ends, thereby allowing the use of already developed adaptors. Further, elimination of repetitive Further, elimination of repetitive DNA sequences with unequal copy numbers, which seemingly are controlled by stresses, in plants might improve the discovery and enrichment of stress-related very unique induced genome changes. Certain restriction enzymes, such as *Mse*I, are known to digest DNA quite frequently in retrotransposons (personal communication, M. van der Merwe) and should be evaluated for their potential in eliminating repeated DNA sequences before the production of RDA representations. A further future approach might also include using bulked amplicons for RDA. Although two different plants were bulked in this study, bulking of a greater number of plants of stressed and non-stressed plants might identify polymorphisms that are restricted to a particular group of individuals. Therefore, bulking a series of samples of the two different types of plants and then executing the subtraction might preferably identify group specific polymorphisms, rather than individual specific polymorphisms.

Annex

A) Materials and methods

Detailed description of techniques introduced into the laboratory.

Section I: Plant transformation and selection

Section II: Molecular techniques

Section I: Plant transformation and selection

Plant transformation

The standard procedure as outlined by Horsch et al. (1985) was followed to transform tobacco with *Agrobacterium tumefaciens* (C58pMP90) carrying the binary vector pKYOCI, which was obtained from L. Jouanin at INRA Versailles, France. The pKYOC-I plasmid encodes the gene for *OC-I* under the control of a double 35S promoter (P70) from cauliflower mosaic virus between the left border (LB) and right border (RB). Further present on the T-DNA are:

1. A Ω leader sequence for gene expression enhancement.
2. The *nptII* gene under the control of a 35S promoter (P35SNPTII) for kanamycin resistance used as a selectable marker for transgenic tissue /plants.
3. An intron-containing *gus* gene (*gus*) encoding β-glucuronidase (P35S GUSint) under the control of a 35S promoter as a specific and easily detectable reporter for plant transformation.

For plant transformation, tobacco leaf disks were cut from a fully expanded leaf of tobacco plant (*Nicotiana tabacum* L., var Samsun) using a scalpel and disks were submerged into the *Agrobacterium* solution. After blotting away with filter paper the excess of bacterial culture, disks were co-cultivated with bacteria for two days in the dark on a regeneration medium containing MS salts (Murashige and Skoog, 1962), 8 g/l bactoagar (Difco Laboratories, Detroit, USA) 20 g/l sucrose and 1 mg/ml BAP (6-Benzylaminopurine). After co-cultivation, the tobacco disks were transferred to an identical tissue culture medium, but containing 500 mg/l cefotaxime to inhibit further *Agrobacterium* growth and 150 mg/l kanamycin for selection of transgenic material cultivation. Plates carrying the leaf disks were placed into a growth room with a growth temperature of 25°C and a 16/8 h light/dark cycle. Leaf disks were transferred to new MS medium containing BAP, kanamycin and cefotaxime

after 4 weeks. This process was repeated until shoots derived from the leaf disks.

For root formation of selected, putative genetically engineered shoots, the shoots were placed on a medium containing half-strength MS medium, 100 mg/l kanamycin and 300 mg/l cefotaxime. Thirteen rooted and engineered shoots that grew on the kanamycin-containing medium and expressed *gus* were transferred to soil, grown in the greenhouse and tested for *OC-I* expression.

Analysis of genetically modified plant material

Histochemical GUS assay

Histochemical assay to screen for expression of β-glucurodinase (*gus*) activity in genetically modified plants was carried out according to the method of Jefferson et al. (1987). For analysis, leaf tissue was incubated in reaction buffer containing 50 mM NaH_2PO_4 (pH 7), 0.01% Tween 20, 10 mM Na_2EDTA and 0.3% (w/v) 5-bromo-4-chloro-3-indolyl glucuronide as a substrate for the enzyme. Tissue was incubated at 37°C overnight and screened for blue staining indicating *gus* activity under a stereo-microscope. A short treatment with methanol was carried out before visual examination to intensify the blue staining.

Detection of OC-I expression by PAGE

For *OC-I* determination, 0.2 g or 0.4 g of leaf material was homogenized in 10 mM Tris-HCl (pH 8) and after centrifugation for 10 minutes in a micro-centrifuge the extract was heated for 10 minutes at 75 minutes to denature proteins (Masoud et al., 1993). After heat treatment, the extract was centrifuged for 10 minutes to remove denatured protein and the supernatant

containing heat-stable *OC-I* was freeze-dried and dissolved in 0.02 ml of distilled water. For analysis, the *OC-I* containing extract was separated by polyacrylamide gel electrophoresis (PAGE) in the presence of sodium dodecyl sulfate (SDS) on a 15% gel as outlined by Laemmli (1970). Protein gels to detect *OC-I* expression were stained with Coomassie Brilliant Blue R 250 and the molecular mass markers were from Bio-Rad (Hercules, CA).

Detection of OC-I expression by immunoblotting

Western blots of denaturing PAGE to detect *OC-I* expression in genetically modified plants were done as described by Sambrook et al. (1989). *OCI*-antiserum, which has been raised against recombinant purified *OC-I* produced in rabbit, was obtained from D. Michaud at Lavall University, Canada. The antiserum was used as the primary antibody to detect *OC-I* on Hybond C extra membranes (Amersham Life Science, UK). Anti-rabbit IgG horseradish peroxidase conjugate (Amersham Life Science, UK) was used as the secondary antibody. The protein was detected with the help of the ECL KIT (Amersham Life Science, UK) through the initiation of a photo-reaction and fluorescence detection on a Biomax MR film.

Detection of OC-I expression by cysteine proteinase assay

Total cystatin activity in genetically modified tobacco plants was measured according to the method of Barrett et al. (1982). A plant extract (10 µl) was diluted in 500 µl of a solution containing 0.1% Brij 35 and 250 µl of a proteinase reaction buffer. For temperature equilibration and activation of the enzyme, the solution was placed at 30°C for 1 minute and after equilibration, 250 µl of a 20 µM of the cysteine proteinase substrate Z-Phe-Arg-Nmec was added to release after proteinase action the fluorescent compound 7-amino-4-methylcoumarin. After incubation for 10 minutes at 30°C, 1 ml of monochloroacetate stopping reagent was added and the fluorescence of the

free aminomethylcoumarin was determined in a fluorometer using 370 nm for excitation and 460 nm for emission.

Growth and selection of genetically modified plants

From all tested plantlets that derived from the transformation process, three plants were finally selected for self-fertilization because of their low endogenous cysteine proteinase activity, expressing *gus* and showing a band with the predicted size for *OC-I* on a SDS-PAGE and by immunoblotting. After seed collection (F_1 generation), 40 seeds of all three lines were tested for being genetically engineered by placing them on a half-strength MS medium containing 100 mg/l of kanamycin. Seeds that germinated and produced rooted dark-green plantlets on the antibiotic-containing medium were again tested for expression of *gus*, endogenous cysteine proteinase activity and presence of showing a band with the predicted size for *OC-I* on a SDS-PAGE. Three plants expressing *gus* and with low endogenous cysteine proteinase activity were selected. These plants were further tested by SDS-PAGE after heat treatment of plant extracts to remove the majority of proteins and concentrating the plant extract by freeze drying to confirm *OC-I* expression by detection of a protein band with a predicted size of about 12 kDa for *OC-I*. Three plants were finally selected based on their level of *OC-I* expression, which represented lines T4/3-1, T4/3-2 and T4/5. These plants were then self-fertilised again to produce the F_2 generation. Finally, 40 seeds produced by each of the three plants representing the different genetically engineered lines were again tested on a medium containing half-strength MS and 100 mg/l kanamycin to determine that at least 80% of seeds germinate and produce normal seedlings on the kanamycin containing medium. Two genetically modified plants growing on a kanamycin-containing medium and expressing *gus* of line T4/5 were randomly selected for representational difference analysis.

Section II: Molecular techniques

Plant material and growth

Transgenic tobacco plants (*Nicotiana tabacum* L.) used for molecular analysis derived from seeds supplied by Prof. Kunert at FABI/UP (genetically modified seeds of the cultivar 'Samsun'), the John Innes Institute, Norwich, UK (wild-type Samsun seeds) and the ARC Cotton and Tobacco Research Institute at Rustenburg, South Africa (wild-type 'Samsun' and CDL23 seeds). For genome analysis, seeds were germinated *in vitro* in a medium containing half-strength MS salts, 2% sucrose and 0.8% agar adjusted to pH 5.8. Plantlets were grown in a growth cabinet at 25°C in a 16h/8h light-dark cycle at 50 μmol m^{-2} s^{-1} light intensity to a height of about 10 cm and then harvested.

Genomic DNA isolation

Genomic DNA from tobacco plants to carry out the representational difference analysis (RDA) technique was extracted according to the method of Gawel and Jarret (1991). Tobacco leaves (1.5 g) were pre-chilled at -80°C and crushed in liquid nitrogen with a mortar and pestle. A pre-heated extraction buffer (10 ml) containing 100 mM Tris-HCl (pH 8); 1.4 mM NaCl; 20 mM EDTA; 0.1% mercaptoethanol and 2% CTAB was added to crushed leaf material. The mixture was incubated at 65°C for 30 min followed by addition of a chloroform:isoamyl alcohol (24:1) mixture (8 ml) and incubation of the homogenate for 15 min. After centrifugation for 5 min at 10 000 x g at room temperature, the aqueous phase was filtered through Miracloth to remove remaining cellular debris and an equal volume of ice-cold isopropanol was added to precipitate the DNA. The DNA was collected by centrifugation for 10 min at 4°C and the pellet was washed with 70% ethanol, dried and re-suspended in 250 µl sdH_2O (sterile distilled water). Any contaminating RNA was removed by addition of 2.5 µl of a 10 µg/ml stock solution of RNase and incubation at 37°C for 30 min. DNA was recovered by the addition of 1/10

volume of 3M sodium acetate (pH 6.8) and 2 volumes of 96% ethanol to the DNA containing solution as outlined by Sambrook et al. (1989) and finally dissolved in 200 µl sdH_2O.To test for 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 gel with ethidium bromide for 15 min, DNA quality was determined on a white/UV-transilluminator, photographed with a Grab-IT system (Vacutec, USA) and the DNA concentration of samples was visually determined using 4 different λ DNA amounts (25 ng, 50 ng, 100 ng and 250 ng DNA) for comparison.

Polymerase chain reaction (PCR)

Standard DNA amplifications by PCR were carried out in a 100 µl reaction mixture containing 500 mM KCL, 25 mM $MgCl_2$, 100 mM Tris-HCL, (pH 8.3), 25 mM dNTPs and 5U TaKaRa Taq DNA polymerase (Takara, City, Japan) in a GeneAmp PCR 9600 system (Perking Elmer, Palo Alto). Primers for PCR were designed using the online tools of Molecular Biology Shortcuts (MBS), program "Oligos and Primers" (www.mbshortcuts.com/biotools/index.htm). Primers used in this study were commercially purchased from MWG -Biotech AG (Germany). The standard PCR program consisted of 94°C (5 min) to denature the DNA. This was followed by 35 to 42 cycles of amplification consisting of denaturing DNA at 94°C (1 min), primer annealing at 55°C or higher depending on the primer pair (1 min), and extension of the DNA chain at 72°C (2 min). Extension at the last cycle was at 72°C for 7 min, and optional soak period at 4°C. Amplification products were separated on a 1.5 % agarose gel, stained with ethidium bromide and visualized under UV light.

Southern blot analysis

For Southern blot analysis, the general outline by Sambrook et al. (1989) was followed. Total isolated DNA was digested with a respective restriction

enzyme, digested DNA was run on a 1% agarose gel in TAE buffer and then blotted onto a membrane. For probe labeling, the Gene Image random prime labeling kit was used (Amersham Life Science, UK). Labeled probes were hybridized to blotted DNA, which was pre-hybridized and hybridized at 60°C in a hybridization buffer containing 5% SSC, 0.1% SDS and 20-fold dilution of the liquid block provided overnight and washed at 60°C using a 1% SSC and 0.1% SDS solution followed by incubation in a liquid blocking solution as recommended by the supplier. Membranes were then incubated with a 5000-fold diluted anti-fluorescein-AP conjugate to obtain a fluorescence signal. After washing, fluorescence signals on the membrane were detected using a Gene Images CDP-Star detection kit (Amersham Life Science, UK). The membranes were finally exposed to Hyperfilm ECL (Amersham Pharmacia Biotech, UK) and the films developed which was followed by exposure to an X-ray film.

DNA sequencing

DNA sequencing analysis was carried out with the dideoxy chain termination method developed by Sanger et al. (1977). Recombinant plasmids were sequenced by primer walking using fluorescent dye terminators and AmpliTaq in a cycle sequencing protocol according to the recommendations of the manufacturer on a ABI377 automatic DNA sequencer (PE Applied Biosystems). Correctness of DNA sequences was confirmed by GATC Biotech AG (Germany). Sequence comparisons and database searches were done with the basic local alignment search tool for fast database searching (BLAST). BLAST emphasize regions of local alignment to detect relationships among sequences which share only isolated regions of similarity. The program, blastn was used to compare a nucleotide query sequence against a nucleotide sequence database. The database nr contain all non-redundant GenBank + EMBL + DDBJ + PDB sequences (but no EST, STS, GSS, or HTGS sequences).

Representational difference analysis (RDA)

Restriction endonuclease digestion of genomic DNA

For RDA, the technique outlined by Lisitsyn et al. (1993) with some modifications was followed. Figure A.1 outlines the single steps involved in the process.

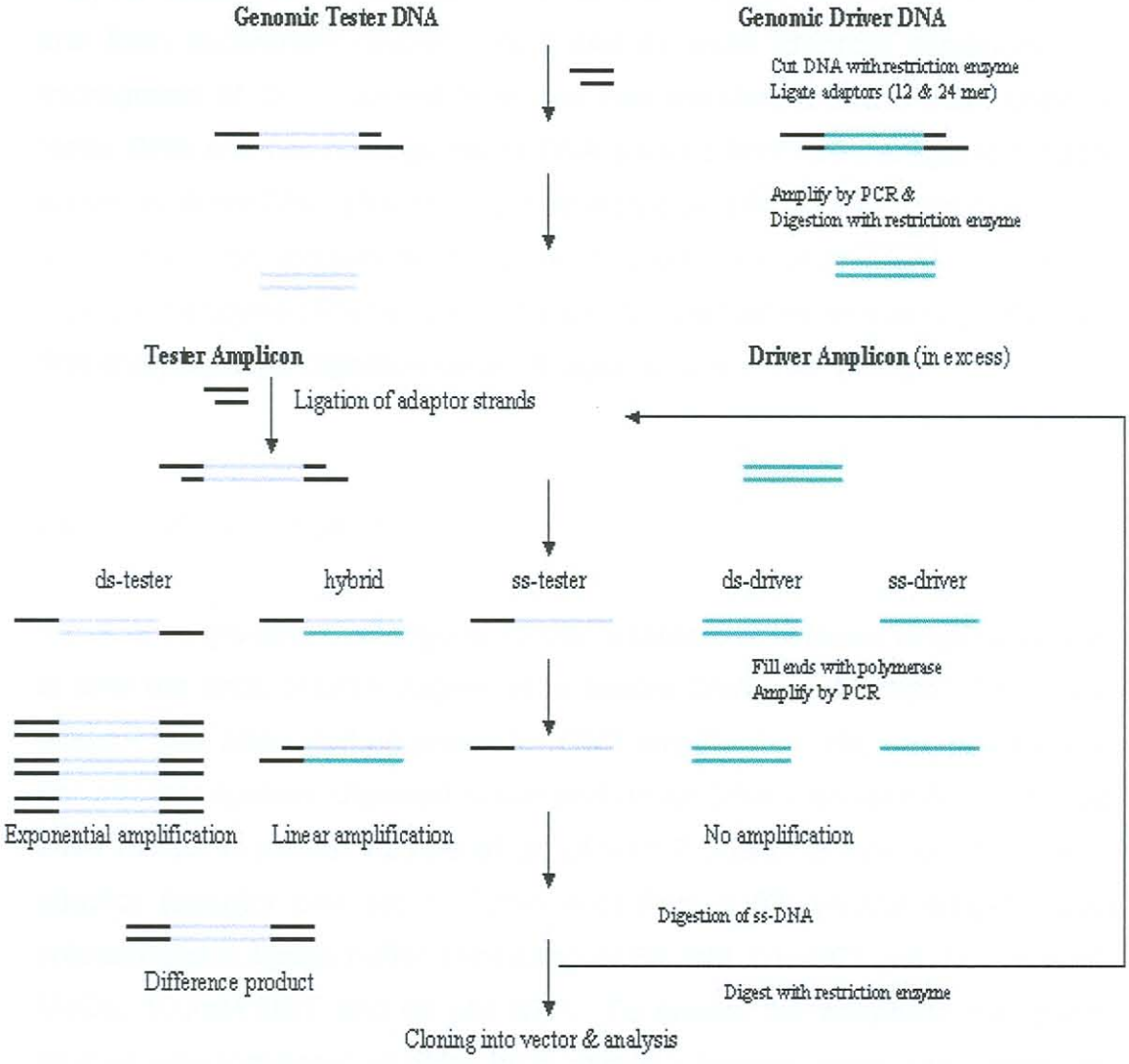

Figure A.1: Steps involved in the DNA subtraction procedure of Representational Difference Analysis.

To produce the representations for RDA, total cellular DNA was isolated from 1 g of tobacco leaves (Gawel and Jarret, 1991) and digested with the restriction enzymes *Hind*III (methylation-insensitive) and *Hpa*II (methylation-sensitive). For production of amplicons (representations), two micrograms of DNA derived from one non-modified tobacco plant used as driver DNA and two micrograms of DNA derived from two genetically modified tobacco plants as tester DNA after mixing their individual DNAs (1:1), were digested in the first an third experiments (experiment 1 and 3) with 80 units of either the enzyme *Hpa*II (methylation-sensitive) or the enzyme *Hind*III. In the second and forth experiment (experiment 2 and 4) under identical conditions, two micrograms of DNA derived from one non-modified tobacco plant used as tester DNA and two micrograms of DNA derived from two transgenic tobacco plants as driver DNA after mixing their individual DNAs (1:1) were digested in 50 µl digestion mixture at 37°C for 1 h with 80 units of an appropriate restriction enzyme (Roche, Switzerland). Before further processing, DNA was first analyzed after digestion on a 1% agarose gel in TAE buffer.

Ligation of DNA adaptors

A pair of single-stranded oligonucleotide adaptors of different length was used to alter the ends of DNA fragments to enable DNA amplification. The longer adaptor was also used as primer for DNA amplification after adaptor ligation. For adaptor ligation, digested tester and driver DNA (between 0.5 and 1 µg) were mixed in a total volume of 30 µl with 7.5 µl of a 12-mer and 24-mer adaptor (adaptor pair set 1, Table A.1) from a 62 pmol/µl adaptor stock solution and a ligase buffer consisting of 66 mM Tris-HCl (pH 7.6); 6.6 mM $MgCl_2$; 10 mM DDT and 66 µM ATP. To anneal the adaptors, the ligation mixture was incubated at 55°C for 5 min in a heating block after which, the block was immediately placed into a cold room for approximately 1h until the temperature dropped in the ligation mixture to 10-15°C. The reaction tubes were incubated on ice for 3 min after which 4 µl (1U/µl) of T4 DNA ligase (Amersham Life Science, UK) was added to the mixture and then incubated overnight at 16°C for ligation.

Amplification of tester and driver DNA

For preparation of tester and driver amplicons by PCR, ligated DNA was diluted with 500 µl TE buffer containing 10 mM Tris-HCl (pH 8) and 0.1 mM EDTA. For DNA amplification, a PCR tube containing an PCR amplification mixture (100 µl), which contained 40 ng of ligated DNA; 372 pmol of the 24-mer adaptor (adaptor pair set 1, Table A.1); 10 mM dNTP's (4 µl); 25 mM $MgCl_2$ (6 µl) and PCR buffer consisting of 50 mM KCl; 10 mM Tris-HCl (pH 8.3); 1.5 mM $MgCl_2$ and 0.001% w/v gelatine, was placed into a pre-warmed (72°C) thermocycler (GeneAmp PCR System, Perkin Elmer, USA). To fill-in the 3'-recessed ends of the ligated fragments 6 units of Taq DNA polymerase (Amersham Life Science, UK) were added to the PCR amplification mixture. DNA amplification by PCR was carried out for 25 cycles (11 sec at 95°C; 2:07 min at 72°C) with the last cycle followed by a DNA extension period for 10 min at 72°C. Approximate total amount of DNA of amplified tester and driver amplicon was determined on a 1.5% agarose gel in TAE buffer with sheared herring sperm DNA as a standard. Amplified DNA was phenol/chloroform purified and after ethanol precipitation (Sambrook et al., 1989) amplicon DNA was dissolved in TE buffer to obtain a DNA concentration of about 0.5µg/µl.

To cleave adaptors from amplified DNA, driver DNA and tester DNA (40 µg) were digested for 1 h at 37°C with the initially selected restriction enzyme (20 units enzyme/µg DNA). Yeast tRNA (10 µg) was added to digested DNA, which was phenol/chloroform purified, ethanol precipitated and finally dissolved in 70 µl of sdH_2O.

The tester amplicon DNA (1µg) from which adaptors were cleaved was then ligated to a second adaptor pair (adaptor pair set 2; Table A.1) following the procedure outlined under "Ligation of DNA adaptors". Ligated tester amplicon DNA was then amplified following the procedure outlined under "Amplification of DNA by PCR and adaptor removal" but with the addition of 10 extra DNA amplification cycles (35 cycles).

Kinetic enrichment of DNA

Ligated tester DNA was diluted to 50 ng/µl in a total volume of 70 µl with TE buffer (10 mM Tris-HCl, pH8; 0.1 mM EDTA). For hybridization, diluted tester DNA (4 µl) was mixed with driver amplicon DNA (8 µl) and then 10 M ammonium acetate (3 µl) solution and 96% ethanol (38 µl) were added to the two DNAs and mixed with DNAs by sucking and blowing using an Eppendorf pipette. The mixture was chilled at -70°C for 10 min followed by an incubation period of 2 min at 37°C. DNA was precipitated by centrifugation for 10 min at 13 000 x g and the DNA containing pellet was washed twice with 70% ethanol and dried. The DNA pellet was resuspended in 4 µl EE buffer containing 30 mM EPPS (N-(2-hydroxyethyl piperizine)-N-(3-propene sulfonic acid) (pH 8) and 3 mM EDTA. The DNA was overlaid with 35 µl of sterile mineral oil and the sample was incubated at 98°C for 4 min to denature the DNA. A 5 M sodium chloride solution (1 µl) was directly injected into the DNA drop and the mixture was incubated at 67°C overnight.

The mineral oil was removed and tRNA (10 µg) was added to hybridized DNA and the DNA sample was diluted by adding 100 µl TE buffer to the mixture. To fill the adaptor ends, diluted hybridized DNA (20 µl) was added to 180 µl standard PCR reaction mixture as outlined under "Polymerase chain reaction (PCR)". The solution was divided into 2 separate PCR tubes and 1µl of Taq DNA polymerase was added in each tube. The solution was incubated at 72°C for 5 min after which 5 µl of 24-mer primer (adaptor pair set 2; Table A.1) was added to the solution. Ten cycles of PCR (1 min at 95°C and 3 min at 70°C) were performed using an extension at 70°C for 10 min after the last cycle. To evaluate the effectiveness of hybridization step, 20 µl of the hybridization mixture was amplified for an additional 20 cycles of amplification and any amplification products were visualized on a 2% agarose gel in TAE buffer. If amplification products were visible, 20 µl of the hybridization mixture was digested with 20 units of mung bean nuclease at 30°C for 30 min. The reaction was stopped by the addition of TE buffer (160 µl). The digested product was amplified in a standard PCR reaction mixture containing 6 µl of

the 24-mer primer (adaptor pair set 2; Table A.1). Amplified DNA subtraction products were purified with phenol/chloroform and precipitated with ethanol and finally dissolved in 100 µl of sdH_2O.

For a second round DNA subtraction/kinetic enrichment the DNA subtraction products (5 µg) was digested with 100 units of an appropriate restriction enzyme in a total volume of 100 µl. The DNA was phenol/chloroform purified after addition of tRNA (10 µg), ethanol precipitated and resuspended in sdH_2O to obtain a concentration of 20 µg DNA/ml. DNA (100 ng) was ligated to a third set of adaptors (adaptor pair set 3; Table A.1) in a total volume of 30 µl as described above. To ligated DNA 50µl of sdH_2O containing tRNA (20 µg/ml) was added so that the mixture (80 µl) contained about 100 ng of DNA. DNA hybridization and kinetic enrichment was carried out with 50 ng ligated DNA (40 µl) and an appropriate amount of driver amplicon DNA as described above. For a third round DNA subtraction/kinetic enrichment the procedure was repeated but using a fourth adaptor pair set or reusing the first adaptor pair set.

Cloning of DNA subtraction products

RDA subtraction products were treated with appropriate restriction enzymes to remove ligated adaptors, separated on a 1.5% agarose gel in TAE buffer and visualized by ethidium bromide staining. DNA fragments were eluted from the agarose gel and purified using a Sephaglas BandPrep Kit following the protocol given by the supplier (Pharmacia Biotech, USA). Purified DNA fragments were cloned into the *Eco*RV vector pMOSBlue according to protocol of the supplier (Amersham Life Science, UK) with a ligation buffer (20 µl) consisting of 66 mM Tris-HCl, pH 7.6, 6.6 mM $MgCl_2$, 10 mM DDT, 66 µM ATP, and 2 U T4 DNA ligase. Ligation was done at 16°C overnight and MOSBlue competent cells (Amersham Life Science, UK) were transformed with ligated plasmid DNA by heat shock treatment of cells for 40 seconds at 42° in a standard procedure as outlined by Sambrook et al. (1989).

Transformed cells were plated onto LB (Luria Bertani) agar plates containing 10 g/l NaCl, 10 g/l tryptone and 5g/l yeast extract. Plates were supplemented with 100 µg/ml ampicillin, 50 µl 10% X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and 10 µl 100 mM IPTG (isopropyl-β-D-thiogalactopyranoside) to allow blue/white selection of colonies. White colonies containing the cloned DNA fragments were randomly picked and after plasmid purification from these colonies according to the method outlined by Sambrook et al. (1989), cloned DNA fragments were analyzed after restriction enzyme digest with *Bam*HI and *Hind*III by gel electrophoresis on a 1% agarose gel in TAE buffer.

Table A.1: Representational difference analysis adaptors.

Adaptor Pair Set	Name	Sequence
1	R Hind 24	5' AGC ACT CTC CAG CCT CTC ACC GCA 3'
	R Hind 12	5' AGC TTG CGG TGA 3'
2	J Hind 24	5' ACC GAC GTC GAC TAT CCA TGA ACA 3'
	J Hind 12	5' AGC TTG TTC ATG 3'
3	N Hind 24	5' AGG CAG CTG TGG TAT CGA GGG AGA 3'
	N Hind 12	5' AGC TTC TCC CTC 3'
1	R Hpa 24	5' AGC ACT CTC CAG CCT CTC ACC GAC 3'
	R Hpa 11	5' CGG TCG GTG AG 3'
2	J Hpa 24	5' ACC GAC GTC GAC TAT CCA TGA AAC 3'
	J Hpa 11	5' CGG TTT CAT GG 3'
3	N Hpa 24	5' AGG CAA CTG TGC TAT CCG AGG GAC 3'
	N Hpa 11	5' CGG TCC CTC GG 3'
4	S Hpa 24	5' ACT TCT ACG GCT GAA TTC CGA CAC 3'
	S Hpa 12	5' CGG TGT CGG AAT 3'

Construction of a genomic DNA library

Library construction

Two micrograms of genetically modified genomic tobacco plant DNA was digested with 60 units of the restriction enzyme, *Bam*HI and ligated into the *Bam*HI site of a pre-digested λZAP Express vector, which is part of the ZAP Express Predigested Vector Kit (Stratagene, USA). Packaging extracts were used to package the recombinant lambda phage following the instruction of the manufacturer (Gigapack III Gold Packaging Extract; Stratagene, USA). Of the resulting library, 3.0 x 10^5 plaque forming units (pfu) were plated onto NZY agar plates containing 5 g/l NaCl, 2 g/l $MgSO_4x7H_2O$, 5 g/l yeast extract, 10 g/l casein hydrolysate and 15 g/l agar (pH 7.5), using XL1-Blue MRF' bacteria strain as a phage host and incubated overnight at 37°C.

Library amplification

The library was amplified to prepare a large, stable quantity of a high-titer stock of the library. Aliquots of the library suspension containing 5 x 10^4 pfu of bacteriophage were plated out on 150 mm NZY agar plates and incubated overnight at 37°C. The plates were overlaid overnight with SM buffer consisting of 5.8 g/l NaCl, 2g/l $MgSO_4x7H_2O$, 1 M Tris-HCl (pH 7.5) and 2% gelatine to allow the phage to diffuse into the SM buffer. The bacteriophage suspension from each plate was then pooled into a sterile container and cell debris was removed by centrifugation for 10 min at 500 x g. The supernatant was recovered and transferred to a sterile polypropylene tube.

Plaque lifting

The library was plated out at 50 000 pfu/plate on large 150 mm NZY agar plates and incubated overnight at 37°C. A nitrocellulose membrane

(Stratagene, USA) was placed onto each NZY agar plate for 2 minutes to transfer the phage particles to the membrane. The plates were chilled at 4°C for 1 h before placement of membranes onto the agar to prevent the agar from sticking to the nitrocellulose membrane. A needle was used to prick through the membrane and agar for orientation. The membrane was denatured in a solution of 1.5 M NaCl and 0.5 M NaOH for 2 min, which was followed by neutralization for 5 min in 1.5 M NaCl and 0.5 M Tris-HCl, pH 8. The membrane was rinsed for 30 sec in a solution containing 0.2 M Tris-HCl (pH 7.5) and 2 x SSC solution buffer. The DNA was finally cross-linked to the membrane using an UV transilluminator.

Library screening

The genomic DNA library was screened by Southern blot analysis. Three DNA probes constructed from respective DNA subtraction products were labelled with a Gene Images random prime labelling kit (Amersham Life Science, UK) and used for screening. Any positive clones were picked and excised from the ZAP express vector as a recombinant pBK-CMV phagemid plasmid (Stratagene, USA). *In vivo* excision of the pBK-CMV phagemid vector was provided with the help of the ExAssist helper phage, which contains an amber mutation to prevent replication of the phage genome in the non-suppressing *E. coli* strain, *XLOLR*, supplied with the kit. Dilutions of the excised pBK-CMV phagemid vector were mixed with 200 µl *XLOLR* cells and incubated at 37°C for 15 min. After addition of 300 µl NZY broth, the mixture was incubated at 37°C for 45 min. Cell mixtures were plated onto LB plates containing 50 µg/ml kanamycin and incubated overnight at 37°C. Plasmids of individual colonies were confirmed to contain inserts by digestion of plasmid DNA with the restriction enzyme *Bam*HI restriction and detection of DNA fragments by gel electrophoresis on a 1% agarose gel in TAE buffer.

Tail PCR

Amplification of flanking sequences

To determine flanking sequences of RDA subtraction products in the genome, a two-step PCR reaction technique was applied using for amplification a biotinylated primer and degenerated primers. Amplified biotinylated DNA fragments were isolated with a Dynabeads M-280 Streptavidin system (Sorensen et al., 1993). Figure A.2 shows the procedure applied. All PCR reactions were carried out using the PCR protocol outlined under "Polymerase chain reaction" but with 42 cycles of amplification and primer annealing at 62°C. Genomic DNA was used as template in the first PCR reaction, whereas an aliquot from the first PCR reaction was used as a DNA template in the second PCR reaction.

Isolation of amplification products

After the first PCR reaction using a biotinylated primer, amplified, biotinylated DNA fragments were isolated by mixing 40 µl of the PCR mixture with 40 µl of 200 µg pre-washed Dynabead M280-streptavidin as recommended by the supplier (Dynal Biotech, Norway). Biotinylated DNA fragments were removed from the mixture using a Dynal magnetic particle concentrator. All binding and washing steps were done in the presence of a binding and washing buffer consisting of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 2 M NaCl. After incubation for 15 min to remove the biotinylated DNA fragments from the mixture and washing in buffer, the Dynabead-bound DNA fragments were "melted" in 8 µl of 100 mM NaOH for 10 min. The supernatant containing the non-biotinylated strands was then neutralised with 4 µl of 200 mM HCl and 1 µl 1 M Tris-HCl, pH 8. After filling up to 30 µl with sdH_2O, 2 µl of the mixture was used as a DNA template in a second PCR reaction using a specific primer pair for amplification. Amplified and agarose gel-purified DNA fragments bands were finally cloned into the vector pMOSBlue and the

sequence of the cloned DNA fragments was finally analysed on an automated DNA sequencer.

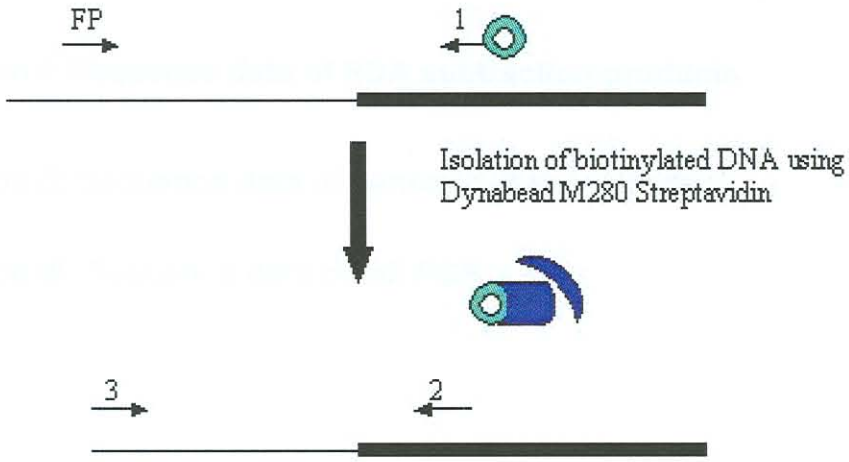

Figure A.2: Isolation of flanking DNA sequences in the genome by a two-step polymerase chain reaction and amplification of unknown DNA sequences (thin line) flanking a known DNA region (broad line). Arrows indicate biotinylated primer (primer 1), specific sequence primers (primers 2 and 3) and degenerated primers (primer FP). Green circle indicates biotin coupled to the 5' of the primer and both blue square and half moon indicate beads with streptavidin covalently bound to their surface.

B) Sequence data

General outline of all obtained DNA sequence data.

Section I: Sequence data of RDA subtraction products

Section II: Sequence data of genomic library clones

Section III: Sequence data of tail PCR

Section I: Sequence data of RDA subtraction products

```
                10         20         30         40         50         60
Hp12    GGAGGAGGCT AGGNTTAGCA CGAAAGATGG TTATCGGTTC AAGAACGTAA GGTGTCCCTG

                70         80         90        100        110        120
        CTTTGTCAGG GTAAGAAGGG GTAGAGAAAA TGCCTCGAGC CAATGTTCGA ATACCAGGCG

               130        140        150        160        170        180
        CTACGGCGCT GGAGTAACCC ATGCCGTACT CCCAGGAAAA GCTCGAACGA CTTTGAGCAA

               190        200        210        220        230        240
        GAGGGTACCT GTACCCGAAA CCGACACAGG CGGTAGGTAG AGAGTACCTA GGGGCGCGAG

               250        260        270        280        290        300
        ACAACTCTCT CTAAGGAACT CGGCAAAATA GCCCCGTAAC TTCGGGAGGA GGGGTGCCTC

               310        320        330
        CTCACAAAGG GGGTCGCAGT GACCAGGCCC G
```

```
                10         20         30         40         50         60
Hp14    CGGGCCTGTC GGCCAAGGCT ATAAACTCGT TGAATACATC AGTGTAGCGC GCGTGCGGCC

                70         80         90        100        110        120
        CAGAACATCT AAGGGCATCA CGGACCTGTT ATTGCCTCAA ACTTCCGCGG CCTAAAAGGC

               130        140        150        160        170        180
        CGTAGTCCCT CTAAGAAGCT GGCCGCGAAG GGATACCTCC GCATAGCTAG TTAGCAGGCT

               190        200        210        220        230        240
        GAGGTCTCGT TCGTTAACGG AATTAACCAG ACAAATCGCT CCACCAACTA AGAACGGCCA

               250        260        270        280        290        300
        TGCACCACCA CCCATAGAAT CAAGAAAGAG CTCTCAGTCT GTCAATCCTT ACTATGTCTG

               310        320        330        340        350        360
        GACCTGGTAA GTTTCCCCGT GTTGAGTCAA ATCAAGCCGC AGGCTCCACT CCTGGTGGTG

               370        380        390        400        410        420
        CCCTTCCGTC AATTCCTTTA AGTTTCAGCC TTGCGACCAT ATTCCCCCCA GAACCCAAAA

               430        440
        ACTTTGATTT CTCATAAGGT GCCG
```

```
                10         20         30         40         50         60
Hi30    AGCTTCATCC TATGGAATGA TTTCCCAAAA CTCCAATGTG AAAAGCGCGT CGGCATCCCG

                70         80         90        100        110        120
        TAGGCCCCAA ACGTGTAAAT GACATCCGAC GGACAAACCT CATAACGCAT CGGATAAAAA

               130        140        150        160        170        180
        TCGAGCCCGA AATCCCTGGC CAATTGTGTC GTCTCTTCAA TACTTTGTTC AAGCGCGGCC

               190        200        210
        ATCTCTTCGT TGTTCATGGA TAGTCGACGT CGGT
```

Section II: Sequence data of genomic library clones

```
                  10         20         30         40         50         60
Hp12L1T3  GGGCCGNTTC TCTTCNNGCC CCGGTTTTAG CAATGGGAAA ATCAAATGGA GCACCTAACA

                  70         80         90        100        110        120
          ACGCATCTTC ACAGACCAAG AACTACCGAG ATCGCCCCTT TCATTCTGGG GTGACGGAGG

                 130        140        150        160        170        180
          GATCCGTACC ATTCGAGCCG TTTTTTTCTT GACTCGAAAT GGGAGCAGGT TTGAAAAAGG

                 190        200        210        220        230        240
          ATCTTAGAGT GTCTAGGGTT GGGCCAGGAG GGTCTCTTAA CGCCTTCTTT TTTCTTCTCA

                 250        260        270        280        290        300
          TCGGAGTTAT TTCACAAAGA CTTGCCAGGG TAAGGAAGAA GGGGGGAACA AGCACACTTG

                 310        320        330        340        350        360
          GAGAGCCGCA AGTACAACGG AGAAGTTGTAT GCTGCGTTC GGGAAGGATG AATCCGCTCC

                 370        380        390        400        410        420
          CGAAAAGGAA TCTATTGATT CTCTCCCAAT TGGGTTGGAC CCGTAGGTGC CGATGATTTA

                 430        440        450        460        470        480
          CTTCACGGGC CGAGGTCTCT GGGTTCAAGT CCAGGATGGG CCCAGCTTGC GCCAGGGAAA

                 490        500        510        520        530        540
          AGAATAGAAG AAGCATCTGA CTACTTCATG CATGCTCCAC TTTGGCTCGG GGGGGATATA

                 550        560        570        580        590        600
          AGCTCAAGTT TGGTAAGAAC CTCCGCTCTT GCAATTGGGG GTCCGTTGCC GATTACCNGG

                 610        620        630        640
          TTTGGGATGT CTAAANTTGT CCANGCCGGT AAATGGAATA AGTATCT
```

```
                  10         20         30         40         50         60
HP12L1T7  GATCCCTCTT GTTCCTGTTT AGTCCCCTTC ATTTCGGAAG CTGTTTCTAT TTCTACATCT

                  70         80         90         100        110        120
          CTTTCTTCCT CACTTTCCAC CCTTTCTTCT GTTTTTGAGG CTTCTTTTAG TTTCTTAGTA

                  130        140        150        160        170        180
          AGAATGGGTG AGGGTATTCT GCCTAAATAG TAGACACAGG TAATAAATAA GAGAATACTA

                  190        200        210        220        230        240
          AAGATCCGAG CCATAGAATT TCTCAATTCT AACACAAGGT ACTTATTAGA TCGAATGTAC

                  250        260        270        280        290        300
          TTATTCGATC TAATAGAATG ATTTTGCCGT ATCCAGACTA ATACCAATCC AAGCCATTTC

                  310        320        330        340        350        360
          ATGAATAAAA TGTGACCAAT TAACCAACCA ACAAAACCAC TTGTTACAAA TAAGATCTTG

                  370        380        390        400        410        420
          CTGTTGCATC GAAAGAGATA AATGTTGACT AATCTGGCTA ACATTGAACT TGGTAAAATG

                  430        440        450        460        470        480
          AAATGGTTGA ATAATTGAAA AATGAGATTA TTCAGGAATA CACATTGAAT GCTGAGATTA

                  490        500        510        520        530        540
          CGCATTGAAT TTCTGGTAGT AGATCCATAA TCAAAAAAGT GTTTGTGATT GTTCCAGAAG

                  550        560        570        580        590        600
          AAATGAAACA AAAGATATGG TAGAGCTAGG ACAGTTATTG NATGAGGTCT ACCCATGCTA

                  610        620        630        640        650
          GATGCAGANG CGCATAATAG AATCGATATG ACATCATGAG CTGCCCCGTA AT
```

```
                    10          20          30          40          50          60
Hp12L2T3  GATCCGGGCG GTCCGGGGGG GACCACCACG GCTCCTCTCT TCTCGAGAAT CCATACATCC

                    70          80          90         100         110         120
          CTTATCAGTG TATGGACAGC TATCTCTCGA GCACAGGTTT AGCAATGGGA AAATAAAATG

                   130         140         150         160         170         180
          GAGCACCTAA CAACGCATCT TCACAGACCA AGAACTACGA GATCGCCCCT TTCATTCTGG

                   190         200         210         220         230         240
          GGTGACGGAG GGATCGTACC ATTCGAGCCG TTTTTTTCTT GACTCGAAAT GGGAGCAGGT

                   250         260         270         280         290         300
          TTGAAAAAGG ATCTTAGAGT GTCTAGGGTT GGGCCAGGAG GGTCTCTTAA CGCCTTCTTT

                   310         320         330         340         350         360
          TTTCTTCTCA TCGGAGTTAT TTCACAAAGA CTTGCCAGGG TAAGGAAGAA GGGGGGAACA

                   370         380         390         400         410         420
          AGCACACTTG GAGAGCGCAG TACAACGGAG AGTTGTATGC TGCGTTCGGG AAGGATGAAT

                   430         440         450         460         470         480
          CGCTCCCGAA AAGGAATCTA TTGATTCTCT CCCAATTGGT TGGACCGTAA GTGCGATGAT

                   490         500         510         520         530         540
          TTACTTCACG GGCGAGGTCT CTGGTTCAAG TCCANGATGG CCACTGCGCC CGGGAAAAAA

                   550         560         570         580         590         600
          TAAAANAAGC ATCTGACTAA TTCATGCATG CTCACTTGGG TCGGGGGGAT ATACTCAATT

                   610         620
          GGTAAANCTC CGCTCTTGNA ATTGGG
```

```
                  10         20         30         40         50         60
Hp12L2T7  GATCCCTCTT GTTCCTGTTT AGTCCCCTTC ATTTCGGAAG CTGTTTCTAT TTCTACATCT

                  70         80         90        100        110        120
          CTTTCTTCCT CACTTTCCAC CCTTTCTTCT GTTTTTGAGG CTTCTTTTAG TTTCTTAGTA

                 130        140        150        160        170        180
          AGAATGGGTG AGGGTATTCT GCCTAAATAG TAGACACAGG TAATAAATAA GAGAATACTA

                 190        200        210        220        230        240
          AAGATCCGAG CCATAGAATT TCTCAATTCT AACACAAGGT ACTTATTAGA TCGAATGTAC

                 250        260        270        280        290        300
          TTATTCGATC TAATAGAATG ATTTTGCCGT ATCCAGACTA ATACCAATCC AAGCCATTTC

                 310        320        330        340        350        360
          ATGAATAAAA TGTGACCAAT TAACCAACCA ACAAAACCAC TTGTTACAAA TAAGATCTTG

                 370        380        390        400        410        420
          CTGTTGCATC GAAAGAGATA AATGTTGACT AATCTGGCTA ACATTGAACT TGGTAAAATG

                 430        440        450        460        470        480
          AAATGGTTGA ATAATTGAAA AATGAGATTA TTCAGGAATA CACATTGAAT GCTGAGATTA

                 490        500        510        520        530        540
          CGCATTGAAT TTCTGGGTAG TTAGATCCAT AAATCAAAAA AGTGTTGGTG ATTGGTCCNG

                 550        560        570        580        590        600
          AANAAATGAA ACAAAAGATT TGGGTANAGC TTAGGACAGT TATTGGTNTG AGGTCTTACC

                 610        620        630        640        650        660
          CAATGCTAAA TGCANAAGGC GCTTAATNNA TCNATNNTGA ACATCATGAN CGGCCCCGTA


          ATNAAAC
```

```
                  10          20          30          40          50          60
Hp12L3T3  GATCCTATAT TTTTAAATTT TTATTTATGT GTTATAAATT ATTTTAGTAT TTTAAATTAA

                  70          80          90         100         110         120
          TTTTCAGAAA CCAGTTACTA TTTTTTATAA ATTAGAAAGG GAAAATGGCT ATTTAAATTT

                 130         140         150         160         170         180
          TAACCAAATT GGCTATCAAA TCTAACCCAA TTCCCTGGCC CAATTTCTAA TTAAACCCGA

                 190         200         210         220         230         240
          GCCTAACCCT TTTAAACCCT ACCCAAACCC GGATCCCCAC CTACCCCATT TAATCTAGGC

                 250         260         270         280         290         300
          CGTTGATCAT TCAGATCAAC GACCCACCAT TCCACCTGCC TAAAATAAAC CCAAACGACC

                 310         320         330         340         350         360
          CCCTTAACCT AAATCATTTC CACCAACCCG CTGCCCTTGA ATCCCCTTCC TCTCTAATCC

                 370         380         390         400         410         420
          TCTCTGCAAC CCCACTCAAA CCCTATCCGC TACCATCCAA ACTAACCCTA ATCCCCTTCG

                 430         440         450         460         470         480
          ATTCTCACCC AATCCATGGA CTCACATGGT TGTTTGAGAC GAGTACCAGT CTCTTATGTC

                 490         500         510         520         530         540
          TCTTGGTTGC CTGTTTTCGT GATTTCATGG AAAGATCTCA AAAGGATCTA GTCCAGTCTT

                 550         560         570         580         590         600
          TGCTCAATTT CTATCTATGG TCTTTTCCCG GCCATTCTAT CTATGGTCTT ACACATTCAA

                 610         620         630         640         650         660
          ACTCTTTTTC TTTCTCTACT ACTTGTGCTA CTGCCTTGTC TAGTAGGCTG AAAGCCAAGA

                 670         680         690         700         710         720
          CTAGACTGTG GAATTTTGAT TGCTTTACTT TCCTTTCCTG CACTTTGCTT CTTCATTGGT

                 730         740         750         760         770         780
          ATGTCCCTAG TTTAATTAGG AACTCACCTA TGTGTNCCTT GNTTTCATCA CTATCTTTCT

                 790         800         810         820         830         840
          GACTGACTAT GGCTATGTNG TGCACNTTNG TTACTGTTAC TAGATGCTAA ATTTGCCTNT

                 850         860         870         880         890         900
          TGTCAATGAA TAGAGNTACT GTCCGTATGC CTACACTGAT TAGNCTGTAG CTTTGGGTTG

                 910         920         930         940         950         960
          TNGGCNTGCC ACCCCCGTGA ACGANNCCCT GNGNNNGGGG GCTCTCTANT TGGTTNGAAC

                 970         980         990        1000
          TAGTAATNCC TGTACTCTAN GNCTGCTCGC CAGCCTTTGC CGCCAGA
```

```
                10         20         30         40         50         60
Hp12L3T7  GATCCACTAG TGTAGTTCCC GGCCAGACGA TTTTTGTGAA TTTGTGCAAT TAGCAAATAA

                70         80         90        100        110        120
          ACAAGTAAGC CTCAAGTAAT AGGGGATTTA AGTTTAAATA CACAAGTGTT TGAAAATAAT

               130        140        150        160        170        180
          TATTGAAAGG TGAAATTTTG AAAAGAGTTA TAATCTAAGG CATGCTTATG AATATAAAAG

               190        200        210        220        230        240
          GGGGTGTCCT AGGTTTGTTT GTAATATGGA TCATATCAAT GCAATACCCG GTATGACACT

               250        260        270        280        290        300
          CCTCAGAAGA GGGGATACAC GTGGTATTAG CGCACCGATC ATTATATCCA TATCTACCCT

               310        320        330        340        350        360
          TTCACGCCCC GTGAAGGTAA TTAAAGCGAG GGTTGGTCTC GACCCCTATT GCATGTTGTT

               370        380        390        400        410        420
          ACTCGTCCCA TTTCTATCAG TCCCGGGGGA ATTTAGGACT CTTATTCCTA TAAGAAGGAG

               430        440        450        460        470        480
          GTTCTAGTCA GATCCTCAGG TTTAAAGGAA AAAATACTAA AGCGACATAC AAAAACATAT

               490        500        510        520        530        540
          AAGACTGCAC TTAGAGGGGA AAACATATAA GCAAGTAATA GGCTCATGCA TACCTCCACA

               550        560        570        580        590        600
          AATAATGCAC ATAGACAGCA TGACTAATAC ACAACTAAGG TCTGAATTTA AAATCCTAAA

               610        620        630        640        650        660
          GCAGGGTGTT TGAGTTGTTG CANCAGAACC AGATTTATTA CATGACTCAG ATAAGAAATC

               670        680        690        700        710        720
          TGAATCAGGC CTGCCTACTG GTTGTAACAG TTGATATTAA ACAAGGCAGT TCCATTTTNT

               730        740        750        760        770        780
          TTAANTATTA CCTANGCTTG CCTAGCNCAA GCNGANCANA NTTTTANATG TNAAAANTTG

               790
          GAGTNTCATA CT
```

```
                    10          20          30          40          50          60
Hp14L1T3  TCGTTGAATA CATCAGTGTA GCGCGCGTGC GGCCCAGAAC ATCTAAGGGC ATCACAGACC

                    70          80          90         100         110         120
          TGTTATTGCC TCAAACTTCC GCGGCCTAAA AGGCCGTAGT CCCTCTAAGA AGCTGGCCGC

                   130         140         150         160         170         180
          GAAGGGATAC CTCCGCATAG CTAGTTAGCA GGCTGAGGTC TCGTTCGTTA ACGGAATTAA

                   190         200         210         220         230         240
          CCAGACAAAT CGCTCCACCA ACTAAGAACG GCCATGCACC ACCACCCATA GAATCAAGAA

                   250         260         270         280         290         300
          AGAGCTCTCA GTCTGTCAAT CCTTACTATG TCTGGACCTG GTAAGTTTCC CCGTGTTGAG

                   310         320         330         340         350         360
          TCAAATTAAG CCGCAGGCTC CACTCCTGGT GGTGCCCTTC CGTCAATTCC TTTAAGTTTC

                   370         380         390         400         410         420
          AGCCTTGCGA CCATACTCCC CCCGGAACCC AAAAACTTTG ATTTCTCATA AGGTGCCGGC

                   430         440         450         460         470         480
          GGAGTCCTAA AAGCAACATC CGCCGATCCC TGGTCGGCAT CGTTTATGGT TGAGACTAGG

                   490         500         510         520         530         540
          ACGGTATCTG ATCGTCTTCG AGCCCCCACT TTCGTTCTTG ATTAATGAAA ACATNCTTGG

                   550         560         570         580         590         600
          CAAATGCTTT CCAGTTGGTC GTCTTTTATA AATTCAAGAA TTTCACCTNT GCTATNAAAT

                   610         620         630         640
          ACNAATGCCC CGACTGTCCT GGTAATAATA CTTCGATCCC NAAGG
```

```
                    10          20          30          40          50          60
Hp14L1T7  CCCATGCTCA TATTTGACGA TCGATTTGCA CGTCAGAACC GCTGCGAGCC TCCACCAGAG

                    70          80          90         100         110         120
          TTTCCTCTGG CTTCACCCTA TACAAGCATA GTTCACCTTC TTTCGGGTCC AACCCTATAT

                   130         140         150         160         170         180
          GCTCTTACTC AAATCCATCC GAGAACATCA GAATCGGTCG ATGATGCGCC GAAGCTCTCA

                   190         200         210         220         230         240
          CCTACGTTCA CGTTCATTAC GCGCTGGGGT TTTACACCCA AACACTCGCA CATAAGGTTG

                   250         260         270         280         290         300
          ACTCCTTGGT CCGTGTTCCA AGACGGGTCG CTGATGACCA TTACGCCAAC ATCCTTGCCG

                   310         320         330         340         350         360
          AAGCGCGGTC CTCAGTCTGC CGAATGGTAT TATGCAAAGG GCTATAACAC TCCCGAGGGA

                   370         380         390         400         410         420
          GCCACATTCC CTAAGCCTTT CTCCCAAACA ACAAACTGAT GTTGGCCTGT ACTGACAGAG

                   430         440         450         460         470         480
          TAAACAAGTC CGAAAACAAG CAAAACTGAC AGAAACAAGT CTGGTCATAG GCGCTTCCTT

                   490         500         510         520         530         540
          TCAACAATTT CACGTGCTGG TTAACTCTCT TTTCAAAGTG CTTTTCATCT TTCGATCACT

                   550         560         570         580         590         600
          CTACTTGNGC GCTATCGGTC TCTACCGGTA TTTAGCTTTA GAAGAGATAT ACCTTCCATT

                   610         620         630         640         650         660
          TAGAGCAGCA NTTCCNAACT ACTCACTCGT TGAAGGACTA TACCAAAGGT TGGTGTCAAC

          CGG
```

```
                  10         20         30         40         50         60
Hp14L2T3  AACAAAAGCT GGAGCTCGCG CGCCTGCAGG TCGACACTAG TGGATCCCAC CAATCAGCTT

                  70         80         90        100        110        120
          CCTTACGCCT TACGGGTTTA CTCGCCCGTT GACTCGCACA CATGTCAGAC TCCTTGGTCC

                 130        140        150        160        170        180
          GTGTTTCAAG ACGGGTCGAA TGGGGAGCCC ACAGGCCAGC GTCCGGAGCG CGCAGATGCC

                 190        200        210        220        230        240
          GAAGCACGCC GGAGGCGCGC GCTGCCTACC ACAATCAAGG AGACGGCGTT CCACGGGCGT

                 250        260        270        280        290        300
          ATCGAAAGCC CGGGCTTTGG CCGCCCCCCC AATCCACGCT GGTCCACGCC CCGAGTCGAT

                 310        320        330        340        350        360
          CGGCGGACCG GCTCGTCACC GTTCCACATC CGACCGGGGC GCATCGCCGG CCCCCATCCG

                 370        380        390        400        410        420
          CTTCCCTCCC GACAATTTCA AGCACTCTTT GACTCTCTTT TCAAAGTCCT TTTCATCTTT

                 430        440        450        460        470        480
          CCCTCGCGGN ACTTGTTCTC TATCGGTCTC TCGCCCGTAT TTAACCTTGG ACGGAATTCA

                 490        500        510        520        530        540
          CCGCCCGATT TGGGCTGNAT TCCAAACAAC CCGACTCGTA GACAGCGCCT TCGTGGTGCG

                 550        560        570        580        590        600
          ACAGGGTNCN GGCACAACGG GGCTNTTACC CTNTNTGGCG CCCCTTTCAA GGGACTTGGG

                 610        620        630        640        650
          CCCGGTCCGC CGTTGAGGAC CTTNTTCAAA CTACAATTTG AACGGNGGAG NCC
```

```
                  10          20          30          40          50          60
Hp14L2T7  TCGTTGAATA CATCAGTGTA GCGCGCGTGC GGCCCAGAAC ATCTAAGGGC ATCACAGACC

                  70          80          90         100         110         120
          TGTTATTGCC TCAAACTTCC GCGGCCTAAA AGGCCGTAGT CCCTCTAAGA AGCTGGCCGC

                 130         140         150         160         170         180
          GAAGGGATAC CTCCGCATAG CTAGTTAGCA GGCTGAGGTC TCGTTCGTTA ACGGAATTAA

                 190         200         210         220         230         240
          CCAGACAAAT CGCTCCACCA ACTAAGAACG GCCATGCACC ACCACCCATA GAATCAAGAA

                 250         260         270         280         290         300
          AGAGCTCTCA GTCTGTCAAT CCTTACTATG TCTGGACCTG GTAAGTTTCC CCGTGTTGAG

                 310         320         330         340         350         360
          TCAAATTAAG CCGCAGGCTC CACTCCTGGT GGTGCCCTTC CGTCAATTCC TTTAAGTTTC

                 370         380         390         400         410         420
          AGCCTTGCGA CCATACTCCC CCCGGAACCC AAAAACTTTG ATTTCTCATA AGGTGCCGGC

                 430         440         450         460         470         480
          GGAGTCCTAA AAGCAACATC CGCCGATCCC TGGTCGGCAT CGTTTATGGT TGAGACTANG

                 490         500         510         520         530         540
          ACGGTATCTG ATCGTCTTCG AGCCCCCAAC TTTCGTTCTT GATTAATGAA AACATCCTTG

                 550         560         570         580         590         600
          GCAAATGCTT TCCAGTTGTT CGTCTTTCAT AAATNCAAGA ACTTCACCTN TTGACTATGA

                 610         620         630         640
          AATACCAATG CCCCCGACTG NCCCTTGTTA ATCATTACTT CCATCCCA
```

```
                  10         20         30         40         50         60
Hp14L3T3  AACAAAAGCT GGAGCTCGCG CGCCTGCAGG TCGACACTAG TGGATCCATT GGAGGGCAAG

                  70         80         90        100        110        120
          TCTGGTGCCA GCAGCCGCGG TAATTCCAGC TCCAATAGCG TATATTTAAG TTGTTGCAGT

                 130        140        150        160        170        180
          TAAAAAGCTC GTAGTTGGAC TTTGGGATGG GCCGGCCGGT CCGCCTTAGG TGTGCACCGG

                 190        200        210        220        230        240
          TCGTCTCGTC CCTTCTGCCG GCGATGCGCT CCTGGCCTTA ATTGGCCGGG TCGTGCCTCC

                 250        260        270        280        290        300
          GGCGCTGTTA CTTTGAAGAA ATTAGAGTGC TCAAAGCAAG CCTACGCTCT GTATACATTA

                 310        320        330        340        350        360
          GCATGGGATA ACATTATAGG ATTTCGGTCC TATTACGTTG GCCTTCGGGA TCGGAGTAAT

                 370        380        390        400        410        420
          GATTAACAGG GACAGTCGGG GGCATTCGTA TTTCATAGTC AGAGGTGAAG TTCTTGGATT

                 430        440        450        460        470        480
          TATGAAAGAC GAACAACTGC GAAAGCATTT GCCAAGGATG TTTTCATTAA TCAAAGAACC

                 490        500        510        520        530        540
          AAAGTTNGGG GCTCGAAAAC GATCAAATAC CGTCTAATCT NAACCATAAA CCATGCCCNA

                 550        560        570        580        590        600
          CCAAGGATCN GCGGATGTTG CTTTTAAGAC TCCCCCGCNC CTTATGAAGA AAACNAAGTT

                 610        620        630        640        650        660
          TTTGGGTTCC CGGGGGAGTT TGGTCNCNAG GNTNNAACTT NAAGGAAATG CCGGAAGGGC
```

```
                  10          20          30          40          50          60
Hp14L3T7  ATCCCCGGCG CGCTCGCTCG CGTGCGTGAC GGGTGATTAA CGAACCCCGG CGCGGAAAGC

                  70          80          90         100         110         120
          GCCAAGGAAT ACTAAATTGA AAGCCTGCCT CTCGCGCCCC GTTCGCGGTG CGCGCGGGGA

                 130         140         150         160         170         180
          ACTTGTGCTT CTTTTGAAAC ACAAACGACT CTCGGCAACG GATATCTCGG CTCTCGCATC

                 190         200         210         220         230         240
          GATGAAGAAC GTAGCGAAAT GCGATACTTG GTGTGAATTG CAGAATCCCG TGAACCATCG

                 250         260         270         280         290         300
          AGTCTTTGAA CGCAAGTTGC GCCCGAAGCC ATTAGGCCGA GGGCACGTCT GCCTGGGCGT

                 310         320         330         340         350         360
          CACGCATCGC GTCGCCCCCC GCACTCCGCG CCCAGAATCA TGGACGCGGT GGTGTCGCGG

                 370         380         390         400         410         420
          GGCGGATACT GGCCTCCCGT GCGCCCCGAG CACGCGGTTG GCCTAAATGC GAGTCCACGG

                 430         440         450         460         470         480
          CGACGGACGT CACGACAAGT GGTGGTTGAA ACTCAACTCT CGTAATGTCG CGGCTCCNGC

                 490         500         510         520         530         540
          CCGTCGCACG TTTGGGCTTC ACGACCCTTG TTGCGCTTAN GCGCTTCGAC CGCGACCCCA

                 550         560         570         580         590         600
          AGTCAGGCGG GACTACCCGC TGAGTTTAAG CATATNAATA AGCGGAGGAA AAGAAATTAC

                 610         620         630         640         650         660
          AAGGATTCCC TATNACGGGA ACGAACCCGG AACAGCCACC TTANAATCGG GCNGGTTCGC

          CGTTC
```

```
                  10         20         30         40         50         60
Hi30L1T3  GATCCTAACC TCTAAAATAA TGTTTTCTAA CAGGCTTGTG GGGTGTGCCA GCATTTCCCA

                  70         80         90        100        110        120
          TTGCTGGGTA TGCCCACTAG CCTGTCCTTG CCTCTTTGCC ACTTCCCCTT TCCCCGTCTC

                 130        140        150        160        170        180
          AGAACTCTGC ACTTACACTC ACTCTTAGCT TCTAAGTTCT GCCCCCCTTC TATGAGCCTT

                 190        200        210        220        230        240
          GCCTAGGGAC CTCGAGTTCC TTCTAAACTT GGACACTTGA GGGCTGGCCC TTCCACACTG

                 250        260        270        280        290        300
          CAGTATGACT TAATACTGCA ATACATTTGG ATGAGAGCAC TGCCCGGAGT CCATATGAGG

                 310        320        330        340        350        360
          CTCTTAGGGA GCTCTGACAC ATCCAAATGA GAGAAAGGCT TTGGATCTTG ATCTTGGGAG

                 370        380        390        400        410        420
          TTGGTTTACT TCATACTTCA GACAGGAAGT CTGAATCAAG CTCTCCTTGG TTGTAATTTT

                 430        440        450        460        470        480
          CAGTTTTCGG ATGTATTTTC TCTATTTTAT TTTTGGAATG TAATAACTTT GTAATAAACT

                 490        500        510        520        530        540
          TTTGGGGTGA TTAGTGAAAA GGGAGGGGTA ACCATGCATG CAAGGGGTAG ATGTCCTGCT

                 550        560        570        580        590        600
          CATAGGGTTT TCTGCACTTT TGCATTTATA CATATACTCT GCCTATAGTT CCTGCACTTC

                 610        620        630        640        650        660
          TACAATCATG CAGATGTGCT GCTCATAGGG TTTTCTGCAC TTCTGCATTT ATACAGATAC

                 670        680        690        700        710        720
          CCTGCCTATA ATTTCTGCAC TTCTGCATCA TGTAGATGTC TGNTANNGGN TTCTGCCTCT

                 730        740        750        760        770        780
          GATTATAGAN CCTGCTTNGT TTGCTTGATA GGAAGCCTGN ATGGGTCGNT NNATCAAAAC

                 790        800        810        820        830        840
          TGCCCCTATC NTCNAACTNN GATGNTCCAA ACGTAATGTG ANGNATGTTA CAANATAGAN


          TCTA
```

```
                  10          20          30          40          50          60
Hi30L1T7  GATCCCAATC ATCCTCTATT TCATTCATGT GAACACCTCC ACCCTTATGG TCTGATAGAG

                  70          80          90         100         110         120
          GGTTGTTGCG GACATTCNAA GTAGGCTCCT TTGCAACAAT GATCTTGTTA TCAATCAAAG

                 130         140         150         160         170         180
          TCTGGATCTT GTCTTTCANA GAGCGGCATT CATCAATGGN ATGCCCTTTT ATGCCNGAAT

                 190         200         210         220         230         240
          GGNATGCATA NGANTTATTT GGATTAACCC ACTGNNAAAG GNTTTCANGG GNTATNNCAN

                 250         260         270         280         290         300
          GGATNAGGGT GACCTAATCG GCAGGGTTGA GCCTTTCATA CAGCTGGTCA ATCGGTTCAG

                 310         320         330         340         350         360
          CAATGGCGGT ATACTGTTTG GGGGGTCTGC GATCAAAATT TGGTCGAGGT CTANGAAAGT

                 370         380         390         400         410         420
          TTTGACGTAT GGGAGGTGAC TGATAGTGGG ATGGCTGAGC ATTGTAGGCT TGGTAGACAT

                 430         440         450         460         470
          GTGCGGGTTG GGAATATTTG GGTAAGGGNA GGTTGTTATG TNAGGANGGN GGGGG
```

```
                  10          20          30          40          50          60
Hi30L2T3  GATCCTCCTT CCTGGAAAGA GCATCAGTTG CAATGTTGTC TTTGCCCTTT TGTAAGAGAT

                  70          80          90         100         110         120
          GGTATAATCA TAACCTAATA GTTTGATTAA CTATTTTTGT TGACTGGGAG TGGTAATTTT

                 130         140         150         160         170         180
          ATGTTCTAGT AGATATTTGA GAGTGTGATG GTCAATCTTG ATTACAAAGT GTCTCCCTAG

                 190         200         210         220         230         240
          TAGGTAAGGT CTCCATTTTT GGACTGCAGT ACCCAATGCT AGCAGCTCCC TCTCATAGGC

                 250         260         270         280         290         300
          TGATAGTGCT TCGTTTTTCT CTGATAAGCC TTTGCTGAAG AAAGCTATAG GTCTGCTATC

                 310         320         330         340         350         360
          TTGAGCCAGT ACTACTCCAA TGCCAGAACC AGAAACATCA GTTTCTACTA CAAACTCTTT

                 370         380         390         400         410         420
          GTTGAAACCA GGTAGTACCA GAACATGTGT TGTAGTGAGT GCCTTTTTCA AGTCTTCAAC

                 430         440         450         460         470         480
          TGTCTGAGTG ACAGCAGAGT TCCAGATGAA ATTCCCTTGT TTTAGCAGAT CATGCAAGGG

                 490         500         510         520         530         540
          TCTTGCAATA ATACCATACC CCCTTATACA CCTTCTATAG TACCTAGTGA GACCTAAGAA

                 550         560         570         580         590         600
          ACCCCTCAAG CTCTTTTAGG TAGTTGGTTG AGGCCAGTTC ACTACTACTT CTACCTTCTG

                 610         620         630         640         650         660
          ATGGTCCATA GCAACTCTCT CTTCAGAGAT TACATGGCCT AAGTAATATA TATTTGTTTC

                 670         680         690         700         710         720
          ACTCCAAAAG CACACTTGTT TTTCTTGACA TAGAGCACAT GTGATCTCAT ACTTGAAAAG

                 730         740         750         760         770         780
          CTTCCTCATG TGAATTAGGT GATCTGCCAG CTNGCCTGTN TTNAGGNTGT CNTCAAGAAA

                 790         800         810         820         830         840
          CCAGACTGNT TTTTNNGGAG TCTGAAATTC TATTCTTNGG TTTGNATNNC NGGGNNCTTT

                 850
          TTAGNCNAGG GCTGAC
```

```
                10          20          30          40          50          60
Hi30L2T7  GATCCTCCTT CCTGGAAAGA GCATCAGTTG CAATGTTGTC TTTGCCCTTT TGTAAGAGAT

                70          80          90         100         110         120
          GGTATAATCA TAACCTAATA GTTTGATTAA CTATTTTTGT TGACTGGGAG TGGTAATTTT

               130         140         150         160         170         180
          ATGTTCTAGT AGATATTTGA GAGTGTGATG GTCAATCTTG ATTACAAAGT GTCTCCCTAG

               190         200         210         220         230         240
          TAGGTAAGGT CTCCATTTTT GGACTGCAGT ACCCAATGCT AGCAGCTCCC TCTCATAGGC

               250         260         270         280         290         300
          TGATAGTGCT TCGTTTTTCT CTGATAAGCC TTTGCTGAAG AAAGCTATAG GTCTGCTATC

               310         320         330         340         350         360
          TTGAGCCAGT ACTACTCCAA TGCCAGAACC AGAAACATCA GTTTCTACTA CAAACTCTTT

               370         380         390         400         410         420
          GTTGAAACCA GGTAGTACCA GAACATGTGT TGTAGTGAGT GCCTTTTTCA AGTCTTCAAC

               430         440         450         460         470         480
          TGTCTGAGTG ACAGCAGAGT TCCAGATGAA ATTCCCTTGT TTTAGCAGAT CATGCAAGGG

               490         500         510         520         530         540
          TCTTGCAATA ATACCATACC CCCTTATACA CCTTCTATAG TACCTAGTGA GACCTAAGAA

               550         560         570         580         590         600
          ACCCCTCAAG CTCTTTTAGG TAGTTGGTTG AGGCCAGTTC ACTACTACTT CTACCTTCTG

               610         620         630         640         650         660
          ATGGTCCATA GCAACTCTCT CTTCAGAGAT TACATGGCCT AAGTAATATA TATTTGTTTC

               670         680         690         700         710         720
          ACTCCAAAAG CACACTTGTT TTTCTTGACA TAGAGCACAT GTGATCTCAT ACTTGAAAAG

               730         740         750         760         770         780
          CTTCCTCATG TGAATTAGGT GATCTGCCAG CTNGCCTGTN TTNAGGNTGT CNTCAAGAAA

               790         800         810         820         830         840
          CCAGACTGNT TTTTNNGGAG TCTGAAATTC TATTCTTNGG TTTGNATNNC NGGGNNCTTT

               850
          TTAGNCNAGG GCTGAC
```

```
                  10          20          30          40          50          60
Hi30L3    GATCCATCCG GCCCAAGGCG GAAGGCATGG GCTATAGCAC ACAAAAATTT GGGACTCGCG

                  70          80          90         100         110         120
          CGAATTTGTA GTTTTATGGC TGTAAAATGC CAAAAGATAT AATGTGGTAA TTTAGAGTGG

                 130         140         150         160         170         180
          GTGGAAATTG TTTCTTATGC CGTATTTGAT ATCCGGGACA AATATTAGGC GATTCCTCGA

                 190         200         210         220         230         240
          CGGATCCATC CTGGCCCAAG GCGGAAGGCA TGGGCTATAG CACACAAAAA TTTGGGACTC

                 250         260         270         280         290         300
          GCGCGGATTT GCAGTTTTAT GGCTATAAAA TGCCAAAAAA TATAATTTGG TTATTTCGAG

                 310         320         330         340         350         360
          ATGGTGGAAA TGGTTTCTTA GGCCGTGTTT GATGTCCGGG ACAAATATTA GGCGATTCCA

                 370         380         390         400         410         420
          GGACGGATCC ATCCGGGCCC AAGGCGGAAG GCATGGACTT TAGCACACGA AAATATGGGA

                 430         440         450         460         470         480
          CTCGCGCGGA TTTGCAGTTT TATGGCTGTA AAATGCCAAA AAATATAATT GGTTATTTCG

                 490         500         510         520         530         540
          AGATGGTGGA AATGGTTTCT TAAGCCGTAT TTGATGTCCG GGACCAATAT TAGGCGATTC

                 550         560         570         580         590         600
          CAAGACGGAT CAATCCGGGA CCAAGGCGGA AGGCATGGGC TATAGCACAC AAAAATTTGG

                 610         620         630         640         650         660
          GACTCGCGCG GATTTGCAAT TTTATGGCTA TAAAATGCCA AAAAATATAA TTCGGTCATT

                 670         680         690         700         710         720
          TCGGGGCGGT GGNAATTGTT TCTTAGGCCG TATTTGATGT CCGAGATAAA TATTAGGCAN

          TNCNGAA
```

```
                  10         20         30         40         50         60
Hi30L4T3  GATCCTGTGT TAGTTCATTA TAGGGATACC ACCCTTTAGA AGGAGAAGAC ACTGTTTGAA

                  70         80         90        100        110        120
          ATTATAGAAG ATGGGGTCCT CATATATCAA GGACGATTAT GTGTCCCTAA TGTTGCAGCG

                 130        140        150        160        170        180
          CTGCGTCAGT AGGTTATGGG GGAAACTCAC TCTTCTCGTT ATTCTATCCG CCCTGGGTCA

                 190        200        210        220        230        240
          ACAAAGATGT ATCATGACAT TAAGGAGGTG TACTGGTGGA ATGACATAAA GAAGAACATT

                 250        260        270        280        290        300
          GTTGAGTTTG TCGCTCAGTG TCCTAGTTGC CAACAGGTGA AGATAGAGCA CCAGAAACCT

                 310        320        330        340        350        360
          GGAGGGCTAA TGTAGACTAT AGAGATCCAG ACATGGAAAT GAGAGGTGAT AAACATGAAC

                 370        380        390        400        410        420
          TGTCATGGGT TTACCTCATT CTTATCGTAA GTTCGATTCC ATATAGGTAA TAGTCAATAG

                 430        440        450        460        470        480
          GCCCACGCAG TCAGCTCATT TCCTACCGGT CAGATCTATA TATACAATAG AAGATTAGGC

                 490        500        510        520        530        540
          AAAGTTATAT ATTAAGAGAT AGTGCGCTAT CGGAGTCCAG TATCTATTAT ATTTGCCATG

                 550        560        570        580        590        600
          GGCCAGTTAC ACACATTTTG AGTATTCAAG AGNCNGGACT AGGATTACCA CTTATTCACT

                 610        620        630        640        650        660
          GTACAGCAGG CATAACCTAG TGTCACTGGT GTAAGGACCT GNNGATCGNG CGTNNCTGTN

                 670        680        690        700        710        720
          GGCGCAANNN CATGTGGNAG AGCNTGNCGN TTGCGANNTC AGAAAATNGT GANNAGTCAA


          CATAAA
```

```
                10         20         30         40         50         60
Hi30L4T7  GATCCTCCTT AATAGAACCC ATGATCTCCG ATGGGTCTAA CTCTATAGTA TACCTTATAT

                70         80         90        100        110        120
          CTCTTTATAA CACCTTTGCA GCCTTCTGAC TGGTTCTAGC GGCTGTAGTC TTTGGAGTAT

               130        140        150        160        170        180
          CGCTGAAAAC ATAACACGTC GTTAAGAACA TGAATCCTTA TATCGCACGA TGTAAGATAA

               190        200        210        220        230        240
          GAAGAGAGGA TAACATCCTC TATGCCCCGT AGCCTCCTGT CTATAAGTGT GGTGCACAAC

               250        260        270        280        290        300
          ACACCTATAA ACAAGACTCT ACTAGACACG GTCTGTAGAC AACCCTAGGA CAGAACTGCT

               310        320        330        340        350        360
          CTGATACCAC TTTTGTCACA ACCCAAACCA ATGGGCCACG ACGGATGCCC GACTCCTACC

               370        380        390        400        410        420
          TGTCAGACAC CCCTAAGCAT GCTTCTAAGA TATAAACCTG AATAACATAT GCTGAATTAC

               430        440        450        460        470        480
          GAGAATAATA TACATGAAGG AAACCTGCCC AAAAGACATA TATACATATA CGTGCAACAT

               490        500        510        520        530        540
          ACGTAGGGCT AGCCGACAAG GCTGCTATAG ACGACTATGT ACCATAAAAT TGNAGTCGGA

               550        560        570        580        590        600
          GGCCACATCT ATCCACTAGA CATCTTCNCA GACTCTATGA ANTATNGACA AACGGGTGGC

               610        620        630        640        650
          CGNATCATTT TTCANTTGCA ACANCNTCGC ANGNCCNTTG GCGCTAGCGT CCGAGAGA
```

Section III: Sequence data of tail PCR

```
Hi30RB          10         20         30         40         50         60
        GCTCGGGAAT GCCGACGCGC TTTTCACATT GNCAGTTTTG GGAAATCATT CCATAGGATG

                70         80         90        100        110        120
        AAGCTTGTTC ATGGATAGTC GACGTCGGTA TCCATATGAC TAGTAGATCC TCTAGAGTCG

               130        140        150        160        170        180
        ACCTGCAGGC ATGCAAGCTT TCCCTATAGT GAGTCGTATT AGAGCTTGGC GTAATCATGG

               190        200        210        220        230        240
        TCATAGCTGT TTCCTGTGTG AAATCGTTAT CCGCTCACAA TTCCACACAA CATACGAGCC

               250        260        270        280        290        300
        GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC ATTAATTGCG

               310        320        330        340        350        360
        TTGCGCTCAC TGCCCCCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA TTAATGAATC

               370        380
        GGCCAGCGCC CCACNCGAAN TCCTGGAC
```

References

Barrett A.J. and Kirschke H. 1982. Cathepsin B, cathepsin H, and cathepsin L. Methods in Enzymology 80: 535-561.

Jefferson R.A., Kavanagh T.A. and Bevan M.W. 1987. Gus fusions: ß-glucuronidase as a sensitive and versatile gene marker in higher plants. EMBO Journal 6: 3901-3907.

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

Horsch R., Fry J.E., Hoffmann N.L., Eichholtz A., Rogers S.G. and Fraley R.T. 1985. A simple and general method for transferring genes into plants. Science 227: 1229-1231.

Laemmli U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

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

Masoud S.A., Johnson L.B., White F.F. and Reeck G.R. 1993. Expression of a cysteine proteinase inhibitor (oryzacystatin-I) in transgenic tobacco plants. Plant Molecular Biology 21: 655-663.

Murashige T. and Skoog F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiologia Plantarum 15: 473-497.

Sambrook J., Fritsch E.F., Maniatis T. 1989. Molecular cloning: A laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Sorensen A.B., Duch M., Jorgensen P. and Pedersen F.S. 1993. Amplification and sequence analysis of DNA flanking integrated proviruses by a simple two-step polymerase chain reaction method. American Society of Microbiology 67: 7118-7124.