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.

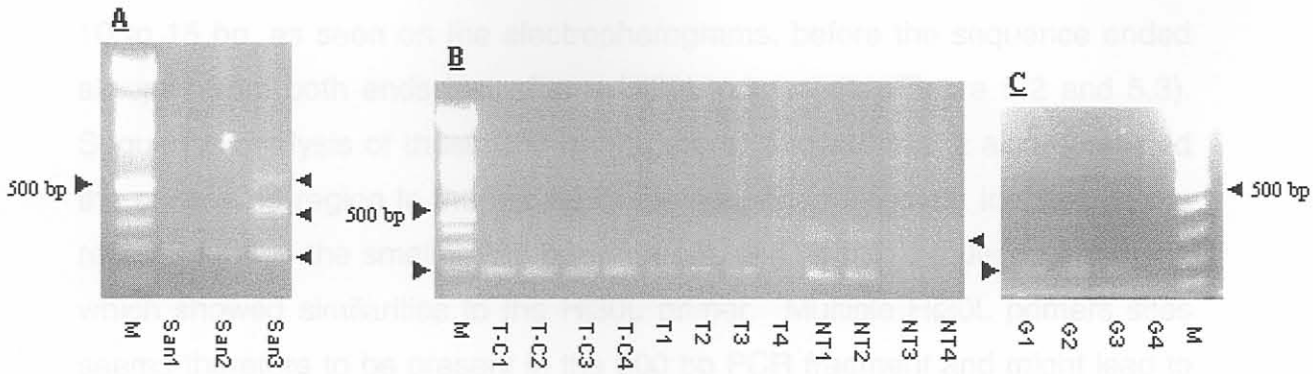

Figure 5.1: PCR amplification of genomic target DNAs with primers designed for subtraction product Hi30. (A) PCR amplification with genomic template DNA derived from tobacco cultivar '*Samsun/San1191'* (San 1-3). Lane M represents a 100 bp marker. (B) PCR products after 60 amplification cycles with genomic 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 .

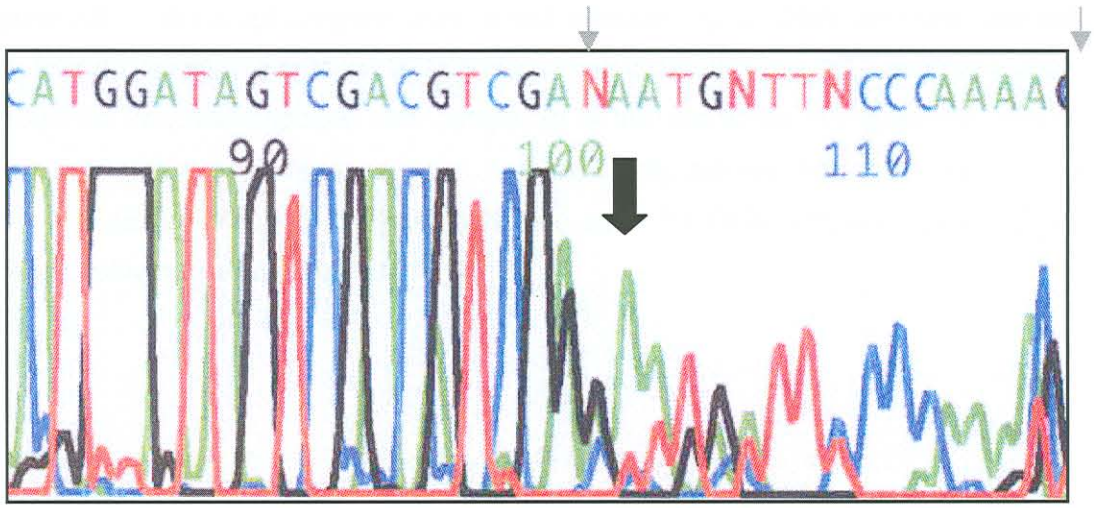

Figure 5.2: Electropherogram from direct sequencing of DNA template derived from tobacco cultivar '*Samsun/UK*' (NT1) showing the end of the DNA fragment when sequenced from the one side. Black arrow indicates the decrease in peak heights. Grey arrows indicate Hi30L primer site. A similar electropherogram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

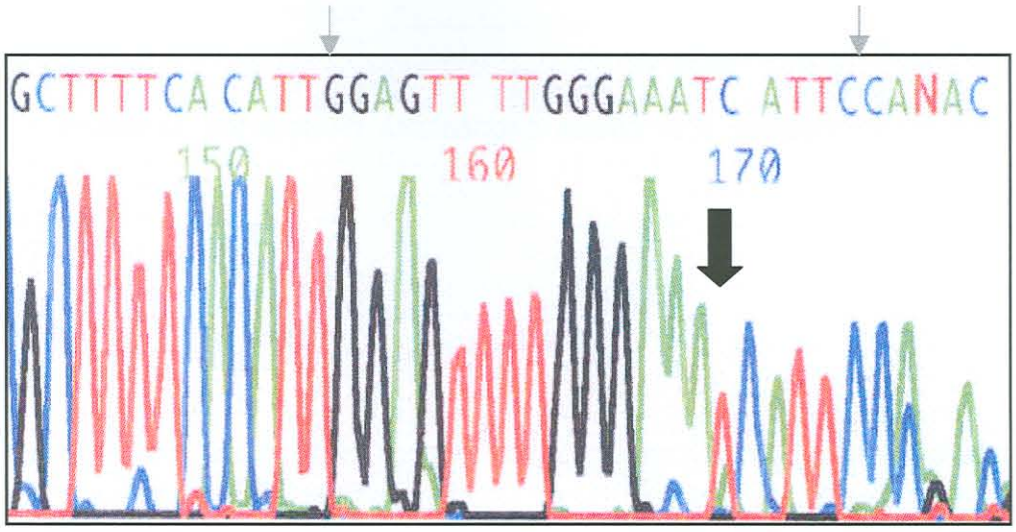

Figure 5.3: Electropherogram from direct sequencing of DNA template derived from tobacco cultivar '*Samsun/UK*' (NT1) showing the end of the DNA fragment when sequenced from the opposite side compared to the sequence in Figure 5.2. Black arrow indicates the decrease in peak heights. Grey arrows indicate 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.

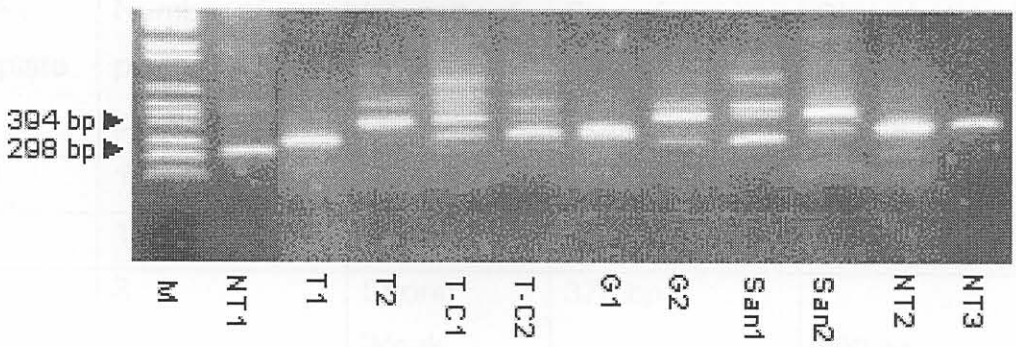

Figure 5.5: PCR amplification of genomic target DNAs with primers designed for flanking sequence of the Hi30 subtraction product. PCR amplification with genomic template DNA derived from tobacco cultivar '*Samsun/UK*' (NT1-3), with NT1 representing the original non-modified tobacco plant used in the 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).

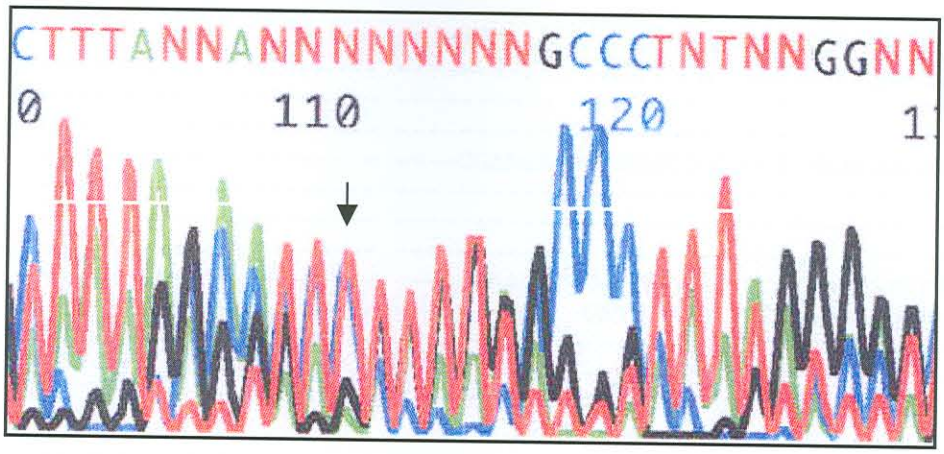

Figure 5.6: Electropherogram from sequencing of DNA template derived from a genetically modified tobacco plant (*'Samsun/UK'*) containing the *gus* gene insert (G1). Arrow indicate an example of a double peak.

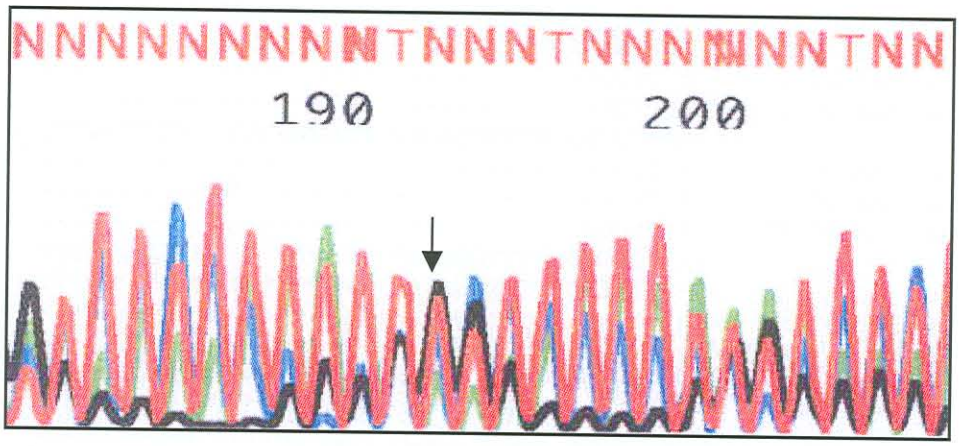

Figure 5.7: Electropherogram from sequencing of DNA template derived from a genetically modified tobacco plant (*'Samsun/UK'*) containing the *gus* gene insert (G2). Arrow indicate an example of a double peak

```
                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).

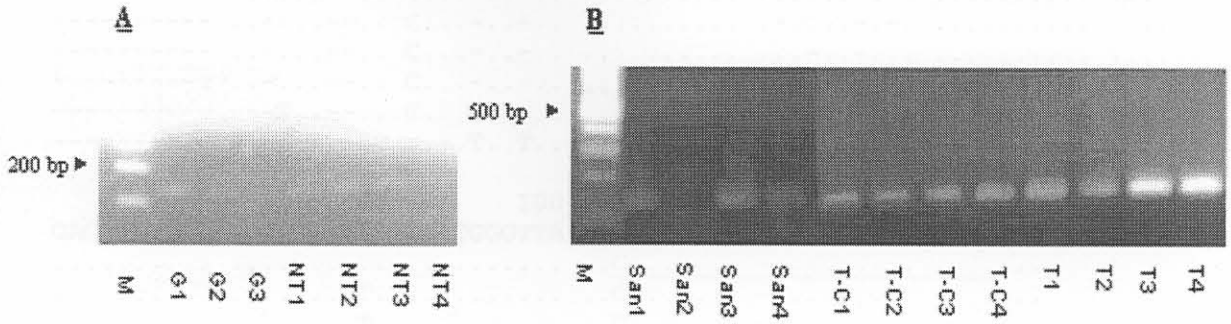

Figure 5.9: Amplification of genomic target DNA using PCR with primers designed for Hp12 with a single base pair change as identified in subtraction product Hp12. (A) DNA amplification with genomic DNA-derived from either non-modified wild-type individual tobacco plants of cultivar '*Samsun/UK*' (NT1 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).

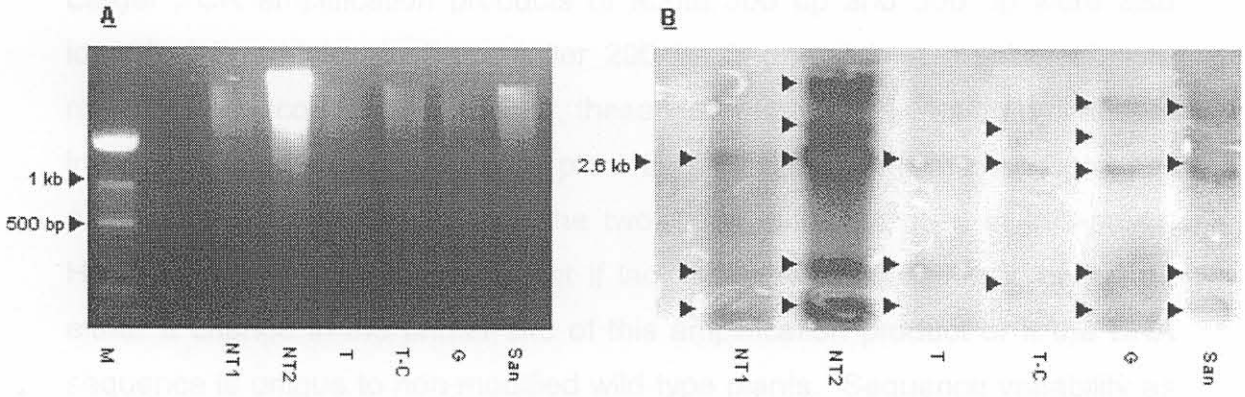

Figure 5.11: (A) Genomic *Bam*HI-digested DNA after separation on an agarose gel and stained with ethidium bromide from non-modified wild-type *'Samsun/UK'* plants (NT1 and NT2). NT1 represents the original non-modified tobacco plant DNA used in the RDA procedure; genetically modified plant DNA carrying 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.