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.

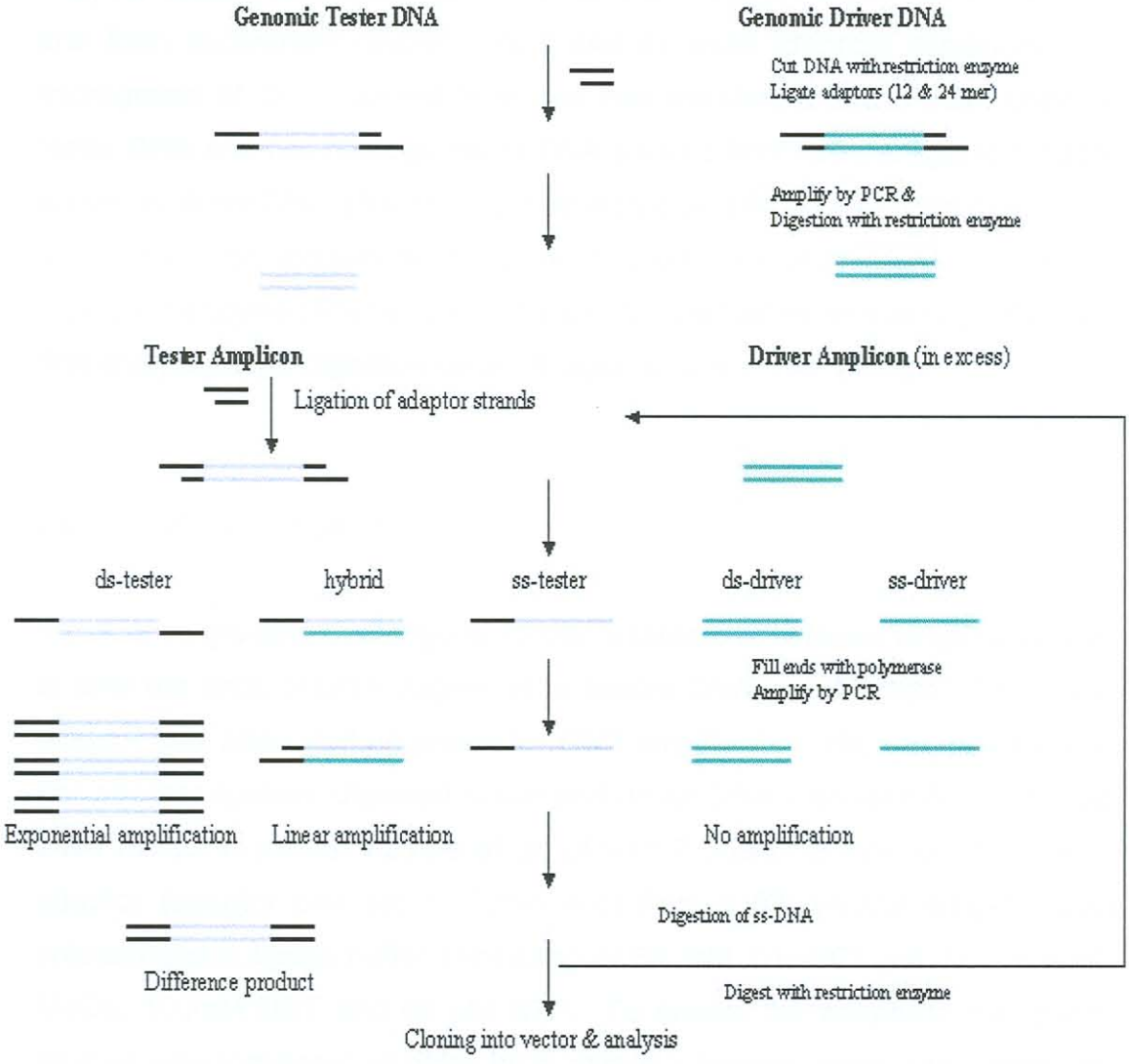

Figure A.1: Steps involved in the DNA subtraction procedure of Representational Difference Analysis.

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.

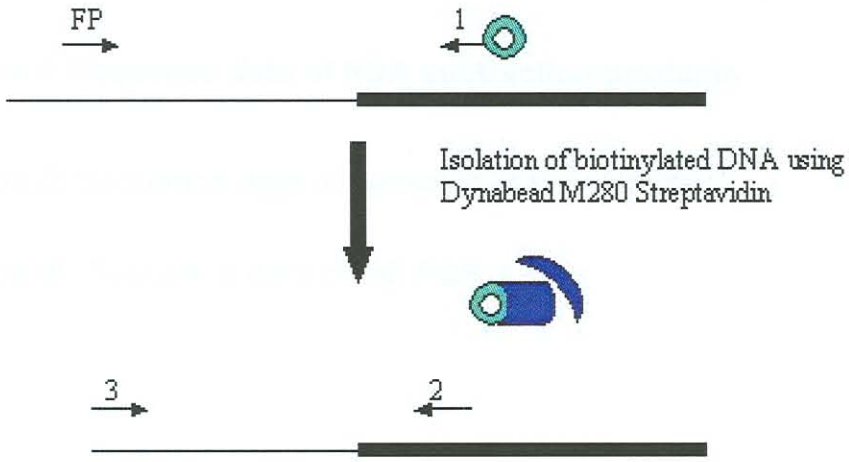

Figure A.2: Isolation of flanking DNA sequences in the genome by a two-step polymerase chain reaction and amplification of unknown DNA sequences (thin line) flanking a known DNA region (broad line). Arrows indicate biotinylated primer (primer 1), specific sequence primers (primers 2 and 3) and degenerated 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.