FUNCTIONAL ANALYSIS OF TTG1 HOMOLOGUES IN COTTON

Honours 2001

John Humphries

Discipline of Genetics

Department of Molecular Biosciences

Adelaide University

Contents

1. Introduction

Cotton fibres are single epidermal cells developed from the ovule. While a number of genes expressed exclusively in cotton fibre development have been isolated, the genes responsible for differentiation of fibre cells have not yet been identified (Basra, 1999). Arabidopsis leaf trichomes could potentially serve as a model for investigating the genetic mechanism controlling cotton fibre differentiation, as the two developmental processes have several features in common. Arabidopsis trichome development is well characterised both genetically and molecularly, and provides an excellent model for plant cell differentiation for several reasons (Marks et al., 1991). These reasons include ease of observation, relative simplicity of the developmental process, and viability of plants with mutant trichomes. One gene shown to be required for Arabidopsis trichome initiation is TRANSPARENT TESTA GLABRAl (TTGl) (Koornneef, 1981). This gene has recently been cloned and shown to encode a 341 amino acid protein with four WD40 repeats (Walker et al., 1999). The TTGl locus regulates several pathways in *Arabidopsis* (in addition to leaf trichome formation), including anthocyanin pigmentation, root hair formation, and seed coat mucilage production. Due to the role played by TTG 1 in *Arabidopsis* trichome initiation, it was thought homologues of *TTGl* might exist in cotton to control fibre initiation. Four putative *TTG1* homologues have been isolated from cotton, designated TTG(I), TTG(ll), TTG①P1 and pTTG(c)E2 (Orford, unpublished).

The aim of this project was to test for functional homologues of the *Arabidopsis TTG1* (*AtTTG1*) gene in cotton. Four putative *TTG1* homologues have previously been isolated from cotton, designated TTG(I), TTG(II), TTG①P1 and pTTG(c)E2 (Orford, unpublished). Two separate assays were performed to test each of these genes for functional homology to *Arabidopsis TTG1*, which has roles in trichome development and anthocyanin production. The first was a transient expression of these *TTG1*-like genes in a *Matthiola incana ttg1* mutant. Wild-type *M. incana* has purple petals, but the *ttg1* mutant has white petals, due to the anthocyanin production pathway being disrupted (see Figure 1) (A. Walker, pers. comm.).

Figure 1: **Wild-type *M. incana* flower (left) and *ttg1* mutant (right).**

Each cotton *TTG1*-like gene was transiently expressed in the mutant *M. incana via* particle bombardment. Restoration of the purple petal phenotype in some cells would demonstrate that complementation of the mutation has occurred, indicating that the cotton gene used in the bombardment was a functional homologue of *TTGI*. The *Arabidopsis TTG1* gene was also used in this assay as a positive control.

The second test involved a stable transformation of the *Arabidopsis ttg1* mutant, to determine if the cotton homologues could complement the mutant phenotype. *Arabidopsis* with a mutant *ttg1* gene exhibits no trichome formation and a transparent testa phenotype (Szymanski *et al.*, 2000). Mutant plants were to be transformed with the four *TTG1*-like cotton genes via *Agrobacterium*-mediated transformation, and progeny selected for successful transformation, and studied for restoration of trichome formation and normal testa colour. As with the *M. incana* assay, the *Arabidopsis TTG1* gene will be used as a positive control.

2. Particle Bombardment of *Matthiola*: a transient assay

2.1 Formation of constructs for particle bombardment

The four cotton *TTG1*-like genes were cloned into the vector pART7 (Gleave, 1992) (Figure 2) such that their expression would be driven by the constitutive cauliflower mosaic virus (CaMV) 35S promoter. This promoter confers high-level expression in all plant tissues (Harpster *et al.*, 1988), and hence would enable transient expression of the *TTG1*-like genes after particle bombardment.

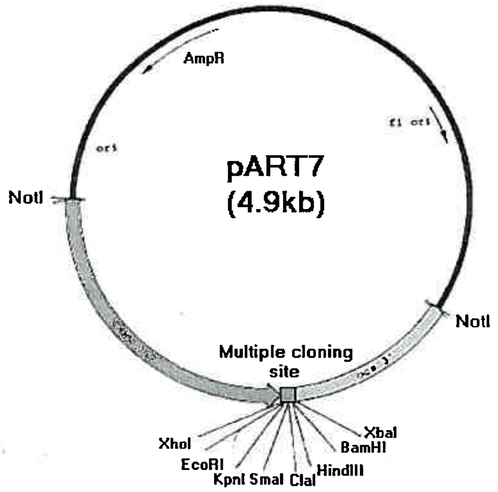


Figure 2: **pART7, the 4.9 kb primary cloning vector with ampicillin resistance (Amp^R), multiple cloning site (orange), *NotI* restriction sites, cauliflower mosaic virus 35S promoter (green) and 3' octopine synthase terminator region (blue) indicated (Adapted from Gleave, 1992).**

The original vector containing three of the four *TTG1*-like genes (TTG(I), TTG(II) and pTTG(c)E2) was the plasmid pBluescript SK(-), whilst TTG①P1 was in pGEM-T Easy.

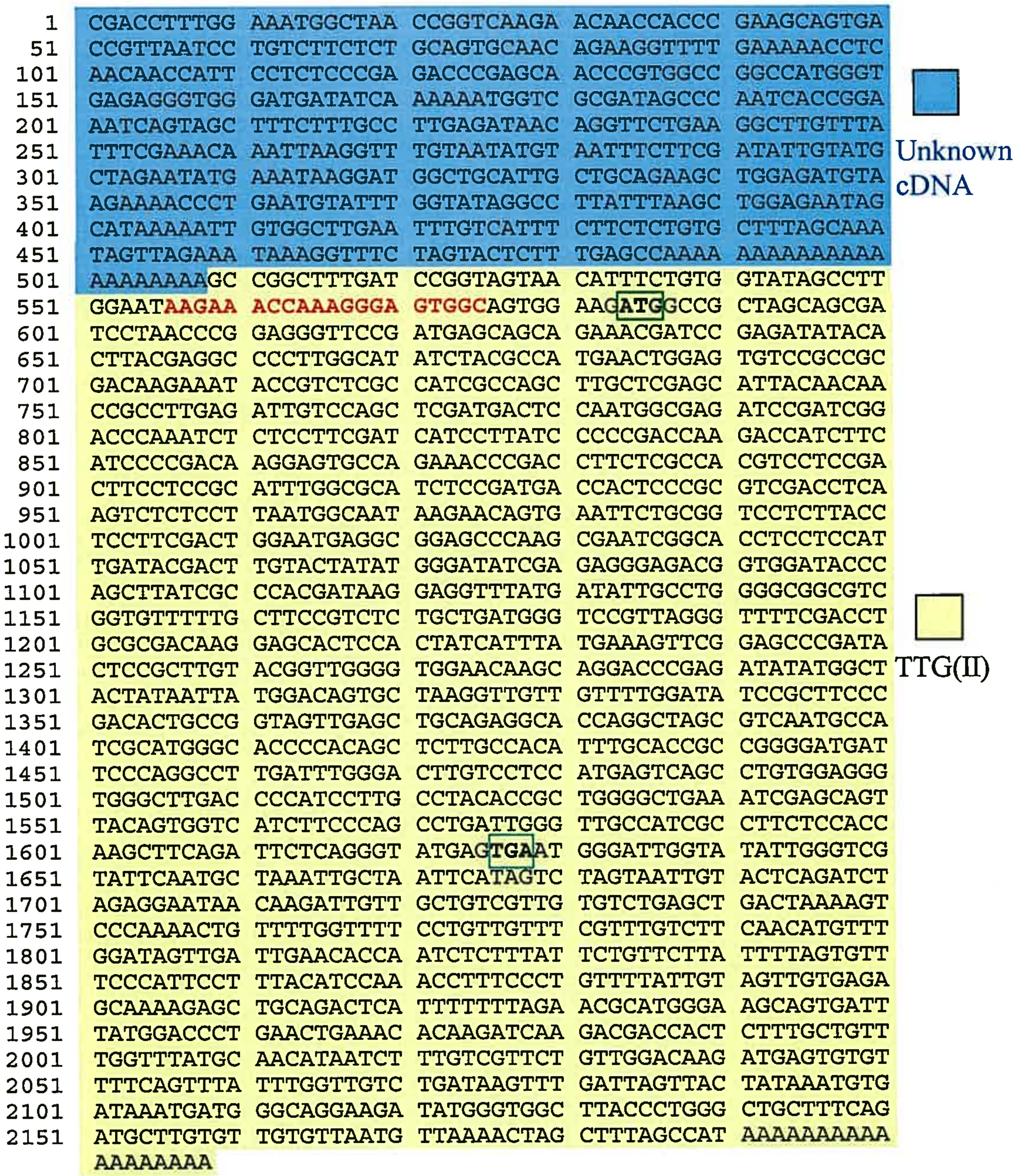
The multiple cloning site of pART7 contains *Xho*I, *Eco*RI, *Kpn*I, *Sma*I, *Hind*III, *Cla*I, *BamH*I and *Xba*I restriction sites to facilitate cloning of the gene of interest downstream of the 35S promoter (Figure 3).

35S ACGCTCGAGGAATTCGGTACCCCGGGTTCGAAATCGAAATCGATAAGCTTGGATCCTCTAGA **ocs**

*Xho*I *Eco*RI *Kpn*I *Sma*I *Cla*I *Hind*III *Bam*HI *Xba*I

<u>Figure 3:</u> Multiple cloning site of pART7

Directional cloning into pART7 was accomplished by performing a double digest on the clones containing TTG(I), TTG①P1, and pTTG(c)E2. However, TTG(II) in pBluescript contained no suitable restriction enzyme sites for cloning into pART7 because the original clone was chimeric. A 508 bp region of unknown cDNA was inserted 5' of TTG(II) in this chimeric clone (see Figure 4). This region of unknown cDNA would preferably be excluded from cloning into pART7. Therefore, instead of a direct insertion by digestion into pART7, PCR was performed on TTG(II) to first clone it into pGEM-T Easy.



```
   1  CGACCTTTGG AAATGGCTAA CCGGTCAAGA ACAACCACCC GAAGCAGTGA
  51  CCGTTAATCC TGTCTTCTCT GCAGTGCAAC AGAAGGTTTT GAAAAACCTC
 101  AACAACCATT CCTCTCCCGA GACCCGAGCA ACCCGTGGCC GGCCATGGGT
 151  GAGAGGGTGG GATGATATCA AAAAATGGTC GCGATAGCCC AATCACCGGA
 201  AATCAGTAGC TTTCTTTGCC TTGAGATAAC AGGTTCTGAA GGCTTGTTTA
 251  TTTCGAAACA AATTAAGGTT TGTAATATGT AATTTCTTCG ATATTGTATG
 301  CTAGAATATG AAATAAGGAT GGCTGCATTG CTGCAGAAGC TGGAGATGTA
 351  AGAAAACCCT GAATGTATTT GGTATAGGCC TTATTTAAGC TGGAGAATAG
 401  CATAAAAATT GTGGCTTGAA TTTGTCATTT CTTCTCTGTG CTTTAGCAAA
 451  TAGTTAGAAA TAAAGGTTTC TAGTACTCTT TGAGCCAAAA AAAAAAAAAA
 501  AAAAAAAAGC CGGCTTTGAT CCGGTAGTAA CATTTCTGTG GTATAGCCTT
 551  GGAATAAGAA ACCAAAGGGA GTGGCAGTGG AAGATGGCCG CTAGCAGCGA
 601  TCCTAACCCG GAGGGTTCCG ATGAGCAGCA GAAACGATCC GAGATATACA
 651  CTTACGAGGC CCCTTGGCAT ATCTACGCCA TGAACTGGAG TGTCCGCCGC
 701  GACAAGAAAT ACCGTCTCGC CATCGCCAGC TTGCTCGAGC ATTACAACAA
 751  CCGCCTTGAG ATTGTCCAGC TCGATGACTC CAATGGCGAG ATCCGATCGG
 801  ACCCAAATCT CTCCTTCGAT CATCCTTATC CCCCGACCAA GACCATCTTC
 851  ATCCCCGACA AGGAGTGCCA GAAACCCGAC CTTCTCGCCA CGTCCTCCGA
 901  CTTCCTCCGC ATTTGGCGCA TCTCCGATGA CCACTCCCGC GTCGACCTCA
 951  AGTCTCTCCT TAATGGCAAT AAGAACAGTG AATTCTGCGG TCCTCTTACC
1001  TCCTTCGACT GGAATGAGGC GGAGCCCAAG CGAATCGGCA CCTCCTCCAT
1051  TGATACGACT TGTACTATAT GGGATATCGA GAGGGAGACG GTGGATACCC
1101  AGCTTATCGC CCACGATAAG GAGGTTTATG ATATTGCCTG GGGCGGCGTC
1151  GGTGTTTTTG CTTCCGTCTC TGCTGATGGG TCCGTTAGGG TTTTCGACCT
1201  GCGCGACAAG GAGCACTCCA CTATCATTTA TGAAAGTTCG GAGCCCGATA
1251  CTCCGCTTGT ACGGTTGGGG TGGAACAAGC AGGACCCGAG ATATATGGCT
1301  ACTATAATTA TGGACAGTGC TAAGGTTGTT GTTTTGGATA TCCGCTTCCC
1351  GACACTGCCG GTAGTTGAGC TGCAGAGGCA CCAGGCTAGC GTCAATGCCA
1401  TCGCATGGGC ACCCCACAGC TCTTGCCACA TTTGCACCGC CGGGGATGAT
1451  TCCCAGGCCT TGATTTGGGA CTTGTCCTCC ATGAGTCAGC CTGTGGAGGG
1501  TGGGCTTGAC CCCATCCTTG CCTACACCGC TGGGGCTGAA ATCGAGCAGT
1551  TACAGTGGTC ATCTTCCCAG CCTGATTGGG TTGCCATCGC CTTCTCCACC
1601  AAGCTTCAGA TTCTCAGGGT ATGAGTGAAT GGGATTGGTA TATTGGGTCG
1651  TATTCAATGC TAAATTGCTA ATTCATAGTC TAGTAATTGT ACTCAGATCT
1701  AGAGGAATAA CAAGATTGTT GCTGTCGTTG TGTCTGAGCT GACTAAAAGT
1751  CCCAAAACTG TTTTGGTTTT CCTGTTGTTT CGTTTGTCTT CAACATGTTT
1801  GGATAGTTGA TTGAACACCA ATCTCTTTAT TCTGTTCTTA TTTTAGTGTT
1851  TCCCATTCCT TTACATCCAA ACCTTTCCCT GTTTTATTGT AGTTGTGAGA
1901  GCAAAAGAGC TGCAGACTCA TTTTTTTAGA ACGCATGGGA AGCAGTGATT
1951  TATGGACCCT GAACTGAAAC ACAAGATCAA GACGACCACT CTTTGCTGTT
2001  TGGTTTATGC AACATAATCT TTGTCGTTCT GTTGGACAAG ATGAGTGTGT
2051  TTTCAGTTTA TTTGGTTGTC TGATAAGTTT GATTAGTTAC TATAAATGTG
2101  ATAAATGATG GGCAGGAAGA TATGGGTGGC TTACCCTGGG CTGCTTTCAG
2151  ATGCTTGTGT TGTGTTAATG TTAAAACTAG CTTTAGCCAT AAAAAAAAAA
      AAAAAAAA
```

Figure 4: TTG(II) insert sequence in pBluescript

Blue shading represents unknown cDNA, yellow shading represents TTG(II) region. Binding site of primer ttg2bp1 is signified by red. Start and stop codons of TTG(II) are indicated by green boxes.

2.1a Isolation and cloning of TTG(II)

In order to exclude the 508 bp segment of unknown cDNA prior to ligation into pGEM-T Easy, a primer was designed to bind a 20 bp region 3' of this unknown section, but 5' of the initiation codon of TTG(II) (Figure 4). This primer contained an engineered *Bam*HI site in order to excise the TTG(II) gene from this point in the next cloning step, excluding the section of unknown cDNA. The primer was designated ttg2bp1, and its sequence was as follows:

BamHI Binding region
ttg2bp1: 5' GGGGATCCAAGAAACCAAAGGGAGTGGC 3'

PCR was carried out on the TTG(II) clone in pBluescript, using the ttg2bp1 primer and the reverse sequencing primer (RSP) of pBluescript to amplify a 1.9 kb region containing the TTG(II) gene (Figure 5). The conditions for the PCR were as follows; Initial denaturation at 94°C for 2 minutes, followed by 30 cycles of denaturation at 94°C for 30 seconds, primer annealing at 53°C for 30 seconds, and extension at 72°C for 2 minutes, followed by final extension at 72°C for 4 minutes.

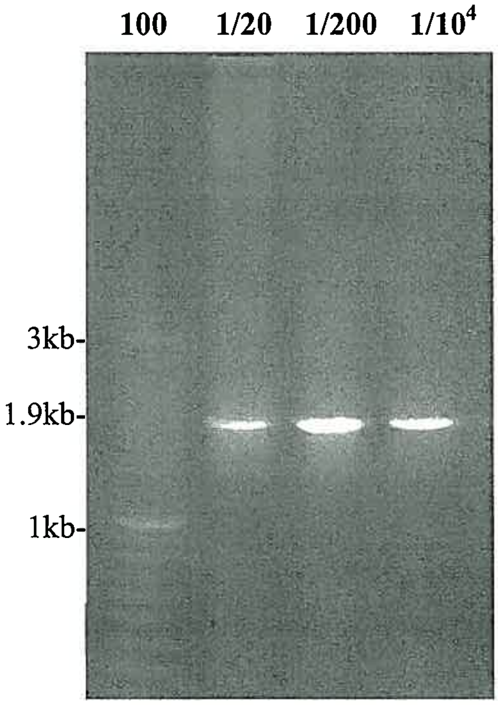


Figure 5: **PCR of TTG(II) gene in pBluescript using RSP and ttg2bp1 primers. Above each lane the dilution of alkaline lysis miniprep used as the template is indicated. Amplification of the required 1.9 kb band is displayed in each lane, with different dilutions having apparently little effect**

As pBluescript possesses the same antibiotic resistance gene (Amp^R) as pGEM-T Easy, the PCR bands were required to be extracted from the gel before ligation into pGEM-T Easy. This was performed to reduce the chance of recovering TTG(II) in pBluescript. Gel extraction was performed over three PCR bands using the QIAGEN gel extraction kit. After eluting in 50 µl 1 X TE and ethanol precipitation, the DNA was resuspended in 10 µL dH_2O. Approximately 200 ng of DNA (3 µL) was added to the ligation reaction with 50 ng of pGEM-T Easy. After heat-shock transformation at 42°C using $CaCl_2$-competent cells (DH5αF') **(Sambrook, 1989),** blue/white selection was carried out on an ampicillin plate (supplemented with ITPG and X-gal) to identify recombinants. Alkaline lysis minipreps (Sambrook, 1989) were performed on three white colonies, and these were digested with *Eco*RI to check for the presence of the TTG(II) insert. If the TTG(II) gene was ligated into pGEM-T Easy correctly, an *Eco*RI digest was expected to give fragments of 2.9 kb, 1.5 kb and 400 bp (Figure 6A). The digests produced the expected fragments, as shown in Figure 6B.

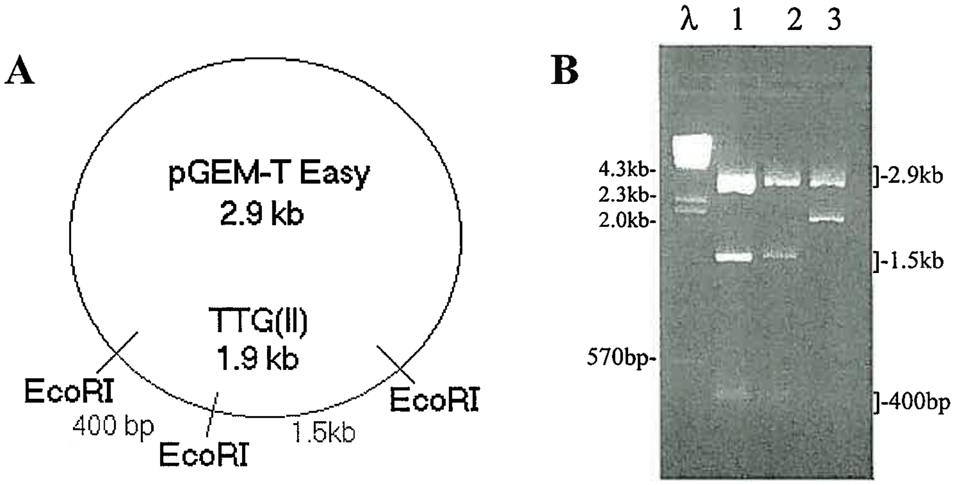


Figure 6: *EcoRI* digestion of TTG(II) in pGEM-T Easy.

(A) Expected fragment sizes for *EcoRI* digest of TTG(II) inserted correctly into pGEM-T Easy.

(B) Agarose gel of *EcoRI* digestion of the three colonies tested (1-3)

Figure 6B demonstrates that samples 1 and 2 contain the TTG(II) PCR product in pGEM-T Easy, as they display the expected fragment sizes. The fragments generated from sample 3 are inconsistent with expectation. Plasmid DNA from clone #2 was sequenced with the SP6 and T7 primers, which confirmed that TTG(II) had been correctly ligated into pGEM-T Easy.

2.1b Cloning of cotton *TTG1*-like genes into pART7

The next step was to transfer the four cotton *TTG1*-like genes from their respective vectors (TTG(I) and TTG(c)E2 in pBluescript, TTG(II) and TTG①P1 in pGEM-T Easy) into the pART7 plasmid. This was achieved by excising each gene by the restriction digests shown in Figure 7. pART7 was digested with the appropriate enzymes to provide the desired sticky ends for cloning of each of these inserts. The enzymes used for digestions are detailed in Figure 7.

Insert	**Excised with**	**Expected size**	**Digest pART7 with**
TTG(I)	*BamHI* and *Hind*III	1.3 kb	*BamHI* and *Hind*III
TTG(II)	*BamHI*	1.9 kb	*BamHI*
TTG①P1	*Sma*I and *Spe*I	1.7 kb	*Sma*I and *Xba*I*
pTTG(c)E2	*Xba*I and *Kpn*I	1.9 kb	*Xba*I and *Kpn*I

* Digestion with *Spe*I and *Xba*I provides compatible ends for ligation.

<u>Figure 7:</u> Excision of each *TTG1*-like gene from their respective plasmid by digestion.

After a three-hour digestion at 37°C, each of the four pART7 digests was phenol-chloroform extracted and ethanol precipitated to purify the plasmid DNA. pART7 digested for TTG(II) insertion (*Bam*HI single digest) was CIP-treated to suppress re-ligation of the vector. This was not necessary for the remaining three vector digests, as these were prepared as double digests, and the two restriction sites used were incompatible.

Gel extraction of each *TTG1*-like gene from the original vector was required because both vectors containing the genes of interest (pGEM-T Easy and pBluescript) possess ampicillin resistance, as does pART7. Approximately 1 µg of each construct was digested (as detailed in Figure 7) to excise the desired fragment from the vector (see Figure 8).

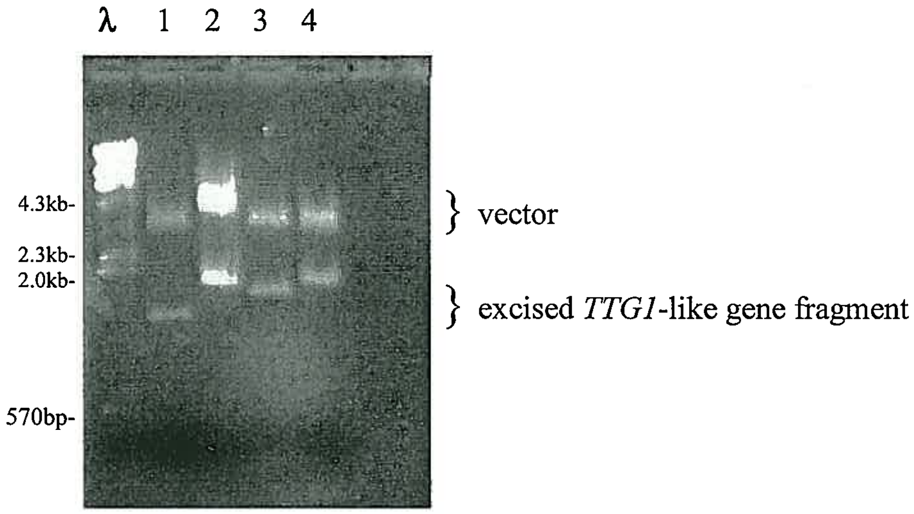


Figure 8: Cloning digests of the four *TTG1*-like genes

1) **TTG(I) in pBluescript (excised band 1.3 kb)**
2) **TTG(II) in pGEM-T Easy (1.9 kb)**
3) **TTG①P1 in pGEM-T Easy (1.7 kb)**
4) **pTTG(c)E2 in pBluescript (1.9 kb)**

After extracting the lower bands of each digest using Qiagen Gel Extraction Kit, and eluting in 30 µL Tris-Cl, 2 µL of each was run on a gel to ensure sufficient DNA was present (Figure 9)

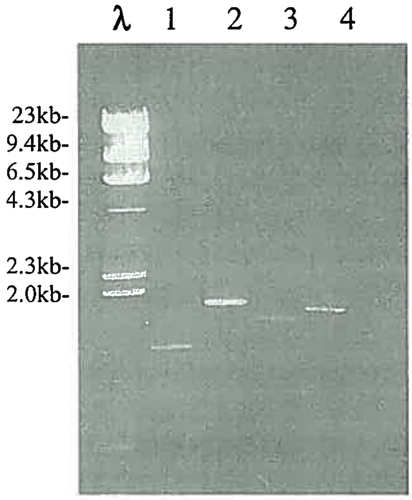


Figure 9: Gel fragment preparations to be ligated into pART7

Lane 1: TTG(I) (1.3) kb
Lane 2: TTG(II) (1.9) kb
Lane 3: TTG①P1 (1.7) kb
Lane 4: pTTG(c)E2 (1.9) kb

For ligation into pART7, 5 µL (approximately 200 ng) of each fragment was added to 50 ng of the appropriately digested pART7 vector, plus T4 DNA ligase and buffer, with incubation overnight at 16°C. Ligation reactions were also performed on the four pART7 digests alone (no insert DNA) as a negative control. These ligations were transformed into $CaCl_2$-competent cells using the heat-shock method. Un-ligated pART7 (digested for each of the four inserts) was also transformed to check the levels of undigested plasmid present. When plated onto ampicillin, the un-ligated plasmids resulted in almost no colonies on the four plates, indicating a low level of undigested pART7. Re-ligated pART7 prepared for the four inserts resulted in a small number of colonies, demonstrating that the level of pART7 digested by only one of the enzymes in each pair was low, and that CIP-treatment of *Bam*HI digested plasmid had been successful. Ligations with the insert present resulted in 10-fold the number of colonies compared to the negative controls; an indication that all four ligations had been successful.

To test colonies for the correct insert on the TTG(I) and pTTG(c)E2 ligation plates, a rapid lysis boiling prep was performed using 1.5 ml of culture grown from five TTG(I) colonies and ten pTTG(c)E2 colonies. PCR was performed on 12 colonies on the TTG(II) ligation plate, and 20 colonies from the TTG①P1 plate to test for the desired insert.

Verification of TTG(I) integration into pART7

*Not*I and *Cla*I digests were performed on the five colony preps to check for correct TTG(I) ligation into pART7. *Cla*I linearises the construct by cutting within the multiple cloning site of pART7 (Figure 10). The expected fragment size resulting from *Cla*I digestion would therefore be 6.2 kb if the insert was present, or 4.9 kb if the vector alone was present. pART7 has two *Not*I restriction sites in pART7, such that *Not*I digestion would produce fragment sizes of 2.0 kb and 2.9 kb. If the TTG(I) insert had been correctly ligated into the pART7 MCS, the expected fragments sizes would be 3.3 kb and 2.9 kb. (Figure 10).

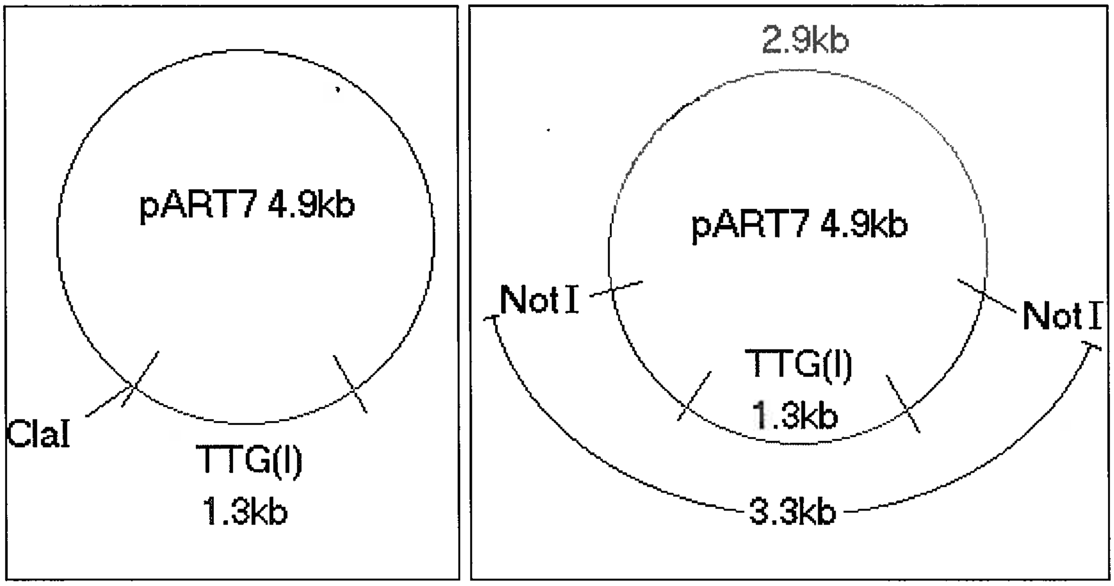


Figure 10: ***Cla*I (left) and *Not*I (right) digestion of TTG(I) in pART7, with expected fragment sizes shown.**

Approximately 200 ng of DNA from each rapid lysis boiling prep was digested for 2 hours at 37°C with *Not*I or *Cla*I. The results of these digests are shown below (Figure 11). pART7 alone was digested as a control.

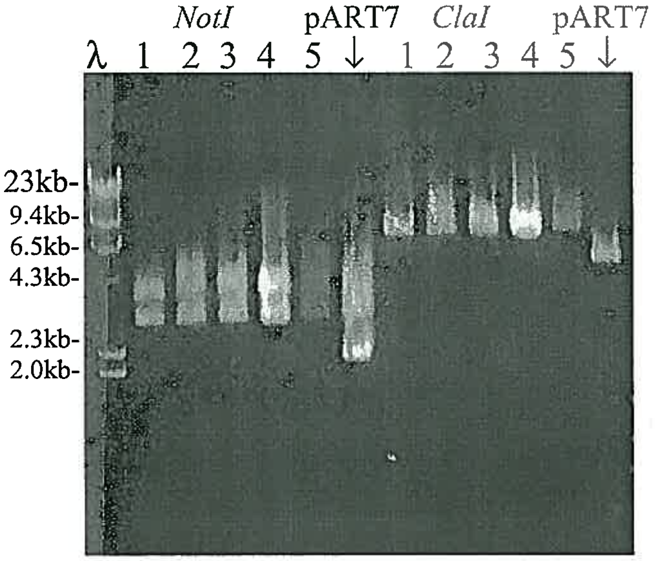


Figure11: ***Not*I (left) and *Cla*I (right) digestion of TTG(I)/pART7 ligation samples 1-5.**

Figure 11 indicates that TTG(I) was correctly inserted into pART7 in all five samples, as the 2 kb band from the *NotI* digest of pART7 alone has 'increased' in size to approximately 3.3 kb, although sample 5 did not show clear bands. The *ClaI* digest, where all five samples show an increase in size compared to pART7 alone, confirms the insertion of TTG(I) into pART7. Plasmid DNA from clone #1 was subsequently sequenced, which confirmed that the ligation was successful (see Section 2.2 and Appendix A).

Verification of TTG(II) integration into pART7

To test for the presence of the TTG(II) insert in pART7, PCR was performed on 12 colonies. As the integration of TTG(II) into pART7 was not a directional cloning step, PCR was also used to test for the correct orientation of TTG(II). This was accomplished using the internal TTG(II) primer, pTTG4 and the external primer pART7A, which binds the vector upstream of the multiple cloning site of pART7 (see Section 2.2). If the insert were in the correct orientation in the vector, these primers would amplify a 650 bp region. However, if the insert were in the incorrect orientation, both primers would extend in the same direction, and there would be no amplification.

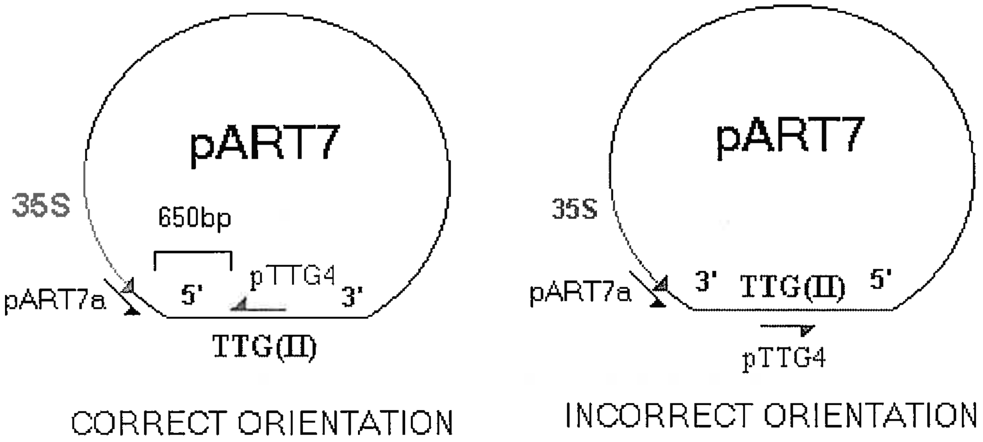


Figure 12: **Correct (left) and incorrect (right) orientation of TTG(II) in pART7, with the 650 bp region to be amplified indicated in the correct orientation.**

The results of the PCR of samples 1-12 are shown in Figure 13.

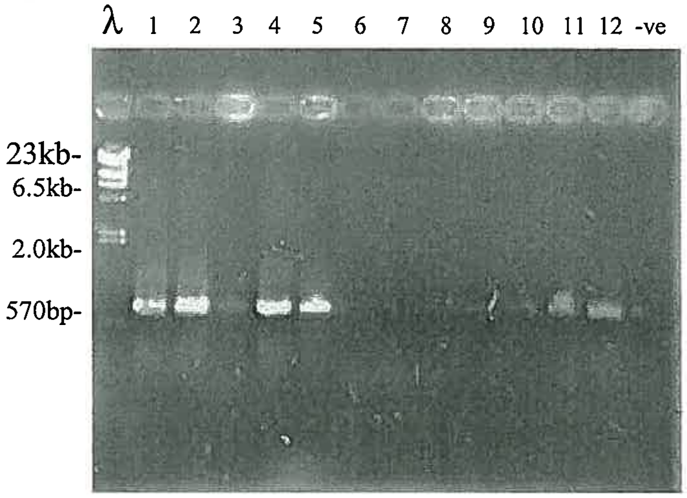


Figure 13: Agarose gel of PCR using pART7a and pTTG4 primers to amplify a 650 bp region with TTG(II) in the correct orientation in samples 1-24. No DNA was added to the negative control.

The results in Figure 13 show that samples 1, 2, 4, 5, 11 and 12 contained the TTG(II) insert in the correct orientation, as all generated the expected 650 bp band on the gel. Sample 2 was sequenced to confirm the ligation had been successful (see Section 2.2 and Appendix B).

Verification of TTG①P1 integration into pART7

PCR was performed on 20 colonies from the TTG①P1/pART7 ligation plate, using internal primers pTTG2B and pTTG1 (Orford, unpublished), which amplify a 570 bp region of the TTG①P1 gene (Figure 14).

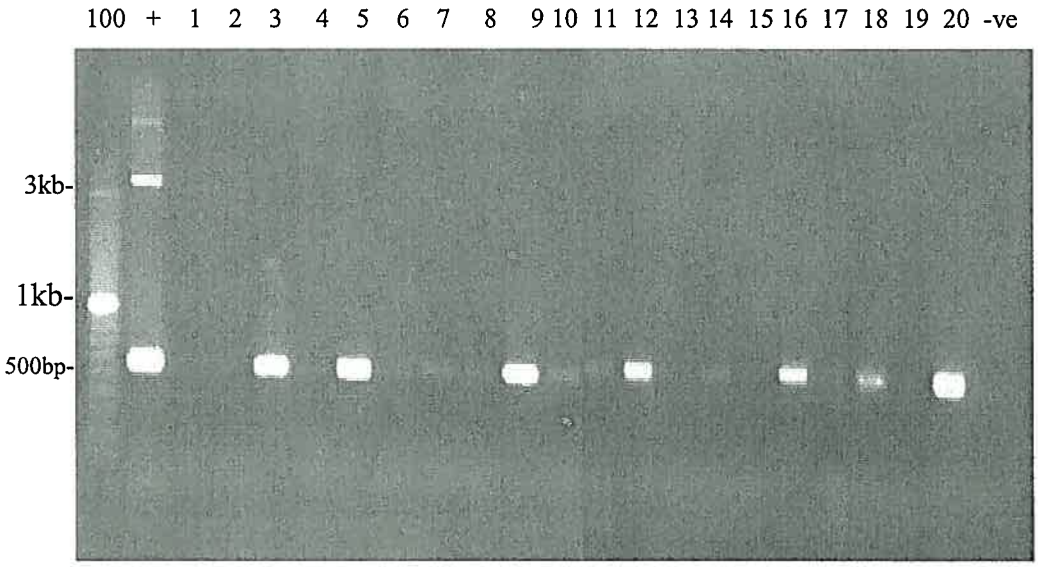
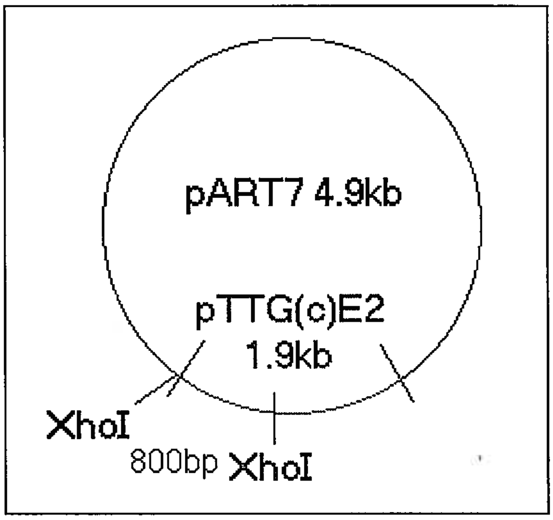


Figure 14: **Agarose gel of PCR of TTG①P1/pART7 ligation colonies (samples 1-20) with primers pTTG2B and pTTG1. The original pGEM-T Easy clone containing the TTG①P1 gene was amplified with the same primers as a positive control. No DNA was added to the negative control.**

This result demonstrates that several clones (3, 5, 9, 12, 16, 18, 20) contained the TTG①P1 insert, since they all displayed the amplified 570 bp region. Sample 5 was subsequently sequenced to confirm the successful ligation (see Section 2.2 and Appendix C).

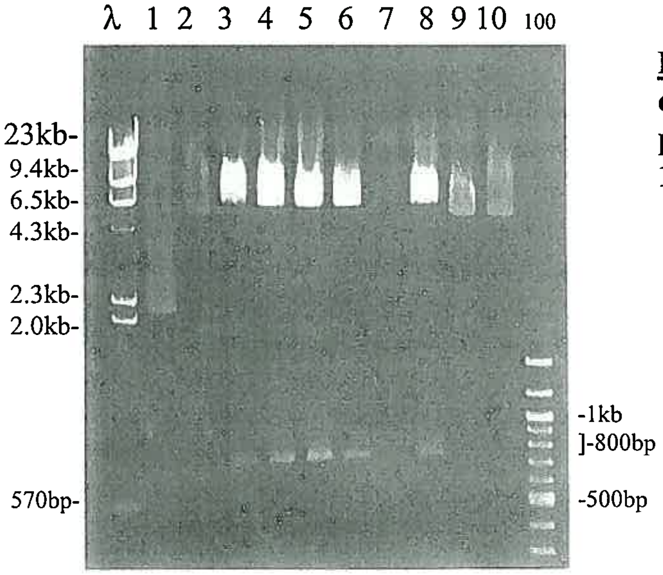
Verification of pTTG(c)E2 integration into pART7

A rapid lysis boiling prep was performed on 10 colonies resulting from the pTTG(c)E2 ligation into pART7. Each was digested with *Xho*I, which cuts within the pTTG(c)E2 gene, and in the MCS of pART7 (see Figure 15).



<u>Figure15:</u> *Xho*I digestion of pART7 with pTTG(c)E2 inserted.

If the pTTG(c)E2 insert had been successfully ligated into pART7, a *Xho*I digest of the construct would be expected to give a 6.0 kb and 800 bp fragment. The pART7 vector alone would show a single 4.9 kb fragment. The results of the digest are shown in Figure 16.



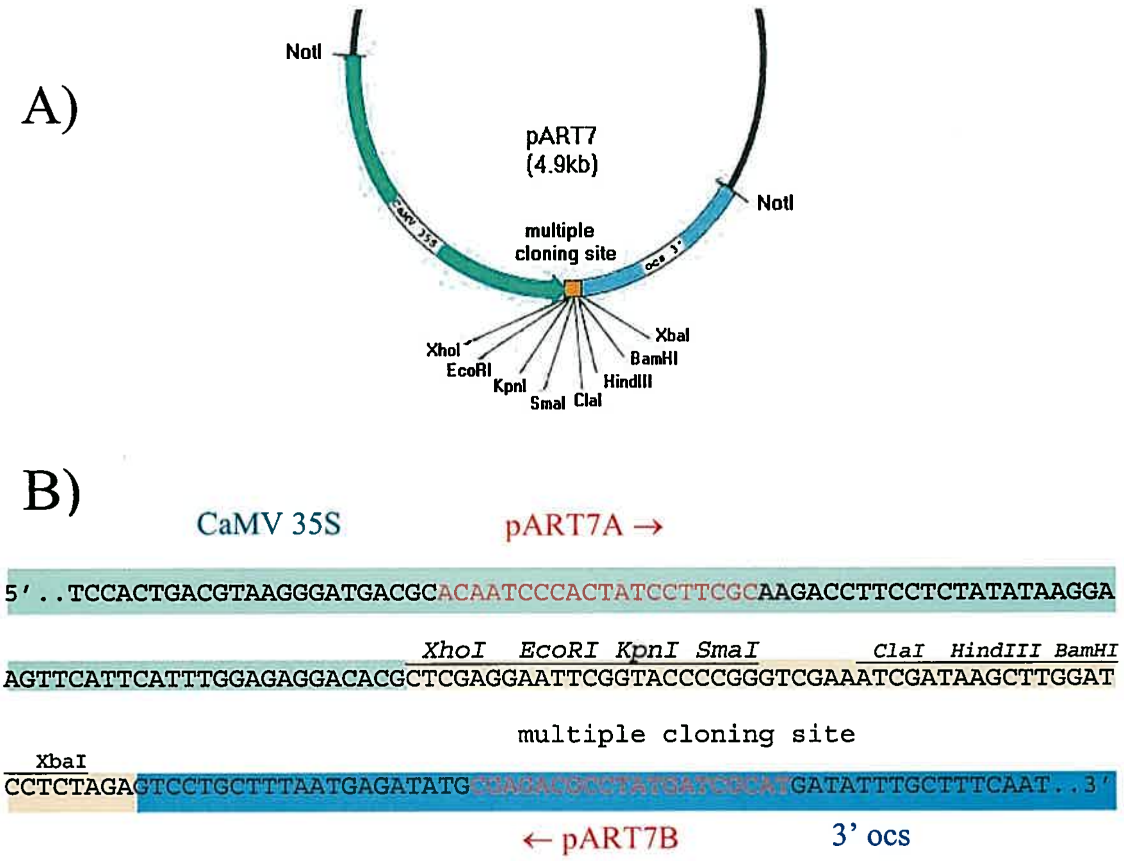
<u>Figure 16:</u> Agarose gel of *Xho*I digest of potential pTTG(c)E2 in pART7 constructs (samples 1-10).

Figure 16 shows that the *Xho*I digest of clones 3, 4, 5, 6 and 8 produced the expected fragments of approximately 6 kb and 800 bp, while samples 2, 9 and 10 reflected the pART7 vector only, with a single band at 4.9 kb. Sample 8 was subsequently sequenced, to confirm the ligation had occurred correctly (see Section 2.2 and Appendix D).

2.2 Sequence Analysis of constructs for particle bombardment

After template DNA purification using the Bio-rad Plasmid Miniprep kit, each pART7 construct was sequenced using internal primers that were specific for each inserted gene (Orford, unpublished). All of these sequences matched their original gene sequence (Orford, unpublished), as expected. However, these sequences showed only that the gene of interest was present, not how it was orientated into pART7, as none of the vector is sequenced in these reactions.

To gain information about the nature of the genes' integration into the vector, external primers were designed to bind upstream and downstream of the multiple cloning site of pART7 (designated pART7A and pART7B) (see Figure 17B).



<u>Figure 17:</u> (A) Diagram of pART7 multiple cloning site (orange) flanked by 35S CaMV promoter (green) and the 3' octopine synthase terminator region (blue). (B) Sequence of pART7 MCS and flanking primer sites of pART7A and pART7B (shown in red).

Using either of these primers (pART7A or pART7B), the gene of interest can be sequenced in the context of the pART7 vector. These primers have an additional advantage in that they can be used in PCR to test for presence or size of an insert in the MCS of pART7 without being specific for the gene of interest.

Appendices A-D detail the data obtained by sequencing each of the four constructs with pART7A and pART7B in conjunction with internal primers specific for each of the *TTG1*-like cotton genes (Figure 18). These sequences were as expected, confirming that all four genes had correctly ligated into the MCS of pART7.

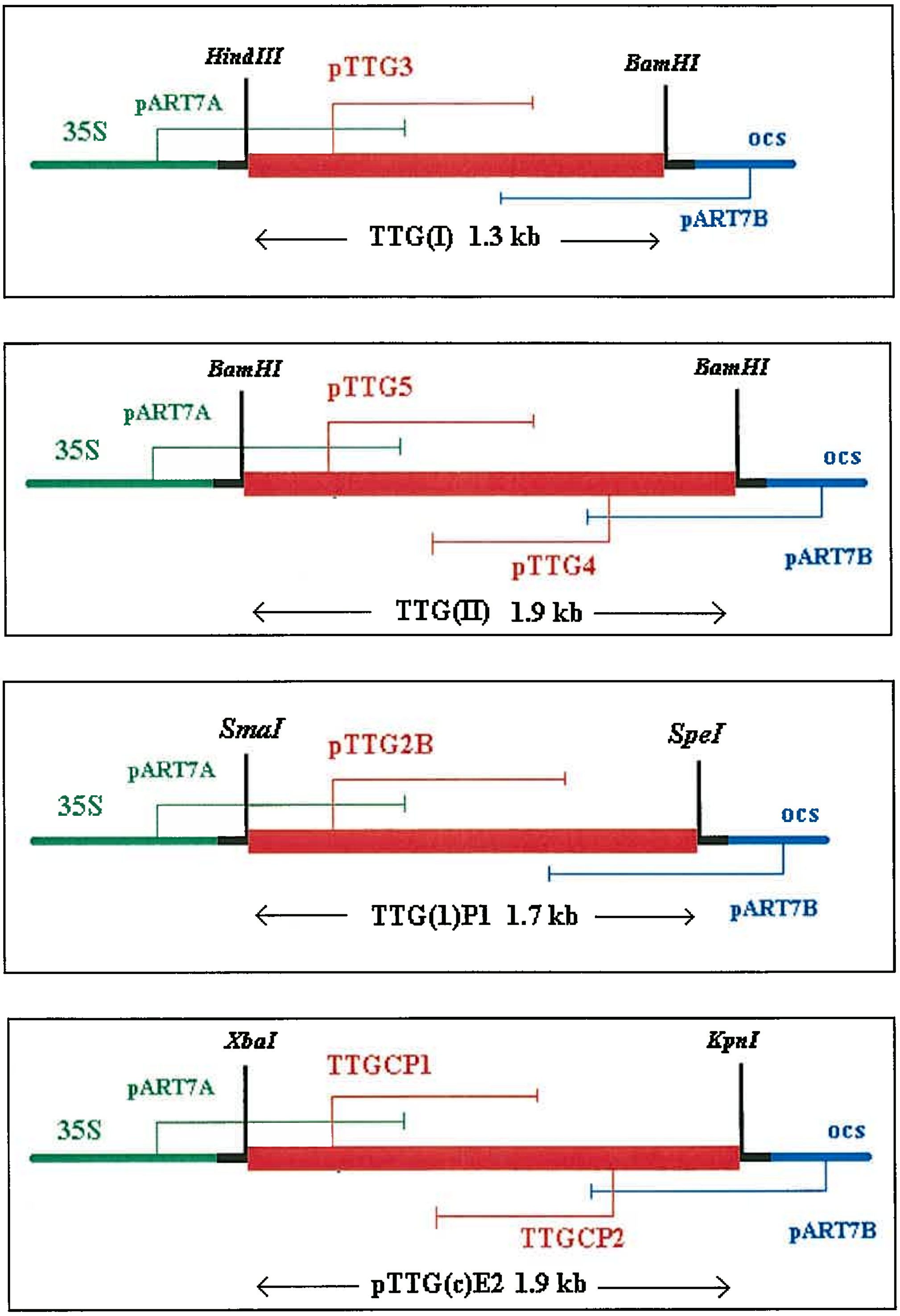


Figure 18: Sequencing scheme for each clone. The two pART7 primers (pART7A and pART7B) flanking the MCS were used to sequence each clone, while internal primers for each gene (shown in red) were used to complete gaps in the sequence. Restriction sites used for ligation into pART7 are indicated.

2.3 Particle Bombardment

Introduction

The principle of particle bombardment is to use high-velocity microprojectiles to penetrate outer cell walls in order to introduce genetic material into living cells. DNA is coated onto the microprojectile (gold particles to be used in this investigation), and accelerated to velocities of approximately 300 m/s to penetrate intact plant cell walls (Oard, 1993). The particle injection gun accelerates particles directly in a stream of low pressure helium (Vain *et al.*, 1993). The process is performed in a vacuum chamber to reduce the air friction of particles and tissue damage.

M incana mutants were used for particle bombardments. The white petal phenotype of *ttgl* mutants was tested for complementation by bombarding petals with cotton *TTG1*-like gene constructs, and inspecting for anthocyanin induction within transformed cells. Chalcone synthase (*chs*) mutants were also bombarded as a negative control. Chalcone synthase is involved early in the anthocyanin biosynthesis pathway, while TTGl is involved in later regulatory steps (Graham, 1998). The *chs* mutant in *M. incana* also produces the white petal phenotype. Transformation of the *chs* mutant with *AtTTG1* is not expected to produce purple spots of anthocyanin and therefore provides an ideal negative control for this assay.

M. incana bombardment procedures

A selection of flowers for transient expression assays was removed from *Matthiola incana ttg1* (line 17) or *chs* (line 18) mutants, and were placed onto a sterile petri dish containing Murashige and Skoog (MS) media and 0.75% (w/v) agar until bombardment (Figure 19B). The *M. incana* plants used were sterile, due to chromosome rearrangements (Lesley, 1973). There was therefore an experimental advantage to using these plants, as cells of the petal still expand, but cell division stops once petals have formed. The *ttg1* homologue within the mutant line 17 (*ttg1* mutant) has been sequenced and shown to contain a lesion (arginine substitution for a tryptophan) in one of the WD-40 repeats (Figure 33) (A. Walker, pers. com.).

Each of the pART7 constructs was purified using the QIAGEN midi-plasmid kit, yielding DNA concentrations ranging from 2 to 4 µg/µl. 1 micron gold particles were used as the vehicle for the bombardment. For each round of transformation (eight bombardments), 8 mg of gold was prepared by washing in cold ethanol and sterile water, before resuspending thoroughly in a final volume of 100 µl of water. 10 µg of plasmid DNA was then added to the gold suspension, followed by the addition of 40 µl 0.1M spermidine and 100 µl 2.5M $CaCl_2$ with gentle vortexing to precipitate the DNA onto the gold. The DNA-coated gold suspension was pipetted onto each of eight Swinney filter holders (Figure 19A), and allowed to dry for 10 minutes. The filter with absorbed gold particles was then screwed into the particle injection gun, and a *M. incana* flower placed approximately 25 cm beneath it (Figure 19C). The chamber was then evacuated to 90 psi and petals bombarded with the DNA-coated gold particles *via* helium injection for 0.12 milliseconds. Bombarded flowers were then incubated on MS plates at 25°C for approximately 40 hours before examination for purple spots of anthocyanin.

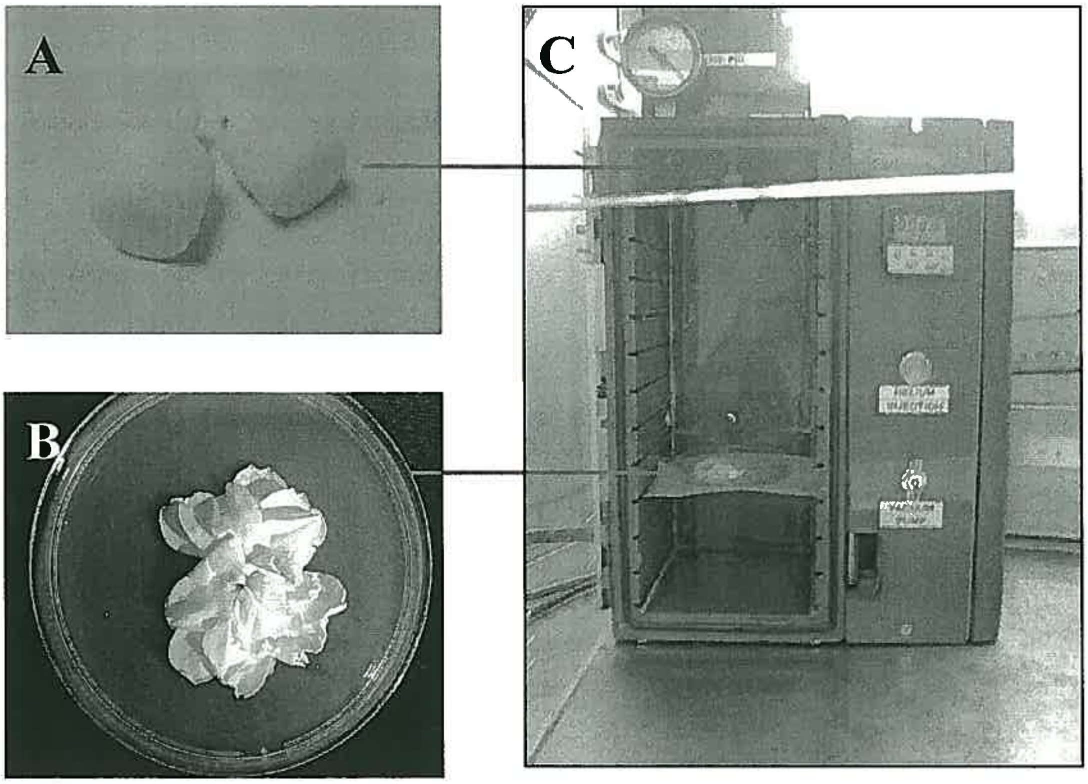


Figure 19: Particle bombardment procedure. (A) Swinney filter holders onto which DNA-coated gold particles are placed. (B) A *Matthiola* flower placed on MS agar. (C) The particle injection gun.

2.4 Transient functional assay: Positive control

To ensure that the particle preparation and bombardment conditions were suitable, bombardment was first performed using the *Arabidopsis TTG1* (*AtTTG1*) gene as a positive control (cloned into pART7; A. Walker, pers. comm.). *AtTTG1* bombardment of *M. incana ttg1* mutant petals has previously been shown to produce purple spots of anthocyanin (A. Walker, pers. comm.).

Particle bombardment was performed on four *M. incana ttg1* mutant flowers, and four *chs* mutants as a negative control Purple spots of anthocyanin were observed by the naked eye on the four *ttg1 M. incana* mutant flowers approximately 40 hours after bombardment (Figure 20 A-D). The anthocyanin spots increased in intensity until approximately 60 hours after transformation. No spots were observed on any of the four negative controls (Figure 20 E).

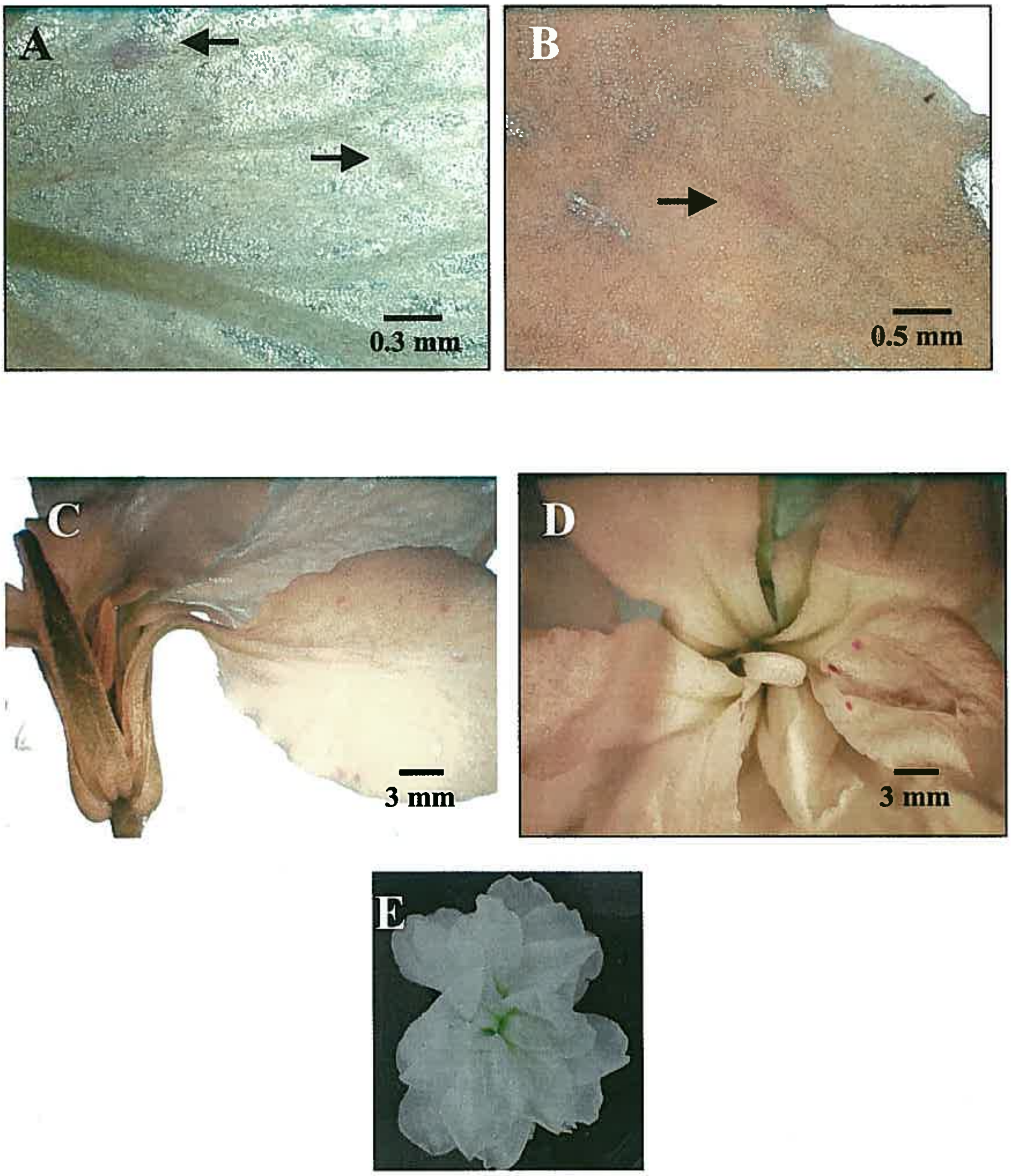


Figure 20: Four *M. incana* ttg1 flowers (A-D) 48 hours after transformation with At*TTG1* displaying purple spots of anthocyanin, and *chs* mutant (E) with negative result.

2.5 Particle bombardment of Matthiola with cotton *TTG1*-like genes

Once the bombardment conditions were shown to be suitable for this assay (see Section 2.4), biolistic transformation was performed on *M. incana ttg1* mutants with each of the four cotton *TTG1*-like genes to test for complementation of the mutant phenotype. In each experiment, two different genes could be bombarded. In the first experiment, TTG(I) and TTG(II) constructs were each used to bombard three sample flowers from the *M. incana ttg1* mutant with the *chs* mutant as a negative control. After approximately 40 hours at 25°C, purple spots of anthocyanin were visible on two of the three *ttg1* flowers bombarded with TTG(I) (Figure 21). None of the petals bombarded with TTG(II), nor either of the *chs* mutant negative controls, showed any sign of purple anthocyanin production.

Figure 21: Two of the M. incana *ttg1* mutant flowers transformed with cotton gene TTG(I) .

In the next series of bombardments (performed several days later), cotton genes TTG①P1 and pTTG(c)E2 were tested for functional homology to *TTG1*, again by bombarding each onto three *ttg1 M. incana* mutant flowers with a *chs* mutant flower as a negative control. After approximately 40 hours incubation at 25°C, purple spots were observed on all three *ttg1* mutant flowers bombarded with TTG①P1, but on none of those transformed with pTTG(c)E2 (Figure 22). No anthocyanin produciton was observed on either of the *chs* mutant flowers.

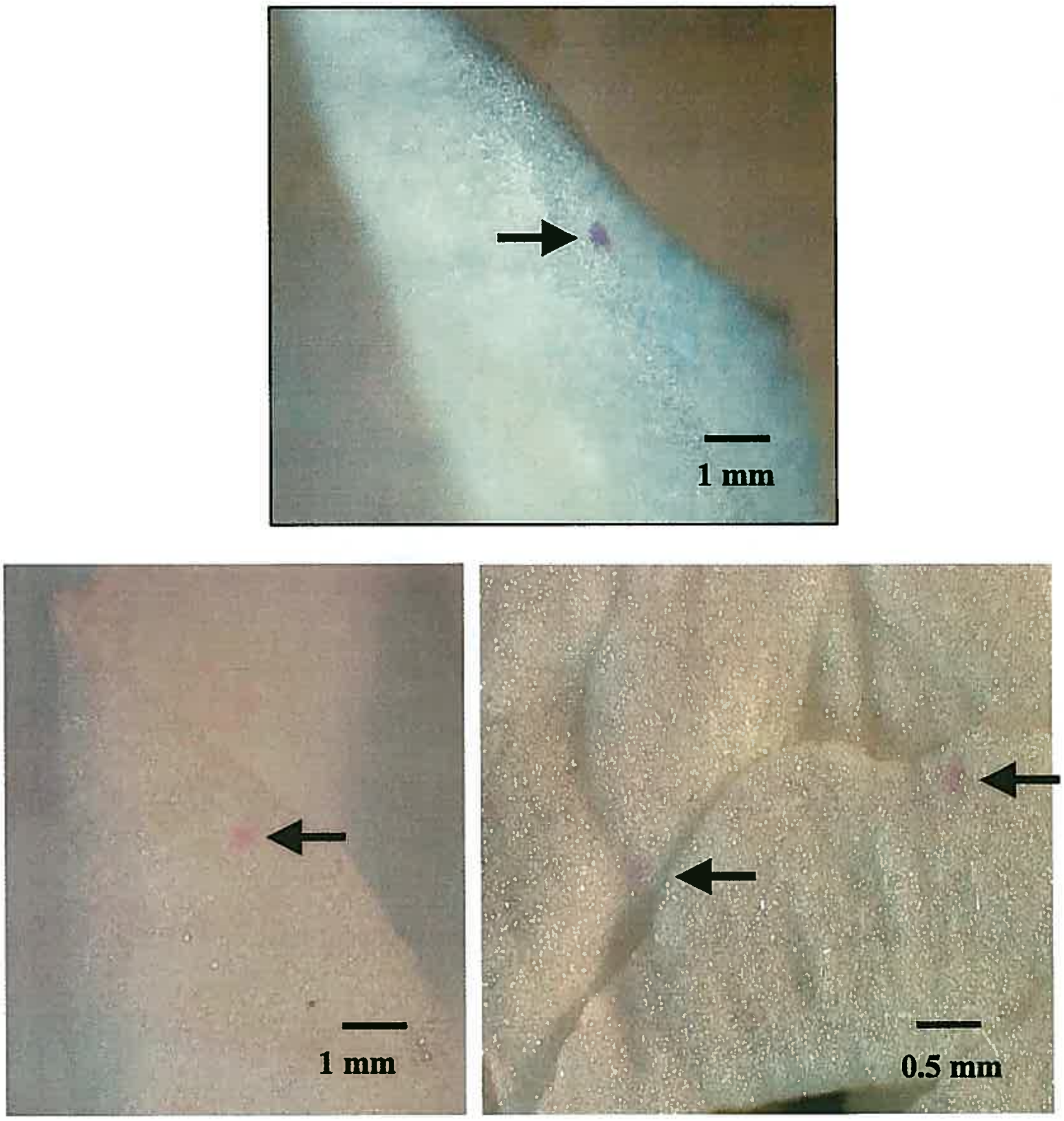
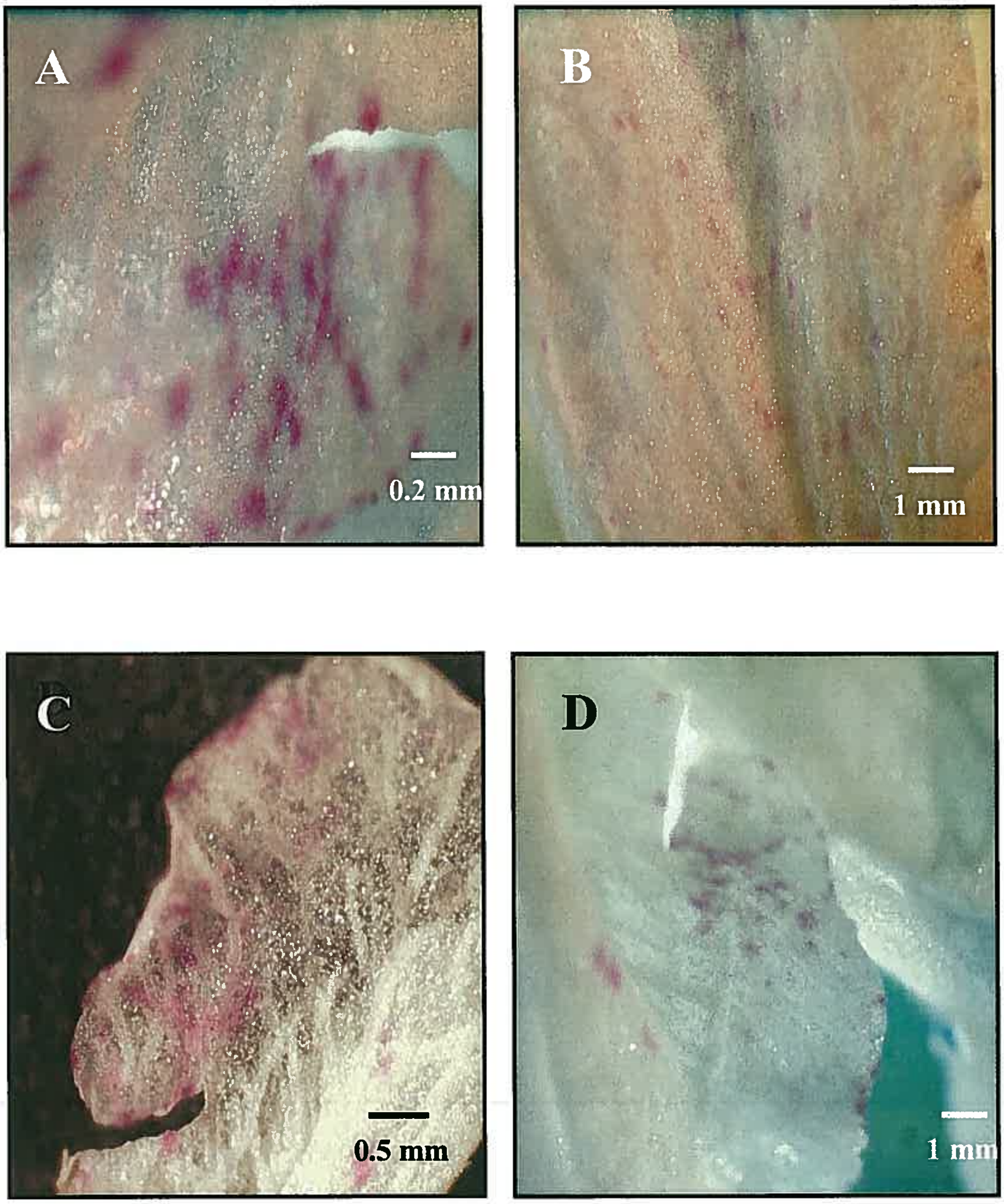


Figure 22: **Purple patches of anthocyanin from TTG①P1 bombardment of three *ttg1 M. incana* mutant flowers.**

It was observed that in both rounds of bombardments, the first gene construct transformed (TTG(I) and TTG①P1) produced purple patches of anthocyanin, whereas the second gene bombarded gave a negative result. To rule out the possibility that the results obtained were a consequence of favourable conditions for gold particles prepared and bombarded first in each pair, the transformations were repeated, but the order of bombardment was reversed.

Optimization of particle bombardment

In the replicate experiments, the flower being transformed was placed approximately 5 cm closer to the particle gun than in previous bombardments in an attempt to improve efficiency of transformation to enable results to be viewed more easily. This second round of experiments verified the results of the original bombardments, with TTG(I) and TTG①P1 producing purple patches of anthocyanin on bombarded petals. Some petals were observed to undergo a 50−fold increase in the number of purple patches, compared with the original bombardments (see Figure 23). The intensity of anthocyanin spots also increased in the repeated bombardments, suggesting that the decrease in distance between the particle injection gun and the flower being transformed considerably increased the efficiency.



<u>Figure 23:</u> Increased intensity and frequency of purple anthocyanin patches. Petals bombarded by TTG(I) (A and B) and TTG①P1 (C and D).

Transformation produced multi-cellular spots

The purple patches observed were not single transformed cells, but groups of cells, allowing visualisation with the naked eye. It has been shown previously that complementation of a *M. incana* bHLH mutant (line 19, A. Walker, pers. comm.) by particle bombardment results in only single cells containing visible anthocyanin. Therefore, the multi-cellular spots observed complementing *ttg1* mutants is unlikely to be due to clusters of cells being hit by gold particles, or diffusion of anthocyanin between cells, as this could have equally occurred when complementing the bHLH mutant. The presence of multi-cellular spots in the *ttg1* mutants may indicate that cell-cell communication has occurred. The TTG1 protein itself may move between cells, or may regulate a biosynthetic intermediate that is transported between cells.

Variation in anthocyanin patches

The size of the anthocyanin spots varied between bombardments, as did the intensity of the pigmented areas. These variations could be caused by a number of factors. The stage of flower development, or varying flower shape, may affect the receptiveness of the target plant tissue to express foreign DNA, or susceptibility to damage may vary between flowers. Additionally, a range of transcription factors is required for a flower to be competent for anthocyanin production, so a lack or decrease in purple patches in a petal may simply reflect a low level of certain transcription factors. There may be variation from shot-to-shot in the amount of DNA-coated gold particle added to each filter, the amount of DNA precipitated onto the particles, and small variations in particle acceleration from the injection gun.

One petal only of flower displayed purple patches

In some cases, a single petal of a *Matthiola* flower showed strong anthocyanin production, while surrounding petals of the flower remained unchanged (Figure 25). This may simply be due to a petal being in a prominent position at the time of bombardment. However, this situation appeared to occur primarily in flowers that were just opening, indicating that cells may only be competent for transient gene expression and anthocyanin production during a specific time period of petal development.

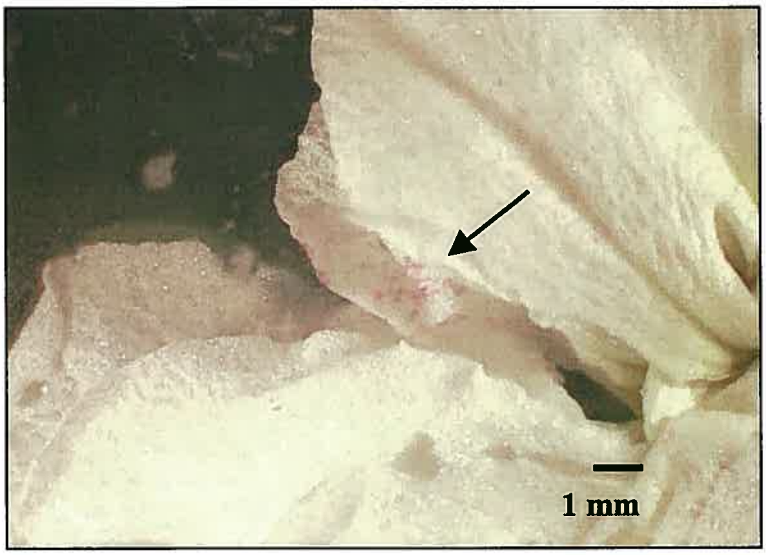


Figure 24: **Example of a single petal transformed, where surrounding petals show no purple anthocyanin.**

Detached petals showed no anthocyanin induction

In addition to complete flowers, detached petals on MS agar were also bombarded with the four cotton genes (Figure 25). Anthocyanins were not induced in any of the detached petals, when bombarded with any of the cotton TTG1-like genes. This may have been due to damage caused to the single petals by the bombardment, or the reduced capacity of the detached petals to survive for over 40 hours on the MS agar. Senescence signals from the flower may be required for anthocyanin production to occur.

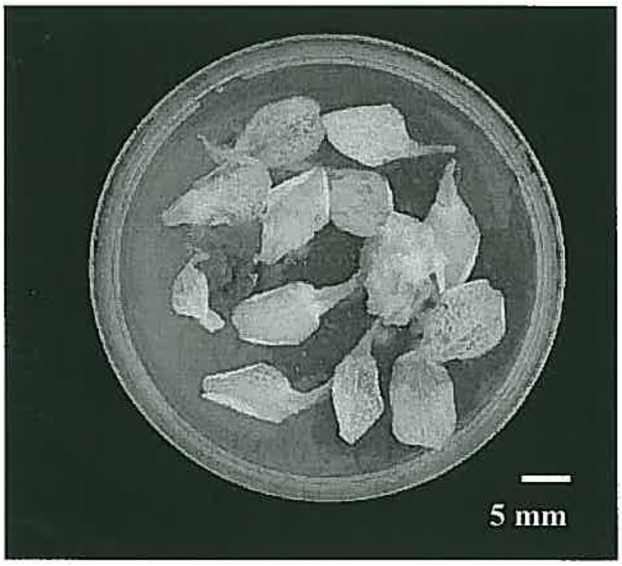


Figure 25: **Detached petals bombarded with cotton TTG1-like genes produced negative results.**

Anthocyanin patches frequently observed along edges of petals

In some cases, purple patches of anthocyanin were seen in increased intensity along the edges of petals (Figure 26).

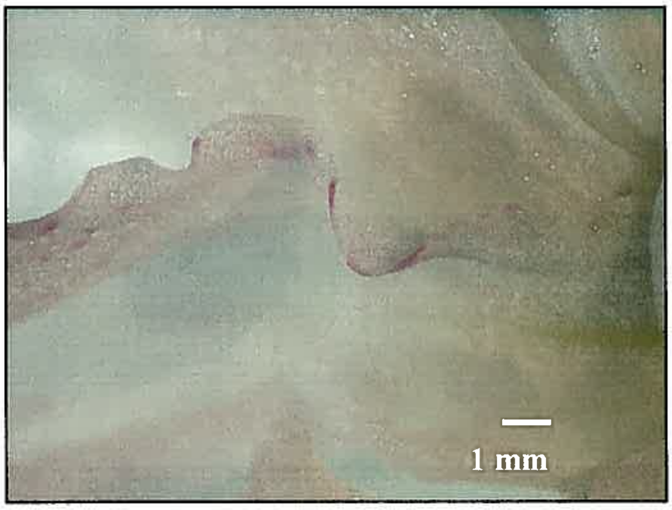


Figure 26: **Purple anthocyanin localised along edge of *M. incana ttg1* petal**

The presence of these anthocyanin clusters may have been caused by the petal edges being prominently exposed at the time of bombardment. Alternatively, these clusters may have been due to the expansion of epidermal cells outwards to the tip of the petal as the flower develops. While spots in the middle of petals would spread out and diffuse as cell expansion occurs, anthocyanin patches near the edge of petals would accumulate at these boundaries, causing an increase in anthocyanin intensity. Another possible explanation is that cells located at the edge of petals may be more susceptible to transformation by particle bombardment.

In summary, two of the four cotton *TTG1*-like genes, namely TTG(I) and TTG①P1, were able to complement a *ttg1* mutation in *Matthiola incana*, thereby restoring anthocyanin production in petals. It is possible that proteins produced by the two cotton genes act in a similar way to *Arabidopsis* TTG1 and may therefore play a pivotal role in trichome, or cotton fibre, development.

3 *Agrobacterium*-mediated transformation of *Arabidopsis*: whole-plant transformation

3.1 Introduction

The bacterial plant pathogen *Agrobacterium tumefaciens* carries a large tumour inducing (Ti) plasmid, a segment of which is designated T-DNA. The T-DNA segment is transmitted by *A. tumefaciens* into individual plant cells, where it penetrates the plant cell nucleus. The T-DNA is then integrated randomly into the genome, where it is stably incorporated and inherited. *Agrobacterium*-mediated transformation takes advantage of this natural plant transformation system to stably transform genes of interest into particular plant species. *A. tumefaciens* is 'dis-armed' by removing oncogenic genes responsible for the formation of tumours, and in their place virtually any gene of interest can be inserted. This results in the stable integration of the desired gene within the plant genome after infection with this altered strain of *Agrobacterium.*

In the *Agrobacetrium*-mediated transformation method, *A. tumefaciens* carrying a cloned gene of interest is applied directly to mature flowering *Arabidopsis* plants. The flowers of the plant to be transformed are submerged into an *Agrobacterium* solution harbouring vectors which contain the gene of interest. This method introduces the bacteria into intercellular spaces within the plant tissue, so infection of the plant can occur.

In investigating functional homology to *AtTTG1* of the four cotton *TTG1*-like genes, it is the function of trichome initiation which is most relevant to cotton fibre development. It is for this reason that the putative cotton *TTG1* homologues were tested for complementation of the *ttg1* mutant phenotype in a second assay in *Arabidopsis*. Additionally, much of the variation observed in the *M. incana* functional assay is likely to be caused by the bombardment process, and a stable transformation is required for stronger evidence of functional homology to *AtTTG1*. Restoration of trichome formation and complementation of the transparent testa phenotype in the seed coat of *Arabidopsis ttg1* mutant will confirm the results from the transient assay.

3.2 Preparation of constructs

Whilst the pART7 constructs were suitable for particle bombardment, a further cloning step was required for *Agrobacterium*-mediated transformation of *Arabidopsis*. For *Agrobacterium*-mediated delivery of a gene into plants, the DNA is cloned into a binary vector, in this case pART27 (Gleave, 1992). This binary vector contains the RK2 minimal replicon (Schmidhauser *et al.*, 1985) for replication in both *E.coli* and *A. tumefaciens*. pART27 also contains a kanamycin resistance gene, so transformed progeny of plant can be selected

To clone each cotton *TTG1*-like gene into pART27, the expression cartridge of pART7 (comprising the CaMV 35S promoter, the gene of interest, and the 3' transcriptional terminator region of the octopine synthase gene) was transferred as a *NotI* fragment into the binary vector. (Figure 27) (Gleave, 1992).

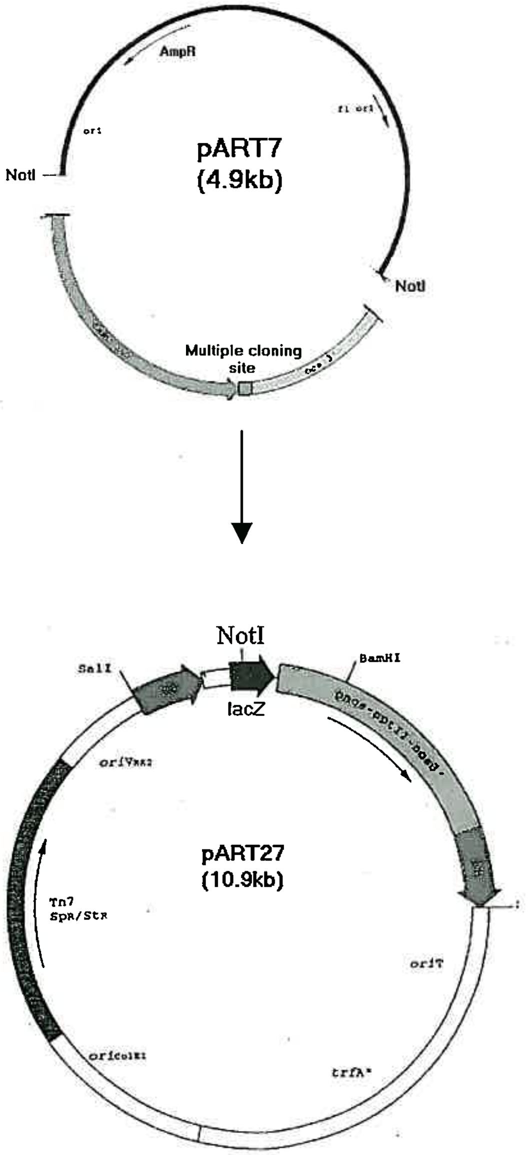


Figure 27: **Ligation of pART7 expression cassette into the binary vector pART27 containing right and left borders (red arrows), kanamycin resistance gene (green) and *lacZ* region with *NotI* site. (Adapted from Gleave, 1992).**

A *NotI* digestion was performed on 2 μg of each of the four cotton *TTG1*-like genes and *AtTTG1* in pART7. The size of the *NotI* fragment to be ligated into pART27 ranged from approximately 3.2 kb for the *Arabidopsis ttg1* gene insert up to 4 kb for the pTTG(c)E2 insert. An agarose gel of the *NotI* digest of the five constructs shows the 2.9 kb backbone of pART7, and the larger expression cartridge (Figure 28).

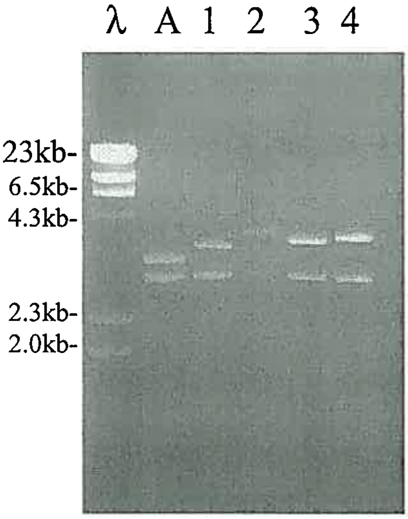


Figure 28: ***NotI* digest of *Arabidopsis TTG1* gene (A) and the four cotton *TTG1*-like genes (1-4) in pART7. The lower band is the 2.9 kb segment remaining of pART7, while the higher band represents the expression cartridge to be ligated into pART27.**

In this case there was no need for a gel extraction, as pART27 has the kanamycin resistance gene, whilst pART7 has ampicillin resistance. As a consequence of this, only cells containing ligations of the pART7 expression cartridge into pART27 (or pART27 re-ligations) would survive on kanamycin plates. pART27 also contains a *lacZ* region for blue/white selection of colonies, so re-ligation of the vector was easily identified.

The fragments shown in Figure 27 were purified by phenol/chloroform extraction, followed by an ethanol precipitation, and resuspension in a final volume of 15 μL. Approximately 500 ng (5 μl) of these purified fragments was added to 50 ng of *NotI*-digested pART27 plasmid (which had also been purified by a phenol/chloroform extraction and CIP-treated) for the ligation reaction at 16°C overnight.

Initially heat-shock at 42°C was performed to transform the ligations into $CaCl_2$-competent cells. However, this method failed to generate any colonies when plated onto kanamycin plates. This lack of colonies was most likely due to the large size of the pART27 plasmid (10.9 kb). With the inserts correctly ligated into pART27, constructs to be transformed had an expected size of approximately 15 kb. Heat-shock transformation was subsequently attempted using cells (DH5αF') prepared by the method described by Inoue *et al.* (1990), which are capable of a much higher efficiency of transformation (2×10^8 transformants per μg of DNA compared to 1×10^4 for $CaCl_2$-competent cells (Hengen, 1996)). 1 ng of uncut pART27 was transformed using these competent cells as a positive control, and generated approximately 1000 colonies, indicating that a much higher transformation efficiency had been achieved. The five constructs were then transformed using this method. White colonies generated from these ligations of the four cotton *TTG1*-like genes and *AtTTG1* into pART27 were selected. Alkaline lysis minipreps were performed on each, which were digested with *NotI* to test for the presence of the expression cassette. Each gene of interest was shown to have been inserted correctly, with the *NotI* digest displaying a 10.9 kb band for the pART27 plasmid, and the expected size of the expression cassette (Figure 29). The orientation of the insert into pART27 did not matter, as the 35S promoter from pART7 would drive expression of the gene of interest regardless of the orientation.

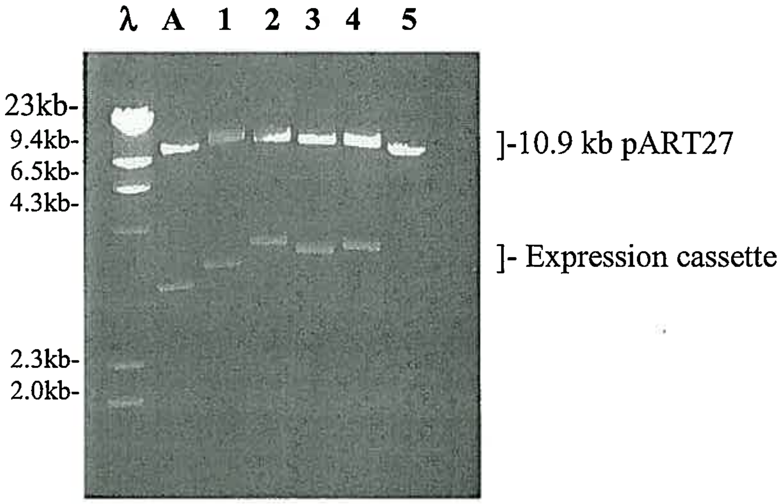
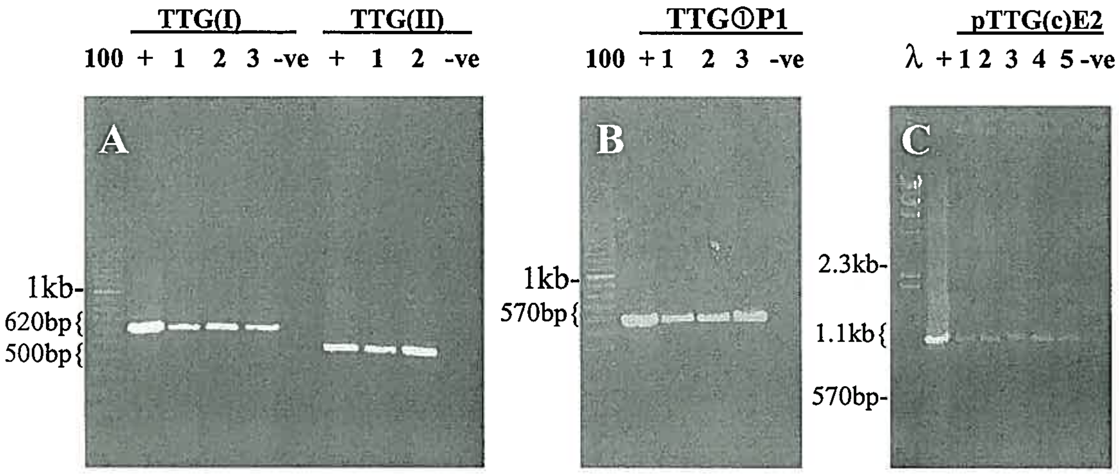


Figure 29: ***NotI*** **digests of the five pART27 constructs (*Arabidopsis TTG1* (lane A) and four cotton genes (lanes 1-4)), and pART27 alone (lane 5).**

3.3 Transformation of Agrobacterium

Each pART27 construct was purified with the QIAGEN midi-plasmid kit In preparation for transformation of *Arabidopsis*, these plasmids were transferred to *Agrobacterium* by electroporation (Walkerpeach and Velten, 1994). 200 ng of plasmid DNA was added to 40 µl of electro-competent cells, which were then pulsed at specified settings and left to recover in 1 ml of Luria broth for 4 hours prior to plating onto YEB agar with kanamycin. An *Agrobacterium* miniprep was performed on colonies formed after incubation at 28°C for approximately 72 hours. PCR was then carried out on each miniprep using internal primers specific for each gene to ensure the plasmid had indeed been transformed into the *Agrobacterium*. The results of the PCR (Figure 30) showed that each construct had been successfully transformed into *Agrobacterium*.



<u>Figure 30:</u> PCR of *Agrobacterium* minipreps using gene-specific primers.

A) TTG(I) samples 1-3 (expected size 620 bp) and TTG(II) samples 1 and 2 (500 bp)
B) TTG①P1 samples 1-3 (570 bp)
C) pTTG(c)E2 samples 1-5 (1.1 kb).

Positive controls (+) for each gene were obtained by amplifying the original clone known to contain the gene of interest with the same primers as the samples tested. No DNA was added to the negative controls (-).

As no internal primers for the *AtTTG1* gene (positive control) were available, six *Agrobacterium* minipreps of *AtTTG1*/pART27 were digested with *NotI* to demonstrate that this construct had been successfully transformed into *Agrobacterium*. Fragment sizes of 3.1 kb (insert) and 10.9 kb (vector) were expected for a positive result (see Figure 31).

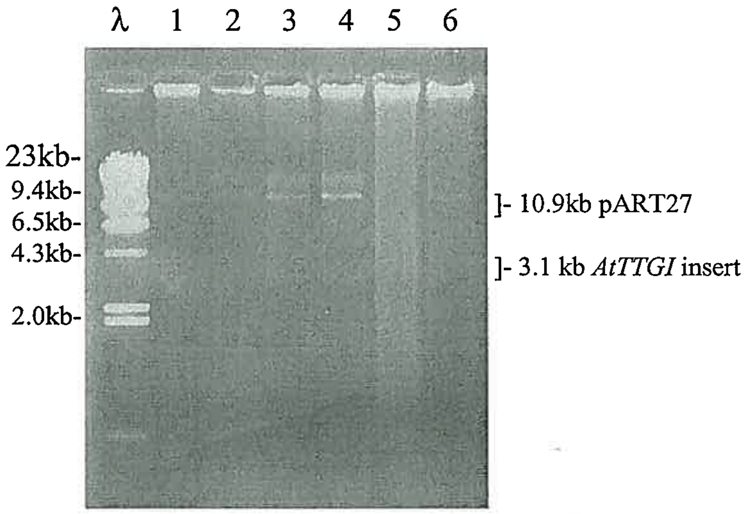


Figure 31: ***NotI* digestions of *Agrobacterium* midipreps containing AtTTG1/pART27 construct (samples 1-6).**

While *Agrobacterium* minipreps rarely provide high yields of DNA (as can be seen from the low intensity of the bands in Figure 29), the agarose gel showed that samples 3 and 4 contained the *AtTTG1*/pART27 construct. A quantity of un-digested plasmid or genomic DNA can be seen above the 10.9 kb pART27 band in these samples. The five constructs in *Agrobacterium* were kept as glycerol stocks until transformation of *Arabidopsis* was performed.

3.4 Growth of *Arabidopsis* plants

Arabidopsis thaliana ttg1 mutant seeds (line N89) were supplied by the Nottingham (UK) Arabidopsis Stock Centre (NASC). The seeds were sterilised in a sodium hypochlorite solution, and then planted in small pots of autoclaved soil. The plants were grown at room temperature in natural daylight until flowering.

As only a small number of seeds were obtained from NASC, the first plants grown were originally intended for the harvesting of more seeds for the *Agrobacterium*-mediated transformations. However, a combination of slow growth and low survival rate of these plants resulted in only five or six plants flowering in time for transformation to be performed within the time-course of the project (Figure 32).

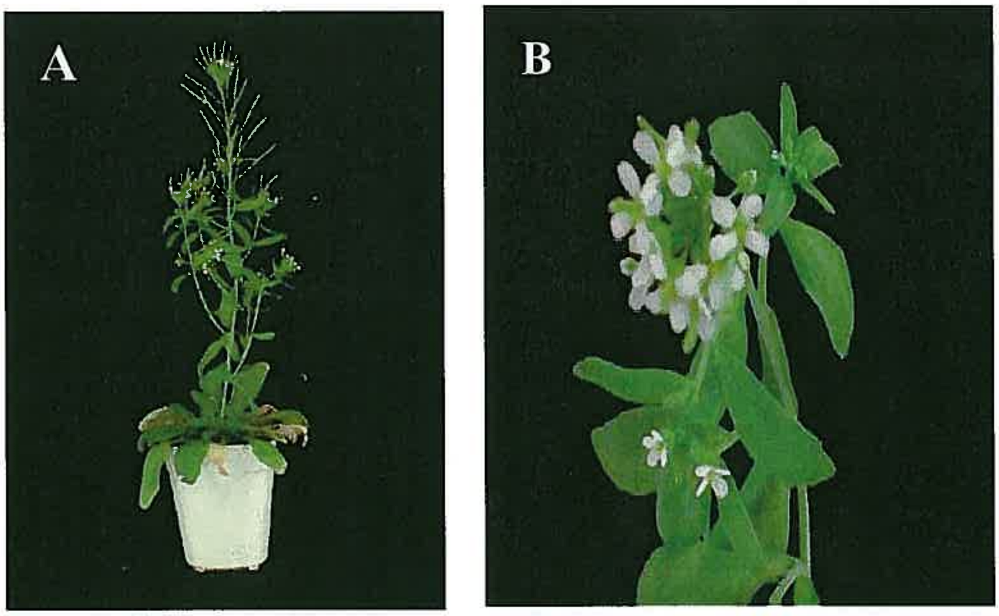


Figure 32: ***Arabidopsis ttg1*** **mutant suitable for transformation, with bolt length of approximately 10 cm (A), and opened flowers (B).**

3.5 Transformation of mutant Arabidopsis with pART27 constucts

As a result of having so few plants available, only one gene construct was used to transform all the plants. The cotton gene with the highest homology to *AtTTGI*, TTG①P1 was selected to transform four *Arabidopsis ttg1* mutant plants. TTG①P1 had also been shown to complement the *Matthiola ttg1* mutant phenotype, and was therefore considered most likely to complement the mutant *Arabidopsis*.

In preparation for the transformation, a single *Agrobacterium* colony containing TTG①P1 in pART27 was used to inoculate 25 ml of liquid YEB with kanamycin. The night before transformation, 10 ml of this culture was used to start a 500 ml YEB liquid culture (plus kanamycin), which was grown overnight to OD_{600}=1.5. The bacteria were harvested by centrifugation of the culture, followed by resuspension in 500 ml of infiltration medium (Bent *et al.*, 1998). The *Agrobacterium* suspension was then transferred to a large beaker, and a surfactant (Silwet L-77) was added to enhance infiltration. The plants were inverted and flowers submerged in the *Agrobacterium* suspension for about 5 seconds to allow introduction of *Agrobacterium* into plant tissue. After transformation, the plants were kept in humid conditions with minimal watering for two days. Approximately one week after the first transformation, the plants were dipped into a fresh Agrobacterium solution to increase the chance of obtaining positive transformants.

Due to time limitations, seeds are not yet mature on these transformed *Arabidopsis* plants, so no results are presently available.

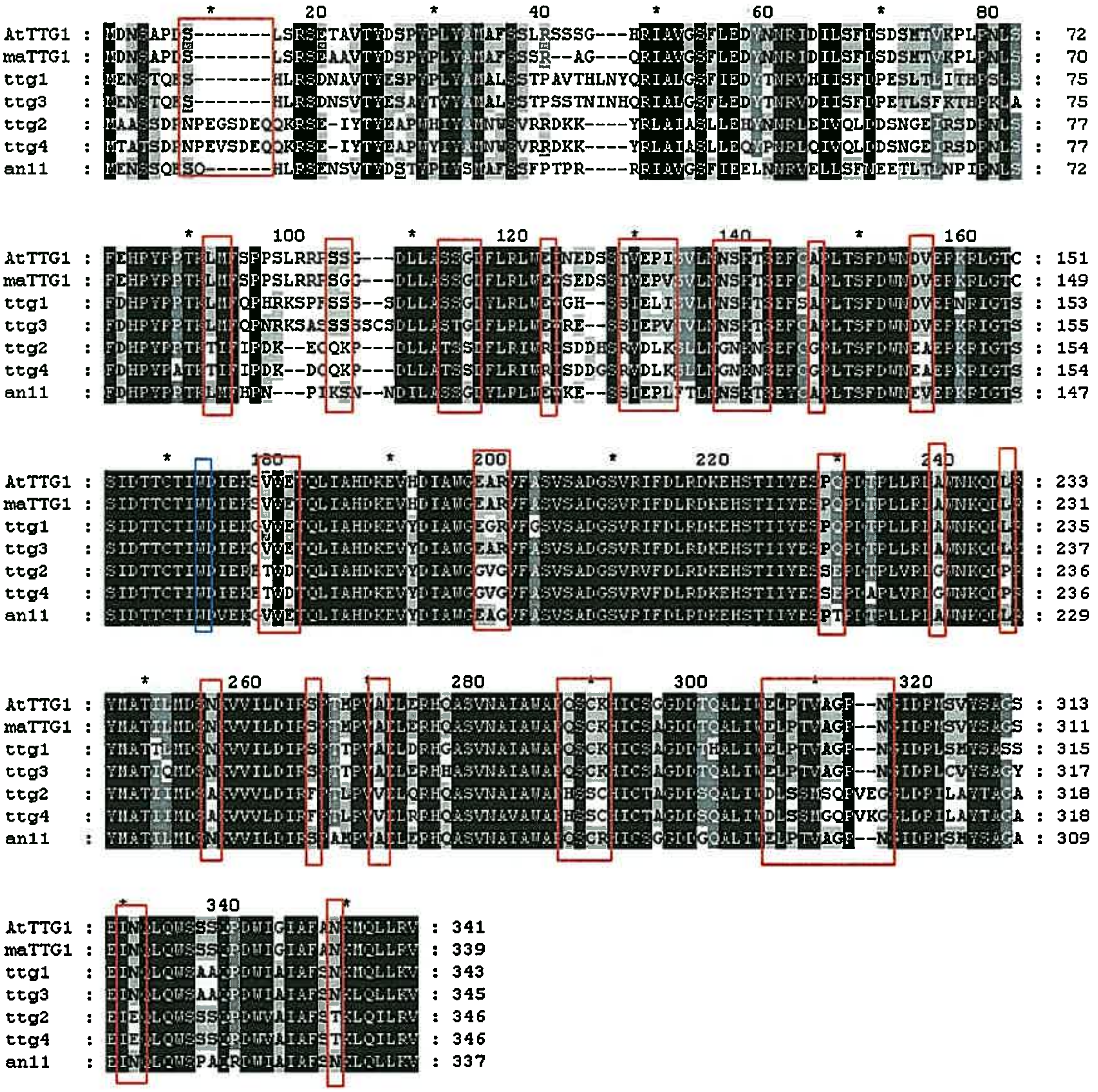
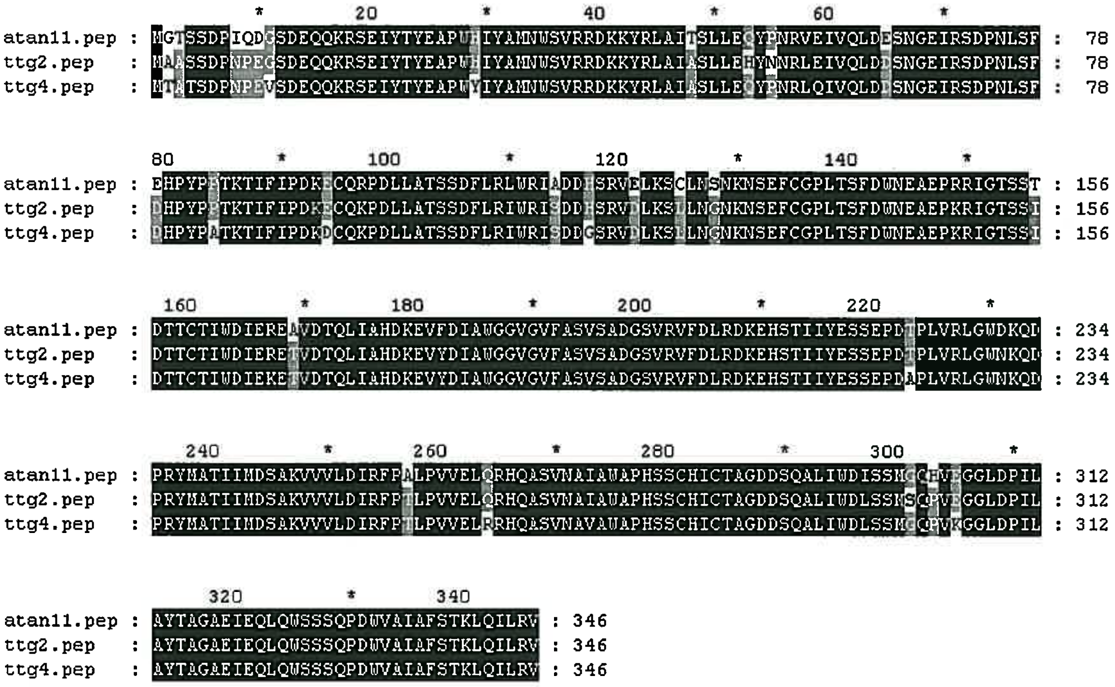


Figure 33: Conceptual translation of four cotton *TTG1*-like genes (ttg1-4) compared with *Arabidopsis* and *Matthiola* TTG1. Variable regions where TTG1 and TTG①P1 are similar to AtTTG1 to the exclusion of TTG(II) and pTTG(c)E2 are shown in red boxes. The arginine of the WD-40 repeat that is mutated in the *Matthiola ttg1* mutant is indicated by a blue box.

The particle bombardment results therefore correlate with the protein alignment, as those genes displaying functional homology (TTG(I) and TTG①P1) possess the highest sequence similarity to AtTTG1.

BlastP searches on the two cotton TTG1-like proteins that did not complement the *Matthiola ttg1* mutant (TTG(II) and pTTG(c)E2) indicated that they possess high homology to the *Arabidopsis* protein ATAN11a (de Vetten *et al.*, 1997) (Figure 34).



```
                          *        20         *        40         *        60         *
atan11.pep : MGTSSDPIQDGSDEQQKRSEIYTYEAPWHIYAMNWSVRRDKKYRLAITSLLEQYPNRVEIVQLDESNGEIRSDPNLSF :  78
ttg2.pep   : MAASSDPNPEGSDEQQKRSEIYTYEAPWHIYAMNWSVRRDKKYRLAIASLLEHYNNRLEIVQLDDSNGEIRSDPNLSF :  78
ttg4.pep   : MTATSDPNPEVSDEQQKRSEIYTYEAPWYIYAMNWSVRRDKKYRLAIASLLEQYPNRLQIVQLDDSNGEIRSDPNLSF :  78

             80         *       100         *       120         *       140         *
atan11.pep : EHPYPTTKTIFIPDKECQRPDLLATSSDFLRLWRIADDHSRVELKSCLNSNKNSEFCGPLTSFDWNEAEPRRIGTSST : 156
ttg2.pep   : DHPYPTTKTIFIPDKECQKPDLLATSSDFLRIWRISDDHSRVDLKSILNGNKNSEFCGPLTSFDWNEAEPKRIGTSSI : 156
ttg4.pep   : DHPYPATKTIFIPDKDCQKPDLLATSSDFLRIWRISDDGSRVDLKSILNGNKNSEFCGPLTSFDWNEAEPKRIGTSSI : 156

               160         *       180         *       200         *       220         *
atan11.pep : DTTCTIWDIEREAVDTQLIAHDKEVFDIAWGGVGVFASVSADGSVRVFDLRDKEHSTIIYESSEPDTPLVRLGWDKQD : 234
ttg2.pep   : DTTCTIWDIERETVDTQLIAHDKEVYDIAWGGVGVFASVSADGSVRVFDLRDKEHSTIIYESSEPDTPLVRLGWNKQD : 234
ttg4.pep   : DTTCTIWDIEKETVDTQLIAHDKEVYDIAWGGVGVFASVSADGSVRVFDLRDKEHSTIIYESSEPDAPLVRLGWNKQD : 234

                 240         *       260         *       280         *       300         *
atan11.pep : PRYMATIIMDSAKVVVLDIRFPALPVVELQRHQASVNAIAWAPHSSCHICTAGDDSQALIWDISSMGCHVEGGLDPIL : 312
ttg2.pep   : PRYMATIIMDSAKVVVLDIRFPTLPVVELQRHQASVNAIAWAPHSSCHICTAGDDSQALIWDLSSNSQPVEGGLDPIL : 312
ttg4.pep   : PRYMATIIMDSAKVVVLDIRFPTLPVVELRRHQASVNAVAWAPHSSCHICTAGDDSQALIWDLSSMGQPVKGGLDPIL : 312

                   320         *       340
atan11.pep : AYTAGAEIEQLQWSSSQPDWVAIAFSTKLQILRV : 346
ttg2.pep   : AYTAGAEIEQLQWSSSQPDWVAIAFSTKLQILRV : 346
ttg4.pep   : AYTAGAEIEQLQWSSSQPDWVAIAFSTKLQILRV : 346
```

Figure 34: Amino acid sequence comparison of cotton proteins TTG(II) and pTTG(c)E2 with *Arabidopsis* protein ATAN-11.

TTG(II) displays 96% amino acid identity to ATAN-11, while pTTG(c)E2 possesses 93% identity.

A maximum parsimony phylogenetic tree (Figure 35) derived from comparisons of these sequences suggests a potential functional relationship, with TTG(II) and pTTG(c)E2 grouping with ATAN-11a, while TTG(I) and TTG①P1 form a monophyletic group with the TTG1 proteins from *M. incana* and *Arabidopsis*.

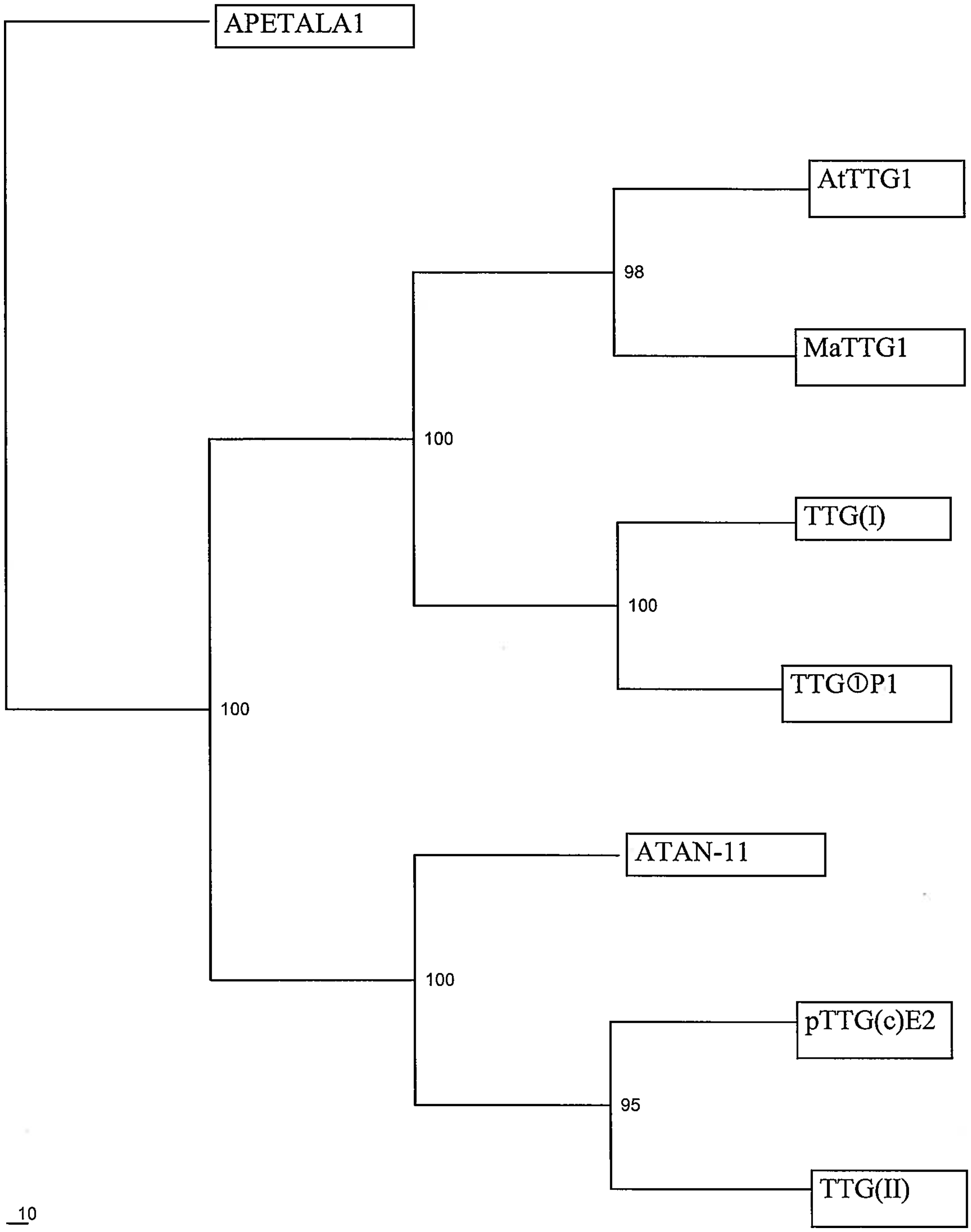


Figure 35: Phylogenetic tree of cotton TTG1-like proteins with *Arabidopsis* and *Matthiola* TTG1, AN-11 of petunia and ATAN-11 of *Arabidopsis*. Accession numbers: APETALA1: Z16421, AtTTG1: AJ133743.1, ATAN-11a: U94748, remaining sequences not submitted.

So far the role of ATAN11a in *Arabidopsis* is largely unknown, although it is also thought to be involved in flower pigmentation. Further investigation into the roles of ATAN11a in *Arabidopsis* may give an insight into possible roles of TTG(II) and pTTG(c)E2 in cotton.

It must be noted that the results obtained in this investigation provide evidence for the replacement of AtTTG1 function by cotton TTG(I) and TTG①P1 in anthocyanin production only. While this may provide an indication of homology in the function of trichome initiation for these cotton genes, it is by no means direct evidence. It remains to be seen whether the complementation of one phenotype of the *ttg1* mutant by these cotton genes will necessarily correlate with a complementation of other phenotypes. The fact that many alleles of *ttg1* have the same pleiotropic phenotype (Walker et al., 1999) indicates the functions are strongly linked, and hence suggests complementation of the anthocyanin pathway may extend to the function of trichome initiation. It is the role of TTG1 in trichome initiation which is most relevant to possible roles in cotton fibre development.

The *R* gene in maize, and *delila* in snapdragon are bHLH proteins implicated in anthocyanin production (Spelt *et al.*, 2000). The R gene in maize has previously been shown to complement both the glabrous and anthocyanin-deficient phenotypes of *Arabidopsis ttg1* mutants (Lloyd *et al.*, 1992). However, the *delila* gene in snapdragon alters pigmentation in transgenic tomato and tobacco, but does not affect trichome deficiency in *ttg1* mutants. Additionally, an11, which controls pigmentation in petunia, is expressed in all tissues, but does not affect hair formation (de Vetten *et al.*, 1997). Thus, the fact that the cotton genes TTG(I) and TTG①P1 can replace TTG1 function in anthocyanin production only hints at a role in trichome initiation. Results from the stable transformation of *Arabidopsis ttg1* mutants would provide evidence for this.

To complete the investigation into complementation of the *ttg1* mutant in *Arabidopsis*, plants transformed with TTG①P1 will be harvested for seeds about four weeks after transformation. Transgenic progeny will then be identified by germination of seeds on selective media (kanamycin). These progeny will then be inspected microscopically for restoration of trichome formation. Seeds will be harvested from plants not transformed to grow more *ttg1* mutant *Arabidopsis* to test remaining cotton TTG1-like genes for complementation.

Functional homology to *TTG1* indicates a possible role for TTG(I) and TTG①P1 in the initiation of cotton fibres. The results also appear to strengthen the link suggested between cotton fibre development and *Arabidopsis* trichomes, providing further support for the use of trichomes as a model for cotton fibre development in general. The link between cotton fibre and trichome initiation will become more definite if a role in cotton fibre initiation is established for these TTG1 homologues in cotton.

To study the influence these genes have on fibre development, expression analysis could be performed on lint and fuzz cotton fibres, to test for any differences in expression of TTG(I) or TTG①P1. A significant increase in gene expression of a cotton *TTG1*-like gene in lint fibres compared to the shorter fuzz fibres would be an excellent indicator for a role in cotton fibre initiation. Additionally, an increase in expression during fibre initiation compared to other stages of fibre development would be indicative of a genes' role in initiating cotton fibres.

Experiments could be performed to identify regions responsible for the complementation of the *ttg1* mutant phenotype in *M. incana*. The variable regions (as identified in Figure 33, red boxes) could be removed from TTG(I) and TTG①P1 to identify which are important for the complementation of the *ttg1* mutant. The region surrounding the site of substitution causing the white-petal phenotype (Figure 33, blue box) appears likely to be significant in TTG1's role in anthocyanin production.

The generation of transgenic cotton may be required to further investigate the role of TTG(I) and TTG①P1 in cotton fibre development. Anti-sense genes have previously been used to reduce levels of specific proteins in cotton (John, 1996). The observation of phenotypic changes in fibres of a cotton plant with reduced TTG(I) or TTG①P1 levels would provide strong evidence of a role for these genes in cotton fibre development, or specifically fibre initiation, depending on the types of changes observed. The tetraploid nature of the cotton genome, and the possibility of gene redundancy, complicates this experiment.

5 Conclusion

By placing four cotton *TTG1*-like genes under the control of a constitutive 35S promoter, each was tested for complementation of the *M. incana* white petal phenotype by particle bombardment. Two of these genes, designated TTG(I) and TTG①P1, produced purple patches of anthocyanin on transformed petals. This result indicates that these two cotton genes are functional homologues of the *TTGI* gene. However, it is only the role played by *TTGI* in anthocyanin production which has been complemented, and not that of its function in trichome initiation. Restoration of trichome formation in *Arabidopsis ttg1* mutants will provide evidence for functional homology in trichome initiation, which can be more strongly linked to the development of cotton fibres.

The results of this project provide putative evidence for a role of two cotton *TTG1*-like genes in fibre initiation, by demonstrating functional homology to *TTG1*. As no regulators of cotton fibre cell growth have yet been isolated, further characterisation of these genes may lead to significant advances in the understanding of cotton fibre initiation.

References

Basra, A. (1999) Cotton Fibres: *Developmental biology, quality improvement and textile processing*, The Haworth Press, USA.

Clough, S. & Bent, A. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *The Plant Journal* **16:** 735-743.

de Vetten, N., Quattrocchio, F., Mol, J. & Koes, R. (1997) The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes & Development* **11:** 1422-1434.

Gleave, A.P., (1992) A Versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Molecular Biology* **20:** 1203-1207.

Graham, T. (1998) Flavonoid and flavonol glycoside metabolism in *Arabidopsis*. *Plant Physiology & Biochemistry* **36:** 135-144.

Harpster, M., Townsend, J., Jones, J., Bedbrook, J. & Dunsmuir, P. (1988) Relative strengths of the 35S califlower mosaic virus, 1′, 2′, and nopaline synthase promoters in transformed tobacco sugarbeet and oilseed rape callus tissue. *Molecular Genomes & Genetics* **212:** 182-190.

Hengen, P. (1996) Preparing ultra-competent *Escherichia coli*. *Trends in Biochemical Sciences* **21:** 75-76.

Inoue, H., Nojima, H., & Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96:** 23-28.

John, M. (1996) Structural characterization of genes corresponding to cotton fiber mRNA, E6: reduced E6 protein in transgenic plants by antisense gene. *Plant Molecular Biology* **30:** 297-306.

Kemp, A., Parker, J. & Grierson, C. (2001) Biolistic transformation of *Arabidopsis* root hairs: a technique to facilitate map-based cloning. *The Plant Journal* **27:** 367-371.

Koorneef, M. (1981) The complex syndrome of *ttg* mutants. *Arabidoposis Information Service* **18:** 45-51.

Lesley, M. (1973) An unstable S chromosome in Diploid *Matthiolo incana*. *Journal of Heredity* **64:** 359-362.

Lloyd, A & Davis, R. (1992) Arabidopsis and nicotiana anthocyanin production Activated by maize regulators R and C1. *Science* **258:** 1773-1775.

Marks, M., Esch, J., Herman, P., Sivakumaran, S. & Oppenheimer, D. (1991) A model for cell-type determination and differentiation in plants. *Society for Experimental Biology* **123:** 77-87.

Neer, E., Schmidt, C., Nambudripad, R. & Smith, T. (1994) The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371:** 297-300.

Oard, J. (1993) Development of an airgun device for particle bombardment. *Plant Cell, Tissue and Organ Culture* **33:** 247-250.

Payne, T., Clement, J., Arnold, D. & Lloyd, A. (1999) Heterologous myb genes distinct from *GL1* enhance trichome production when overexpressed in *Nicotiana tabacum*. *Development* **126:** 671-682.

Sambrook, J., Fritsch, E. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbor, NY, USA.

Schmidhauser, T. & Helinski, D. (1985) Region of broad-host range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. *Journal of Bacteria* **164:** 446-455.

Spelt, C., Quattrocchio, F., Mol, J. & Koes, R. (2000) *anthocyanin1* of Petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. *The Plant Cell* **12:** 1619-1631.

Szymanski, D., Lloyd, A & Marks, M. (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in Arabidopsis. *Trends in Plant Science* **5:** 214-219.

Vain, P., Keen, N., Murillo, J., Rathus, C., Nemes, C. & Finer, J. (1993) Development of the Particle Inflow Gun. *Plant Cell, Tissue and Organ Culture* **33:** 237-246.

Walker, A., Davison, P., Bolognesi-Winfield, A., James, C., Srinivasan, N., Blundell, T., Esch, J., Marks, M., & Gray, J. (1999) The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* **11:** 1337-1349.

Walkerpeach, C. & Velten, J. (1994) *Agrobacterium*-mediated gene transfer to plant cells: cointegrate and binary vector systems. *Plant Molecular Biology Manual* **B1:** 1-19.

Appendix A-D

Appendixes A-D describe sequences obtained of the four TTG1-like cotton genes ligated into pART7. The sequences were obtained in most cases from external pART7 primers either side of the point of insertion (pART7a and pART7b) and also via internal primers within the gene itself, combined to give all or the majority of the gene sequence. For each construct, the expected sequence was compiled (top line in Appendixes A-D), and compared with the sequence data received (bottom line). Following is the region of the construct represented by the colour shading of the sequence:

□	CaMV 35S promoter in pART7
▨	pART7 multiple cloning site
□	TTG-like insert
▩	pBluescript remnant from previous cloning digest
▨	3' Octopine synthase region of pART7

Binding sites of external primers are shown in red.

Insertion restriction sites are shown in blue.

'Blast' searches performed on an unexpected stretch of sequence 3' of TTG(II) (Appendix B) revealed a cloning vector, likely to have been used in previous cloning steps. This explanation is also probable for a similar 12bp region 5' of TTG(I) (Appendix A).

Appendix A: TTG(I) in pART7

```
                        10        20        30        40        50        60
                         .         .         .         .         .         .
35.ttg1.ocs     ATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGAC
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     CAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCAT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCC
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     AAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCGTCTACTCCAAG
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     AATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     ATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     CAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCTTCCTCTATATAAGGAA
part7.1.seq     -------------------------------------AAGACCTTCCTCTATATAAGGAA
                                                     ***********************

35.ttg1.ocs     GTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGTTCGAAATCGAT
part7.1.seq     GTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGTTCGAAATCGAT
                ************************************************************

35.ttg1.ocs     AAGCTTGATATCGAATTCC------------CTAGAGGAAAAAATCCAATGGAGAATTCA
part7.1.seq     AAGCTTGATATCGAATTCCGTTGCTGTCGCCCTAGAGGAAAAAATCCAATGGAGAATTCA
                *******************            *****************************

35.ttg1.ocs     ACTCAGGAATCCCACCTCCGATCCGATAACGCCGTAACCTACGAATCACCCTACCCACTC
part7.1.seq     ACTCAGGAATCCCACCTCCGATCCGATAACGCCGTAACCTACGAATCACCCTACCCACTC
                ************************************************************

35.ttg1.ocs     TACGCCATGGCCTTATCTTCCACGCCCGCCGTCACCCATCTCAACTACCAACGCATCGCT
part7.1.seq     TACGCCATGGCCTTATCTTCCACGCCCGCCGTCACCCATCTCAACTACCAACGCATCGCT
                ************************************************************
```

```
35.ttg1.ocs     CTCGGCAGTTTCATCGAGGATTACACTAACAGAGTCCACATAATCTCTTTCGACCCTGAA
part7.1.seq     CTCGGCAGTTTCATCGAGGATTACACTAACAGAGTCCACATAATCTCTTTCGACCCTGAA
                ************************************************************

35.ttg1.ocs     TCACTTACCCTCATAACCCACCCATCTTTGTCGTTCGACCACCCTTACCCACCCACCAAA
part7.1.seq     TCACTTACCCTCATAACCCACCCATCTTTGTCGTTCGACCACCCTTACCCACCCACCAAA
                ************************************************************

35.ttg1.ocs     CTTATGTTCCAACCCCACCGGAAATCCCCTTTCTCTTCCTCCTCCGACCTCCTCGCTTCA
part7.1.seq     CTTATGTTCCAACCCCACCGGAAATCCCCTTTCTCTTCCTCCTCCGACCTCCTCGCTTCA
                ************************************************************

35.ttg1.ocs     TCCGGTGACTACCTCCGTCTCTGGGAAGTTGGGCACTCTTCTATCGAACTTATCTCCGTT
part7.1.seq     TCCGGTGACTACCTCCGTCTCTGGGAAGTTGGGCACTCTTCTATCGAACTTATCTCCGTT
                ************************************************************

35.ttg1.ocs     CTCGACAACAGCAAAACCAGCGAGTTTTCTGCTCCCTTAACTTCCTTCGATTGGAACGAC
part7.1.seq     CTCGACAACAGCAAAACCAGCGAGTTTTCTGCTCCCTTAACTTCCTTCGATTGGAACGAC
                ************************************************************

35.ttg1.ocs     GTTGAACCCAATAGAATCGGAACATCCAGCATCGACACCACCTGCACCATTTGGGACATT
part7.1.seq     GTTGAACCCAATAGAATCGGAACATCCAGCATCGACACCACCTGCACCATTTGGGACATT
                ************************************************************

35.ttg1.ocs     GAAAAAGGCGTTGTGGAAACCCAATTGATTGCTCACGACAAGGAGGTTTACGACATCGCT
part7.1.seq     GAAAAAGGCGTTGTGGAAACCCAATTGATTGCTCACGACAAGGAGGTTTACGACATCGCT
                ************************************************************

35.ttg1.ocs     TGGGGTGAAGGTAGAGTTTTCGGTTCCGTTTCCGCTGATGGGTCCGTAAGGATTTTCGAT
part7.1.seq     TGGGGTGAAGGTAGAGTTTTCGGTTCCGTTTCCGCTGATGGGTCCGTAAGGATTTTCGAT
                ************************************************************

35.ttg1.ocs     TTAAGGGACAAAGAACACTCAACTATCATCTACGAAAGCCCTCAACCGGACACCCCTTTG
part7.1.seq     TTAAGGGACAAAGAACACTCAACTATCATCTACGAAAGCCCTCAACCGGACACCCCTTTG
                ************************************************************

35.ttg1.ocs     CTAAGATTAGCTTGGAACAAACAAGATTTGAGGTATATGGCTACGACACTAATGGATAGC
part7.1.seq     CTAAGATTAGCTTGGAACAAACAAGATTTGAGGTATATGGCTACGACACTAATGGATAGC
                ************************************************************

35.ttg1.ocs     AATAAAGTTGTGATTTTGGACATAAGGTCGCCTACTACGCCGGTGGCTGAGTTGGATAGA
part7.1.seq     AATAAAGTTGTGATTTTGGACATAAGGTCGCCTACTACGCCGGTGGCTGAGTTGGATAGA
                ************************************************************

35.ttg1.ocs     CATGGAGCCAGTGTCAATGCGATTGCTTGGGCTCCCCAGAGCTGTAAGCACATTTGCTCT
part7.1.seq     CATGGAGCCAGTGTCAATGCGATTGCTTGGGCTCCCCAGAGCTGTAAGCACATTTGCTCT
                ************************************************************

35.ttg1.ocs     GCAGGGGATGATACGCATGCCCTTATTTGGGAGTTACCAACTGTAGCTGGGCCTAACGGA
part7.1.seq     GCAGGGGATGATACGCATGCCCTTATTTGGGAGTTACCAACTGTAGCTGGGCCTAACGGA
                ************************************************************

35.ttg1.ocs     ATCGATCCTTTGTCTATGTACTCGGCTAGTTCCGAAATTAATCAGTTGCAGTGGTCTGCT
part7.1.seq     ATCGATCCTTTGTCTATGTACTCGGCTAGTTCCGAAATTAATCAGTTGCAGTGGTCTGCT
                ************************************************************

35.ttg1.ocs     GCTCAGCCTGATTGGATTGCCATTGCCTTCTCCAACAAAATGCAGCTTCTCAAAGTTTGA
part7.1.seq     GCTCAGCCTGATTGGATTGCCATTGCCTTCTCCAACAAAATGCAGCTTCTCAAAGTTTGA
                ************************************************************

35.ttg1.ocs     GGTTCCCTGCTTGAAATGTATTCAACTCTCATTTTTGTTGCATATACTGCTCACGCTTGT
part7.1.seq     GGTTCCCTGCTTGAAATGTATTCAACTCTCATTTTTGTTGCATATACTGCTCACGCTTGT
                ************************************************************


35.ttg1.ocs     AATTAGGTGGGAGTTAAAAACTGTAGAATCAGTTTTATTGGCTGATAATTTTGTTAGTAC
part7.1.seq     AATTAGGTGGGAGTTAAAAACTGTAGAATCAGTTTTATTGGCTGATAATTTTGTTAGTAC
                ************************************************************
```

```
35.ttg1.ocs     TTTGTAAGACCAGCTTTAATCTACAATTATTTGTTAGGTCCTCACAACAAATATGAGACT
part7.1.seq     TTTGTAAGACCAGCTTTAATCTACAATTATTTGTTAGGTCCTCACAACAAATATGAGACT
                ************************************************************


35.ttg1.ocs     CGAATGCATATTTGGATTTTGACCTGCAGCCCGGGGATCCTCTAGAGTCCTGCTTTAAT
part7.1.seq     CGAATGCATATTTGGATTTTGACCTGCAGCCCGGGGATCCTCTAGAGTCCTGCTTTAAT
                ************************************************************

35.ttg1.ocs     GAGATATGCGAGACGCCTATGATCGCATGATATTTGCTTTCAATTCTGTTGTGCACGTTG
part7.1.seq     GAGATATG----------------------------------------------------
                ********

35.ttg1.ocs     TAAAAAACCTGAGCATGTGTAGCTCAGATCCTTACCGCCGGTTTCGGTTCATTCTAATGA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     ATATATCACCCGTTACTATCGTATTTTTATGAATAATATTCTCCGTTCAATTTACTGATT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GTACCCTACTACTTATATGTACAATATTAAAATGAAAACAATATATTGTGCTGAATAGGT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TTATAGCGACATCTATGATAGAGCGCCACAATAACAAACAATTGCGTTTTATTATTACAA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     ATCCAATTTTAAAAAAAGCGGCAGAACCGGTCAAACCTAAAAGACTGATTACATAAATCT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TATTCAAATTTCAAAAGGCCCCAGGGGCTAGTATCTACGACACACCGAGCGGCGAACTAA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     TAACGTTCACTGAAGGGAACTCCGGTTCCCCGCCGGCGCGCATGGGTGAGATTCCTTGAA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GTTGAGTATTGGCCGTCCGCTCTACCGAAAGTTACGGCACCATTCAACCCGGTCCAGCA
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     CGGCGGCCGGGTAACCGACTTGCTGCCCCGAGAATTATGCAGCATTTTTTTGGTGTATGT
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     GGGCCCCAAATGAAGTGCAGGTCAAACCTTGACAGTGACGACAAATCGTTGGGCGGGTCC
part7.1.seq     ------------------------------------------------------------

35.ttg1.ocs     AGGGCGAATTTTGCGACAACATGTCGAGGCTCAGCAGGACCCGGT
part7.1.seq     ---------------------------------------------
```

Appendix B: TTG(II) in pART7

```
35.ttg2.ocs     ATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGAC
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     CAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCAT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     TGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAA
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     TGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCC
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     AAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     TCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCGTCTACTCCAAG
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     AATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTA
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     ATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACA
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     GTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     CAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     GAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     GACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCTTCCTCTATATAAGGAA
pART7.2.seq     --------------------------------------GNAGTTTTTNNTNTTNTAAGGA

35.ttg2.ocs     GTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGTTCGAAATCGAT
pART7.2.seq     AGTTTTTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGTTCGAAATCGAT

35.ttg2.ocs     AAGCTTGGATCCAAGAAACCAAAGGGAGTGGCAGTGGAAGATGGCCGCTAGCAGCGATCC
pART7.2.seq     AAGCTTGGATCCAAGAAACCAAAGGGAGTGGCAGTGGAAGATGGCCGNTAGCAGCGATCC


35.ttg2.ocs     TAACCCGGAGGGTTCCGATGAGCAGCAGAAACGATCCGAGATATACACTTACGAGGCCCC
pART7.2.seq     TAACCCGGAGGGTTCCGATGAGCAGNAGAAACGATCCGAGATATACACTTACGAGGCCCC

35.ttg2.ocs     TTGGCATATCTACGCCATGAACTGGAGTGTCCGCCGCGACAAGAAATACCGGTCTCGCTC
pART7.2.seq     TTGGCATATCTACGCCATGAACTGGAGTGTTCGTNGCGACAAGAAATACCGGTCTCGCTC
```

```
35.ttg2.ocs      ATCGCCAGCCTTGCTCGAGCATTACAACAACCGCCTTGAGATTGTCCAGCTCGATGACTC
pART7.2.seq      ATCGCCAGCCTTGCTCGAGCATTACANCANCCGCCTTGAGATTGTCCAGCTCGATGACTC

35.ttg2.ocs      CAATGGCGAGATCC-GATC-GGACCC-AAAT-CTCTCCTTCG-ATCATCC-TTATCCCCC
pART7.2.seq      CAATGNCNANATTCCNATCCGGACCCCAAATCNTCTCCNTTGNATCATCCCTTNTCCCCC

35.ttg2.ocs      CGGACCCAAGACACATTCTTCATCCCCGACCAAAGGAGCTGCCCAGAAACCCGACCTTCT
pART7.2.seq      CGGACCCAAGACACATTCTTTATCCCCCCCCAAAGGAGCTGCCCAGAAACTCGACCTTCT

35.ttg2.ocs      CGCCACGTCCTCCGACTTCCTCCGCATTTGGCGCATCTCCGATGACCACTCCCGCGTCGA
pART7.2.seq      CGCCACGTCCTCCGACTTCCTCCGCATTTGGGGCATCTCCGATGACCACTCCCGCGTCGA

35.ttg2.ocs      CCTCAAGTCTCTCCTTAATGGCAATAAGAACAGTGAATTCTGCGGTCCTCTTACCTCCTT
pART7.2.seq      CCTCAAGTCTCTCCTTAATGGCAATAAGAACAGTGAATTCTGCGGTCCTCTTACCTCCTT

35.ttg2.ocs      CGACTGGAATGAGGCGGAGCCCAAGCGAATCGGCACCTCCTCCATTGATACGACTTGTAC
pART7.2.seq      CGACTGGAATGAGGCGGAGCCCAAGCGAATCGGCACCTCCTCCATTGATACGTCTTGTAC

35.ttg2.ocs      TATATGGGATATCGAGAGGGAGACGGTGGATACCCAGCTTATCGCCCACGATAAGGAGGT
pART7.2.seq      TATATGGGATATCGAGAGGGAGACGGTGGATACCCAGCTTATCGCCCACGATAAGGAGGT

35.ttg2.ocs      TTATGATATTGCCTGGGGCGGCGTCGGTGTTTTTGCTTCCGTCTCTGCTGATGGGTCCGT
pART7.2.seq      TTATGATATTGCCTGGGGCGGCGTCGGTGTTTTTGCTTCCGTCTCTGCTGATGGGTCCGT

35.ttg2.ocs      TAGGGTTTTCGACCTGCGCGACAAGCTCGAGCACTCCACTATCATTTATGAAAGTTCGGA
pART7.2.seq      TAGGGTTTTCGACCTGCGCGACAAGCTCGAGCACTCCACTATCATTTATGAAAGTTCGGA

35.ttg2.ocs      GCCCGATACTCCGCTTGTACGGTTGGGGTGGAACAAGCAGGACCCGAGATATATGGCTAC
pART7.2.seq      GCCCGATACTCCGCTTGTACGGTTGGGGTGGAACAAGCAGGACCCGAGATATATGGCTAC

35.ttg2.ocs      TATAATTATGGACAGTGCTAAGGTTGTTGTTTTGGATATCCGCTTCCCGACACTGCACGG
pART7.2.seq      TATAATTATGGACAGTGCTAAGGTTGTTGTTTTGGATATCCGCTTCCCGACACTGAACGG

35.ttg2.ocs      TAGTTGAGCTGCAGAGGCACCAGGCTAGCGTCAATGCCATCGCATGGGCACCCCACAGCT
pART7.2.seq      TAGTTGAGCTGCAGAGGCACCAGGCTAGCCTAGGTGCCATCGCATGGGCACCCCACAGCT

35.ttg2.ocs      CTTGCCACATTTGCACCGCCGGGGATGATTCCCAGGCCTTGATTTGGGACTTGTCCTCCA
pART7.2.seq      CTTGCCACATTTGCACCGCCGGGGATGATTCCCAGGCCTTGATTTGGGACTTGTCCTCCA

35.ttg2.ocs      TGAGTCAGCCTGTGGAGGGTGGGCTTGACCCCATCCTTGCCTACACCGCTGGGGCTGAAA
pART7.2.seq      TGAGTCAGCCTGTGGAGGGTGGGCTTGACCCCATCCTTGCCTACACCGCTGGGGCTGAAA

35.ttg2.ocs      TCGAGCAGTTACAGTGGTCATCTTCCCAGCCTGATTGGGTTGCCATCGCCTTCTCCACCA
pART7.2.seq      TCGAGCAGTTACAGTGGTCATCTTCCCAGCCTGATTGGGTTGCCATCGCCTTCTCCACCA

35.ttg2.ocs      AGCTTCAGATTCTCAGGGTATGAGTGAATGGGATTGGTATATTGGGTCGTATTCAATGCT
pART7.2.seq      AGCTTCAGATTCTCAGGGTATGAGTGAATGGGATTGGTATATTGGGTCGTATTCAATGCT

35.ttg2.ocs      AAATTGCTAATTCATAGTCTAGTAATTGTACTCAGATCTAGAGGAATAACAAGATTGTTG
pART7.2.seq      AAATTGCTAATTCATAGTCTAGTAATTGTACTCAGATCTAGAGGAATAACAAGATTGTTG

35.ttg2.ocs      CTGTCGTTGTGTCTGAGCTGACTAAAAGTCCCAAAACTGTTTTGGTTTTCCTGTTGTTTC
pART7.2.seq      CTGTCGTTGTGTCTGAGCTGACTAAAAGTCCCAAAACTGTTTTGGTTTTCCTGTTGTTTC
```

```
35.ttg2.ocs     GTTTGTCTTCAACATGTTTGGATAGTTGATTGAACACCAATCTCTTTATTCTGTTCTTAT
pART7.2.seq     GTTTGTCTTCAACATGTTTGGATAGTTGATTGAACACCAATCTCTTTATTCTGTTCTTAT

35.ttg2.ocs     TTTAGTGTTTCCCATTCCTTTACATCCAAACCTTTCCCTGTTTTATTGTAGTTGTGAGAG
pART7.2.seq     TTTAGTGTTTCCCATTCCTTTACATCCAAACCTTTCCCTGTTTTATTGTAGTTGTGAGAG

35.ttg2.ocs     CAAAAGAGCTGCAGACTCATTTTTTTAGAACGCATGGGAAGCAGTGATTTATGGACCCTG
pART7.2.seq     CAAAAGAGCTGCAGACTCATTTTTTTAGAACGCATGGGAAGCAGTGATTTATGGACCCTG

35.ttg2.ocs     AACTGAAACACAAGATCAAGACGACCACTCTTTGCTGTTTGGTTTATGCAACATAATCTT
pART7.2.seq     AACTGAAACACAAGATCAAGACGACCACTCTTTGCTGTTTGGTTTATGCAACATAATCTT

35.ttg2.ocs     TGTCGTTCTGTTGGACAAGATGAGTGTGTTTTCAGTTTATTTGGTTGTCTGATAAGTTTG
pART7.2.seq     TGTCGTTCTGTTGGACAAGATGAGTGTGTTTTCAGTTTATTTGGTTGTCTGATAAGTTTG

35.ttg2.ocs     ATTAGTTACTATAAATGTGATAAATGATGGGCAGGAAGATATGGGTGGCTTACCCTGGGC
pART7.2.seq     ATTAGTTACTATAAATGTGATAAATGATGGGCAGGAAGATATGGGTGGCTTACCCTGGGC

35.ttg2.ocs     TGCTTTCAGATGCTTGTGTTGTGTTAATGTTAAAACTAGCTTTAGCCATAAAAAAAAAAA
pART7.2.seq     TGCTTTCAGATGCTTGTGTTGTGTTAATGTTAAAACTAGCTTTAGCCATAAAAAAAAAAA

35.ttg2.ocs     AAAAAAA-----------------------------------------AATTCCTGCAGC
pART7.2.seq     AAAAAAGCCGGCCGCTCGAGTCTCTCTCTTCTCTCTCGACAGCAACGGAATTCCTGCAGC

35.ttg2.ocs     CCGGGGGATCCTCTAGAGTCCTGCTTTAATGAGAATATGCGAGACGCCTATGATCGCATG
pART7.2.seq     CCGGGGGATCCACTAGAGTCCTGCTTTAATGAGAATATG---------------------

35.ttg2.ocs     ATATTTGCTTTCAATTCTGTTGTGCACGTTGTAAAAAACCTGAGCATGTGTAGCTCAGAT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     CCTTACCGCCGGTTTCGGTTCATTCTAATGAATATATCACCCGTTACTATCGTATTTTTA
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     TGAATAATATTCTCCGTTCAATTTACTGATTGTACCCTACTACTTATATGTACAATATTA
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     AAATGAAAACAATATATTGTGCTGAATAGGTTTATAGCGACATCTATGATAGAGCGCCAC
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     AATAACAAACAATTGCGTTTTATTATTACAAATCCAATTTTAAAAAAAGCGGCAGAACCG
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     GTCAAACCTAAAAGACTGATTACATAAATCTTATTCAAATTTCAAAAGGCCCCAGGGGCT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     AGTATCTACGACACACCGAGCGGCGAACTAATAACGTTCACTGAAGGGAACTCCGGTTCC
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     CCGCCGGCGCGCATGGGTGAGATTCCTTGAAGTTGAGTATTGGCCGTCCGCTCTACCGAA
pART7.2.seq     ------------------------------------------------------------
```

```
35.ttg2.ocs     AGTTACGGGCACCATTCAACCCGGTCCAGCACGGCGGCCGGGTAACCGACTTGCTGCCCC
pART7.2.seq     ------------------------------------------------------------



35.ttg2.ocs     GAGAATTATGCAGCATTTTTTTGGTGTATGTGGGCCCCAAATGAAGTGCAGGTCAAACCT
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     TGACAGTGACGACAAATCGTTGGGCGGGTCCAGGGCGAATTTTGCGACAACATGTCGAGG
pART7.2.seq     ------------------------------------------------------------

35.ttg2.ocs     CTCAGCAGGACCCGGTA
pART7.2.seq     -----------------
```

Appendix C: TTG①P1 in pART7

```
                        10        20        30        40        50        60
                         .         .         .         .         .         .
35.ttg3.ocs     ATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGAC
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     CAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCAT
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCC
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     AAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCT
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCGTCTACTCCAAG
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     AATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     ATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     GTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTT
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     CAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     GAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACT
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     GACGTAAGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGA
pART7.3.seq     -------------------------------------AAGACCCTTCCTCTATATAAGGA
                                                     ***********************

35.ttg3.ocs     AGTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGCTATAAACCCT
pART7.3.seq     AGTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCCCGGGCTATAAACCCT
                ************************************************************

35.ttg3.ocs     AAAAGCTCTTCCCAGATTTGGTTGAGAAGAAATGGAGAATTCAACTCAAGAATCCCACCT
pART7.3.seq     AAAAGCTCTTCCCAGATTTGGTTGAGAAGAAATGGAGAATTCAACTCAAGAATCCCACCT
                ************************************************************

35.ttg3.ocs     GAGATCCGATAATTCGGTAACCTACGAATCAGCTTACACAGTTTATGCCATGGCCTTATC
pART7.3.seq     GAGATCCGATAATTCGGTAACCTACGAATCAGCTTACACAGTTTATGCCATGGCCTTATC
                ************************************************************

35.ttg3.ocs     TTCCACGCCTTCCTCCACCAATATCAACCATCAACGCATCGCTCTCGGCAGCTTCCTCGA
pART7.3.seq     TTCCACGCCTTCCTCCACCAATATCAACCATCAACGCATCGCTCTCGGCAGCTTCCTCGA
                ************************************************************
```

```
35.ttg3.ocs      AGATTACACTAACAGAGTCGACATAATCTCATTTGATCCAGAAACCCTCTCCTTCAAGAC
pART7.3.seq      AGATTACACTAACAGAGTCGACATAATCTCATTTGATCCAGAAACCCTCTCCTTCAAGAC
                 ************************************************************

35.ttg3.ocs      CCACCCAAAGCTTGCCTTCGACCACCCTTATCCACCCACCAAGCTCATGTTCCAACCCAA
pART7.3.seq      CCACCCAAAGCTTGCCTTCGACCACCCTTATCCACCCACCAAGCTCATGTTCCAACCCAA
                 ************************************************************

35.ttg3.ocs      TCGAAAATCCGCTTCATCCTCTTCTTCTTGTTCCGATCTTCTCGCTTCAACCGGCGATTT
pART7.3.seq      TCGAAAATCCGCTTCATCCTCTTCTTCTTGTTCCGATCTTCTCGCTTCAACCGGCGATTT
                 ************************************************************

35.ttg3.ocs      CCTGCGTCTCTGGGAAGTTCGAGAATCTTCCATTGAACCTGTCACTGTTCTAAACAATAG
pART7.3.seq      CCTGCGTCTCTGGGAAGTTCGAGAATCTTCCATTGAACCTGTCACTGTTCTAAACAATAG
                 ************************************************************

35.ttg3.ocs      CAAAACCAGCGAGTTTTGTGCCCCGTTAACTTCCTTCGATTGGAACGATGTTGAACCCAA
pART7.3.seq      CAAAACCAGCGAGTTTTGTGCCCCGTTAACTTCCTTCGATTGGAACGATGTTGAACCCAA
                 ************************************************************

35.ttg3.ocs      GAGAATTGGGACTTCCAGCATCGACACAACTTGCACCATTTGGGACATCGAGAAATGCGT
pART7.3.seq      GAGAATTGGGACTTCCAGCATCGACACAACTTGCACCATTTGGGACATCGAGAAATGCGT
                 ************************************************************

35.ttg3.ocs      CGTTGAAACCCAATTGATCGCTCATGATAAAGAGGTTTACGACATTGCTTGGGGTGAAGC
pART7.3.seq      CGTTGAAACCCAATTGATCGCTCATGATAAAGAGGTTTACGACATTGCTTGGGGTGAAGC
                 ************************************************************

35.ttg3.ocs      TAGAGTTTTTGCTTCCGTTTCCGCTGATGGGTCCGTTAGGATTTTCGATTTGAGAGACAA
pART7.3.seq      TAGAGTTTTTGCTTCCGTTTCCGCTGATGGGTCCGTTAGGATTTTCGATTTGAGAGACAA
                 ************************************************************

35.ttg3.ocs      AGAACATTCCACCATCATTTATGAAAGTCCCCAACCCGACACCCCCTTTATTAAGATTGG
pART7.3.seq      AGAACATTCCACCATCATTTATGAAAGTCCCCAACCCGACACCCCCTTTATTAAGATTGG
                 ************************************************************

35.ttg3.ocs      CTTGGAACAAGCAAGATTTGAAGTATATGGCTACCATTCAAATGGATAGCAATAAAGTTG
pART7.3.seq      CTTGGAACAAGCAAGATTTGAAGTATATGGCTACCATTCAAATGGATAGCAATAAAGTTG
                 ************************************************************

35.ttg3.ocs      TGATTTTAGGATATAAGCGTCACCCCACAACCCCCGTTGCTGAGTTGGAGCGGCATCACG
pART7.3.seq      TGATTTTAGGATATAAGCGTCACCCCACAACCCCCGTTGCTGAGTTGGAGCGGCATCACG
                 ************************************************************

35.ttg3.ocs      CTAGTGTCAATGCCATTGCCTGGGCTCCTCAGAGTTGTAAGCATATTTGTTCCGCTGGGG
pART7.3.seq      CTAGTGTCAATGCCATTGCCTGGGCTCCTCAGAGTTGTAAGCATATTTGTTCCGCTGGGG
                 ************************************************************

35.ttg3.ocs      ATGATACCCAGGCTCTTATTTGGGAGTTGCCTACGGTGGCGGGACCTAATGGTATCGATC
pART7.3.seq      ATGATACCCAGGCTCTTATTTGGGAGTTGCCTACGGTGGCGGGACCTAATGGTATCGATC
                 ************************************************************

35.ttg3.ocs      CCTCTGTGTGTTTACTCTGCTGGCTACGAAATTAATCAATTACAGTGGTCTGCTGCGCAA
pART7.3.seq      CCTCTGTGTGTTTACTCTGCTGGCTACGAAATTAATCAATTACAGTGGTCTGCTGCGCAA
                 ************************************************************

35.ttg3.ocs      CCTGATTGGATTGCCATTGCCTTTTCCAACAAATTGCAGCTTCTCAAAGTTTGAGGTTTT
pART7.3.seq      CCTGATTGGATTGCCATTGCCTTTTCCAACAAATTGCAGCTTCTCAAAGTTTGAGGTTTT
                 ************************************************************

35.ttg3.ocs      CACACTGTATTTTCCTGCTTGAAGTCTGGTCTTCTGTTTCATAGACTGCTTAAGCTTGTA
pART7.3.seq      CACACTGTATTTTCCTGCTTGAAGTCTGGTCTTCTGTTTCATAGACTGCTTAAGCTTGTA
                 ************************************************************

35.ttg3.ocs      TTCCCCTTTGACTTGATTTGTGGAATCAGTTTTAGTTAAGCCTGTCAATTCTGACTTGAT
pART7.3.seq      TTCCCCTTTGACTTGATTTGTGGAATCAGTTTTAGTTAAGCCTGTCAATTCTGACTTGAT
                 ************************************************************

35.ttg3.ocs      TTGTGGAATCAGTTTTAGTTTAGCTGTTAATTCTGCCATTACTTTGCACCATTGATGAAG
pART7.3.seq      TTGTGGAATCAGTTTTAGTTTAGCTGTTAATTCTGCCATTACTTTGCACCATTGATGAAG
                 ************************************************************
```

```
35.ttg3.ocs     TAAATGTTCTTCCAAAAATCTGTATTGAAATTCCTTTAGTTAAGTTTATCATCATATGGT
pART7.3.seq     TAAATGTTCTTCCAAAAATCTGTATTGAAATTCCTTTAGTTAAGTTTATCATCATATGGT
                ************************************************************

35.ttg3.ocs     AAACATTACATTCAAGAATTTTTAAGTTTCGTTTGTTTGATGCTTGATTGTCTTTATGTA
pART7.3.seq     AAACATTACATTCAAGAATTTTTAAGTTTCGTTTGTTTGATGCTTGATTGTCTTTATGTA
                ************************************************************

35.ttg3.ocs     GTACAATTGAACAAAACCTCCCCCTTTTAGATAATTTTCATGTGATACAGTTTCAGAAAA
pART7.3.seq     GTACAATTGAACAAAACCTCCCCCTTTTAGATAATTTTCATGTGATACAGTTTCAGAAAA
                ************************************************************

35.ttg3.ocs     GAATTTCACTGGGCATCTAGGTCATCGAATTATATGATCAGTTTCATGTTAAGGAATTTA
pART7.3.seq     GAATTTCACTGGGCATCTAGGTCATCGAATTATATGATCAGTTTCATGTTAAGGAATTTA
                ************************************************************

35.ttg3.ocs     ACTTAACAGGTAAAAGCATCATGGAGGTCCTTGTACTGGGAGTCAACTTGCATTTTGCCC
pART7.3.seq     ACTTAACAGGTAAAAGCATCATGGAGGTCCTTGTACTGGGAGTCAACTTGCATTTTGCCC
                ************************************************************

35.ttg3.ocs     TCACTACTCAAAAAATGGGTAAATTAGTCCCTCTACGATAGAGCAAACTAGAGTCCTGCT
pART7.3.seq     TCACTACTCAAAAAATGGGTAAATTAGTCCCTCTACGATAGAGCAAACTAGAGTCCTGCT
                ************************************************************

35.ttg3.ocs     TTAATGAGATATGCGAGACGCCTATGATCGCATGATATTTGCTTTCAATTCTGTTGTGCA
pART7.3.seq     TTAATGAGATATG-----------------------------------------------
                *************

35.ttg3.ocs     CGTTGTAAAAAACCTGAGCATGTGTAGCTCAGATCCTTACCGCCGGTTTCGGTTCATTCT
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     AATGAATATATCACCCGTTACTATCGTATTTTTATGAATAATATTCTCCGTTCAATTTAC
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TGATTGTACCCTACTACTTATATGTACAATATTAAAATGAAAACAATATATTGTGCTGAA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TAGGTTTATAGCGACATCTATGATAGAGCGCCACAATAACAAACAATTGCGTTTTATTAT
pART7.3.seq     ------------------------------------------------------------


35.ttg3.ocs     TACAAATCCAATTTTAAAAAAAGCGGCAGAACCGGTCAAACCTAAAAGACTGATTACATA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     AATCTTATTCAAATTTCAAAAGGCCCCAGGGGCTAGTATCTACGACACACCGAGCGGCGA
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     ACTAATAACGTTCACTGAAGGGAACTCCGGTTCCCCGCCGGCGCGCATGGGTGAGATTCC
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TTGAAGTTGAGTATTGGCCGTCCGCTCTACCGAAAGTTACGGGCACCATTCAACCCGGTC
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     CAGCACGGCGGCCGGGTAACCGACTTGCTGCCCCGAGAATTATGCAGCATTTTTTTGGTG
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     TATGTGGCCCCAAATGAAGTGCAGGTCAAACCTTGACAGTGACGACAAATCGTTGGGCG
pART7.3.seq     ------------------------------------------------------------

35.ttg3.ocs     GGTCCAGGGCGAATTTTGCGACAACATGTCGAGGCTCAGCAGGACCCGGT
pART7.3.seq    --------------------------------------
```

Appendix D: pTTG(C)E2 in pART7

```
                        10        20        30        40        50        60
                         .         .         .         .         .         .
35.ttg4.ocs     ATGGTGGAGCACGACACTCTCGTCTACTCCAAGAATATCAAAGATACAGTCTCAGAAGAC
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     CAAAGGGCTATTGAGACTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCCAT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     TGCCCAGCTATCTGTCACTTCATCAAAAGGACAGTAGAAAAGGAAGGTGGCACCTACAAA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     TGCCATCATTGCGATAAAGGAAAGGCTATCGTTCAAGATGCCTCTGCCGACAGTGGTCCC
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     AAAGATGGACCCCCACCCACGAGGAGCATCGTGGAAAAAGAAGACGTTCCAACCACGTCT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     TCAAAGCAAGTGGATTGATGTGATAACATGGTGGAGCACGACACTCTCGTCTACTCCAAG
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     AATATCAAAGATACAGTCTCAGAAGACCAAAGGGCTATTGAGACTTTTCAACAAAGGGTA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     ATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCAAAAGGACA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     GTAGAAAAGGAAGGTGGCACCTACAAATGCCATCATTGCGATAAAGGAAAGGCTATCGTT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     CAAGATGCCTCTGCCGACAGTGGTCCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     GAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTCCACT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     GACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGA
pART7.4.seq     -------------------------------------AAGACCCTTCCTCTATATAAGGA
                                                     ***********************

35.ttg4.ocs     AGTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCAAAGGGTGTCCAAAAC
pART7.4.seq     AGTTCATTTCATTTGGAGAGGACACGCTCGAGGAATTCGGTACCAAAGGGTGTCCAAAAC
                ************************************************************

35.ttg4.ocs     GTTTTGGATTTCCTTCAAATTGATTTCCCTGACATGGATGTCATTGGCATCTCTGGTGAG
pART7.4.seq     GTTTTGGATTTCCTTCAAATTGATTTCCCTGACATGGATGTCATTGGCATCTCTGGTGAG
                ************************************************************

35.ttg4.ocs     TATCATTCCCATGAATATATACTTCTTCATTTTCGTCTTCTGAGCTAGACCATTGAACTC
pART7.4.seq     TATCATTCCCATGAATATATACTTCTTCATTTTCGTCTTCTGAGCTAGACCATTGAACTC
                ************************************************************
```

```
35.ttg4.ocs      TTAAGCAAAGGTGAAAGAGCAGGACCTTGAGTTTGACCATGGCTACTATTACTGCAATAG
pART7.4.seq      TTAAGCAAAGGTGAAAGAGCAGGACCTTGAGTTTGACCATGGCTACTATTACTGCAATAG
                 ************************************************************


35.ttg4.ocs      TCAAATCCGAATATTTTATATTTCTGCAGATAAGACACCCAGTTCTTTTATACATATATA
pART7.4.seq      TCAAATCCGAATATTTTATATTTCTGCAGATAAGACACCCAGTTCTTTTATACATATATA
                 ************************************************************

35.ttg4.ocs      TATTTTATGTTTTCTGATTCACTGTTTGAAACGGCATGGGTCCATTTCTTTTCTTTTCTT
pART7.4.seq      TATTTTATGTTTTCTGATTCACTGTTTGAAACGGCATGGGTCCATTTCTTTTCTTTTCTT
                 ************************************************************

35.ttg4.ocs      TTTAAATTTTCGGCTTTTATATTTGAAAAAAAAATTAAATAAATAAATTTAGCTTTAAAA
pART7.4.seq      TTTAAATTTTCGGCTTTTATATTTGAAAAAAAAATTAAATAAATAAATTTAGCTTTAAAA
                 ************************************************************

35.ttg4.ocs      TTTTTGTTATGGAATGAAATAATATAGTATTGTGCGGACTGCGGAGTATATAACGACTAA
pART7.4.seq      TTTTTGTTATGGAATGAAATAATATAGTATTGTGCGGACTGCGGAGTATATAACGACTAA
                 ************************************************************

35.ttg4.ocs      ATATTAAAACAAAAAGCATATATAACCTTTTTGTCATTAAGTAGATGAAAAAGGAATGTC
pART7.4.seq      ATATTAAAACAAAAAGCATATATAACCTTTTTGTCATTAAGTAGATGAAAAAGGAATGTC
                 ************************************************************

35.ttg4.ocs      GTCAAAACACGTTACCTTCCATAATCTTCACACCAATCCAACCTTTCTTTGGTTATCTGC
pART7.4.seq      GTCAAAACACGTTACCTTCCATAATCTTCACACCAATCCAACCTTTCTTTGGTTATCTGC
                 ************************************************************

35.ttg4.ocs      TAAGAGTAAAAAATAGCTGACAGCGGCAGCGTTGGTAATGACGGCCACCAGCGATCCGAA
pART7.4.seq      TAAGAGTAAAAAATAGCTGACAGCGGCAGCGTTGGTAATGACGGCCACCAGCGATCCGAA
                 ************************************************************

35.ttg4.ocs      CCCCCGAAGTTTCCGACGAGCAGCAAAAACGATCGGAGATATACACATATGAAGCTCCCT
pART7.4.seq      CCCCCGAAGTTTCCGACGAGCAGCAAAAACGATCGGAGATATACACATATGAAGCTCCCT
                 ************************************************************

35.ttg4.ocs      GGTACATCTACGCTATGAATTGGAGCGTCCGCCGCGACAAAAAGTACCGACTCGCCATCG
pART7.4.seq      GGTACATCTACGCTATGAATTGGAGCGTCCGCCGCGACAAAAAGTACCGACTCGCCATCG
                 ************************************************************

35.ttg4.ocs      CCAGCCTCCTCGAGCAATACCCTAACCGTCTCCAGATTGTTCAACTCGACGACTCCAATG
pART7.4.seq      CCAGCCTCCTCGAGCAATACCCTAACCGTCTCCAGATTGTTCAACTCGACGACTCCAATG
                 ************************************************************

35.ttg4.ocs      GGTGAGATCCGGTCGGATCCCAACCTCTCCTTCGACCATCCCTATCCCGCTACTAAGACC
pART7.4.seq      GGTGAGATCCGGTCGGATCCCAACCTCTCCTTCGACCATCCCTATCCCGCTACTAAGACC
                 ************************************************************

35.ttg4.ocs      ATCTTCATTCCCGACAAGGATTGCCAGAAACCCGATCTTCTCGCTACCTCCTCCGATTTT
pART7.4.seq      ATCTTCATTCCCGACAAGGATTGCCAGAAACCCGATCTTCTCGCTACCTCCTCCGATTTT
                 ************************************************************

35.ttg4.ocs      CTTCGCATATGGCGGATCTCGGGATGACGGTTCCCGCGTTGACCTCAAATCGCTCCTTAA
pART7.4.seq      CTTCGCATATGGCGGATCTCGGGATGACGGTTCCCGCGTTGACCTCAAATCGCTCCTTAA
                 ************************************************************

35.ttg4.ocs      TGGCAATAAGAACAGTGAATTTTGGCGGTCCTCTCACGTCTTTCGACTGGAACGAGGCGG
pART7.4.seq      TGGCAATAAGAACAGTGAATTTTGGCGGTCCTCTCACGTCTTTCGACTGGAACGAGGCGG
                 ************************************************************

35.ttg4.ocs      AGCCGAAGCGAATCGGGACTTCCTCCATTGATACGACTTGCACTATTTGGGATATCGAGA
pART7.4.seq      AGCCGAAGCGAATCGGGACTTCCTCCATTGATACGACTTGCACTATTTGGGATATCGAGA
                 ************************************************************

35.ttg4.ocs      AGGAAACAGTGGATACCCAGTTAATTGCCCACGATAAGGAGGTTTACGACATCGCTTGGG
pART7.4.seq      AGGAAACAGTGGATACCCAGTTAATTGCCCACGATAAGGAGGTTTACGACATCGCTTGGG
                 ************************************************************
```

```
35.ttg4.ocs     GTGGAGTTGGGGTTTTCGCTTCCGTCTCCGCCGACGGTTCCGTTAGGGTTTTCGATTTAC
pART7.4.seq     GTGGAGTTGGGGTTTTCGCTTCCGTCTCCGCCGACGGTTCCGTTAGGGTTTTCGATTTAC
                ************************************************************

35.ttg4.ocs     GCGACAAGGAACATTCGACGATCATCTACGAGAGTTCGGAGCCGGACGCGCCGCTGGTGC
pART7.4.seq     CCNACAAGGAACATTCGACGATCATCTACGAGAGTTCGGAGCCGGACGCGCCGCTGGTGC
                ************************************************************

35.ttg4.ocs     GGTTGGGGTGGAACAAGCAGGACCCCAGATATATGGCGACCATTATAATGGATAGTGCCA
pART7.4.seq     GGTTGGGGRGGAACAAGCAGGACCCCAGATATATGGCGACCATTATAATGGATAGTGCCA
                ************************************************************

35.ttg4.ocs     AGGTGGTTGTTTTGGATATCCGTTTCCCGACGTTGCCGGTGGTTGAGTTGCGGAGACACC
pART7.4.seq     AGGTGGTTGTTTTGGATATCCGTTTCCCGACGTTGCCGGTGGTTGAGTTGCGGAGACACC
                ************************************************************

35.ttg4.ocs     AGGCGAGCGTCAATGCCGTCGCTTGGGCTCCCCATAGTTCCTGCCACATTTGCACCGCCG
pART7.4.seq     AGGCGAGCGTCAATGCCGTCGCTTGGGCTCCCCATAGTTCCTGCCACATTTGCACCGCCG
                ************************************************************

35.ttg4.ocs     GCGATGATTCTCAGGCGTTCGATTTGGGAATTTGTCTTCGATGGGTCAGCCTGTTAAAGG
pART7.4.seq     GCGATGATTCTCAGGCGTTCGATTTGGGAATTTGTCTTCGATGGGTCAGCCTGTTAAAGG
                ************************************************************

35.ttg4.ocs     TGGGCTTGACCCCATTCTTGCATACACGGCTGGGGCTGAAATTGAACAGTTGCAATGGTC
pART7.4.seq     TGGGCTTGACCCCATTCTTGCATACACGGCTGGGGCTGAAATTGAACAGTTGCAATGGTC
                ************************************************************

35.ttg4.ocs     GTCGGTCTCAAGCCTGATTGGGTGGCCATTGCCTTCTCCACTAAGCTTCAGATTCTAAGG
pART7.4.seq     GTCGGTCTCAAGCCTGATTGGGTGGCCATTGCCTTCTCCACTAAGCTTCAGATTCTAAGG
                ************************************************************

35.ttg4.ocs     GTATGAATTACCGGTGTGAAGTTAGGTTTTTCTGTTTGTTTTGACATTAACTAACAACAT
pART7.4.seq     GTATGAATTACCGGTGTGAAGTTAGGTTTTTCTGTTTGTTTTGACATTAACTAACAACAT
                ************************************************************

35.ttg4.ocs     GCTTGCTTGATCTTCCATTTGAATCGAGTTTCACTTCTGAATTCCTGCAGCCCGGGGGAT
pART7.4.seq     GCTTGCTTGATCTTCCATTTGAATCGAGTTTCACTTCTGAATTCCTGCAGCCCGGGGGAT
                ************************************************************

35.ttg4.ocs     CCACTAGTTCTAGAGTCCTGCTTTAATGAGATATGCGAGACGCCTATGATCGCATGATAT
pART7.4.seq     CCACTAGTTCTAGAGTCCTGCTTTAATGAGATATG-------------------------
                ***********************************

35.ttg4.ocs     TTGCTTTCAATTCTGTTGTGCACGTTGTAAAAACCTGAGCATGTGTAGCTCAGATCCTT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     ACCGCCGGTTTCGGTTCATTCTAATGAATATATCACCCGTTACTATCGTATTTTTATGAA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     TAATATTCTCCGTTCAATTTACTGATTGTACCCTACTACTTATATGTACAATATTAAAAT
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     GAAAACAATATATTGTGCTGAATAGGTTTATAGCGACATCTATGATAGAGCGCCACAATA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     ACAAACAATTGCGTTTTATTATTACAAATCCAATTTTAAAAAAAGCGGCAGAACCGGTCA
pART7.4.seq     ------------------------------------------------------------

35.ttg4.ocs     AACCTAAAAGACTGATTACATAAATCTTATTCAAATTTCAAAAGGCCCCAGGGGCTAGTA
pART7.4.seq     ------------------------------------------------------------
```

```
35.ttg4.ocs      TCTACGACACACCGAGCGGCGAACTAATAACGTTCACTGAAGGGAACTCCGGTTCCCCGC
pART7.4.seq      ------------------------------------------------------------

35.ttg4.ocs      CGGCGCGCATGGGTGAGATTCCTTGAAGTTGAGTATTGGCCGTCCGCTCTACCGAAAGTT
pART7.4.seq      ------------------------------------------------------------


35.ttg4.ocs      ACGGGCACCATTCAACCCGGTCCAGCACGGCGGCCGGGTAACCGACTTGCTGCCCCGAGA
pART7.4.seq      ------------------------------------------------------------

35.ttg4.ocs      ATTATGCAGCATTTTTTTGGTGTATGTGGCCCCAAATGAAGTGCAGGTCAAACCTTGAC
pART7.4.seq      ------------------------------------------------------------

35.ttg4.ocs      AGTGACGACAAATCGTTGGGCGGGTCCAGGGCGAATTTTGCGACAACATGTCGAGGCTCA
pART7.4.seq      ------------------------------------------------------------

35.ttg4.ocs      GCAGGACCCGGT
pART7.4.seq      ------------
```