

# The *TRANSPARENT TESTA GLABRA1* Locus, Which Regulates Trichome Differentiation and Anthocyanin Biosynthesis in Arabidopsis, Encodes a WD40 Repeat Protein

Amanda R. Walker,<sup>a,1</sup> Paul A. Davison,<sup>a,2</sup> Agnese C. Bolognesi-Winfield,<sup>a</sup> Celia M. James,<sup>a,3</sup> N. Srinivasan,<sup>b,4</sup> Tom L. Blundell,<sup>b</sup> Jeffrey J. Esch,<sup>c,5</sup> M. David Marks,<sup>c,6</sup> and John C. Gray<sup>a</sup>

<sup>a</sup> Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

<sup>b</sup> Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, United Kingdom

<sup>c</sup> School of Biological Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska 68588

The *TRANSPARENT TESTA GLABRA1* (*TTG1*) locus regulates several developmental and biochemical pathways in Arabidopsis, including the formation of hairs on leaves, stems, and roots, and the production of seed mucilage and anthocyanin pigments. The *TTG1* locus has been isolated by positional cloning, and its identity was confirmed by complementation of a *ttg1* mutant. The locus encodes a protein of 341 amino acid residues with four WD40 repeats. The protein is similar to AN11, a regulator of anthocyanin biosynthesis in petunia, and more distantly related to those of the  $\beta$  subunits of heterotrimeric G proteins, which suggests a role for *TTG1* in signal transduction to downstream transcription factors. The 1.5-kb *TTG1* transcript is present in all major organs of Arabidopsis. Sequence analysis of six mutant alleles has identified base changes producing truncations or single amino acid changes in the *TTG1* protein.

## INTRODUCTION

The *TRANSPARENT TESTA GLABRA1* (*TTG1*) locus controls many apparently unrelated characters of Arabidopsis (catalogued by Koornneef, 1981), several of which appear to be confined to the epidermal cell layer of different tissues. *ttg1* mutants have a glabrous phenotype, possessing none of the leaf or stem hairs (trichomes) that normally are derived from the meristematic L1 cell layer. Purple anthocyanin pigments are absent from the *ttg1* seed coat, causing the transparent testa phenotype in which the yellow cotyledons are visible through the testa. In wild-type plants, anthocyanins are present in the hypocotyl of seedlings and in the stem and leaves of plants as they age, and they are inducible by many forms of stress, including high light, poor nutri-

ents, or water stress. *ttg1* mutants completely lack anthocyanins in the epidermis and in subepidermal layers of leaves and stems. Mucilage normally found in the cell wall of the seed coat is absent in *ttg1* mutants. Seeds of *ttg1* plants do not require drying and cold treatments to germinate and therefore exhibit an altered seed dormancy when compared with ecotypes, such as Landsberg *erecta* (*Ler*; Koornneef, 1981; Léon-Kloosterziel et al., 1994). This characteristic of *ttg1* mutants may be linked to an altered seed coat structure. The *TTG1* gene appears to have the opposite effect on root hair formation when compared with its effect on leaf hair initiation. In Arabidopsis, root hairs extend from root epidermal cells only in files of cells that contact two underlying cortical cells, whereas in *ttg1* mutants, extra root hairs occur in the atrichoblast cell files (Galway et al., 1994). Under laboratory growth conditions, mutations at the *ttg1* locus do not greatly affect the viability of the plants.

In *ttg1* mutants, the anthocyanin biosynthetic pathway is blocked at the dihydroflavonol-4-reductase (DFR) step, because DFR-encoding transcripts have not been detected in these mutants (Shirley et al., 1995). By contrast, transcripts of the chalcone synthase and chalcone isomerase genes are unaffected. The point of regulation of the pathway by *TTG1* was confirmed by the cloning and analysis of the expression of the Arabidopsis gene encoding flavanone 3-hydroxylase, which catalyzes the enzymatic step between chalcone isomerase and DFR (Pelletier and Shirley, 1996).

<sup>1</sup>To whom correspondence should be addressed. E-mail arw13@cus.cam.ac.uk; fax 44-1223-333953.

<sup>2</sup>Current address: Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.

<sup>3</sup>Current address: Plant Breeding and Biotechnology, Horticulture Research International, East Malling, Kent ME19 6BJ, UK.

<sup>4</sup>Current address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

<sup>5</sup>Current address: Department of Botany, University of Wisconsin, Madison, WI 53706.

<sup>6</sup>Current address: Department of Genetics and Cell Biology, University of Minnesota, St. Paul, MN 55108.

This feature of *ttg1* mutants resembles the effect of the *Delila* mutation of snapdragon (Martin et al., 1991). *Delila* mutants also lack transcripts encoding DFR and subsequent enzymes in the anthocyanin pathway. In maize, a group of closely related genes, including *Booster*, *R*, and *Lc*, regulates the entire anthocyanin biosynthetic pathway from chalcone synthase onward. *Delila* and *R* both encode MYC-like transcription factors (Ludwig et al., 1989; Goodrich et al., 1992). *R* has been shown to activate directly the transcription of several genes encoding anthocyanin biosynthetic enzymes in conjunction with a MYB transcription factor encoded by the *C1* gene in maize (Goff et al., 1992; Lesnick and Chandler, 1998).

The cauliflower mosaic virus 35S promoter was used to express the maize *R* gene in Arabidopsis plants. The phenotypes of these plants complemented all of the mutant phenotypes of *ttg1* plants, suggesting that *TTG1* might encode an Arabidopsis *R* homolog or might regulate an *R*-like gene (Lloyd et al., 1992). Transformation of wild-type Arabidopsis plants with the cauliflower mosaic virus 35S-*R* construct resulted in more anthocyanin in the plant and increased numbers of trichomes on leaves and stems. However, although transgenic tobacco plants expressing the maize *R* gene had flowers with elevated anthocyanin levels, no change in trichome development was observed (Lloyd et al., 1992).

Many genetic loci involved in trichome differentiation and development in Arabidopsis have been described (Hülkamp et al., 1994; Marks, 1997). *TTG1* and *GLABROUS1* (*GL1*) play a key role in the initiation of trichomes; mutations at either locus lead to hairless leaves and stems (Koornneef, 1981; Koornneef et al., 1982). The *GL1* gene encodes a MYB transcription factor (Oppenheimer et al., 1991). Double mutant analysis has suggested that *GL1* and *TTG1* may affect the same point in the pathway leading to trichome formation (Larkin et al., 1994), although *gl1* mutants do not show the same pleiotropic effects as *ttg1* mutants. Another locus that affects both trichome development and pigment levels has recently been identified and isolated. This locus has been designated *TTG2* (C.S. Johnson and D.R. Smyth, personal communication).

Mutations at another *GLABRA* locus, *GL2*, have some features in common with *ttg1*, although *gl2* mutants have a normal anthocyanin content and have rudimentary trichomes, suggesting that *GL2* is not required to determine the fate of these cells (Koornneef, 1981). *gl2* mutants have an increased number of ectopic root hairs, although *gl2* atrichoblast cells resemble those of the wild type more than they do those of the *ttg1* mutant, indicating a role later in the development of the root epidermal cells (Masucci et al., 1996). The effects on the seed coat and mucilage are similar to those of the *ttg1* mutations (Koornneef, 1981). The *GL2* locus encodes another transcription factor, a homeodomain protein with a leucine zipper domain (Rerie et al., 1994). By using in situ hybridization, studies showed that the *GL2* gene is expressed in developing trichomes (Rerie et al., 1994) and in the atrichoblast cell files of roots (Masucci et al.,

1996). *GL2* transcripts are greatly reduced in *ttg1* mutants, suggesting that the *TTG1* gene product is required for normal levels of *GL2* expression (Di Cristina et al., 1996). However, *TTG1* does not influence the pattern of *GL2* expression, which occurs at very low levels in specific tissues in the absence of a functional *TTG1* gene (Hung et al., 1998; Szymanski et al., 1998).

Another MYB-related gene, *CAPRICE* (*CPC*), may have an effect opposite to that of *TTG1* in promoting root hair development (Wada et al., 1997). Plants overexpressing *CPC* have extra root hairs and a reduced number of leaf trichomes. The data from studies of *GL1*, *CPC*, and *R* suggest that the MYB and MYC transcription factors regulate trichome, root hair, and anthocyanin production and that *TTG1* may control directly or indirectly the activation of one or more transcription factors.

The *TTG1* locus has been mapped by Koornneef et al. (1983) to chromosome 5 between *MS1* and *GA3*. We used this information to isolate the *TTG1* locus by using map-based techniques. Restriction fragment length polymorphisms (RFLPs) were used to map crossovers on both sides of the *ttg1-1* allele followed by complementation of the mutant phenotype by the introduction of genomic fragments containing the *TTG1* locus. The *TTG1* locus encodes a WD40 repeat protein. This protein may regulate signal transduction pathways governing leaf hair formation and anthocyanin biosynthesis.

## RESULTS

### Molecular Mapping of *ttg1-1*

Two sets of recombinants between *ttg1-1* and flanking genetic markers were used for fine mapping of the *TTG1* region of chromosome 5. A total of 140 recombinants between the *ttg1-1* and *MS1* loci were analyzed. Between *ttg1* and *GA3*, 299 recombinants were selected covering a 4-centimorgan (cM) interval. Figure 1A shows a large-scale map of the *TTG1* region between *MS1* and *GA3* that was derived from RFLP analysis of the recombinants on both sides of the *ttg1-1* mutation by using a variety of probes. Figure 1B shows a higher resolution map of the region covered by yeast artificial chromosome (YAC) EG20H2 (Grill and Somerville, 1991). RFLPs visible upon probing with the g4556 cosmid (Nam et al., 1989) could be separated from *ttg1-1* only in a single recombinant called Dennis (on the *MS1* side of *ttg1-1*; Figure 1B), suggesting that g4556 was very close to the mutation in *ttg1-1*. Figure 2A shows a DNA gel blot of EcoRI-digested genomic DNA of several recombinants probed with g4556. The pattern of hybridizing bands in Dennis is similar to that of the Wassilewskija (*Ws*) parent.

The cosmid g4556 and YAC EG20H2, which hybridized

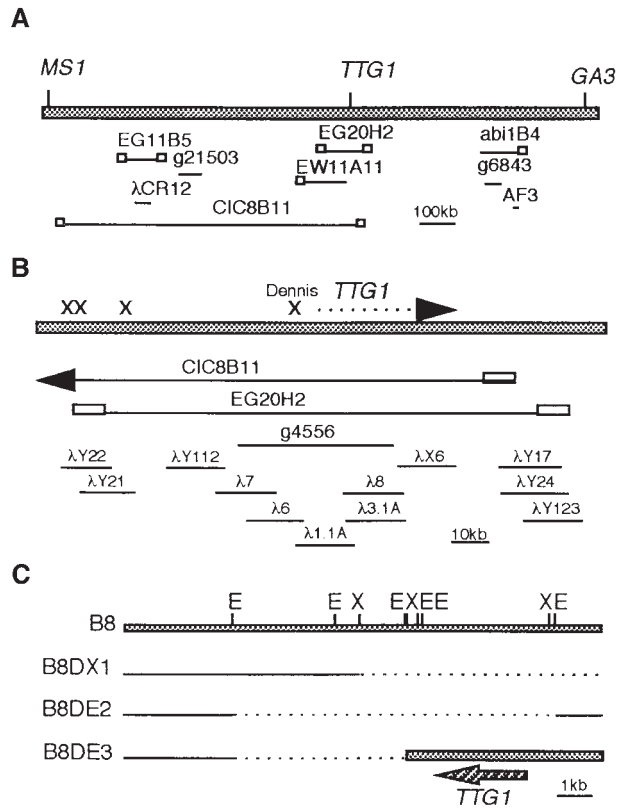
with cosmid g4556, were used to isolate overlapping fragments of *Ler* genomic DNA in  $\lambda$  vectors. The fine-scale map (Figure 1B) shows the order of DNA fragments and the positions of several recombination events close to *TTG1-1*, including the crossover in Dennis. It was not possible to map nearby crossover events on the proximal side of *ttg1-1* due to the lack of RFLPs between the ecotypes used (*Ler* and *Rschew* [RLD1]). Moreover, part of this region from YAC EG20H2 could not be subcloned into bacterial vectors and was not represented by clones from three genomic libraries (including a cosmid library) screened with the YAC DNA.

The fine mapping of the *TTG1* locus was aided by analysis of the *ttg1-13* mutation, which was generated by fast-neutron bombardment from an RLD1 line carrying the *dis2* mutation. The results of a DNA gel blot (Figure 2B) of mutant and parental DNA digested with *EcoRV* and *EcoRI* and probed with  $\lambda$ 8 (Figure 1B) suggest that a deletion of this region could account for the phenotype of the *ttg1-13* mutant, indicating that  $\lambda$ 8 lies close to or contains the *TTG1* locus.

**Complementation of *ttg1* Mutations Using Genomic Fragments**

The entire genomic fragments from overlapping  $\lambda$  clones 1.1A, 8, and X6 (marked in Figure 1B) were inserted into the plasmid binary vector pBINNOT (a derivative of pBIN19 [Bevan, 1984] with a unique *NotI* site in the polylinker) by using *NotI* restriction sites in the  $\lambda$  vector; the plasmids were called pB1.1A, pB8, and pBX6. These constructs were used to transform *Arabidopsis* (the *ttg1-1* mutant in the *Ler* background) by root coculture, with *Agrobacterium* carrying the plasmids (Valvekens et al., 1988). One plant transformed with pB8, derived from  $\lambda$ 8, had trichomes and produced anthocyanin, but it failed to set seed. By contrast, several plants transformed with pB1.1A or pBX6 did not have trichomes. These results suggest that the *TTG1* locus might be contained in the genomic insert in pB8.

Three deletions of pB8 were made by using restriction sites within the genomic sequence and the polylinker of the vector. These deletion constructs (shown in Figure 1C) were used to transform *Arabidopsis* plants, carrying the *ttg1-9* mutation in a Columbia background, via vacuum infiltration (Bechtold et al., 1993), with *Agrobacterium* carrying the plasmids. From the pB8DE3 construct, 60 of 61 kanamycin-resistant transformants produced trichomes. All 19 transformants from the deletion construct pB8DX1 and 36 transformants from pB8DE2 had no or very few trichomes, a phenotype similar to that of plants carrying the *ttg1-9* allele. All transformants from pB8DE3 bearing normal trichomes also exhibited wild-type characteristics of brown seed, seed mucilage, purple coloring of the plant, and normal root hair numbers, indicating that the other phenotypes of the *ttg1* mutant had also been complemented. The DNA region specific to pB8DE3, and therefore the region that was presumed to contain the *TTG1* locus, is as shown in Figure 1C.



**Figure 1.** Maps of the *TTG1* Region.

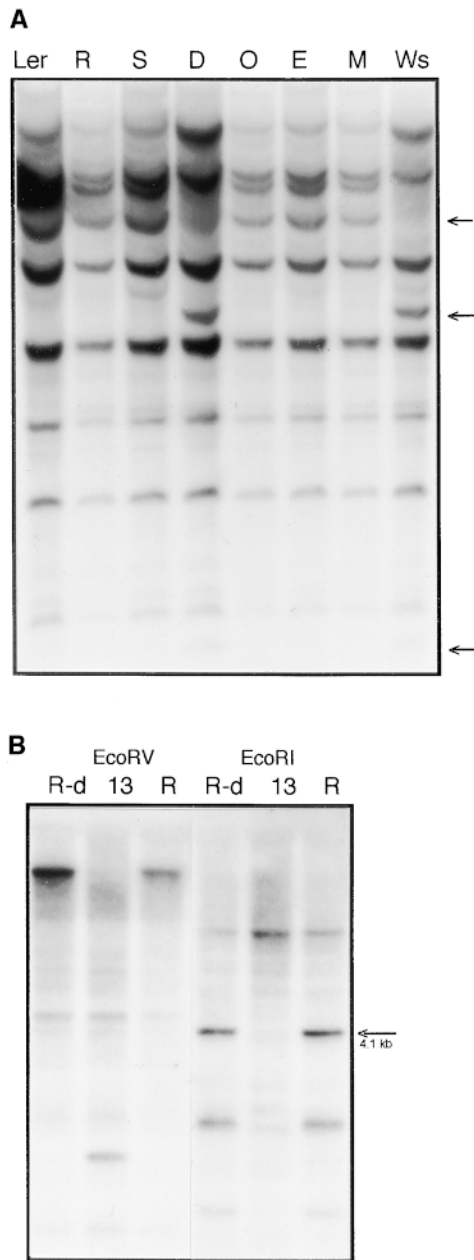
(A) Large-scale map from *MS1* to *GA3* indicating the position of the most important markers used to determine the position of the *TTG1* locus. Boxes show end probes generated from YACs. Cosmid,  $\lambda$ , and plasmid probes for RFLPs were made by labeling whole constructs.

(B) Fine map of the *TTG1* region showing the estimated position of *TTG1* (arrowhead) in relation to the DNA constructs, including YACs EG20H2 and CIC8B11, cosmid g4556, and  $\lambda$  clones (obtained from library screens). The positions of four crossovers close to *TTG1* are indicated by X's, and Dennis (the closest recombinant to *ttg1-1*) is labeled.

(C) Restriction map of the insert from  $\lambda$ 8 showing *EcoRI* (E) and *XbaI* (X) sites used to generate the deletion constructs indicated below. Solid lines represent the DNA remaining in the constructs; dotted lines indicate the deleted regions. Sequencing showed the position and direction of the *TTG1* transcript (arrow). All maps are drawn in the same orientation.

**Analysis of the *TTG1* Region**

Sequence analysis of 5777 bp from pB8DE3 (Figure 1C), using Genmark (Borodovsky and McIninch, 1993) and Net-PlantGene (Hebsgaard et al., 1996), revealed a single gene. Both computer programs rely on the very high A+T content of *Arabidopsis* intergenic regions to predict coding regions.



**Figure 2.** Genomic DNA Gel Blots Used to Locate the *TTG1* Locus.

**(A)** Genomic DNA (1.5  $\mu$ g) from recombinants from the *MS1* side of *TTG1* digested with *EcoRI* and probed with g4556. The three differences in band mobility between the parental lines *Ler* and *Ws* are indicated by arrows. Rachael (R), Simon (S), Dennis (D), Oscar (O), Elaine (E), and Melissa (M) are the six recombinant lines that were used.

**(B)** Genomic DNA (1  $\mu$ g) from *RLD1-dis2* (R-d), *ttg1-13* (13), and *RLD1* (R) was digested with *EcoRV* or *EcoRI* and probed with  $\lambda$ 8. The 4.1-kb *EcoRI* fragment containing the *TTG1* locus is indicated.

This gene was shown to represent a single-copy sequence by comparing DNA gel blot data with cloned genomic restriction fragments (data not shown).

The predicted gene has very high similarity to an Arabidopsis expressed sequence tag (EST) (GenBank accession numbers F20055 and F20056) that, by sequence analysis, starts at base 467 and corresponds to the 3' end of the gene shown in Figure 3. Comparison of the genomic and EST sequences showed an intron of 700 bp, located one base 3' of the predicted termination codon. Other introns were not identified in the gene by using computer analysis or by comparison to the EST. The predicted protein sequence of 341 amino acid residues shows sequence similarity to a large and diverse group of proteins with WD40 repeat motifs (see Discussion). The *TTG1* protein has four WD40 repeats indicated in Figure 3. There is a proline-rich region in which eight of 26 amino acid residues are proline. The amino acid sequence most similar to that of *TTG1* in the databases is the product of the *An11* gene of petunia (de Vetten et al., 1997), which contains 79.5% identical and 87.5% similar amino acid residues. An alignment of *TTG1* with the *AN11* sequence is shown in Figure 4 together with the sequence of the  $\beta$  subunit of a heterotrimeric G protein.

To identify the 5' end of the transcript and to confirm the likely translation initiation codon, we conducted a primer extension experiment with an end-labeled oligonucleotide designed from the complementary strand of the *TTG1* gene. The start of transcription was located 110 bp 5' of the predicted start of translation and is indicated as the beginning of the cDNA sequence (Figure 3). There are no other ATG codons in the 5' untranslated region. A possible TATA box is located 22 bp 5' of the start of transcription (Figure 3). The poly(A) addition site is evident in the sequence of the EST clone and is located at position 1907 (Figure 3). The transcript is predicted to be 1317 nucleotides plus a poly(A) tail.

The expression of the gene in different organs of Arabidopsis was examined by using an RNA gel blot hybridized with a *TTG1* probe (Figure 5). The transcript was estimated to be 1.5 kb and was present in all organs examined, including roots, leaves, stems, meristems, flowers, and flower buds. *TTG1* appears to be highly expressed in floral buds, in which there are only a few trichomes on the sepals and no anthocyanin in the flower petals or stamen tissue.

### Identification and Analysis of *ttg1* Alleles

Several different *ttg1* alleles had been identified previously (Koorneef, 1981). Six new *ttg1* alleles were identified in the Kranz collection (Table 1). Initially, the trichome, anthocyanin, and seed mucilage characteristics were compared with those of the *ttg1-1* mutants and were shown to be similar to this mutant. Lines were then crossed to *ttg1-1* to determine whether they carried mutations at the same locus. In the  $F_1$  generation, all plants lacked trichomes and anthocyanin,

```

-180 ccattatcaatagatatacattaaaagatactgattaatccagaggggttataatctacgcc
-120 gtctccattgattattttccgcttcttggaaaaatccgactgacactgacctcaaaactc
-60 tccctcactttcgtcgtgaaagaccgcaaatctcgaatcgaatcagcaccacacatttcc
1 atggataattcagctccagattcgttatccagatcggaaacccggctcacatagcactca
1 M D N S A P D S L S R S E T A V T Y D S
61 CCATATCCACTCTACGCCATGGCTTCTCTCTCCGCTCATCTCCGGTACAGAAATC
P Y P L Y A M A F S S L R S S S G H R I
121 GCCGTCGGAAGCTTCCTCGAAGATTACAACAACCCGATCGACATTCTCTCTTCGATTCC
41 A V G S F L E D Y N N R I D I L S F D S
181 GATTCATGACCGCTTAAAGCCTCTCCGAACTCTCTCTTCGAGCATCTCTTATCTCCAAACA
D S M T V K P L P N L S F E H P Y P P T
241 AAGCTAATGTTGAGTCTCTCTCTCTCCGCTCTCTCTCCGAGATCTCTCTCGCTTCC
81 K L M F S P P S L R R P S S G D L L A S
301 TCCGGCGATTCTCCGCTCTTGGGAATTAACGAAGATTCAATCAACCCGTCGAGCAATC
S G D F L E D S R I N E D S T V E P T
361 TCGGTTCTCAACAACAGAAAACGAGCGAGTTTGTGCGCCGTTGACTTCTCTCGATTGG
121 S V L N N S K T S E F C A P L T S F D W
421 AACGATGTAGAGCCGAAAGCTCTCGGAATCTGTAGTATTGATACGAGCTGTCAGATTGG
141 N D V E P K R L P N L S F E H P Y P P T
481 GATATGAGAGCTCTGTGTTGAGACTACGCTTATAGCTCATGATAAAGAGGTTCTAGAC
161 D I E K S V V E T Q L I A H D K E V H D
541 ATTGCTGGGAGAAGCTAGGGTTTTCGCATCAGTCTCTGCTGATGGATCCGCTAGGATC
181 I A W G E A R V F A S V S A D G S V R I
601 TTTGATTACGGTATGAAGAACTTACAATCATTACGAGAGCTCTCAGCCGCTGATAGC
201 E D L R L K E I I E S P Q P D
661 CCTTGTAAAGACTTCTCGAACAACAAGATCTTAGATATATGGCTACGATTTTGATG
221 P L L R L A W N K Q D L R Y M A T I L M
721 GATTCATAAGTTGATTCTCGATATTCGTTCCGCGACTATGCCTTTGCTGAGCTT
241 D S N K V I L D I R S P T M P V A E L
781 GAAAGACATCAGGCTAGTGTGATCTATAGCTTGGGCGCTCAGAGCTGTAACAATATT
261 E R H Q A S V N A I A W A P Q S C K H I
841 TGTCTGGTGGTATGATACAGGCTCTTATTGGGAGCTTCCCTACTGTTGCTGGACCC
281 C S G G D D T Q A L I W E L P T V A G P
901 AATGGATTGATCGATCTGCTTTATTCCGCTGGTTCGGAGATTAATCAGTTGCATCGG
301 N G I D P M S V Y S A G S E I N Q L Q W
961 TCTTCTCGACCCGATGGATTGGTATTGCTTTTGCATACAAATCAGCTCCTTAGA
321 S S S Q P D W I G I A F A N K M Q L L R
1021 GTTTGAGgtgagagtttctcttcttctacataattctcatttctgtaggctagattctaa
341 V *
1081 tgaggaagcattgattattggtttagattgtgtgcattacagatagttctctaggtttg
1141 gtaactaaacgttttttctgattcttgatacaaaagccactagagatttgacaactaactcg
1201 ttttagatttacctgaatcaatctctgtttaaatacaactcttggatgtaacacata
1261 aatcaacgttttagtagtcaatataatggctcttattagcagaggtctcaactctgctg
1321 taatggctgtagtgtagtcaatctgtggttctcctaaagttgatgtaacatagta
1381 tgaatggtgacactgctcaggtctcaatcttaccataaactctcagctatcttttt
1441 tttgagacaactcttcttaatttaacgaagccactagctacgtatacaaatatgtttaa
1501 tttaacgaagatctcagaaatgttactgctgactctgctgtaacccctcagaaacata
1561 tagaagtggaattggaactctcagctggttgaacatctttgtagtgctcaggtt
1621 tttgtaactcatttagacaacagcattgctatatacaccgacatagcaactagaaaa
1681 tcaataaccttccitataactactatccattcactgatgtagcagTGACAGATGTGAA
1741 GTGATCAATAAGGATTTAGCATAGACCCGATAAATCGTACTGCGTAAGTAGTGTGG
1801 TTTCCGCTCCCTCTCGCTTTTAGTTCGCGAATGACTCTGTATCTATCTGATGTAACATA
1861 AACTGAATTCATTGATGAACCAATGATACATTATCTTATGTTGtataaaacccaa
1921 ccaggatataatgggtttctggttttagatttggaattggagcttagtacaatgcaa

```

Figure 3. Sequence of the *TTG1* Locus.

The cDNA sequence is shown in capital letters; untranscribed regions and the intron are in lowercase letters. The first capitalized base shown (at position -109) is at the start of transcription, although nucleotide and amino acid numbering is from the start of translation. The protein sequence is given below the nucleotide sequence (GenBank accession number AJ133743). The two amino acids (normally WD) at the end of each repeat are underlined. The asterisk denotes the termination codon. The putative tata box is underlined.

and there was no segregation of the mutant phenotype in the F<sub>2</sub> generation. These new alleles have been designated *ttg1-15* to *ttg1-20* (Table 1).

The DNA sequence of a number of *ttg1* mutant alleles and their parental ecotypes was determined to identify the mutations in the *TTG1* locus. Polymerase chain reaction (PCR) was conducted with genomic DNA, and the PCR products were sequenced directly with the primers used to generate them. Sequence comparison of four ecotypes (*Ler*, *Ws*, *Columbia*, and *RLD1*) showed no nucleotide changes in the coding region, although there were several differences within the intron and in the promoter region (data not shown).

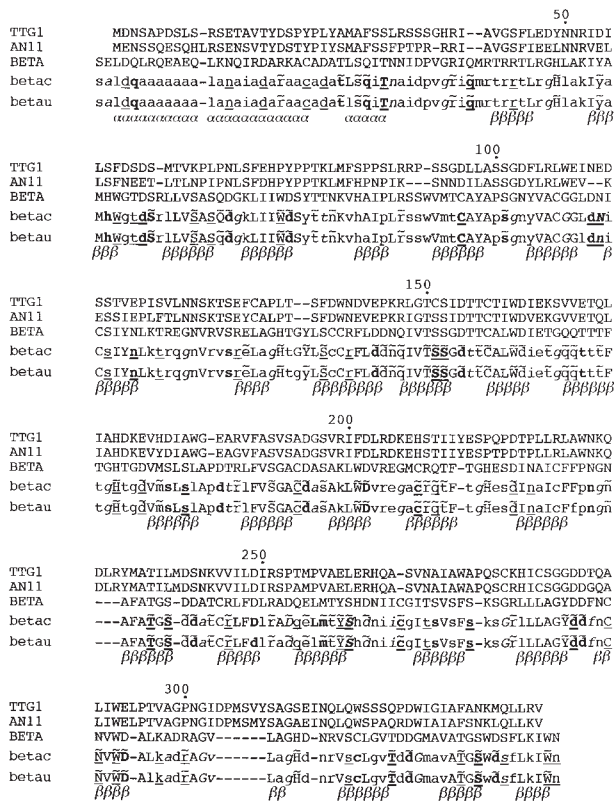
Mutations were identified in all *ttg1* alleles examined, and these are represented in Figure 6A and summarized in Table 1. The mutation in *ttg1-1* is a single C-to-T base transition that introduces a termination codon in place of the codon for glutamine 317. The resulting protein product is truncated, lacking the C-terminal 25 amino acid residues. Termination codons in place of codons for tryptophan 183 in *ttg1-19* and serine 310 in *ttg1-15*, *ttg1-16*, *ttg1-17*, and *ttg1-18* would also lead to truncated translation products. There are two mutations in *ttg1-20*: the same nonsense mutation as in *ttg1-15* and an additional C-to-G base transversion leading to serine 30 being changed to cysteine (Table 1). The only mutation that resulted from a single amino acid substitution was found in *ttg1-9*, in which a C-to-T base transition leads to a change from serine 282 to phenylalanine.

RNA gel blot analysis of total RNA extracted from flower buds of 6-week-old wild-type and mutant plants (Figure 6B) shows that all but one of the mutants described above contain transcripts of the *TTG1* gene. Because *ttg1-15*, *ttg1-16*, *ttg1-17*, and *ttg1-18* all have the same mutation, *ttg1-17* was selected to represent these lines. The lane containing RNA from the deletion mutant *ttg1-13* shows no hybridization band, confirming that the 1.5-kb transcript is a product of the *TTG1* gene.

## DISCUSSION

### The *TTG1* Locus Encodes a WD40 Repeat Protein

The isolation and analysis of the *TTG1* locus have revealed that *TTG1* is a WD40 repeat protein similar to An11 from *petunia* (de Vetten et al., 1997) and that it is not related to MYC transcription factors, as suggested by the results of Lloyd et al. (1992), which showed that constitutive expression of the maize *R* gene in *ttg1* mutants was able to complement the phenotype of *ttg1* mutants. The identification of *TTG1* as a WD40 repeat protein suggests that *TTG1* regulates MYC transcription factors or pathways in which MYC factors are involved.



**Figure 4.** Alignment of Amino Acid Sequences of TTG1, AN11, and the  $\beta$  Subunit of Heterotrimeric G Protein.

The structural features of G- $\beta$  in the complexed form (betac) with G- $\gamma$  and in a form without G- $\gamma$  (betau) are represented at individual residue positions. Because some of the side chains of G- $\beta$  are not well defined in electron density, they are modeled as alanine residues in the crystal structure (Sondek et al., 1996). The residue numbering of TTG1 is shown at the top of the alignment. A key to the structural features in the JOY format (Mizuguchi et al., 1988) is as follows: uppercase letters, the solvent is inaccessible; lowercase letters, the solvent is accessible; italic letters, positive  $\phi$ ; tilde above letters, hydrogen bond to side chain; boldface letters, hydrogen bond to main chain amide; underlined letters, hydrogen bond to main chain carbonyl. Dashes indicate gaps inserted to aid alignment.

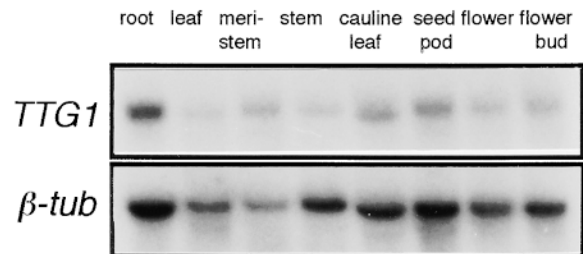
**Structure of WD40 Repeat Proteins**

Proteins containing WD40 repeats are involved in a number of different types of regulatory roles, such as signaling (e.g.,  $\beta$  subunits of heterotrimeric G proteins), cell cycle regulation (e.g., CDC20 and CDC4), transcriptional repression (e.g., yeast TUP1 and Drosophila extra sex combs), vesicular trafficking (e.g., SEC13), and RNA processing (reviewed in Neer et al., 1994). The G- $\beta$  subunits, which contain seven repeats of the WD40 motif, are the best characterized of the WD40 repeat proteins. The crystal structure of the G- $\beta\gamma$  complex shows that the G- $\beta$  subunit adopts a seven-bladed  $\beta$  propeller

fold (Sondek et al., 1996). Each blade is composed of four  $\beta$  strands, and an N-terminal amphipathic  $\alpha$  helix interacts closely with the G- $\gamma$  subunit, which is required for correct folding and function of the G- $\beta$  subunit. The structure of the G- $\beta\gamma$  complex is not significantly altered in the heterotrimeric G- $\alpha\beta\gamma$  protein (Lambright et al., 1996).

A typical WD repeat motif is characterized by GH and WD doublets, which are separated by a characteristic pattern of hydrophobic and hydrophilic residues. The sequence of TTG1 has only four clearcut WD repeat motifs, whereas G- $\beta$  has seven repeats. However, TTG1 may have seven structural repeats; several proteins that display no characteristic WD repeats or have no amino acid sequence similarity to G- $\beta$  have been observed to have a sevenfold symmetric  $\beta$ -propeller shape (Murzin, 1992). For example, the crystal structure of the terminal domain of the clathrin heavy chain shows a seven-bladed  $\beta$  propeller closely resembling G- $\beta$  but with no sequence similarity to WD40 proteins (ter Haar et al., 1998). Murzin (1992) also showed that no strong sequence constraints are necessary for proper  $\beta$ -sheet twist and packing to result in a sevenfold symmetric structure.

An alignment of structural elements of G- $\beta$  and TTG1 (Figure 4) aided in modeling the TTG1 protein. A three-dimensional model of TTG1 (Figure 7A) was generated on the basis of the crystal structure of the G- $\beta$  subunit by using the coordinates of the G- $\beta$  subunit crystal structure of the G- $\beta\gamma$  complex and using the suite of programs called COMPOSER (Blundell et al., 1987; Srinivasan and Blundell, 1993). Whereas the gross structure of TTG1 is likely to resemble that of the G- $\beta$  subunit, there will be some differences. The N-terminal helix of TTG1, if present, will probably differ in its orientation compared with the equivalent helix in the G- $\beta$  structure, which interacts with the helix of the G- $\gamma$  subunit. Deletion of a single residue in the middle of the putative helix in AN11 compared with TTG1 (Figure 4) supports the suggestion that the structure and orientation of this helical region could radically differ from that of the G- $\beta$  subunit. The first two blades do not match the WD40 consensus sequence well but could still fold to resemble propeller blades. How-



**Figure 5.** TTG1 Is Expressed in All Major Organs of Arabidopsis.

RNA gel blot analysis of total RNA (10  $\mu$ g) of different organs of 4-week-old Arabidopsis ecotype Columbia plants probed with the coding region of the TTG1 gene and with a  $\beta$ -tubulin ( $\beta$ -*tub*) probe.

**Table 1.** Mutations Identified in *ttg1* Mutant Alleles

Allele <sup>a</sup>	Origin	Parent	Phenotype <sup>b</sup>	Mutation in <i>TTG1</i>
<i>ttg1-1</i>	M. Koornneef	<i>Ler</i>	Glabra, transparent testa	Q-317→stop codon
<i>ttg1-9</i>	J. Larkin	Columbia	Similar to <i>ttg1-1</i>	S-282→F
<i>ttg1-13</i>	D. Oppenheimer	RLD1	Similar to <i>ttg1-1</i>	Deletion
<i>ttg1-15</i>	N300	Antwerp-1	<i>ttg1</i> -like, pale, branched	S-310→stop codon
<i>ttg1-16</i>	N319	Enkheim-1	<i>ttg1</i> -like, dwarf	S-310→stop codon
<i>ttg1-17</i>	N339	Enkheim-1	<i>ttg1</i> -like, pale	S-310→stop codon
<i>ttg1-18</i>	N372	Enkheim-1	<i>ttg1</i> -like, upright rosette	S-310→stop codon
— <sup>c</sup>	N373	Enkheim-1	<i>ttg1</i> -like, extreme dwarf	ND <sup>d</sup>
<i>ttg1-19</i>	N406	Enkheim-1	<i>ttg1</i> -like	W-183→stop codon
<i>ttg1-20</i>	N420	Enkheim-1	<i>ttg1</i> -like, early flowering	S-30→C, S-310→stop codon
— <sup>e</sup>	N447	Enkheim-1	<i>ttg1</i> -like, dwarf	Q-317→stop codon

<sup>a</sup> Allele numbers were assigned to mutants from the Kranz collection after checking for lack of anthocyanins in the seed coat and throughout the plant, reduced trichome number, and lack of seed mucilage and examining allelism by crossing to *ttg1-1*.

<sup>b</sup> The phenotype has been abbreviated to *ttg1*-like, where glabra and yellow seeds are features listed in the NASC catalog.

<sup>c</sup> Line N373 did not set seed and therefore was not used in crosses. This line was not investigated further.

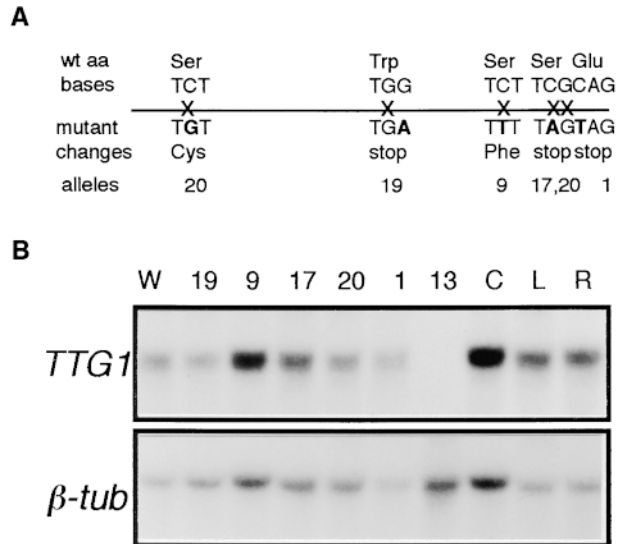
<sup>d</sup> ND, not determined.

<sup>e</sup> Line N447 was shown to be identical to the *ttg1-1* mutant in the *Ler* background by using sequence and gel blot analyses. Probes from the *TTG1* region and from elsewhere in the genome (data not shown) were used for these analyses.

ever, it is possible that the proline-rich region present at the beginning of the propeller fold could result in a disruption in the first two blades. These features, together with deviations in the periodicity of the GH and WD repeat motifs, suggest that the regularity of the repeats in the TTG1 structure is lower than that of the G-β subunit structure. The amino acid residues in the G-β subunit identified as interacting with G-α and G-γ subunits are not conserved in TTG1, suggesting that TTG1 does not interact with partners that are very similar to Gα and Gγ.

There are several sequences in the databases encoding putative proteins that show higher similarity to TTG1 than to the G-β subunit from either plants or animals. An Arabidopsis genomic sequence, located on chromosome 3 and called *atan11b* (de Vetten et al., 1997), could encode a protein showing 72% similarity to TTG1. An Arabidopsis EST called *atan11a* encodes a protein also with 72% similarity to TTG1, although the putative products of *atan11a* and *atan11b* are much more similar to each other (93% similarity). Two nematode genes arrayed in tandem (in cosmid F53C11) may encode proteins with greater similarity to TTG1 than proteins encoded by any locus from yeast, although a yeast gene (*S61017*) encodes a protein that is 60% similar to TTG1. Because TTG1 regulates pathways that involve cell differentiation in Arabidopsis, this type of function also may be required for the TTG1 homologs in other multicellular organisms. In yeast, the TTG1 homolog may be involved in regulating metabolic pathways, as it is in Arabidopsis.

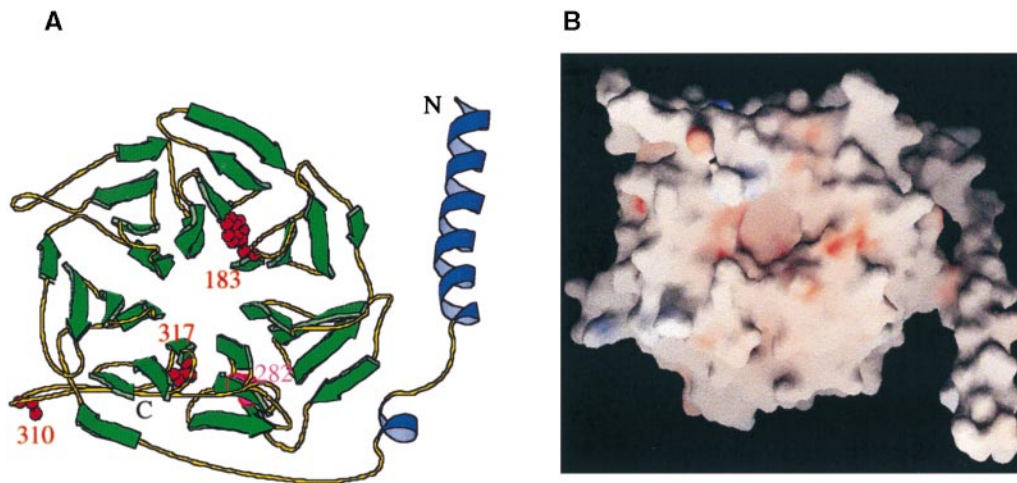
Genes encoding several WD40 proteins, including putative G-β subunits, the constitutive photomorphogenic COP1 protein, and retinoblastoma binding-like proteins, have been isolated from plants (reviewed in Ma, 1994; Ach et al., 1997).



**Figure 6.** Analysis of *ttg1* Mutations.

**(A)** Scheme of the coding region of the *TTG1* gene with the position of mutations (X) indicated. The wild-type (wt) codon is given above together with the amino acid (aa). The base changes in the mutants are in boldface, and the amino acid sequence changes are shown below. The *ttg1* allele numbers are shown below. The stop codon in *ttg1-17* is also present in *ttg1-15*, *ttg1-16*, *ttg1-18*, and *ttg1-20*.

**(B)** RNA gel blot of total RNA (10 μg) isolated from flower buds of mutant alleles probed with the coding region of the *TTG1* gene and the β-tubulin (*β-tub*) gene. Parental ecotypes Ws (W), Columbia (C), *Ler* (L), and RLD1 (R) and mutant alleles with identified mutations are shown.



**Figure 7.** Model of the TTG1 Structure Based on the Crystal Structure of the G $\beta$  Subunit.

**(A)** Ribbon representation of the TTG1 protein. The  $\alpha$  helices,  $\beta$  strands, and loops are shown in blue, green, and yellow, respectively. The positions of the mutations at residue 183 in *ttg1-19*, residue 310 in *ttg1-15*, *ttg1-16*, *ttg1-17*, *ttg1-18*, and *ttg1-20*, and residue 317 in *ttg1-1* that all lead to truncations are shown in red, whereas serine-282, which is changed to phenylalanine in *ttg1-9*, is indicated in pink. The side chains of these four amino acid residues are represented in the ball-and-stick model in red. The N and C termini are labeled. This diagram was produced using the program MOLSCRIPT (Kraulis, 1991).

**(B)** The electrostatic potential surface of the model of TTG1 seen from the opposite side diagrammed in **(A)** and showing the acidic patch at the center. The acidic and basic clusters are shown in red and blue, respectively. The image was produced using the program GRASP (Nicholls et al., 1991).

The COP1 protein from *Arabidopsis* contains three domains and acts as part of a protein complex negatively regulating gene transcription. The WD40 region of COP1 has been shown to be essential for the function of this protein (McNellis et al., 1994) as a negative regulator of photomorphogenesis. The interaction with other proteins in the complex is probably through the WD40 repeat domain. The *PRL1* gene of *Arabidopsis* encodes a WD40 repeat protein involved in the repression of glucose-responsive genes (Németh et al., 1998). A mutation at this locus has pleiotropic effects, including short roots and ectopic root hairs, and altered responses to sucrose, plant growth factors, and temperature. However, sequence comparisons between TTG1 and other plant proteins such as COP1 and PRL1 show that TTG1 is more similar to G- $\beta$  subunits than to other classes of WD proteins.

#### TTG1 and AN11 May Regulate Different Transcription Factors

TTG1 is most similar to the petunia AN11 protein encoded by the *An11* locus, which was isolated by transposon tagging (de Vetten et al., 1997). AN11 controls anthocyanin pigmentation only in flowers; mutations at this locus do not result in changes to trichomes or to anthocyanins in the rest of the plant. AN11 controls anthocyanin biosynthesis in flowers, most probably by regulating AN2, a MYB transcrip-

tion factor similar to C1 from maize. AN2 can complement the *an11* mutation in a transient assay in mutant flowers (de Vetten et al., 1997). In petunia, the transcription factors AN2 and JAF13 or AN1, MYC transcription factors similar to R from maize, are thought to act together in a manner similar to C1 and R to regulate positively the anthocyanin biosynthetic pathway (reviewed in Mol et al., 1996). The regulation of a MYB transcription factor by AN11 in petunia appears to be contrary to the evidence that TTG1 might regulate a MYC transcription factor, based on the expression in *Arabidopsis* of the maize *R* gene (Lloyd et al., 1992). One possible explanation for these apparently conflicting results is that TTG1 regulates the MYB-MYC complex (rather than one component), which then activates transcription. Expression of the heterologous maize *R* gene in *Arabidopsis* might bypass the normal requirement for activation of endogenous transcription factors by TTG1. GL1, the MYB transcription factor of *Arabidopsis* involved in trichome initiation, does not play a role in other pathways in which TTG1 is involved, whereas one MYC factor may function in many pathways. This suggests that other MYB factors might act in conjunction with the R-like MYC to regulate the anthocyanin biosynthesis and root hair initiation pathways.

Dissimilarity between TTG1 and AN11 in the N-terminal region (Figure 4) might possibly result in the regulation of different components, with TTG1 regulating MYC and AN11 regulating MYB factors. However, this seems unlikely because



functional assays using yeast and human homologs of AN11 (de Vetten et al., 1997) suggest that TTG1 could replace AN11 in petunia and that AN11 might be capable of replacing TTG1 function in Arabidopsis. The identification of two WD40 repeat proteins, which regulate anthocyanin biosynthesis and, in the case of the TTG1 protein, several other pathways, suggests that this class of protein may be involved in regulating biosynthetic pathways in other organisms.

From studies of other WD40 proteins, it seems probable that the TTG1 protein does not act directly as a transcription factor but binds to other proteins to promote the initiation of trichomes in leaves and stems. Figure 7B shows a model of the reverse side of TTG1, with a large acidic patch (shown in red) that could be involved in interactions with other proteins. The TTG1 protein might possibly act as part of a DNA binding complex to regulate transcription. However, the TTG1 sequence contains no recognizable nuclear localization signal, although, due to the difficulty in recognizing such signals, a cryptic site might be present. It is also possible that another protein is required to form a complex for nuclear import, as is the case with APETALA3 and PISTILLATA from Arabidopsis (McGonigle et al., 1996). Alternatively, the TTG1 protein may be located only in the cytoplasm, and it may act as part of a signal transduction pathway to downstream transcription factors. Preliminary investigations using  $\beta$ -glucuronidase-TTG1 fusion proteins (data not shown) suggest that TTG1 may be located in the cytoplasm. This is in agreement with the results of a study showing the location of AN11 (de Vetten et al., 1997). In this study, gel blot analysis with proteins of cell fractions showed that AN11 was primarily located in the cytoplasm. Currently, we are generating TTG1-green fluorescent protein fusions to study subcellular localization in vivo under different conditions, including stress, in which activation of the anthocyanin pathway can be controlled.

### Cloning of the *TTG1* Locus

The positional cloning of the *TTG1* locus has provided information about the order of and distances between a number of RFLP markers that is complementary to the data given in the physical maps of the region (Schmidt et al., 1997; Thorlby et al., 1997). Within a 10-cM region, 439 recombinants were analyzed, and the line Dennis contained the only crossover close to the *TTG1* locus. This suggests that recombination rates are reduced close to this gene. Recombination frequencies are known to vary along chromosomes in many species (reviewed in Lichten and Goldman, 1995).

A large number of mutant alleles at the *TTG1* locus have been identified over the years. These alleles originally suggested that all of the phenotypes of the mutant were associated with one locus (Koorneef, 1981). We have shown that single base changes within the coding region of the *TTG1* gene are correlated with mutant phenotypes, verifying Koorneef's postulate. The mutations mainly result from sin-

gle base changes near the 3' end of the coding region, indicating that the C-terminal end of the protein is important for proper functioning of the protein. The *ttg1-1* mutation, which has been used most extensively in experiments pertaining to the *TTG1* locus (e.g., Galway et al., 1994), results in C-terminal truncation of 25 amino acid residues. Plants carrying the homozygous *ttg1-1* mutation have a severe phenotype, with very few trichomes on leaves and no anthocyanin in the testa or plant, indicating that this C-terminal region is vital for the structure and function of TTG1.

The *ttg1-9* mutation is the result of a single base change resulting in the replacement of a serine residue at position 282 with phenylalanine. Serine is conserved at this position in AN11 and in the proteins encoded by *atan11b* and the yeast gene and is replaced by threonine in the proteins encoded by *atan11a* and the two nematode genes, suggesting that a hydroxyl group at this position is important for the structure or function of TTG1 and related proteins. The position of this amino acid residue is marked on the ribbon model (Figure 7A). This residue is positioned by the modeling on the surface of the outer edge of the protein and may be involved in interacting with other proteins as well as maintaining the tertiary structure of TTG1.

The high degree of conservation between TTG1, AN11, and other proteins from nematodes, yeast, and humans (de Vetten et al., 1997) and the array of effects of the *ttg1* mutations together suggest that this family of WD40 repeat proteins plays important and varied roles in many organisms.

## METHODS

### Plant Strains and Growth Conditions

*Arabidopsis thaliana* ecotypes Landsberg *erecta* (*Ler*), Columbia, Rschew (RLD1), and Wassilewskija (*Ws*) were supplied by the Nottingham (UK) Arabidopsis Stock Centre (NASC) together with the *ttg1-1* mutant. The *Ler* line carrying *ms1* and *ttg1-1* was obtained from C. Somerville (Carnegie Institute of Washington, Stanford, CA). The line containing *ttg1-1*, *ga3*, and *ch5* in the *Ler* background was a gift from M. Koorneef (University of Wageningen, Wageningen, The Netherlands). Mutants *ttg1-9* and *ttg1-13* were obtained from J. Larkin (Louisiana State University, Baton Rouge, LA) and D. Oppenheimer (University of Alabama, Tuscaloosa, AL), respectively. Possible *ttg1* alleles were selected by their description (glabrous and yellow seed) from lines in the Kranz collection held at the NASC.

For transformation, five to 10 plants were grown in a peat-based soil mixture of Levington (Ipswich, UK) M3 potting compost-vermiculite-perlite (3:2:1 [v]) in 4-inch pots in a greenhouse at 22°C with supplementary lighting. The soil was covered with muslin, and the seeds, which were mixed with sand, were sprinkled on top. Plantlets were thinned to  $\leq 10$  per pot at 2 to 3 weeks and allowed to grow until flowering bolts were  $\sim 10$  cm tall. Bolts were cut back once, and new bolts were allowed to emerge for several days before infiltration. For crosses and seed production, Arabidopsis was grown in the Arabicon system (Beta Tech, Ghent, Belgium) with 16 hr of light (100  $\mu\text{mol}$  of photons  $\text{m}^{-2} \text{sec}^{-1}$ ) and 8 hr of dark in a growth chamber at

22°C. Plants for DNA extraction were grown in trays with 12 hr of light (100  $\mu\text{mol}$  of photons  $\text{m}^{-2}$   $\text{sec}^{-1}$ ) followed by 12 hr of dark in a growth chamber at 22°C to delay the onset of flowering and obtain more leaf material. All plants received a weekly feed of macronutrients in the form of Tomorite (Levington).

To obtain root material for RNA gel blot analysis, plants were grown for 21 days on 1.2% phytigel (Sigma) containing Murashige and Skoog salts (2.16  $\text{g L}^{-1}$ ; Sigma) with the plates in a nearly vertical position in a growth room as above with 16 hr of light per day.

### Generation of Recombinant Lines

Lines that were generated in a cross between *Ler* carrying *ttg1* and *ms1* and the *Ws* ecotypes and covering a 6-centimorgan (cM) region were selected and analyzed. Lines resulting from recombination events on the proximal side of *ttg1-1* covering a 4-cM interval between *ttg1-1* and *GA3* were selected from a cross of *Ler*, carrying the mutations *ttg1-1*, *ga3*, and *ch5*, and the *RLD1* ecotype.

### Restriction Fragment Length Polymorphism Analysis

DNA was isolated from 4-week-old plant material as described by Dellaporta et al. (1983) followed by CsCl banding to remove RNA and polysaccharides (Walker et al., 1997). The purity of the DNA enabled the detection of small mobility differences of restriction fragments on DNA gel blots. Restriction enzyme digestion of DNA, gel electrophoresis, blotting, and hybridization techniques were as described in Walker et al. (1997).

A variety of probes detecting restriction fragment length polymorphisms (RFLPs) between these pairs of parental lines were used, including cosmid and bacteriophage  $\lambda$  constructs already mapped to this region by Nam et al. (1989) and Chang et al. (1988). End probes from yeast artificial chromosomes (YACs) (Ward and Jen, 1990) that had been mapped between *ms1* and *ga3* (Chapple et al., 1996; Schmidt et al., 1997; Thorlby et al., 1997) were generated by plasmid end rescue and inverse polymerase chain reaction (PCR) to map the location of *ttg1*. Probe AF3, a cDNA mapped to this region in other experiments (Chow et al., 1998), and CR12, a genomic fragment in a  $\lambda$  replacement vector (A.R. Walker, unpublished data), were also used.

Probes were labeled using random primer extension with  $\alpha$ - $^{32}\text{P}$ -dATP, as discussed in Walker et al. (1997). Yeast chromosome preparations (prepared essentially as given in Carle and Olson, 1985) were electrophoresed in 1% low-melting-point agarose (Seaplaque; Flowgen, Lichfield, UK), using a Beckman (Palo Alto, CA) Geneline II pulsed-field electrophoresis system to facilitate recovery of whole YACs for subcloning or labeling for probes used in library screens. YAC end probes were generated by plasmid end rescue and inverse PCR (Bancroft et al., 1992).

### Library Screening

A genomic library in  $\lambda$  DashII (Stratagene, La Jolla, CA) from ecotype *Ler* (Boyce et al., 1994) distributed by the EEC-BRIDGE Arabidopsis DNA Stock Centre (Cologne, Germany) was screened with YAC EG20H2 and cosmid g4556 by using standard techniques (Sambrook et al., 1989). The isolated  $\lambda$  clones were ordered by restriction pattern and

hybridization to fragments of DNA from the region, and the genomic inserts of some were ligated into the NotI site of the pBINNOT vector for transformation of plants.

### Construction of pBINNOT

The pBIN19 vector (Bevan, 1984) was modified to contain a NotI site in the polylinker. To remove an unwanted NotI site, pBIN19 was digested with NotI, treated with the Klenow fragment of DNA polymerase I (Boehringer Mannheim) and deoxynucleotide triphosphates to fill in the site, ligated in a large volume, and digested again with NotI. *Escherichia coli* XL1-Blue (Stratagene) was transformed with the ligation product using electroporation (Dower et al., 1988). Plasmid DNA was isolated from kanamycin-resistant colonies and digested with NotI to check that the original NotI site had been removed. The vector was digested with XbaI and Asp718I, and two annealed oligonucleotides (N1, 5'-GTACCGCGGCCGCAT-3'; and N2, 5'-CTAGATGCGGCCGCG-3') containing a NotI site were ligated into the vector to reconstitute the XbaI and Asp718I sites in the polylinker. The ligated DNA was digested with BamHI to remove parental molecules, and pBINNOT was purified after transformation of *E. coli* XL1-Blue.

### Sequence Analysis

Sequencing of *TTG1* was performed using an ABI 377 sequencer (Perkin-Elmer, Warrington, UK) from subclones of the construct pB8 in pUC18 and pBluescript SK+ (Stratagene). The sequence was completed with two primers designed from a previously obtained sequence. Sequence analysis was conducted at the Human Genome Mapping Project Resource Centre (Hinxton, UK), using the software package of the Genetics Computer Group (Madison, WI); BLAST (Altschul et al., 1990) and FASTA (Pearson and Lipman, 1988) searches were conducted to find similar sequences.

PCR products from the region of the *TTG1* gene were generated with primers designed to give overlapping fragments of  $\sim 700$  bp by using genomic DNA from different *ttg1* mutants and their parental ecotypes as the template and a mixture (10:1) of BioTaq (Biolone, London, UK) and Pfu (Stratagene) polymerases. The PCR products were separated by electrophoresis on a 1.2% agarose gel in  $1 \times$  TAE buffer (Sambrook et al., 1989). The relevant bands were excised, and the DNA was purified using Qiaex II gel kit (Qiagen, Chatsworth, CA) and then sequenced using the PCR primers.

### RNA Analysis

Total RNA from different organs of 4-week-old soil-grown Columbia plants was extracted from 200 mg of tissue using Tripure (Boehringer Mannheim). Roots were gathered from plants grown on vertical agar plates, and RNA was extracted using the same technique. Floral buds of soil-grown plants were used for the mutant analysis. For RNA gel blots, 10  $\mu\text{g}$  of total RNA was fractionated by electrophoresis on a formaldehyde-1.2% agarose gel and blotted onto Gene-screen Plus (Du Pont), according to the manufacturer's instructions. *TTG1* probes were prepared as described by Walker et al. (1997), and hybridization was performed at 42°C in 50% formamide for 16 hr. Blots were washed in  $0.1 \times$  SSC ( $1 \times$  SSC is 0.15 M NaCl and 0.015 M sodium citrate) and 2% SDS at 65°C for 15 min. After exposure to film, blots were stripped by boiling in  $0.1 \times$  SSC and 2%

SDS, and hybridized with a  $\beta$ -tubulin sequence (Z25960; 3' untranslated region of the *TUB5* gene) as a loading control.

Primer extension was conducted as given in Sambrook et al. (1989). An oligonucleotide primer designed to the complementary strand of the *TTG1* sequence starting at base 54 shown in Figure 3 (5'-GTATGTGACGGCGTTTCCGATCTGG-3'; 400 ng) was end labeled using 50  $\mu$ Ci  $\gamma$ -<sup>32</sup>P-ATP (3000 Ci mmol<sup>-1</sup>) (Amersham) and polynucleotide kinase (Boehringer Mannheim). The primer (32 ng) was annealed to 20  $\mu$ g of total RNA extracted from Columbia flower buds and was extended using Moloney murine leukemia virus reverse transcriptase (Promega). RNA from *ttg1-13* plants was used as a negative control.

### Plant Transformation

*Agrobacterium tumefaciens* Agl1 (Lazo et al., 1991) was transformed with constructs in the pBINNOT vector by electroporation (Shen and Forde, 1989). Root transformation of *ttg1-1* material with constructs pB1.1A, pB8, and pBX6 was conducted essentially as described by Valvekens et al. (1988). By using vacuum infiltration (Bechtold et al., 1993), genomic fragments were also introduced into *ttg1-9* plants. Plants were covered for 2 days to maintain humidity, then grown for 4 weeks to produce as much seed as possible. Mutant plants containing *ttg1-10* in a WS background did not transform efficiently. Selection of transformants from seed of plants subjected to vacuum infiltration was conducted on 0.8% agar containing Murashige and Skoog salts (2.16 g L<sup>-1</sup>) and kanamycin (35  $\mu$ g mL<sup>-1</sup>) for 10 days. Kanamycin-resistant seedlings were moved onto fresh kanamycin plates for 10 days and then transferred to soil to set seed.

### Three-Dimensional Modeling

A search in the Brookhaven protein data bank (Bernstein et al., 1977) for sequences similar to that of TTG1 identified the G- $\beta$  subunit of G proteins as the most similar protein. Whereas the G- $\beta$  subunit has typical WD repeats, the percentage of sequence identity between TTG1 and G- $\beta$  subunits is low (~18%), and this resulted in ambiguity in some regions of the alignment. Hence, an alignment of all known WD repeat proteins, including TTG1, was made (data not shown), and these data were used to make the alignment shown in Figure 4. The structure of the G- $\beta$  subunit (Wall et al., 1995; Lambright et al., 1996; Sondek et al., 1996) was used as the basis to construct a model of TTG1.

The suite of programs encoded in COMPOSER (Blundell et al., 1987; Srinivasan and Blundell, 1993) and incorporated into SYBYL (Tripos Inc., St. Louis, MO) was used to generate a three-dimensional model of TTG1 (Figure 7). The structures of most of the conserved regions of G- $\beta$  proteins have been extrapolated to the equivalent regions of TTG1. The variable regions in the alignment, which often involved the insertion or deletion of stretches of amino acids, were modeled by identifying a suitable segment from a known structure in the data bank. A search was made for segments having the desired number of residues and the proper end-to-end distances across the three "anchor" C- $\alpha$  atoms at either side of the putative loop, so that the segment could be fitted to the contiguous conserved regions. Side chains were modeled either by extrapolating from the equivalent positions in the basic structure, where appropriate, or by using rules derived from the analysis of known protein structures (Sutcliffe et al., 1987). Energy minimization of the model helped to rectify the

inconsistencies in the stereochemistry and relieved the short contacts between nonbonded atoms.

### ACKNOWLEDGMENTS

We thank Maarten Koornneef, David Oppenheimer, Chris Somerville, and John Larkin for gifts of seeds; Zoe Wilson and Renate Schmidt for sharing results of chromosome walks; and George Coupland, Ian Furner, and Diane Hatton for helpful discussions. Technical assistance was provided by Martyn Seekings, Sue Sanderson, and Rebecca Rochford. We thank members of the Cambridge Centre for Molecular Recognition for oligonucleotide synthesis and DNA sequencing. Computing facilities were provided by the Human Genome Mapping Project (Cambridge, UK). Libraries were supplied by the EEC-BRIDGE Arabidopsis DNA Stock Centre, and some seed stocks came from the NASC; the EST was obtained from the Arabidopsis Biological Resource Center (Ohio State University, Columbus). This work was funded by research grants from the Agricultural and Food Research Council (Grant Nos. PMB00535 and PMB00658) and the Biotechnology and Biological Sciences Research Council (BBSRC) (Grant No. GER00618) to J.C.G., and a BBSRC Research Studentship to A.C.B.-W. T.L.B. and N.S. thank the Wellcome Trust for support. Some aspects of the work were initiated by J.J.E. and M.D.M. and sponsored by the National Science Foundation (Grant No. IBN 9506192 to M.D.M.).

Received January 20, 1999; accepted April 9, 1999.

### REFERENCES

- Ach, R.A., Taranto, P., and Grissem, W. (1997). A conserved family of WD-40 proteins binds to the retinoblastoma protein in both plants and animals. *Plant Cell* **9**, 1595-1606.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.
- Bancroft, I., Westphal, L., Schmidt, R., and Dean, C. (1992). PFGE-resolved RFLP analysis and long-range restriction mapping of the DNA of *Arabidopsis thaliana* using whole YAC clones as probes. *Nucleic Acids Res.* **20**, 6201-6207.
- Bechtold, N., Ellis, J., and Pelletier, G. (1993). In planta *Agrobacterium*-mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Ser. III Sci. Vie* **316**, 1194-1199.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Beyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The protein data bank: A computer archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
- Bevan, M. (1984). Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* **12**, 8711-8721.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E., and Thornton, J.M. (1987). Knowledge-based prediction of protein structures and design of novel molecules. *Nature* **326**, 347-351.
- Borodovsky, M., and McIninch, J. (1993). Genmark—Parallelgene recognition for both DNA strands. *Comput. Chem.* **17**, 123-133.

- Boyce, J.M., Coates, D., Fricker, M.D., and Evans, D.E. (1994). Genomic sequence of a calnexin homolog from *Arabidopsis thaliana*. *Plant Physiol.* **106**, 1691.
- Carle, G.F., and Olson, M.V. (1985). An electrophoretic karyotype of yeast. *Proc. Natl. Acad. Sci. USA* **82**, 3756–3760.
- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., and Meyerowitz, E.M. (1988). Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **85**, 6856–6860.
- Chapple, R.M., Chaudhury, A.M., Blomer, K.B., Farrell, L.B., and Dennis, E.S. (1996). Construction of a YAC contig of 2 megabases around the *MS1* gene in *Arabidopsis thaliana*. *Aust. J. Plant Physiol.* **23**, 453–465.
- Chow, K.S., Singh, D.P., Walker, A.R., and Smith, A.G. (1998). Two different genes encode ferrochelatase in *Arabidopsis*: Mapping, expression and subcellular targeting of the precursor proteins. *Plant J.* **15**, 531–541.
- Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983). A plant DNA miniprep: Version II. *Plant Mol. Biol. Rep.* **1**, 19–21.
- de Vetten, N., Quattrocchio, F., Mol, J., and Koes, R. (1997). The *an11* locus controlling flower pigmentation in petunia encodes a novel WD-repeat protein conserved in yeast, plants, and animals. *Genes Dev.* **11**, 1422–1434.
- Di Cristina, M., Sessa, G., Dolan, L., Linstead, P., Baima, S., Ruberti, I., and Morelli, G. (1996). The *Arabidopsis* Athb-10 (*GLABRA2*) is an HD-Zip protein required for regulation of root hair development. *Plant J.* **10**, 393–402.
- Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988). High-efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127–6145.
- Galway, M.E., Masucci, J.D., Lloyd, A.M., Walbot, V., Davis, R.W., and Schiefelbein, J.W. (1994). The *TTG* gene is required to specify epidermal cell fate and cell patterning in the *Arabidopsis* root. *Dev. Biol.* **166**, 740–754.
- Goff, S.A., Cone, K.L., and Chandler, V.L. (1992). Functional analysis of the transcriptional activator encoded by the maize *B* gene: Evidence for a direct functional interaction between two classes of regulatory proteins. *Genes Dev.* **6**, 864–875.
- Goodrich, J., Carpenter, R., and Coen, E.S. (1992). A common gene regulates pigmentation pattern in diverse plant species. *Cell* **68**, 955–964.
- Grill, E., and Somerville, C. (1991). Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Mol. Gen. Genet.* **226**, 484–490.
- Hebsgaard, S.M., Korning, P.G., Tolstrup, N., Engelbrecht, J., Rouzé, P., and Brunak, S. (1996). Splice-site predictions in *Arabidopsis thaliana* pre-messenger-RNA by combining local and global sequence information. *Nucleic Acids Res.* **24**, 3439–3452.
- Hülkamp, M., Miséra, S., and Jürgens, G. (1994). Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* **76**, 555–566.
- Hung, C.Y., Lin, Y., Zhang, M., Pollock, S., Marks, M.D., and Schiefelbein, J. (1998). A common position-dependent mechanism controls cell-type patterning and *GLABRA2* regulation in the root and hypocotyl epidermis of *Arabidopsis*. *Plant Physiol.* **117**, 73–84.
- Koornneef, M. (1981). The complex syndrome of *ttg* mutants. *Arabidopsis Inf. Serv.* **18**, 45–51.
- Koornneef, M., Dellaert, S.W.M., and van der Veen, J.H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L) Heynh. *Mutat. Res.* **93**, 109–123.
- Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J., and Feenstra, W.J. (1983). The linkage map of *Arabidopsis thaliana*. *J. Hered.* **74**, 265–272.
- Kraulis, P.J. (1991). Molscript—A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
- Lambright, D.G., Sondek, J., Bohm, A., Skiba, N.P., Hamm, H.E., and Sigler, P.B. (1996). The 2.0 Å crystal structure of a heterotrimeric G protein. *Nature* **379**, 311–319.
- Larkin, J.C., Oppenheimer, D.G., Lloyd, A.M., Paparozzi, E.T., and Marks, M.D. (1994). Roles of the *GLABROUS1* and *TRANSPARENT TESTA GLABRA* genes in *Arabidopsis* trichome development. *Plant Cell* **6**, 1065–1076.
- Lazo, G.R., Stein, P.A., and Ludwig, R.A. (1991). A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology* **9**, 963–967.
- Léon-Kloosterziel, K.M., Keijzer, C.J., and Koornneef, M. (1994). A seed shape mutant of *Arabidopsis* that is affected in integument development. *Plant Cell* **6**, 385–392.
- Lesnick, M.L., and Chandler, V.L. (1998). Activation of the maize anthocyanin gene *a2* is mediated by an element conserved in many anthocyanin promoters. *Plant Physiol.* **117**, 437–445.
- Lichten, M., and Goldman, A.S. (1995). Meiotic recombination hotspots. *Annu. Rev. Genet.* **29**, 445–476.
- Lloyd, A.M., Walbot, V., and Davis, R.W. (1992). Anthocyanin production in dicots activated by maize anthocyanin-specific regulators, *R* and *C1*. *Science* **258**, 1773–1775.
- Ludwig, S.R., Habera, L.F., Dellaporta, S.L., and Wessler, S.R. (1989). *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcription activators and contains the *myc*-homology region. *Proc. Natl. Acad. Sci. USA* **86**, 7092–7096.
- Ma, H. (1994). GTP-binding proteins in plants: New members of an old family. *Plant Mol. Biol.* **26**, 1611–1636.
- Marks, M.D. (1997). Molecular genetic analysis of trichome development in *Arabidopsis*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 137–163.
- Martin, C., Prescott, A., Mackay, S., Bartlett, J., and Vrijlandt, E. (1991). Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. *Plant J.* **1**, 37–49.
- Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D., and Schiefelbein, J.W. (1996). The homeobox gene *GLABRA2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development* **122**, 1253–1260.
- McGonigle, B., Bouhidel, K., and Irish, V.F. (1996). Nuclear localization of the *Arabidopsis* *APETALA3* and *PISTILLATA* homeotic gene products depends on their simultaneous expression. *Genes Dev.* **10**, 1812–1821.
- McNellis, T.W., von Arnim, A.G., Araki, T., Komeda, Y., Miséra, S., and Deng, X.-W. (1994). Genetic and molecular analysis of an allelic series of *cop1* mutants suggests functional roles for the multiple protein domains. *Plant Cell* **6**, 487–500.

- Mizuguchi, K., Deane, C.M., Johnson, M.S., Blundell, T.L., and Overington, J.P. (1988). JOY: Protein sequence-structure representation and analysis. *Bioinformatics* **14**, 617-623.
- Mol, J., Jenkins, G., Schafer, E., and Weiss, D. (1996). Signal perception, transduction, and gene expression involved in anthocyanin biosynthesis. *Crit. Rev. Plant Sci.* **15**, 525-557.
- Murzin, A.G. (1992). Structural principles for the propeller assembly of  $\beta$ -sheets: The preference for seven-fold symmetry. *Proteins* **14**, 191-201.
- Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M., and Goodman, H.M. (1989). Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell* **1**, 699-705.
- Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory protein family of WD-repeat proteins. *Nature* **371**, 297-300.
- Németh, K., Salchert, K., Putnoky, P., Bhalerao, R., Koncz-Kálmán, Z., Stankovic-Stangeland, B., Bakó, L., Mathur, J., Ókrész, L., Stabel, S., Geigenberger, P., Stitt, M., Rédei, G.P., Schell, J., and Koncz, C. (1998). Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in *Arabidopsis*. *Genes Dev.* **12**, 3059-3073.
- Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association—Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
- Oppenheimer, D.G., Herman, P.L., Sivakumaran, S., Esch, J., and Marks, M.D. (1991). A *myb* gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* **67**, 483-493.
- Pearson, W.R., and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.
- Pelletier, M.K., and Shirley, B.W. (1996). Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings. *Plant Physiol.* **111**, 339-345.
- Rerie, W.G., Feldmann, K.A., and Marks, M.D. (1994). The *GLABRA2* gene encodes a homeodomain protein required for normal trichome development in *Arabidopsis*. *Genes Dev.* **8**, 1388-1389.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).
- Schmidt, R., Love, K., West, J., Lenehan, Z., and Dean, C. (1997). Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. *Plant J.* **11**, 563-572.
- Shen, W.-J., and Forde, B.G. (1989). Efficient transformation of *Agrobacterium* sp. by high voltage electroporation. *Nucleic Acids Res.* **17**, 8385.
- Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* **8**, 659-671.
- Sondek, J., Bohm, A., Lambright, D.G., Hamm, H.E., and Sigler, P.B. (1996). Crystal structure of a  $G_A$  protein  $\beta\gamma$  dimer at 2.1 Å resolution. *Nature* **379**, 369-374.
- Srinivasan, N., and Blundell, T.L. (1993). An evaluation of the performance of an automated procedure for the comparative modeling of protein tertiary structure. *Protein Eng.* **6**, 501-512.
- Sutcliffe, M.J., Hayes, F.R.F., and Blundell, T.L. (1987). Knowledge-based modeling of homologous proteins. 2. Rules for the conformations of substituted side chains. *Protein Eng.* **1**, 385-392.
- Szymanski, D.B., Jilk, R.A., Pollock, S.M., and Marks, M.D. (1998). Control of *GL2* expression in *Arabidopsis* leaves and trichomes. *Development* **125**, 1161-1171.
- ter Haar, E., Musacchio, A., Harrison, S.C., and Kirchhausen, T. (1998). Atomic structure of clathrin: A  $\beta$  propeller terminal domain joins an  $\alpha$  zigzag linker. *Cell* **95**, 563-573.
- Thorlby, G.L., Shlumukov, L., Vizir, I.Y., Yang, C.Y., Mulligan, B.J., and Wilson, Z.A. (1997). Fine-scale molecular genetic (RFLP) and physical mapping of a 8.9-cM region on the top arm of *Arabidopsis* chromosome 5 encompassing the male sterility gene, *ms1*. *Plant J.* **12**, 471-479.
- Valvekens, D., van Montagu, M., and van Lijsebettens, M. (1988). *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. USA* **85**, 5536-5540.
- Wada, T., Tachibana, T., Shimura, Y., and Okada, K. (1997). Epidermal cell differentiation in *Arabidopsis* determined by a *Myb* homolog, *CPC*. *Science* **277**, 1113-1116.
- Walker, A.R., Cole, T.E., Torney, K., and Gray, J.C. (1997). Chromosomal locations of five genes encoding photosynthetic electron transfer proteins in *Arabidopsis thaliana*. *Photosynth. Res.* **54**, 155-163.
- Wall, M.A., Coleman, D.E., Lee, E., Iñiguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. (1995). The structure of G-protein heterodimer G- $\alpha$ 1 $\beta$ 1 $\gamma$ 2. *Cell* **83**, 1047-1058.
- Ward, E.R., and Jen, G.C. (1990). Isolation of single-copy sequence clones from a yeast artificial chromosome library of randomly sheared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.* **14**, 561-568.

**The *TRANSPARENT TESTA GLABRA1* Locus, Which Regulates Trichome Differentiation and Anthocyanin Biosynthesis in Arabidopsis, Encodes a WD40 Repeat Protein**

Amanda R. Walker, Paul A. Davison, Agnese C. Bolognesi-Winfield, Celia M. James, N. Srinivasan, Tom L. Blundell, Jeffrey J. Esch, M. David Marks and John C. Gray

*Plant Cell* 1999;11;1337-1349

DOI 10.1105/tpc.11.7.1337

This information is current as of November 26, 2020

<b>References</b>	This article cites 63 articles, 22 of which can be accessed free at: <a href="/content/11/7/1337.full.html#ref-list-1">/content/11/7/1337.full.html#ref-list-1</a>
<b>Permissions</b>	<a href="https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X">https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X</a>
<b>eTOCs</b>	Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>CiteTrack Alerts</b>	Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>Subscription Information</b>	Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a>