Fine-Scale Mapping and Molecular Genetic Analysis of the *Transparent Testa 9* Locus of *Arabidopsis thaliana*

by
Marguerite Indriati Hood

Submitted in partial fulfillment of the requirements for Major Honors in Biology

Houghton College, Houghton, NY
May 2006

Honors Committee

Dr. Matthew K. Pelletier, Chair

Dr. Jacqueline Crisman, Reader

Dr. Irmgard Howard, Reader

Acknowledgements

I would like to express sincere gratitude and appreciation to my professor and chair of my honors committee, Dr. Matthew Pelletier, for providing me with this opportunity and challenge, and allowing me room to grow. His insight, advice and honest evaluation have been instrumental in guiding my way in this endeavor. I might have given up had it not been for his tremendous support and encouragement and his own example of patience and dedication. I would like to thank Dr. Jacqueline Crisman and Dr. Irmgard Howard for their support and willingness to serve on my honors committee. I would also like to thank Dr. Patricia Conklin of SUNY Cortland for providing the necessary *Agrobacterium* strain as well as the many students whose contributions to this work have been invaluable over the years.

Table of Contents

Section		Page Number
I.	Abstract	1
II.	Introduction	1
III.	Literature Review	
	a. Flavonoid Biosynthesis	8
	b. Characterization of Flavonoid Mutants in *Arabidopsis thaliana*	15
	c. *Transparent Testa 9*	24
	d. Fine-Scale Genetic Mapping	27
	e. Methods for the Evaluation of Candidate Genes	35
IV.	Materials and Methods	43
V.	Results	48
VI.	Discussion	60
VII.	References	65
VIII.	Appendix	
	a. Primer Sequences and Reaction Conditions	A-1
	b. Summary of Candidate Gene Analysis	A-3

I. Abstract

Transparent testa 9 is a recessive mutation in *Arabidopsis thaliana* which results in plants that produce small, pale-brown seeds, exhibiting less efficient germination than wild-type seeds. The pale seed-color phenotype is the result of reduced levels of proanthocyanidins in the testa. In order to determine the specific gene affected in *tt9*, its location was mapped using a positional cloning approach which involved analyzing a mapping population resulting from a cross between *tt9* plants in the Landsberg *erecta* (L*er)* ecotype background with wild-type Columbia (Col) plants. Genetic analysis of nearly 2500 F_2 plants has enabled the narrowing of the region containing *tt9* to a 200,000 base pair interval on the third chromosome.

Due to the limited potential of further narrowing the region containing *tt9*, efforts have been focused on evaluating candidate genes found within the known mapping interval. This effort has proceeded on multiple fronts including review of publicly available RNA expression data, comparative RNA expression assays between mutant and wild type and sequencing of predicted coding regions. A reverse genetics approach centered on the evaluation of T-DNA tagged lines with insertions in the genes of interest has also been employed. Efforts continue along this front as well as in a focused complementation endeavor aimed at cloning genes of interest in order to attempt to complement the mutation in *tt9*. To date, however, the gene responsible for the *tt9* phenotype remains elusive.

II. Introduction

Flavonoids are a class of plant secondary metabolites which are derived from phenylalanine, via the general phenylpropanoid pathway, and malonyl-coenzyme A (CoA), a product of the fatty acid pathway. There are nine different subclasses of flavonoids playing

varied roles in plant physiology. Several types of flavonoids, for example, have been shown to play a role in protecting plants against the harmful effects of UV radiation (Bieza and Lois, 2001; Li et al., 1993). The reduced palatability associated with large amounts of certain flavonoids serves as a feeding deterrent, while their antimicrobial properties protect against pathogen attack (reviewed in Winkel-Shirley, 2001a). Flavonoids in the seed coat, in particular the subclass of proanthocyanidins (PAs) or condensed tannins (CTs), have been shown to harden the seed coat, thus acting as a mechanical barrier against imbibition damage and radicle protrusion, both of which protect against premature germination (Leon-Kloosterziel and Koornneef, 1994). Investigation of mutants lacking PAs in the seed coat has offered further confirmation for the role of these compounds in contributing to seed-coat imposed dormancy as well as of their important role in maintaining seed viability during storage (Debeaujon et al., 2000). Beyond their functions in plant physiological processes, flavonoids have also been shown to have a role in animal diets. By their ability to bind proteins flavonoids in certain forage crops seem to offer some protection against cattle bloat, a potentially lethal condition in ruminants caused by rapid digestion of proteins in the rumen leading to the production of excess methane gas and foam (reviewed in Dixon et al., 2005). When incorporated into human diets, PAs in particular have been suggested as important nutraceuticals, aiding perhaps among other things in the prevention of cancer, and reducing the risk of cardiovascular disease (reviewed in Cos et al., 2003).

Much of the early work in flavonoid biosynthesis centered on enzymatic studies primarily conducted in cell culture systems (reviewed in Winkel-Shirley, 2001b). More recently however, the study of mutants defective at some level of the flavonoid pathway has made an invaluable contribution to our current understanding of the genetic basis of flavonoid synthesis. Of

particular interest in *Arabidopsis* have been those mutants that are defective in the synthesis and/or accumulation of condensed tannins, the compounds within the seed coat which are oxidized to give wild type seeds their brown color. *Transparent testa* mutants, named for their characteristic lack of, or diminished pigmentation in the seed coat, make up one of the major classes of mutants with diminished PA synthesis in *Arabidopsis*. Many of the first mutants in this class were isolated in a forward genetics screen of mutants derived from chemical (ethylmethanesulphonate, EMS) mutagenesis, fast neutron bombardment or X-ray irradiation. Screening of the resulting mutants led to the identification of 31 mutants with altered seed coat color and/or anthocyanin production representing mutations at 10 distinct genetic loci (Koornneef, 1990). Other mutants in this class, as well as mutations allelic to known loci and several related *tds* (*tannin-deficient seed*) mutations, have been isolated by screening T-DNA insertion lines for reduced dormancy or abnormal seed color (Shirley et al., 1995; Debeaujon et al., 2001; Abrahams et al., 2002; Nesi et al., 2000, 2001, 2002). Although 21 distinct *transparent testa* loci have been described, which are comprised of *tt1-tt19*, *ttg (transparent testa glabra) 1* and *ttg2,* several of the *transparent testa* mutants were recently found to be allelic to other known loci (I. Debeaujon, personal communication). Other mutants with altered flavonoid biosynthesis include *banyuls, aha10* (*auto-inhibited* H^+*-ATPase isoform 10*) and six distinct *tds* mutants.

To date, analysis of mutants with altered seed-color phenotypes has led to the identification of eighteen genetic loci involved in flavonoid biosynthesis. Among the genes affected are several structural genes encoding enzymes necessary for flavonoid synthesis or accumulation, as well as several genes involved in signal transduction and regulation of transcription (Table 1). The gene responsible for *transparent testa 9* (*tt9*), which is the focus of

the current project, has not yet been characterized. *Transparent testa 9,* generated by EMS mutagenesis and in the L*er* background, bears similarity to other *transparent testa* mutants in that it has pale brown seeds indicating reduced levels of PAs in the testa (Shirley et al., 1995). Unlike many other *tt* mutants however, *tt9* does not appear to demonstrate reduced dormancy as compared to wild-type, and seeds of these plants tend to germinate less efficiently (Debeaujon et al., 2000).

Mutant Locus	**Encoded Protein**
tt1	WIP Subfamily of Plant Zinc Finger Proteins (Sagasser et al., 2002)
tt2	R2R3 MYB Domain Protein (Nesi et al., 2001)
tt3	Dihydroflavonol 4-Reductase (DFR) (Shirley et al., 1992)
tt4	Chalcone Synthase (CHS) (Shirley et al., 1995)
tt5	Chalcone Isomerase (CHI) (Shirley et al., 1992)
tt6	Flavanone 3-Hydroxylase (F3H) (Pelletier, 1997)
tt7	Flavonoid 3`Hydroxylase (F3`H) (Koornneef et al., 1982)
tt8	Basic Helix-Loop-Helix Domain Protein (Nesi et al., 2000)
tt9	Unknown
tt10	Laccase-like enzyme (Pourcel et al., 2005)
tt11	Leucoanthocyanidin dioxygenase (LDOX) – allelic to TT18 (I. Debeaujon, personal communication)
tt12	Multidrug and Toxic Compound Extrusion (MATE) Family Transporter (Debeaujon et al., 2001)
tt13	Allelic to AHA10 (I. Debeaujon, personal communication)
tt14	Allelic to TT19 (I. Debeaujon, personal communication)
tt15	Cloned, unpublished results (I. Debeaujon, personal communication)
tt16	Arabidopsis BSISTER MADS Domain Protein (Nesi et al., 2002)
tt18	Leucoanthocyanidin dioxygenase (LDOX) – allelic to TDS4
tt19	*Arabidopsis* Glutathione S-Transferase (GST) Gene Family (Kitamura et al., 2004)
ttg1	WD40 Repeat Protein (Walker et al., 1999)
ttg2	WRKY Transcription Factor (Johnson et al., 2002)
tds4	LDOX – allelic to TT18 (Abrahams et al., 2003)
aha10	P-type ATPase superfamily, H^+-ATPase (Baxter et al., 2005)
banyuls	Anthocyanidin Reductase (ANR) (Xie et al., 2003)

Table 1: *Transparent Testa* Loci and Other Loci Involved in Flavonoid Biosynthesis.

Once a mutant such as *tt9* has been isolated by these means, it is necessary to determine which of the nearly 30,000 genes in *Arabidopsis* has been disrupted. One common and effective approach in forward genetics applications is map-based or positional cloning. This involves

setting up a cross between the mutant line and a wild-type line in a different ecotype background, then designing molecular markers to take advantage of polymorphic sequences between the two ecotypes. More specifically, this type of chromosome walking relies upon recombination rates between tightly linked molecular markers with the goal of narrowing down the genetic interval within which the target gene is located (reviewed in Lukowitz et al., 2000). It is not expected that through positional cloning alone, the precise identity of the defective gene can be discovered. Rather, once an interval has been narrowed sufficiently, other methods such as sequencing and ultimately complementation must be employed to determine which of the genes is responsible for the particular phenotype observed (reviewed in Lukowitz et al., 2000). This method is favorable for its relative technical ease, however difficulties arise when the nature of recombination rates, combined with the availability of exploitable molecular markers is such that the interval cannot easily be narrowed to contain a manageable number of genes for carrying out sequencing and/or complementation. In such cases, other methods to analyze genes within the mapping interval are required.

Candidate gene approaches have been employed in the identification of several *TT* loci. An approach such as this involves examining the predicted genes within the mapping interval known to contain the locus of interest for evidence that may suggest that a given gene is defective in the mutant. Laboratory methods for analyzing candidate genes are varied; however RNA expression assays and sequencing hold considerable potential for elucidating the identity of the target gene. Several publicly available databases offer investigators the ability to determine whether genes within the mapping interval are expressed in tissues or at developmental stages which one might predict based on the phenotype of the mutant (Meyers et al., 2004; Zimmermann et al., 2004). Comparison of mutant versus wild type expression patterns for

candidate genes such as achieved with quantitative Reverse Transcriptase PCR (RT-PCR) may uncover changes in the expression of a gene directly altered in the mutant plant. In other cases, the expression levels of other genes known to be involved in flavonoid synthesis may be assayed, in which case a change in expression level may suggest that the target gene is involved in the regulation of other flavonoid pathway genes.

Sequencing can be quite advantageous in that it has the potential to uncover specific alterations in the DNA sequence of a given gene. In this way, the mutant sequence can be quickly analyzed against the wild-type sequence, and any changes can be evaluated to predict the effect such a mutation may cause in the ultimate gene product. In this way, even point mutations can be easily identified. This method may be limited in terms of cost-effectiveness, however, particularly if the mapping interval in question remains fairly broad.

Reverse genetics approaches have also proven fairly successful in efforts aimed at uncovering the genes involved in a variety of physiological processes. In the case of flavonoid biosynthesis, for example, screens of Transfer-DNA (T-DNA)-tagged lines displaying reduced dormancy characteristics or altered seed color led to the characterization of several genes required for the normal accumulation of flavonoid derivatives (Abrahams et al., 2002; Abrahams et al., 2003; Debeaujon et al., 2001; Nesi et al., 2000, 2001, 2002). This method is particularly amenable to investigations of particular gene families since it is theoretically possible to obtain insertion lines for each member of a particular family. With the near saturation of the *Arabidopsis* genome with insertions of this type, and the availability of seed stocks harboring these insertions, investigators may also use this approach in the final stages of a forward genetics study, such as described above, by analyzing lines harboring insertions in candidate genes within the mapping region for the particular phenotype observed in their mutant. The rationale is that if

in screening T-DNA insertion lines one should find a line displaying the same phenotype of the mutant, it is possible to know precisely which gene is disrupted in the T-DNA tagged line, and consequently it may be inferred that the same gene is defective in the mutant plant. While this method would require further confirmation that the T-DNA tagged insertion is allelic to the mutant, such as by complementation (gene rescue) described below, the evidence generated would be compelling, aiding in focusing the efforts of the initiative, particularly if the mapping interval were fairly narrow.

Once the list of candidate genes has been refined to a single strong candidate using methods such as those described above, it must be confirmed that the gene is responsible for the mutant phenotype observed. Although it is generally considered compelling and acceptable evidence if it can be shown that two or more independently obtained allelic insertion lines display the mutant phenotype, the ultimate confirmation is obtained through complementation of the defective gene with a wild-type construct. This method, which is also termed "gene rescue," uses *Agrobacterium tumefaciens* as the vehicle with which a foreign DNA molecule may be inserted into the host genome. Once expressed within the host, if the inserted, wild-type construct is shown to complement or "rescue" the defective gene in the host by restoring the mutant to the wild-type phenotype, it is then confirmed that the gene in question was the gene responsible for the mutant phenotype.

This work presents the methods employed in the effort to map and clone the gene responsible for *transparent testa 9* as well as the results of this investigation to date. Positional cloning was used initially in narrowing the genetic interval containing *tt9*. Together, ten different genes have been analyzed for differences in RNA expression and/or differences in nucleotide

sequence. T-DNA tagged lines harboring insertions in genes within the mapping interval have also been screened for the presence of the *tt9* phenotype.

III. Literature Review

Flavonoid Biosynthesis

The flavonoid biosynthetic pathway is one of the most highly studied pathways of plant secondary metabolism. Studies in various species have led to the characterization of the main enzymes involved in flavonoid biosynthesis as well as, more recently, the specific genes involved and the diverse factors regulating this pathway. This work has been greatly facilitated by studies of mutants defective at various levels of the pathway. Ultimately, gaining insight into the functions of flavonoid compounds and the mechanisms involved in their production and regulation may provide the key toward successful manipulation of the pathway to take advantage of the diverse positive functions that flavonoids may offer.

Flavonoids have a variety of roles in higher plants, several of which have been studied extensively. Despite the highly conserved nature of the central flavonoid pathway, investigations in multiple species have revealed that the specific roles may differ from one species to another. Two functions that have been studied in *Arabidopsis* in particular are the role of flavonoids in protection against UV radiation and in seed-coat imposed dormancy, although flavonoids have also been implicated in these roles as well as in conferring protection against microbial pathogens, insect pests and larger herbivores, in acting as allelochemicals involved in signaling between plants and microbes, and in male fertility in other species (reviewed in Dixon et al., 2005; Winkel-Shirley 2001a).

sequence. T-DNA tagged lines harboring insertions in genes within the mapping interval have also been screened for the presence of the *tt9* phenotype.

III. Literature Review

Flavonoid Biosynthesis

The flavonoid biosynthetic pathway is one of the most highly studied pathways of plant secondary metabolism. Studies in various species have led to the characterization of the main enzymes involved in flavonoid biosynthesis as well as, more recently, the specific genes involved and the diverse factors regulating this pathway. This work has been greatly facilitated by studies of mutants defective at various levels of the pathway. Ultimately, gaining insight into the functions of flavonoid compounds and the mechanisms involved in their production and regulation may provide the key toward successful manipulation of the pathway to take advantage of the diverse positive functions that flavonoids may offer.

Flavonoids have a variety of roles in higher plants, several of which have been studied extensively. Despite the highly conserved nature of the central flavonoid pathway, investigations in multiple species have revealed that the specific roles may differ from one species to another. Two functions that have been studied in *Arabidopsis* in particular are the role of flavonoids in protection against UV radiation and in seed-coat imposed dormancy, although flavonoids have also been implicated in these roles as well as in conferring protection against microbial pathogens, insect pests and larger herbivores, in acting as allelochemicals involved in signaling between plants and microbes, and in male fertility in other species (reviewed in Dixon et al., 2005; Winkel-Shirley 2001a).

Although the precise mechanisms involved in protection against ultraviolet light are not fully understood, it is clear that flavonoids play an important role, particularly in protecting plants against the damaging effects of UV-B (280-320nm) radiation. Through investigation of both flavonoid deficient mutants (*transparent testa* [*tt*] *4*, *tt5* and *tt6*) and of other UV tolerant or UV sensitive mutants it appears that there is a direct correlation between the accumulation of flavonoid derivatives and sensitivity to UV radiation (Bieza and Lois, 2001; Li et al., 1993). While earlier investigations showed that UV radiation could induce expression of chalcone synthase (CHS) – the enzyme catalyzing the first committed step in flavonoid biosynthesis – thereby suggesting a role for flavonoid derivatives in protective mechanisms against UV radiation, evaluation of flavonoid deficient mutants provided more tangible evidence for the role of specific phenolic compounds in the physiological response to UV light (Kreuzaler et al., 1983; Li et al., 1993).

Another important role for flavonoids in plant physiology is the involvement of proanthocyanidins in seed-coat hardening which ultimately affects their dormancy and germination characteristics. *Arabidopsis,* like many other plants, produces seeds which are initially dormant, a characteristic which allows the seeds in the natural environment to survive periods unfavorable to growth and to germinate only once the environmental conditions are amenable to survival of the plant. In order to break dormancy, seeds must often undergo either stratification in the cold, or a period of dry storage known as after-ripening. Germination in *Arabidopsis* is defined by the penetration of the endosperm and testa by the radicle or root tip (Figure 1). Through studies of seed-shape mutants in *Arabidopsis* the testa has been shown to play a role in imposing dormancy, and thus preventing premature germination, by providing a mechanical barrier against radicle penetration (Leon-Kloosterziel et al., 1994).

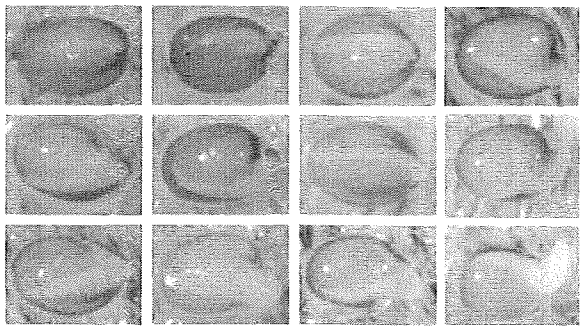


Figure 1: *Arabidopsis* Seed Germination. The frames depict the process of germination in *Arabidopsis* in which the radicle emerges from the seed after penetrating the endosperm and testa (Mosier and Nonogaki, 2003).

The role of PAs in seed-coat imposed dormancy has been shown by studying mutants with reduced pigmentation of the seed coat indicating a change and/or reduction in PA composition. With the exception of *tt9*, *transparent testa* mutants have been shown to display reduced dormancy as compared to wild-type, which was assayed by measuring germination rates after various periods (2-27 days) of after-ripening (Debeaujon et al., 2000). Further characterization of these mutants revealed that they had increased permeability to tetrazolium salts suggesting a general increased permeability of the testa. This may account for the reduced dormancy characteristic in that it suggests the seeds may be more permeable to an endogenous inhibitor (which would therefore be more readily released from the seed) or to an exogenous stimulant of germination such as water or oxygen (Debeaujon et al., 2000). Finally, histological characterization which revealed defects in the cell layers making up the testa, with otherwise normal tissue morphology, confirmed that the testa itself imposes a physical restriction to germination (Debeaujon et al., 2000).

Apart from their role in plant physiology, flavonoids also appear to be important nutraceuticals, conferring a wide range of benefits when incorporated in animal diets. Proanthocyanidins in particular appear to hold potential as antimicrobial, antiviral, antifungal or

antiprotozoal agents. They have also been implicated in the prevention of cancer and cardiovascular disease as well as in preventing cattle bloat, a potentially lethal condition, in ruminants (for review see Cos et al., 2003; Dixon et al., 2005; Nichenametla et al., 2006). The nutritional benefits afforded by flavonoids result, among other things, from their antioxidant properties as well as from their ability to complex metal ions and to inhibit certain enzymatic activities. Although a detailed discussion of these mechanisms is beyond the scope of this work, it may be of interest to point out several structural features which lend to the antioxidant properties observed in flavonoids, since these compounds are among the most potent plant antioxidants. The presence of one or more of the following structural features appears to greatly enhance the antiradical activity of phenolic compounds; (1) an *o*-diphenolic group, (2) a 2-3 double bond conjugated with the oxo function at the 4 position and (3) hydroxyl groups in positions 3 and 5 (Figure 2; reviewed in Bravo, 1998). Among the popular studies in this arena are those which point to the nutritional benefits associated with common foods such as chocolate. Increased plasma concentrations of epicatechin, for example, have been shown to increase antioxidant potential in the blood plasma following consumption of semi-sweet chocolate (Rein et al., 2000).

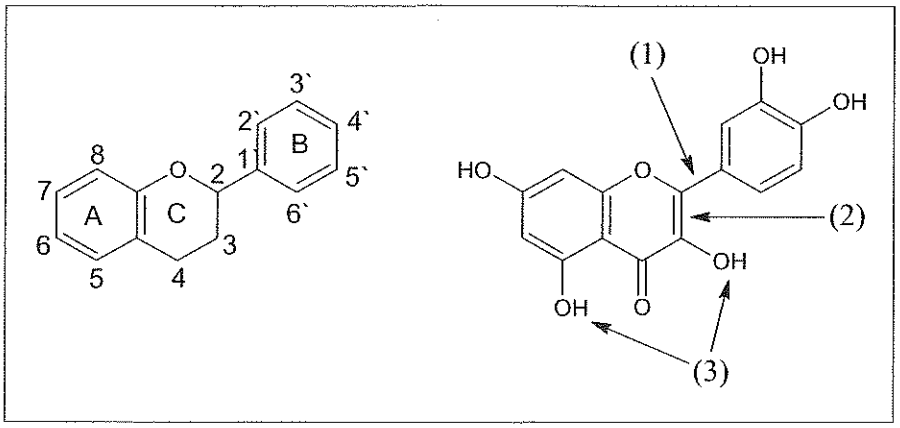


Figure 2: Common Numbering System of Flavonoids and Structure of Quercetin Indicating Structural Features Lending to Antioxidant Activity. See text for description.

An understanding of the function of flavonoids in plants as well as the benefits they may afford to animals necessitates an understanding of the flavonoid biosynthetic pathway itself. Flavonoid biosynthesis derives from the general phenylpropanoid pathway in which several enzyme-catalyzed steps convert the amino acid phenylalanine to 4-coumaroyl-Coenzyme A (CoA) (Figure 3). The enzyme which catalyzes the first committed step in flavonoid biosynthesis, chalcone synthase (CHS), uses 4-coumaroyl-CoA and malonyl-CoA (derived from fatty acid synthesis) as substrates to synthesize chalcone (Figure 4). Subsequent steps in the pathway will vary according to the characteristic end-products produced in the species examined. Legumes, for example, are known to accumulate isoflavonoids via a pathway that diverges in the first step of synthesis in which both chalcone synthase and chalcone reductase function to produce trihydroxychalcone. Once again, through a series of enzyme-catalyzed steps, chalcone is converted to the desired end-product, in this case isoflavonoids (reviewed in Winkel-Shirley, 2001a).

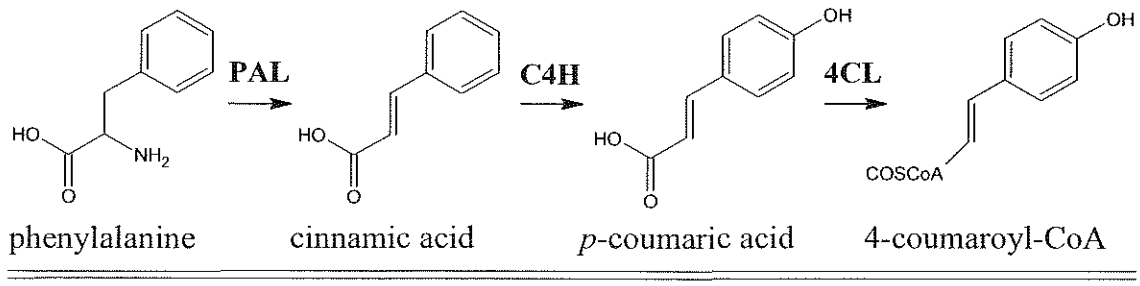


Figure 3: General Phenylpropanoid Pathway. PAL = Phenylalanine ammonia-lyase; C4H = cinnamate-4-hydroxylase; 4CL = 4-coumaroyl:CoA-ligase (adapted from Winkel-Shirley, 2001a).

There are nine different subclasses of flavonoids known to be synthesized in higher plants, six of which – the chalcones, flavones, flavonols, flavandiols, anthocyanins, and condensed tannins - are fairly ubiquitous. The three remaining classes are the aurones, the isoflavonoids, synthesized mainly in legumes as described above and the phlobaphenes, found

primarily in maize, sorghum and gloxinia (reviewed in Winkel-Shirley, 2001a). The precise details regarding the production of these latter products is beyond the scope of this work and will not be dealt with herein.

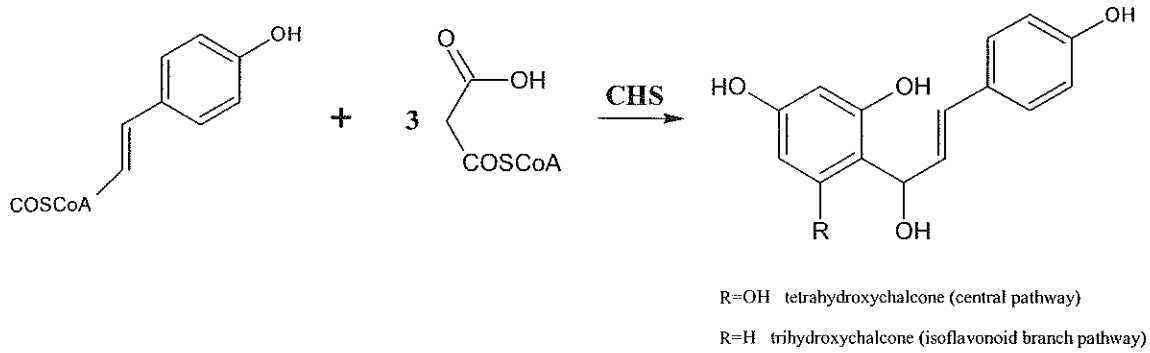


Figure 4: First Committed Step in Flavonoid Biosynthetic Pathway. CHS = Chalcone Synthase.

Of particular relevance for this work are the elements of the pathway important in *Arabidopsis*. These elements include the central pathway, consisting of the conversion of chalcone ultimately to the flavonols or leucoanthocyanidins (flavan-3, 4-diols) as well as the proanthocyanidin (flavan-3-ol) and anthocyanin branch pathways, which are illustrated below (Figure 5). Although much of the early work in identifying the key enzymes in flavonoid biosynthesis were carried out in other species, *Arabidopsis* has proven to be useful in elucidating some of the complex mechanisms involved in regulation of this pathway (for review of work in other species, see Winkel-Shirley, 2001b).

Several of the first flavonoid genes to be characterized in *Arabidopsis* were those encoding enzymes involved in the early steps of the central pathway. Homologues of these enzymes had previously been isolated in other species, although the genes encoding such enzymes were not necessarily known. Isolation of the particular genes involved in flavonoid biosynthesis in *Arabidopsis* was greatly facilitated by study of *transparent testa* mutations which were characterized by their altered seed color phenotype.

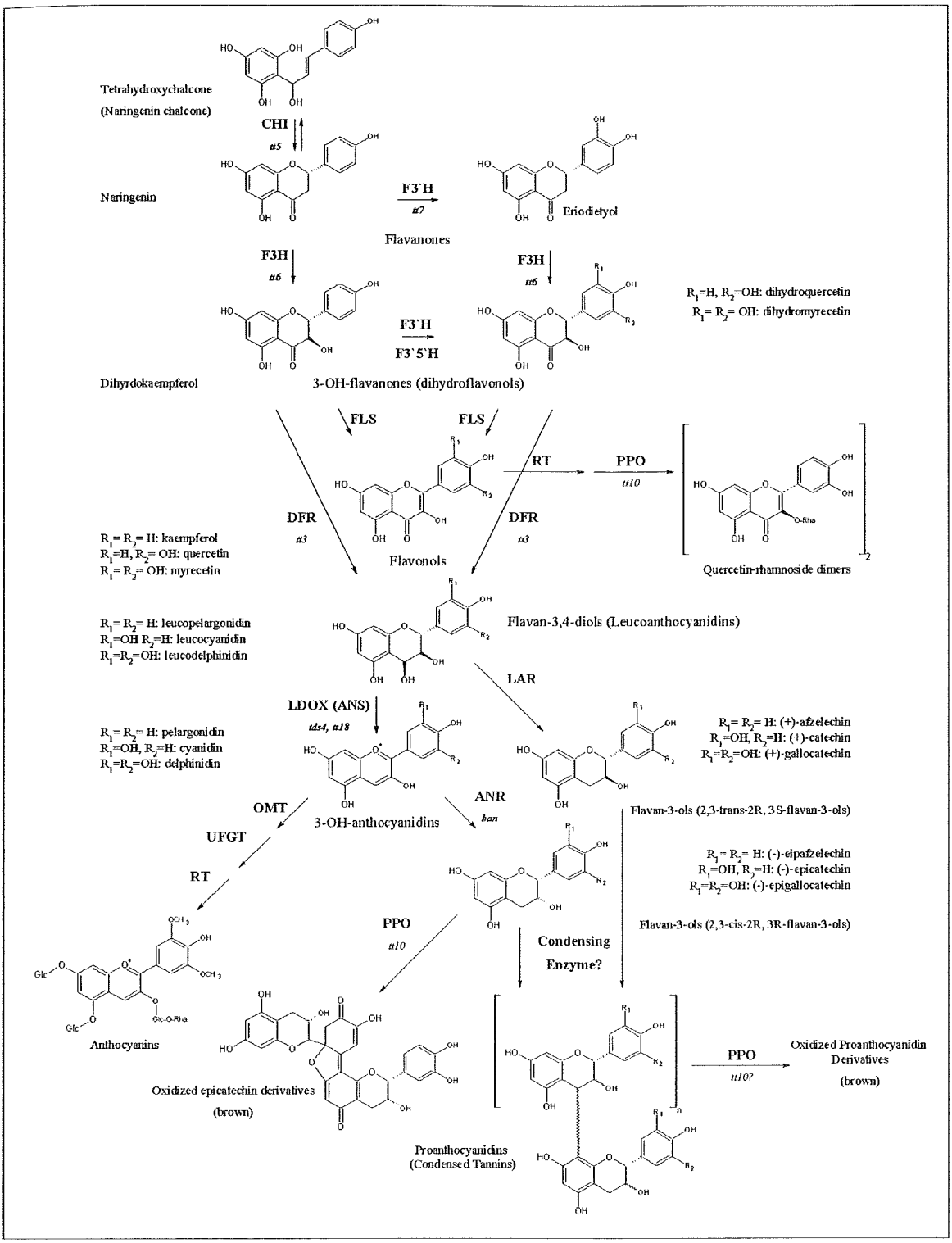


Figure 5: Central Flavonoid Biosynthetic Pathway with Anthocyanin and Condensed Tannin Branch Pathways. CHI = chalcone isomerase, F3`H = flavanone-3`-hydroxylase, F3H = flavanone-3-hydroxylase, F3`5`H = flavanone-3`,5`-hydroxylase, FLS = flavonol synthase, DFR = dihydroflavonol 4-reductase, LAR = leucoanthocyanidin reductase, LDOX = leucoanthocyanidin dioxygenase, ANS = anthocyanidin synthase, ANR = anthocyanidin reductase, OMT = *O*-methyltransferase, UFGT = UDPG-flavonoid glucosyl transferase, RT = rhamnosyl transferase, PPO = polyphenol oxidase (Pourcel et al., 2005; Winkel-Shirley, 2001a; Xie et al., 2003).

Characterization of Flavonoid Mutants in *Arabidopsis thaliana*

The *Arabidopsis* flavonoid 3`-hydroxylase gene was determined to be encoded by the *TT7* locus through biochemical characterization (Koornneef et al., 1982) in which *tt7* plants were shown to be defective in the step controlling 3` hydroxylation on the B ring of the flavonoid skeleton. These mutants were shown to accumulate pelargonidin and kaempferol as a result of this defect, rather than producing cyanidins and quercetin which are found in wild-type *Arabidopsis* (Figure 6; Koornneef et al., 1982).

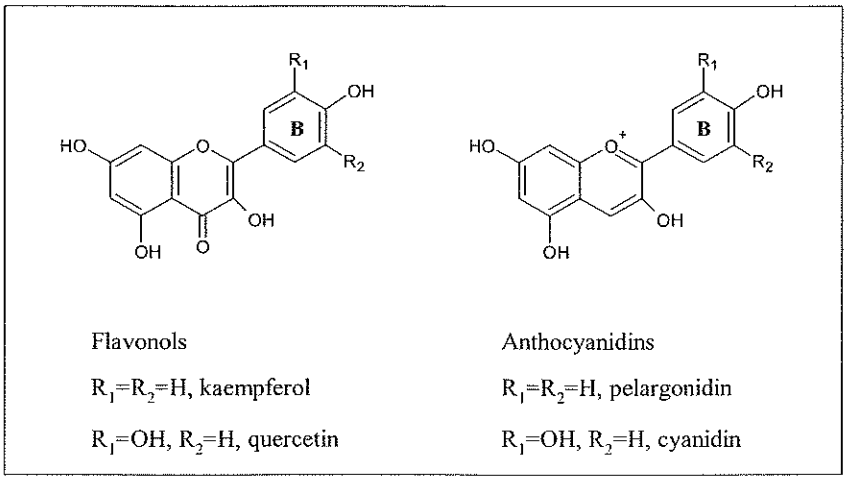


Figure 6: Comparison of Flavonols and Anthocyanidins Which Accumulate in *tt7* and Wild-Type Plants. Due to a lack of flavonol 3` hydroxylation activity in *tt7*, these plants accumulate kaempferol and pelargonidin, rather than quercetin and cyanidin, which are found in wild-type plants (Koornneef et al., 1982).

Chalcone synthase was initially mapped to the *tt4* locus by Restriction Fragment Length Polymorphism (RFLP) mapping, suggesting that the *tt4* gene may encode CHS (Chang et al., 1988). This was confirmed by sequencing of the CHS gene from two distinct *tt4* alleles which showed several single nucleotide polymorphisms when compared with the wild-type sequence (Shirley et al., 1995). RFLP mapping was also used in characterizing chalcone isomerase (CHI)

and dihydroflavonol 4-reductase (DFR) by initially mapping these genes to the short arm of chromosome three near the *tt5* and *tt6* loci and to the long arm of chromosome five near the *tt3* locus respectively (Shirley et al., 1992). Biochemical characterization of *tt5* had previously revealed that this mutant lacked flavonols and anthocyanidins, an observation which would seem consistent with a defect in CHI since CHI catalyzes an early step in flavonoid biosynthesis. Similar characterization of *tt3* indicated that this mutant lacks anthocyanidins, but contains normal levels of flavonols, which is consistent with a defect in a late pathway enzyme such as DFR. DNA blot analysis in which genomic DNA from *tt5* and *tt6* was hybridized with a CHI probe revealed that *tt6* appeared to have a wild-type copy of the gene for CHI, whereas *tt5* appeared to have an altered form of this gene. Further analysis revealed that *tt5* in fact contained a rearrangement of the CHI gene. Hybridization of genomic DNA from a *tt3* mutant with a DFR probe revealed that DFR was not present in this mutant. Together these data confirmed that CHI is encoded by *tt5* and that DFR is encoded by *tt3* in *Arabidopsis* (Shirley et al., 1992).

Flavanone 3-hydroxylase (F3H), which catalyzes the conversion of flavanones to dihydroflavonols (3-OH-flavanones), was found to be encoded by *TT6* in *Arabidopsis*. Expression of this gene was reported to be similar to that of *CHS* and *CHI*, indicating that *F3H* is an early gene in the flavonoid pathway in *Arabidopsis* (Pelletier and Shirley, 1996). *F3H* mapped to the *tt6* locus, and it was found that these mutants produced significantly fewer transcripts for *F3H* than wild-type, indicating that these plants were most likely defective in the gene encoding F3H. Sequencing of *F3H* from *tt6* plants revealed several mutations in the coding region for this gene, indicating that *TT6* is *F3H* (Pelletier, 1997).

Most flavonoid enzymes found in *Arabidopsis* are encoded by single-copy genes – a feature which sets *Arabidopsis* apart from other species and lends to its amenability for studying

this pathway. Flavonol synthase (FLS), however, differs from all other known *Arabidopsis* flavonoid enzymes in that it is not present as a single-copy. This was first shown using DNA blot analysis in which multiple bands were observed when genomic DNA was hybridized with a probe for *FLS* (Pelletier et al., 1997). Sequencing of the *Arabidopsis* genome has since substantiated these results revealing that FLS may be encoded by six distinct loci (B. Winkel-Shirley, personal communication).

Catalyzing the conversion of leucoanthocyanidins to 3-OH-anthocyanidins, leucoanthocyanidin dioxygenase (LDOX) is encoded by *Tannin Deficient Seed-4* (*TDS-4*), as determined by complementation of *tds4* mutants with *LDOX* constructs (Abrahams et al., 2003). Originally thought to compete with a leucoanthocyanidin reductase (LAR) at the branch point of the proanthocyanidin and anthocyanin pathways, it was found that in *Arabidopsis* LDOX is in fact common to the synthesis of both end products (Abrahams et al., 2003; Xie et al., 2003). Biochemical data suggest that proanthocyanidins in *Arabidopsis* are composed exclusively of epicatechin, derived from the reduction of 3-OH-anthocyanidins, rather than catechin, which would be produced by the reduction of leucoanthocyanidins by LAR. This, together with analysis of double mutants which indicate that *tds4* is epistatic to *banyuls* (*ban*, discussed below), provides evidence for the involvement of LDOX at this level of the flavonoid pathway and in particular for the involvement of LDOX in both PA and anthocyanin synthesis, which had not previously been described (Abrahams et al., 2003).

First postulated as a negative regulator of flavonoid biosynthesis based on the accumulation of high levels of anthocyanins in developing and mature seeds of *ban* mutants, *BANYULS* was later proposed as a leucoanthocyanidin reductase (LAR) based on sequence similarity to DFR, which was proposed to compete with LDOX for a common substrate (Albert

et al., 1997; Devic et al., 1999). Biochemical analysis of the *ban* mutant later revealed, however, that *BAN* actually encodes an anthocyanidin reductase which converts 3-OH-anthocyanidins to their corresponding 2,3-*cis*-flavan-3-ols (Xie et al., 2003). This finding was particularly significant in that it proposed a mechanism by which the 2,3-*cis* conformation characteristic of the (-)-epicatechin building blocks of *Arabidopsis* condensed tannins could be generated from the predominantly 2,3-*trans* precursors.

The mechanisms governing the polymerization of epicatechin derivatives and subsequent browning of the otherwise colorless PAs are not fully understood. The recent characterization of *TT10* has provided some insight into this process. *Arabidopsis* seeds darken during desiccation from yellow to brown. In wild-type seeds, this process is usually relatively rapid in that wild-type seeds usually acquire their brown coloration during the maturation process before being released from the parent plant. *tt10* mutants differ in that they produce yellow seeds which tend to slowly darken during storage periods after harvest. Sequencing of mutant alleles as well as functional complementation of *tt10-2* indicated that *TT10* encodes a laccase-like enzyme. Biochemical evidence suggests that this enzyme is involved in oxidative polymerization of flavan-3-ols to yellow to brown pigments (dehydrodiepicatechin A), which differ from the colorless PAs by the placement of their interflavan linkages (Figure 7). *TT10* may also be involved in the oxidation of PAs to their brown derivatives; however this has not been proven (Figure 7). This process may involve another uncharacterized enzyme, or may be a spontaneous process occurring without the aid of an enzyme catalyst.

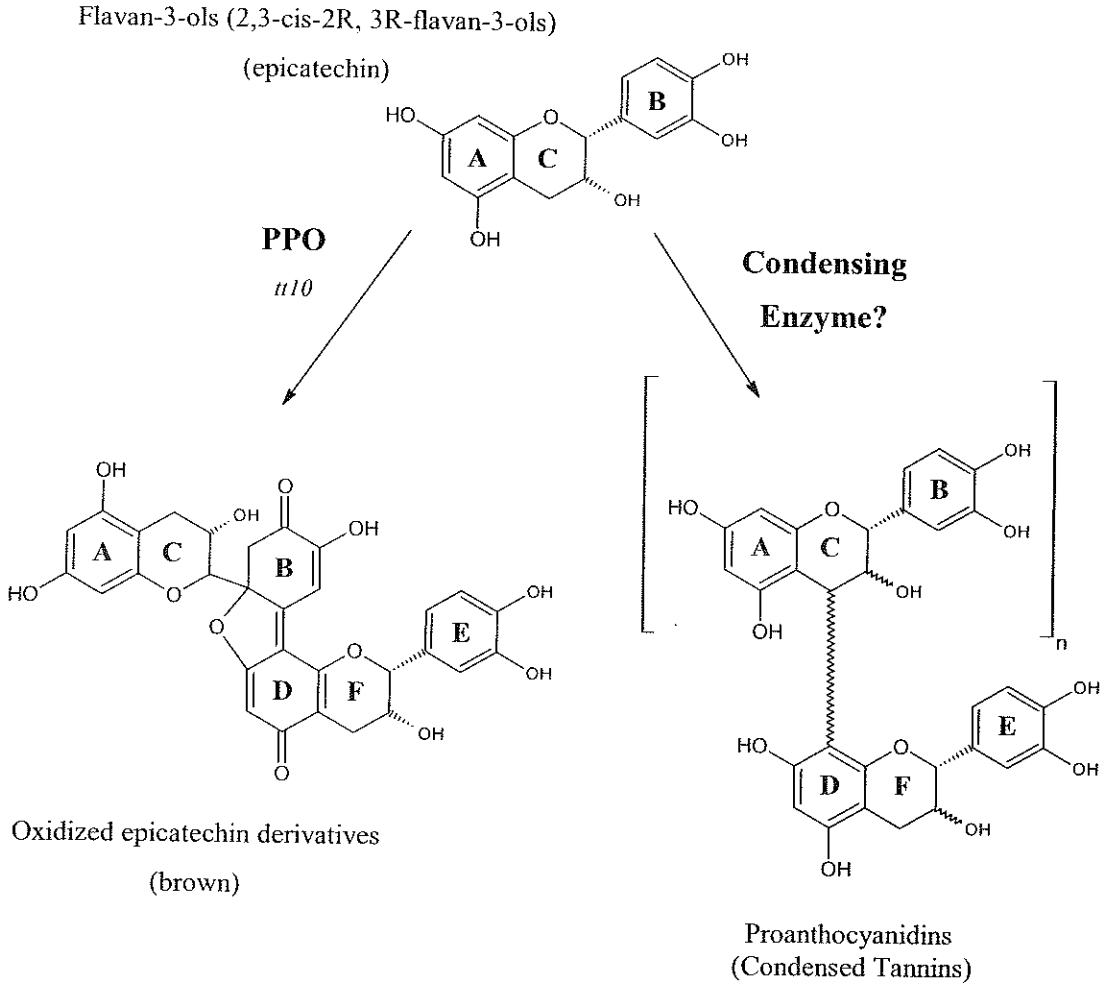


Figure 7: Structural Comparison Between Oxidized Epicatechin Derivatives and Proanthocyanidins. *TT10*, which encodes a polyphenol oxidase, catalyzes the oxidative polymerization of epicatechin monomers to produce yellow to brown pigmented products. Through an unknown mechanism, polymerization can also occur through condensation of epicatechins to produce colorless proanthocyanidins which may be subsequently oxidized to brown pigments. The involvement of *TT10* in the latter process has been suggested; however it has not been shown. Note the differences in the nature and location of the interflavan linkages in each of the two derivatives. PPO = polyphenol oxidase (Pourcel et al., 2005).

Several new *tt* mutant loci were identified through reverse genetics approaches mainly involving screening of T-DNA insertion collections for plants with seed-color phenotype or reduced dormancy. Plants from these collections contain transferred (T)-DNA inserts which disrupt specific genes, often resulting in null alleles for the genes in question. This method has proven instrumental in the characterization of several regulatory elements involved in flavonoid biosynthesis. Flavonoid biosynthesis structural genes have been broadly sub-divided into early

and late genes based on their patterns of temporal expression. In *Arabidopsis*, *CHS*, *CHI* and *F3H* have been shown to be coordinately regulated, and are classified among the early genes in flavonoid synthesis (Pelletier et al., 1996). In addition to these, FLS has also been shown to be encoded by an early gene in at least one case, while leucoanthocyanidin dioxygenase (*LDOX*) and *DFR* are classified as late genes (Pelletier et al., 1997; Shirley et al., 1995). This difference in temporal expression suggests that these genes are regulated by different mechanisms. Early reports supported this idea, in particular suggesting a regulatory role for *TT8* and *TTG1* in the expression of DFR or DFR and LDOX respectively (Pelletier et al., 1997; Shirley et al., 1995).

Characterization of *tt8, ttg1* and *tt2* mutants has provided greater insight into the regulatory mechanisms involved in the expression of structural genes active late in the flavonoid biosynthetic pathway. In addition to *LDOX* and *DFR* mentioned above, *BANYULS* (*BAN*) has also been shown to be coordinately regulated by *TT8, TT2* and *TTG1* (Baudry et al., 2004). *Transparent testa and glabrous-1* (*ttg1*), which is named for its characteristic diminished pigmentation in the seed coat and absence of leaf and stem hairs (trichomes), was the first of the three regulatory mutants to be characterized. It was first suggested as a regulatory element when mutants defective in this gene demonstrated the pleiotropic effects mentioned above, which failed to segregate and therefore appeared to be the result of a mutation at a single genetic locus (Koornneef, 1981). Although ectopic expression of the maize *R* gene was previously shown to complement *ttg1* mutants, suggesting that this gene may, like the *R* gene, encode a MYC transcription factor, recent characterization of this mutant has revealed that the *TTG1* locus encodes a WD40 repeat protein similar to AN11, a regulator of anthocyanin synthesis in petunia (Walker et al., 1999). Its more distant similarity to the ß subunits of heterotrimeric G proteins,

together with the complementation data previously obtained, suggest that *TTG1* may encode a signaling molecule involved in regulating MYC transcription factors.

TT8 and *TT2* were also proposed as encoding regulatory elements, particularly with the observation that mutants defective for either of these genes do not produce transcripts for *BANYULS*. The *TT8* gene was found to encode a basic helix-loop-helix (bHLH) domain protein, an ortholog of the maize *R* transcription factors (Nesi et al., 2000). The *TT2* gene was later found to encode an R2R3 MYB domain protein which was shown to be the primary regulator of the tissue-specific expression of *BAN in planta* (Nesi et al., 2001). While ectopic *TT2* expression can induce the production of *BAN* in tissues which do not normally accumulate PAs, this only occurred where *TTG1* was also expressed and further required the expression of a functional copy of the TT8 protein (Baudry et al., 2004; Nesi et al., 2001). The cooperation between R2R3-MYB and bHLH transcription factors is a phenomenon which has previously been described as being necessary for flavonoid biosynthesis in *Zea mays* (reviewed in Baudry et al., 2004). Using yeast two and three-hybrid systems and expression studies in cultured *Arabidopsis* protoplasts, it was found that *TT2*, *TT8* and *TTG1* are all required for the expression of *BAN* in the plant (Baudry et al., 2004). The yeast two- and three-hybrid systems in particular, which assay the extent to which two or three independent proteins interact *in vivo*, demonstrated the ability of the proteins encoded by *TT2, TT8* and *TTG1* to interact to form a ternary complex (Baudry et al., 2004).

In addition to the regulatory mutants described above, *TT16* and *TT1* have also been shown to encode regulatory elements in *Arabidopsis*. *TT16*, which encodes the ARABIDOPSIS BSISTER (ABS) MADS domain protein, appears to be a tissue-specific regulator of flavonoid biosynthesis (Nesi et al., 2002). Mutants in this background produce normal levels of flavonoid

derivatives in vegetative tissues as well as in the chalazal area of the seed, suggesting that *TT16* is involved only in regulating pigmentation in the seed body. These mutants also show structural aberrations in the endothelial layer, suggesting a role for this gene in the specification of cells in this tissue layer. Furthermore, *TT16* appears to be required for the expression of *BANYULS* in the seed body, adding another dimension to the level of complexity surrounding the regulation of proanthocyanidin synthesis. Similar patterns of *BAN* expression were observed in *tt1* mutants (Nesi et al., 2002). *TT1* has also been characterized as having tissue-specific effects including a role in seed-coat development. Cloning of this gene revealed that *TT1* defines the novel WIP subfamily of plant zinc-finger proteins; however its precise target or activity has not yet been characterized (Sagasser et al., 2002).

The *transparent testa* phenotype, as discussed above, may arise from defects in enzymes catalyzing various steps of the pathway, or from defects in regulatory genes involved in the expression of structural genes. It has also been found that this phenotype may arise from defects in genes encoding proteins involved in proper sequestration of PAs in the vacuoles of the seed-coat endothelium. *TT12*, for example, shows reduced PA deposition in vacuoles of the endothelium indicated by histological sectioning and staining with vanillin-HCl which stains red in the presence of PA and PA precursors. Based on sequence similarity to other proteins, *TT12* is thought to encode a multi-drug and toxic compound extrusion (MATE) family transport protein putatively involved in transporting PAs or PA precursors into the endothelial vacuoles (Debeaujon et al., 2001).

Transparent Testa 19 also appears to play a role in the sequestration of, in this case, both anthocyanins and PAs in the seed-coat endothelium. *TT19* shares approximately 50% similarity with *AN9* which encodes a Glutathione-S-Transferase (GST) in petunia, and represents one of

the 47 GST genes in *Arabidopsis* (Kitamura et al., 2004). The precise mechanism by which *TT19* acts is not fully understood. Although affinity of TT19 for anthocyanins has been demonstrated, and this protein appears to be involved in the transport of these compounds into the vacuole in vegetative tissues, it appears that it may play a different role in the accumulation of PAs in the testa. AN9, which is involved in the transport of anthocyanins in petunia, is able to complement the anthocyanin pigmentation in *tt19*, but not the brown pigmentation of the seeds, further suggesting separate roles for *TT19* in the anthocyanin and PA pathways, roles which are not necessarily shared by *AN9* (Kitamura et al., 2004).

As discussed above, flavonoids, and proanthocyanidins in particular, have a variety of important functions in plant physiology. They also appear to provide important nutritional benefits when incorporated into animal or human diets. Of particular interest in the agricultural community is the role that PAs in forage crops play in reducing the incidence of cattle bloat, mentioned above. This is a potentially lethal condition in which proteins are rapidly digested in the rumen resulting in the production of excess methane gas and foam. PAs alleviate this problem by binding and sequestering proteins in the rumen such that they are resistant to microbial digestion (reviewed in Dixon et al., 2005). This role for PAs has led to attempts at manipulating the flavonoid biosynthetic pathway in forage crops in order to induce the production of PAs in vegetative tissues in which they do not typically accumulate. However, such attempts have been met with difficulty. In one study in which several positive regulators of PA and/or anthocyanin synthesis were ectopically expressed in *Arabidopsis* the resulting plants either failed to accumulate PAs in the desired tissues, or when three such regulators were expressed at high levels (*TT2, PAP1* – a positive regulator of anthocyanin synthesis – and the

maize *Lc* MYC transcription factor), they produced PAs but proved to be unviable, dying within 2 weeks of germination (Sharma and Dixon, 1995).

Transparent Testa 9

The experiment described above highlights the extent to which a thorough understanding of all aspects of flavonoid synthesis and regulation are required in order to facilitate the successful manipulation of the pathway. As stated previously, investigations in multiple species have led to considerable progress in this light. It is in part for this reason that while *Arabidopsis* itself is not a target for biotechnological manipulation for agricultural or nutritional benefit, it serves as a useful model in further elucidating the regulatory and enzymatic mechanisms involved in this pathway. In this light, it is expected that isolating the gene responsible for *transparent testa 9* and further characterizing the mutant will contribute to a broader understanding of flavonoid biosynthesis in general.

The mutant *transparent testa 9* was among the 31 mutants isolated following EMS mutagenesis or irradiation of plants in the Landsberg *erecta* ecotype background described above (Koornneef, 1990). It was mapped to chromosome three near the *GL1* marker by linkage analysis following a cross with the *GL1* marker line, one of several markers described in Koornneef and Hanhart (1983) (Shirley et al., 1995). *Transparent testa 9* has been described as having pale brown or grayish-brown seeds which are significantly reduced in size and weight as compared to wild-type (Debeaujon et al., 2000; Shirley et al., 1995). While a reduction in seed size is common among the *transparent testa* mutants, it is most pronounced in *tt9* (Debeaujon et al., 2000).

Despite the conspicuous decrease in seed-coat pigmentation observed in *tt9*, it appears to accumulate normal levels of anthocyanins as assessed by observation of visible pigmentation in 3-day-old seedlings – a particular developmental stage in which wild-type plants accumulate high levels of anthocyanins (Shirley et al., 1995). Thin layer chromatography (TLC) has been used to assess the presence of anthocyanidins which result from the hydrolysis of anthocyanins and tannins. In such assays, *tt9* showed nearly wild-type levels in some tissues; however such levels were decreased in leaves and absent in seeds. These results may be contrasted with those, e.g. of *tt6*, in which anthocyanidin levels were reduced in all tissues. The TLC analysis confirms the visible pigmentation patterns described above and provides evidence for a tissue-specific involvement of *tt9*, such as a tissue-specific regulatory element or differentially expressed structural gene (Shirley et al., 1995).

It appears that *tt9* also accumulates wild-type levels of flavonols in seeds and flowers as reported following UV fluorescence assays as well as by TLC (Shirley et al., 1995). Wild-type seeds and flowers do not fluoresce under UV light, likely due to attenuation of UV light by flavonoid pigments in these tissues. For comparison, *tt4*, which has a complete absence of flavonoid pigments in the seed coat fluoresces bright blue when irradiated with UV light. These differences in fluorescence are likely due to the type of flavonoid intermediates produced, with the flavonols kaempferol and quercetin being particularly important in attenuating UV light (Shirley et al., 1995). *tt9* accumulates normal levels of these flavonoid derivatives which make up part of the central flavonoid pathway (Shirley et al., 1995).

The biochemical data available on *tt9* suggest a role for this mutant late in the flavonoid biosynthetic pathway. Given that *tt9* plants accumulate normal levels of derivatives from the central pathway, namely kaempferol and quercetin, *tt9* likely acts at least after this step. More

specifically, it is likely that *tt9* is involved in the proanthocyanidin pathway in particular, since *tt9* plants accumulate normal levels of anthocyanins, which are derived from a branch point late in flavonoid synthesis which may in fact compete with PA synthesis for the common 3-OH-anthocyanidin substrates. As stated above, it also appears that *tt9* may have a tissue-specific activity.

Further evidence for *tt9* acting late in flavonoid synthesis is provided by the observation that *tt9* plants tend to accumulate approximately wild-type levels of PAL, CHS, CHI and DFR mRNA, all of which are structural genes involved in the central pathway, before the branch point to proanthocyanidin synthesis (Shirley et al., 1995). In data reported by Shirley et al. (1995) there was a slight decrease in CHS observed in some cases, however results were not reproducible and therefore do not suggest an actual reduction in this transcript.

An interesting feature observed in *tt9* relates to this mutant's dormancy and germination characteristics. Most *tt* mutants display reduced dormancy characteristics relating to the role of condensed tannins in seed coat hardening; however, as described above, *tt9* deviates from this pattern. This stands in contrast to the observation that *tt9* appears to behave like other *tt* mutants with regard to seed longevity, showing a general decrease in viability during long-term storage which is also thought to be related to testa hardening (Debeaujon et al., 2000). In this mutant, not only is dormancy not apparently reduced, *tt9* in fact seems to germinate less efficiently than wild type. *Arabidopsis* seeds typically require a period of time in which certain physiological changes must occur (known as after-ripening) in order to germinate efficiently. Debeaujon et al. (2000) assessed dormancy characteristics by comparing germination rates between *transparent testa* mutants and wild-type plants following varying periods of after-ripening. This group reported that following nine days of after-ripening there was a very slight increase in germination in *tt9* as

compared to wild-type, however with longer periods of after-ripening (18 and 27 days) *tt9* seeds showed significantly reduced rates of germination compared to wild-type under identical conditions (Debeaujon et al., 2000).

Fine-Scale Genetic Mapping

Once a mutation such as *tt9* is identified, a forward genetics approach can be employed to uncover the function of the specific gene. Positional or map-based cloning is a common method employed in such an approach, which has been greatly facilitated by the sequencing of the *Arabidopsis* genome in 2000 (reviewed in Lukowitz et al., 2000). The first step in this method involves crossing the mutant plant, which is in a particular ecotype background with a wild-type plant from another ecotype background (Figure 8). Columbia (Col) and Landsberg *erecta* (L*er*) ecotypes are commonly used for this purpose since the complete genome sequence for the Col ecotype, which was obtained by the *Arabidopsis* Genome Initiative (AGI), is available publicly and a significant amount of sequence information, including a large collection of Single Nucleotide Polymorphisms (SNPs) and insertions/deletions (InDels) between Col and L*er*, has been published by Monsanto Company (Rhee et al., 2003; Jander et al., 2002; The *Arabidopsis* Genome Initiative, 2000).

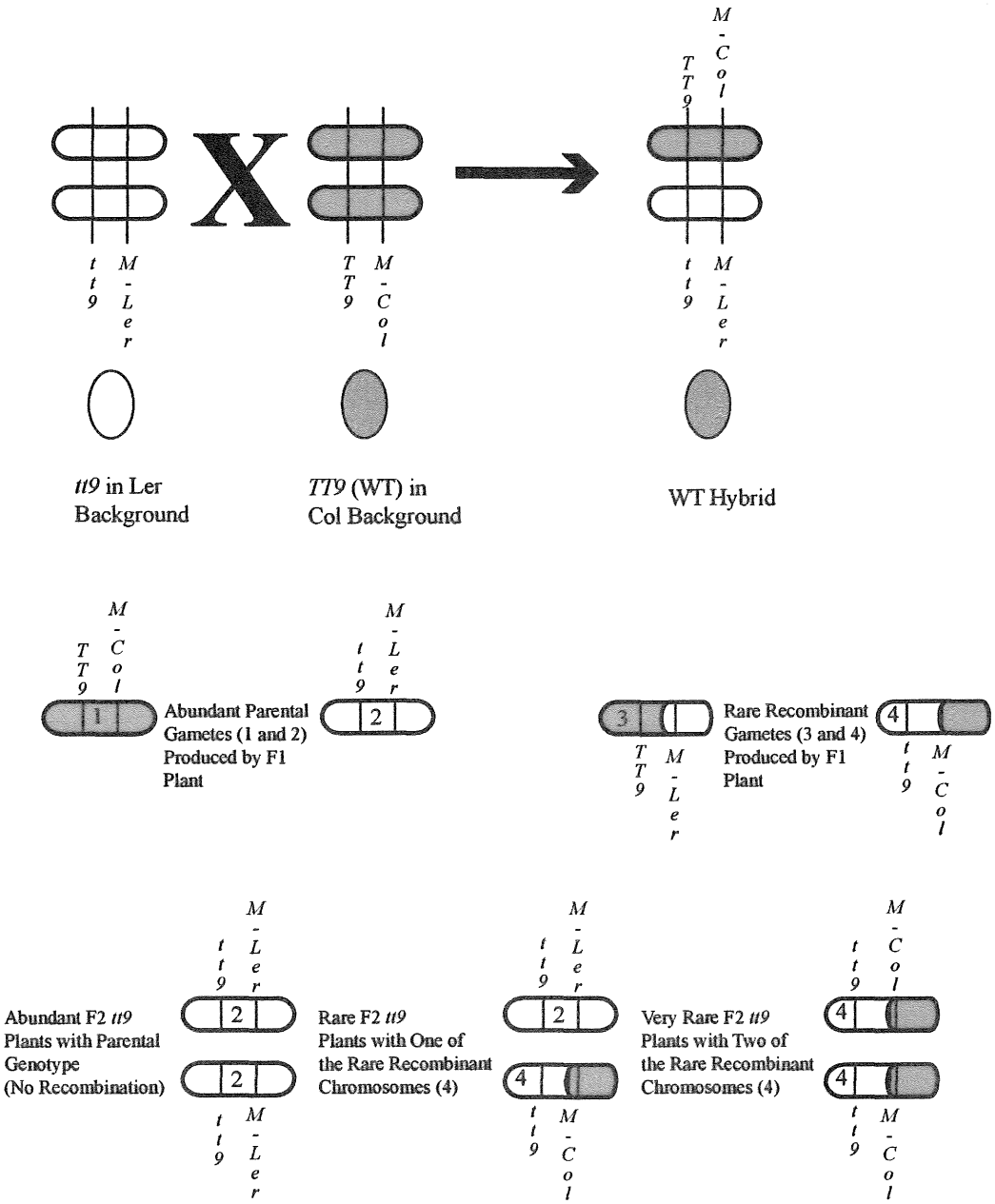


Figure 8: Illustration of a Mapping Cross Between *tt9* and Wild-Type Columbia. In the F_1 generation, all of the plants are hybrids with regard to genotype, and wild-type with regard to seed-color phenotype. The gametes produced by the F_1 plants will consist primarily of those containing a parental chromosome in which no recombination has occurred between the *TT9* locus and the linked marker. Rarely, a recombination event will occur such that the resulting gamete will contain a recombinant chromosome with a different combination of alleles than was observed in either parent. The gametes will come together to form the F_2 generation when the plant is allowed to self-fertilize. If one were to determine the genotype of all the plants that are phenotypically *tt9* with regard to a particular molecular marker linked to the *TT9* locus, it would be found that the majority of the plants are homozygous L*er* for the marker. Occasionally, plants would be found which are heterozygous at the molecular marker, indicating that one of the chromosomes is recombinant. Depending on how tightly linked the marker is to the mutant locus, one might also observe double recombinants which are homozygous Col with respect to the marker assayed. *TT* = *Transparent Testa*; WT = wild-type; L*er* = Landsberg *erecta*; Col = Columbia, M = hypothetical marker.

Self-fertilization in the F_1 plants generated from the initial cross produces the F_2 population which is actually used for mapping analysis. Mapping is carried out by genotyping individual plants from the F_2 generation at various molecular markers designed to exploit polymorphic loci between the two parent ecotypes. The genotype at the molecular marker is compared to the genotype of the plant with regard to the mutation, in order to determine if recombination has occurred between the marker and the mutant locus. In the case of a recessive mutation with a visible phenotype, it is possible to discern which plants are homozygous for the mutant allele without further analysis. Lines which are phenotypically wild-type, however, must be carried out to the next generation in order to determine with certainty their genotype at the mutant locus.

To illustrate the principles involved in mapping as described above, consider an example from the present work. The mutant *tt9*, which is in a L*er* ecotype background, was crossed to a wild-type Col plant to generate a mapping population. Because *tt9* is a recessive mutation, all plants from the F_2 (mapping) population which show the *tt9* phenotype must be homozygous L*er* at the *tt9* locus. Therefore, if a mapping population line was tested at a particular marker and found to be heterozygous, it could be inferred that one recombination event (one recombinant chromosome) exists between the marker tested and *tt9*. Recombination between the marker and the *tt9* locus in phenotypically wild-type plants can also be determined; however as mentioned above, in order to determine the genotype of the F_2 plant with regard to *tt9*, the F_3 generation must be scored to determine whether they segregate for the mutation, which would indicate that the F_2 parent was heterozygous at the *tt9* locus (Figure 9).

Heterozygous F_2 Parent ($TT9^C/tt9^L$)

F_3	$TT9^C$	$tt9^L$
$TT9^C$	$TT9^C/TT9^C$	$TT9^C/tt9^L$
$tt9^L$	$TT9^C/tt9^L$	$tt9^L/tt9^L$

Figure 9: Offspring Resulting from a Heterozygous F_2 Plant When Allowed to Self-Fertilize. By analyzing the F_3 generation and finding that the offspring segregate for the mutation, it is possible to determine that the F_2 parent was heterozygous. If all of the F_3 offspring were phenotypically wild-type, the F_2 parent would have to have been homozygous Col (WT) with regard to *TT9*. The superscripts indicate the ecotype background associated with the wild-type or mutant allele. L=L*er*, C=Col.

Continuing with the illustration above, if the F_2 lines were scored at a molecular marker some distance from *tt9* (for example 100kb from *tt9*) it would be expected that because the marker and the *tt9* locus are linked, the majority of plants analyzed would have the same genotype (Col, L*er* or Heterozygous [Het]) at the marker, as observed at the mutant locus (Figure 8). Occasionally, however, recombination occurs during gamete formation, such that the resulting offspring possess a different combination of alleles from that which was present in either parent. Since recombination is a relatively rare occurrence, and since the rate of recombination between two loci is directly proportional to the distance between the two loci, one can effectively determine the location of the mutation (to within several 10kb) by testing adjacent markers and proceeding in the direction of those for which fewer recombinants are observed.

Several methods exist for developing PCR-based molecular markers of which those based on Simple Sequence Length Polymorphisms (SSLPs) are perhaps the most straight-forward. In this case primers are designed to amplify regions containing relatively large (generally greater than 10 base pairs, although some smaller ones have proven successful) insertions/deletions in the allele of one of the parent ecotypes (Figure 10). With such markers,

the polymorphic region and an additional 200-300bp of flanking sequence are amplified via PCR and separated by electrophoresis in high percentage agarose gels to provide resolution of the different product sizes generated for the two possible alleles. Typically, the allele is amplified from Col and L*er* DNA, as well as a 1:1 mixture of Col and L*er* DNA (Het), all of which are separated in the gel alongside the mapping lines to serve as controls. The PCR products from each of the mapping lines can then be compared to the controls and scored accordingly.

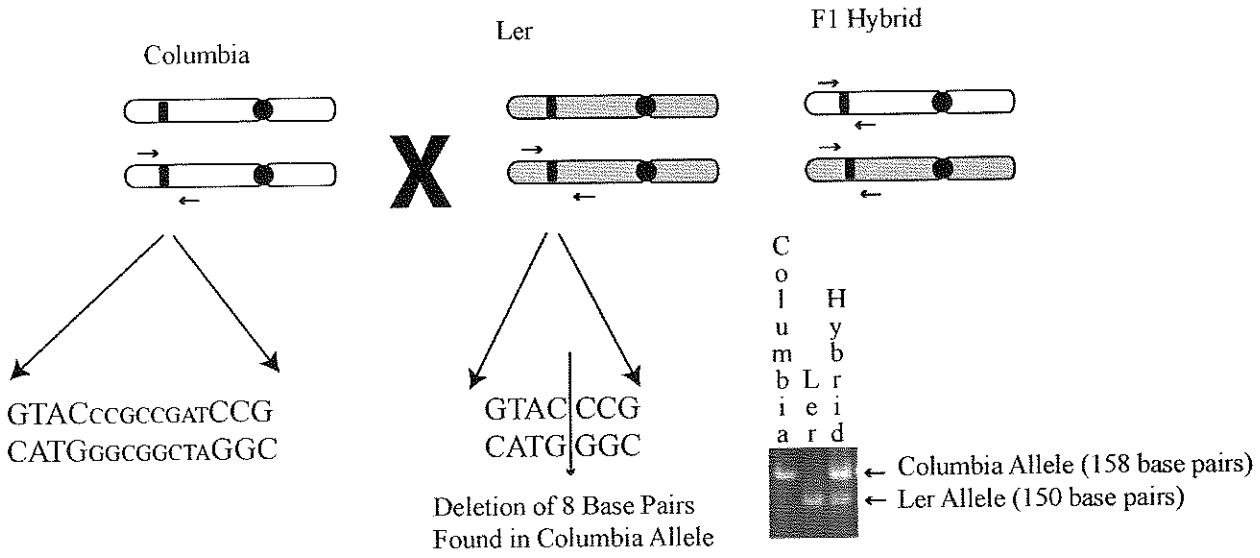


Figure 10: Schematic Representation of SSLP-Based Molecular Marker. In this example, the Col allele contains an 8bp insertion (smaller letters) which is absent in the L*er* allele. By PCR-amplifying a ~150bp fragment containing the polymorphic locus using DNA from Col, L*er* or hybrid plants and separating the resulting fragments by agarose gel electrophoresis (lower right), it is possible to determine the genotype of the individual at this particular polymorphic locus. The small arrows indicate the location where the primers would bind for the PCR.

It is estimated that a polymorphism exists between Col and L*er* in approximately four to eleven positions every 1000bp (Chang et al., 1988; Konieczny and Ausubel, 1993). However, many of these polymorphisms are not large insertions/deletions which can be resolved easily by electrophoresis. Cleaved Amplified Polymorphic Sequence (CAPS) markers and the related derived CAPS (dCAPS) markers allow single nucleotide polymorphisms to be exploited for genotyping purposes by taking advantage of polymorphisms which disrupt a naturally occurring

restriction endonuclease (RE) recognition sequence or an introduced, or derived, RE site respectively. In the case of CAPS markers, primers are designed to amplify the region containing the polymorphic locus such that following amplification by the PCR, the products can be digested with an appropriate enzyme such that one of the alleles will be left intact, while the other will be cut to produce two smaller fragments (Konieczny et al., 1993). dCAPS markers employ a similar principle except that the primers are designed to incorporate a mismatched nucleotide such that a restriction enzyme site will be introduced into one of the alleles, but not the other due to the presence of a SNP (Figure 11; Neff et al., 1998). The development of dCAPS markers is facilitated by use of the dCAPS Finder software developed by Neff et al. (1998). Although these types of markers are beneficial in that they allow an investigator to potentially further saturate his/her map by exploiting the SNPs as well as the larger SSLPs, they require an additional RE digestion step which makes this method more laborious. Furthermore, not every SNP is amenable to this technique, and often the RE sites generated require the use of rare enzymes which limit the cost effectiveness of the method.

A final option for detecting polymorphisms involves amplifying the region containing the polymorphism, as above, but rather than relying upon the ability to resolve size differences, natural or derived, the PCR products are simply sequenced to determine the genotype of the plant at the given marker. This method is best suited for SNPs since Col and L*er* alleles are easily distinguished based on the particular nucleotide present at the polymorphic locus (Figure 12A). Heterozygous lines, which are often indicated by an ambiguous call at the polymorphic locus, may be easily discerned by examining the sequence electropherogram (Figure 12B). Although this represents a fairly reliable technique, given the cost of individual sequencing reactions, this method is not particular cost effective for genotyping many individual lines.

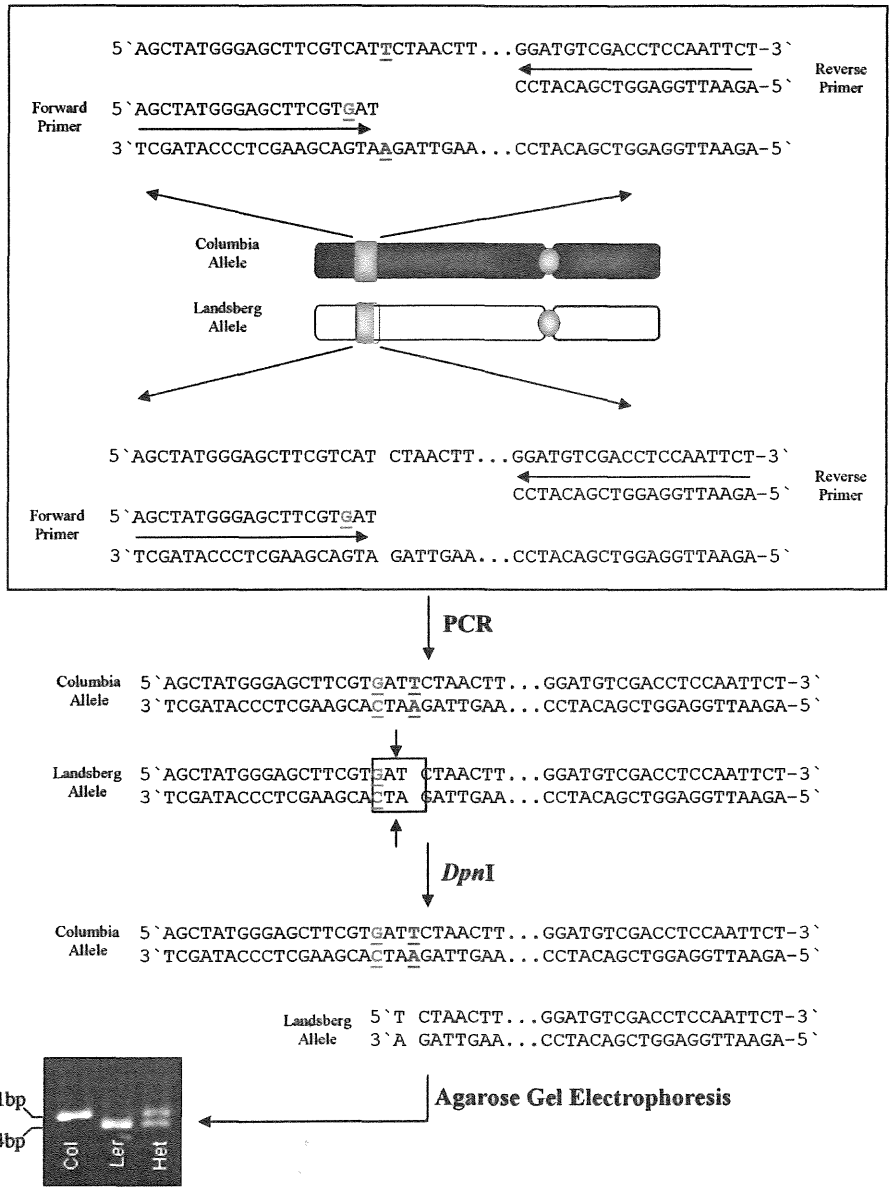
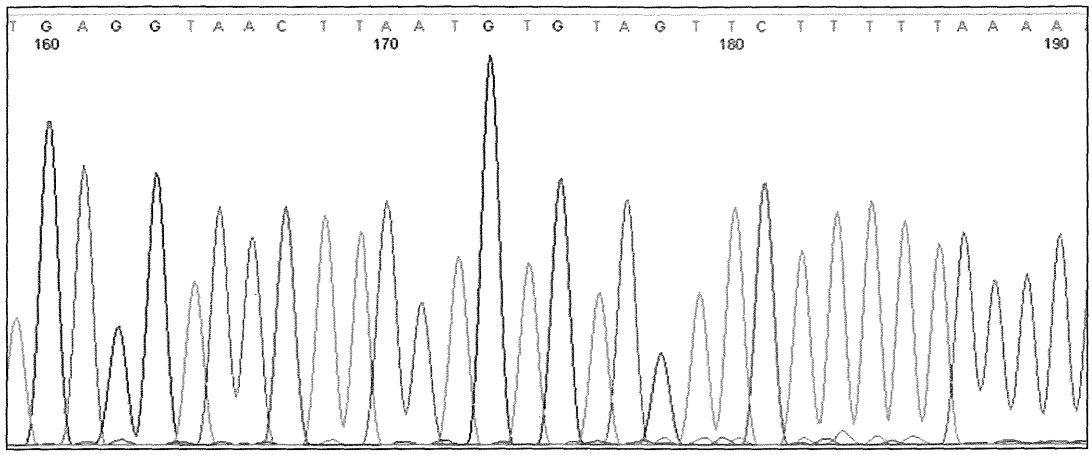
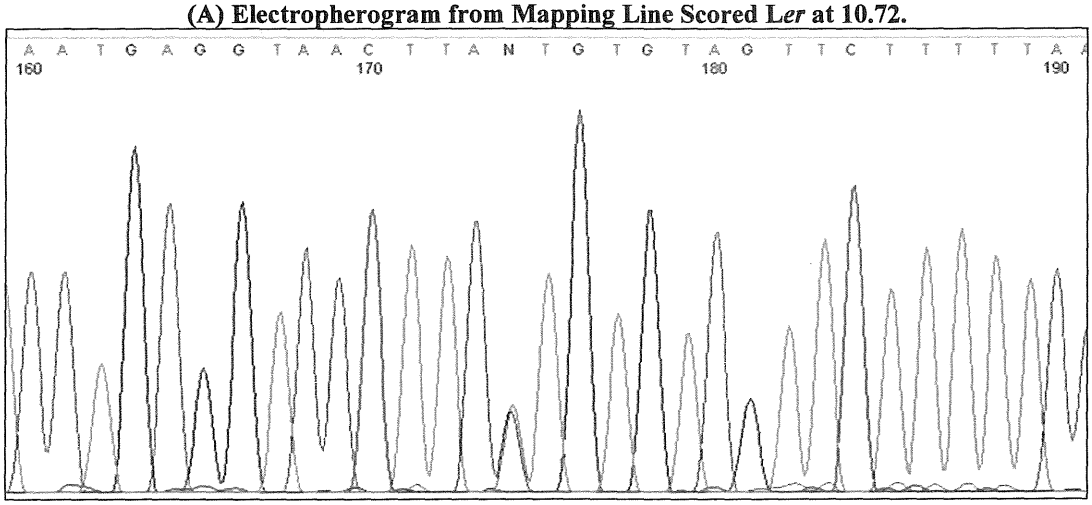


Figure 11: Schematic Overview of dCAPS Marker Design. In the first frame, a heterozygous individual is illustrated (note that only one of the five pairs of chromosomes is depicted). The nucleotide sequences represent the sequences from Col and L*er* alleles which will be amplified by PCR for the 10.67DpnI marker, which is located at 10.67Mbp on chromosome 3. The underlined bases in the forward primer represent a mismatched nucleotide which will be incorporated into the final PCR product. The underlined nucleotides in the Col and L*er* sequences represent the single nucleotide polymorphism for which this marker was designed. Following PCR each of the alleles will be amplified and will incorporate the mismatched nucleotide. Since the primers do not extend past the SNP, these polymorphisms will still be apparent in the resulting PCR product. The box indicates the recognition sequence for *Dpn*I, while the arrows indicate where this enzyme will cut the PCR product to yield a blunt end. The PCR products are digested with *Dpn*I. Since only the L*er* allele has the correct recognition sequence, the digest will yield different fragment sizes for Col and L*er*. The agarose gel depicts the results after separating Col (two Col alleles), L*er* (two L*er* alleles) and Het (one Col and one L*er* allele, as depicted above) controls by electrophoresis (lower left).



(A) Electropherogram from Mapping Line Scored L*er* at 10.72.



(B) Electropherogram from Heterozygous Control

Figure 12: Sequence Electropherograms Illustrating Polymorphism at 10.72Mbp. The polymorphism found at 10.72Mbp consists of a T-A substitution at position 174 above. In the L*er* allele (A), a T is clearly present at position 174. However in the Het (B), although there is an ambiguous call (N) at position 174, it is clear that there are in fact two distinct peaks corresponding to the T (red peak) present on the L*er* allele and the A (green peak) present on the Col allele, both of which are expected to be present in a Het line.

In order to maximize the cost-effectiveness and efficiency of a positional cloning endeavor, a mutant is generally first mapped to a relatively broad chromosomal location, often by use of linkage analysis with specific marker lines. This is particularly effective if the mutant is crossed with a marker line conferring a visible phenotype for rapid detection. Once the gene has been mapped to within a relatively narrow interval and flanking markers have been identified within 10% recombination of one another, it is no longer necessary to analyze all mapping lines with every marker within the mapped region. Instead, all of the mapping lines are genotyped at

the flanking markers such that plants harboring a chromosome in which recombination has occurred between the flanking markers may be identified rapidly, and all other plants may be discarded as they are no longer useful for narrowing the genetic map. Subsequent analysis can then be performed on only a small subset of the population thus significantly decreasing the cost and labor intensity of the procedure. It is important to note that with such a mapping effort, any plants that are identified as wild-type phenotypically (when mapping a recessive mutation) must be carried out to the F_3 generation in order to determine the genotype of the particular plant at the mutant locus (methods reviewed in Lukowitz et al., 2000).

Methods for the Evaluation of Candidate Genes

The goal of a mapping initiative is to narrow the mutant locus to within several thousand base pairs containing a small set of candidate genes such that they may be analyzed and the precise gene responsible for the mutation may be identified. There exist inherent limitations in map-based cloning approaches, which primarily center on the availability of robust molecular markers and exploitable polymorphic loci, as well as on the nature of recombination frequencies which are likely not static throughout the genome. It is possible, in particular with emerging information resources, to use a modified candidate gene approach to identify the specific gene responsible for the observed mutation. In some cases, a simple look at the putative protein products and the characteristics of the mutant may provide sufficient direction for further analysis. This may prove to be misleading however, and other more versatile methods may be required.

The evaluation of candidate genes is meant to uncover evidence which would set a particular candidate over and above all of the others. Examination of RNA expression data, both

in terms of normal, wild-type expression in tissues or during developmental stages of interest may provide some insight for further analysis. Investigation of GeneChip or Microarray data may be particularly useful at this stage of the project. cDNA microarrays allow for rapid RNA expression analysis for every gene within the genome. In this technique, fluorescently-labeled cDNA is reverse-transcribed from RNA which may be isolated, for example, from different tissues, from particular tissues under specific growth conditions, or at particular developmental stages. The cDNA is hybridized to a GeneChip(s) containing single-stranded DNA together representing the entire genome. Based on where RNA was found to hybridize and the intensity of the fluorescent signal, it is possible to determine, respectively, which genes are expressed under the particular conditions assayed and the level of expression. This type of investigation has been facilitated by the availability of public RNA expression databases. Microarray data compiled from several independent laboratories, for example, is accessible by the public at Genevestigator (https://www.genevestigator.ethz.ch/at/) in which investigators may search for expression patterns of their gene(s) in various tissues, at different developmental stages, in response to certain environmental stimuli or stresses or even in correlation with other genes of interest (Zimmermann et al., 2004). The data have been compiled and normalized to facilitate comparison of expression data obtained from independent experiments and to allow for general trends to be observed in the expression patterns of the genes queried.

It may also be helpful to compare the expression patterns of candidate genes between the mutant and wild-type. This type of comparative analysis may uncover a difference in the expression level for a particular gene in the mutant which may indicate disruption of the gene assayed, or a disruption in another gene which regulates the function of the one assayed. Real-time reverse transcriptase PCR (real-time RT-PCR) is one means by which this type of analysis

may be carried out. The principle of real-time RT-PCR is the same as that for traditional RT-PCR in that RNA expression is assayed by first generating cDNA from the RNA in an isolated sample, then carrying out PCR with gene-specific primers to amplify the cDNA – essentially amplifying the "signal"- such that even weakly expressed genes may be assayed.

The real-time method has certain advantages over traditional analysis in that the use of a fluorescent dye in the reaction allows the investigator to monitor the progress of the reaction by sampling the fluorescence after every cycle. Ideally, one would like to determine the point at which the signal increases to a statistically significant level (threshold) above the baseline signal. The threshold occurs during the exponential phase of the reaction when amplification is most efficient, and the cycle in which the threshold is crossed is known as the threshold cycle (C_T). The number of cycles required to reach threshold is indicative of the amount of RNA present in the original sample. It may be difficult to accurately quantify differences in expression level, particularly if such differences are subtle, with traditional RT-PCR since it is more difficult and labor intensive to monitor the progress of the reaction, and thus to determine precisely when the threshold cycle is reached. Real-time analysis allows for more accurate quantification of RNA expression levels by allowing for the precise determination of the threshold cycle (reviewed in Gachon et al., 2004).

An additional means for evaluating candidate genes involves a modified reverse-genetics approach in which plant lines harboring T-DNA insertions within particular genes are analyzed for evidence of the mutant phenotype. This method has gained more widespread use with the generation and availability of T-DNA insertion mutants for nearly every gene in *Arabidopsis*. The general methods employed in generating the T-DNA insertion collections currently available are essentially the same (Alonso et al., 2003; Sessions et al., 2002). *Agrobacterium tumefaciens*

harboring a vector containing the T-DNA is used to infect wild-type plants. Because the transfer and integration of the T-DNA (for which the mechanisms are discussed below) into the host genome is a random process, transformants obtained in this way may have the T-DNA inserted, theoretically, anywhere in their genome. Thousands of transformants may be generated in this way and then screened to determine the precise location of the insertion by determining the nature of the sequences flanking the insertion. This may be done using thermal asymmetric interlaced (TAIL)-PCR which uses primers specific to the T-DNA coupled with degenerate primers in order to amplify sequences flanking the left border of the T-DNA, or other high-throughput methods for recovering the insertion-site (Alonso et al., 2003; Sessions et al., 2002; Singer and Burke, 2003). Ultimately sequencing of the resulting products and Basic Local Alignment Search Tool (BLAST) comparison with the published genome sequence enables the location of the insertion site to be determined.

Several collections of T-DNA insertion mutants currently exist such that investigators may obtain lines with insertions, in this case, in candidate genes they wish to evaluate. The T-DNA is thought to target the maternal chromosome set and integrates as a single copy into the host genome (Bechtold et al., 2000). This means that the primary transformants will be hemizygous for the insertion and must be allowed to self-fertilize in order to obtain homozygous lines. Because T-DNA integration often results in null alleles, in order to evaluate whether the T-DNA line resembles the mutant in question, a homozygous line must be analyzed. A homozygous line which displays the phenotype of the mutant in question would provide strong evidence that the gene disrupted by the T-DNA (which is known) is the same gene disrupted in the mutant plant.

Ultimately, in a study aimed at elucidating the particular gene involved in producing the mutant phenotype, positive confirmation for the candidate gene's identity with the mutant locus must be obtained. The most widely accepted means of confirming that a particular candidate gene is disrupted in the mutant is through complementation of the mutant phenotype with a wild-type copy of the candidate gene. This has been facilitated by advancements in *Agrobacterium*-mediated plant transformation technologies which are typically employed to generate transgenic plants expressing the gene of interest.

Agrobacterium tumefaciens is a gram-negative bacterium which has been known as the causative agent in crown gall tumors in plants for nearly a century. The discovery that tumor formation was the result of gene transfer from *Agrobacteria* to the host plants led to further characterization of the mechanisms involved in this gene transfer, and ultimately to the development of transformation protocols which harness the gene-transfer capabilities of *Agrobacteria* for use in biotechnological applications (reviewed in de la Riva et al., 1998). Two components of the *Agrobacterium* system are of particular importance in its application as a tool for transformation. The first element is the Transferred (T)-DNA which is ultimately integrated into the host genome and the second is the cluster of *virulence* (*vir*) genes which encode some 35 proteins required for gene transfer. Normally, the T-DNA and the *vir* genes are found together on one tumor-inducing (Ti) plasmid (reviewed in Hellens, Mullineaux and Klee, 2000).

The process of T-DNA transfer involves essentially five major steps; (1) bacterial colonization involving the physical attachment of the bacteria to the plant cell; (2) induction of the bacterial *vir* genes; (3) formation of the T-DNA transfer complex; (4) T-DNA transfer to the plant cell; and (5) integration of the T-DNA into the host genome. Following attachment of the bacteria to the plant cell, the *vir* gene products are activated through several *trans*-activating

elements and one possible *cis*-acting element, ultimately producing the T-DNA transfer complex through which the T-DNA is transferred to the host cell. Several *vir* gene products associate with the T-DNA, some of which contain nuclear localization signals (NLS) which are involved in targeting the T-DNA to the nucleus. The precise mechanism by which the T-DNA becomes integrated into the host genome is not fully understood, although it is thought to occur by illegitimate recombination with the involvement of some of the *vir* proteins as well as DNA repair mechanisms within the cell (reviewed in de la Riva et al., 1998).

The genes found within the T-DNA are involved in the formation of tumors; however they are unnecessary for the actual transfer and integration of the T-DNA into the host. This has meant that other genes can be inserted to replace the native genes from the T-DNA which both disarms the T-DNA preventing the formation of tumors, as well as provides a vehicle for transporting genes of interest into the host plant (reviewed in Hellens, Mullineaux and Klee, 2000). The development of binary Ti vectors has greatly facilitated the use of *Agrobacterium*-mediated transformation. In the case of binary vectors, the Ti plasmid is resident within the particular strain of *Agrobacteria* being used and harbors the *vir* genes required for gene transfer. The T-DNA is found on a separate vector which can be manipulated more easily to contain the genes which are to be transferred into the plant (reviewed in Hellens, Mullineaux and Klee, 2000).

There are several features important for effective binary vectors. They typically contain an origin of replication (*ori*) with a broad host range which permits its replication both in *E. coli* and in *Agrobacterium*. In some cases, this broad host range *ori* is supplemented with one which will produce higher copy numbers in *E. coli* such as *Col*E1. The size of the vector is an important factor in the efficiency of the transformation system, such that smaller vectors are preferred. For

this reason, there has been a progressive reduction in size of the vectors used for this purpose. A small vector size is achieved in the pGreen vector series used in this project by dividing the relatively small, broad host-range pSa *Ori* and pSa replicase gene (*repA*) into separate plasmids. The *repA* gene is found on a separate plasmid which is co-transformed with the plasmid containing the T-DNA with the gene of interest into *Agrobacteria.* Like any vector system, binary Ti vectors must also contain antibiotic resistance genes for selection in bacteria. In this system, however, the T-DNA itself must also carry a selectable marker, in this case conferring herbicide resistance for selection *in planta* (reviewed in Hellens, Mullineaux and Klee, 2000).

Several different binary vectors as well as "vector families" are currently available for use in *Agrobacterium*-mediated transformation, which can be tailored for specific applications (reviewed in Hellens, Mullineaux and Klee, 2000). Although in the case of complementation assays, it would be ideal to transfer a gene along with its native promoter, in some cases factors such as the size of the gene may limit the efficiency with which this can be achieved. It is possible to obtain vectors for this system which contain constitutive promoters (e.g. 35S-CaMV) which can drive the expression of, for example, a cDNA clone. The drawback to this method is that it is possible to obtain a false positive in that it is possible that the transgene is capable of complementing the mutant phenotype without actually being the gene disrupted in the mutant. If, for example, the transgene encoded a protein with high similarity to that encoded by the gene disrupted in the mutant, it is possible that the foreign protein, whose expression is driven by a strong, constitutive promoter, will perform the function lacking in the mutant and therefore mask the nature of the underlying mutation (reviewed in Østergaard and Yanofsky, 2004).

Once a clone has been generated of the gene of interest and has been inserted into an appropriate vector for transformation, mutant plants are dipped into media containing

Agrobacteria with the T-DNA construct for transformation. It has been found that reasonable transformation efficiencies can be obtained by simply dipping plants in media containing *Agrobacteria* (Clough and Bent, 1998). Several factors were found to be important in increasing the transformation efficiency when using this method, which include the concentration of sucrose in the dipping media (5 percent) and the concentration of the surfactant Silwet L-77 (0.05%). Repeating the dipping procedure (with the same plants) was shown to improve transformation efficiency, as well as covering the dipped plants for one day after dipping to maintain humidity. Finally, the stage of development of the plants to be dipped was found to be important, with plants having numerous immature inflorescences producing the most transformants (Clough and Bent, 1998). This observation is consistent with the finding that female reproductive tissues are the target for the T-DNA transfer, and that transformation is limited by the accessibility of the bacteria to the ovules, which is facilitated at early stages of floral development (Desfeux, Clough and Bent, 2000).

Following transformation, seeds from the dipped plants are collected and selected with the appropriate herbicide. Transformants are then grown and scored according to whether complementation of the mutant phenotype is observed. Complementation would confirm that the candidate gene is in fact the gene which is disrupted in the mutant plant barring any complications with false positives as described above. Although this method is typically employed once a single candidate gene has been isolated, it is also amenable to a "shotgun cloning" approach where several candidate genes may be tested for complementation based either on map position or other data which would suggest that the gene may be disrupted in the mutant.

The characterization of genes involved in important physiological processes in *Arabidopsis* has been greatly facilitated by a refining of the methods involved in mapping and cloning such genes. Large-scale collaborative efforts, such as the sequencing of the *Arabidopsis* genome and the generation of near-saturation insertion-mutant lines, have provided significant resources for basic scientists carrying out work in this field. In the present work, methods like those described above have been employed in mapping the *tt9* locus in *Arabidopsis* and in the molecular genetic characterization of candidate genes within the mapped interval.

IV. Materials and Methods

Plant materials and growth conditions

Unless otherwise noted, seeds were suspended in 0.1% agarose without prior sterilization and stratified in the dark at 4°C for 2-4 days in order to break dormancy and synchronize germination. Seeds were sown in moistened potting media (Scotts Metro Mix 360) in individual baskets (AraSystem, Lehle Seeds) and cultivated under controlled conditions in a growth chamber set to 20°C and 70 percent relative humidity under continuous white light. Seeds harvested in this lab were placed, unless otherwise noted, at 4°C for long-term storage.

Genetic mapping of the *tt9* locus

A mapping population was previously generated in this lab by crossing a *tt9* plant in the Landsberg *erecta* (L*er*) ecotype background with a wild type plant in the Columbia (Col) background. Plants from the F_1 generation were allowed to self-fertilize and the resulting seeds were routinely stratified and sown as described above. Tissue was collected from rosette leaves of all plants approximately four to five weeks after germination and stored at -70°C. DNA was isolated from tissue harvested from all the plants which were scored *tt9* (based on lighter seed

The characterization of genes involved in important physiological processes in *Arabidopsis* has been greatly facilitated by a refining of the methods involved in mapping and cloning such genes. Large-scale collaborative efforts, such as the sequencing of the *Arabidopsis* genome and the generation of near-saturation insertion-mutant lines, have provided significant resources for basic scientists carrying out work in this field. In the present work, methods like those described above have been employed in mapping the *tt9* locus in *Arabidopsis* and in the molecular genetic characterization of candidate genes within the mapped interval.

IV. Materials and Methods

Plant materials and growth conditions

Unless otherwise noted, seeds were suspended in 0.1% agarose without prior sterilization and stratified in the dark at 4°C for 2-4 days in order to break dormancy and synchronize germination. Seeds were sown in moistened potting media (Scotts Metro Mix 360) in individual baskets (AraSystem, Lehle Seeds) and cultivated under controlled conditions in a growth chamber set to 20°C and 70 percent relative humidity under continuous white light. Seeds harvested in this lab were placed, unless otherwise noted, at 4°C for long-term storage.

Genetic mapping of the *tt9* locus

A mapping population was previously generated in this lab by crossing a *tt9* plant in the Landsberg *erecta* (L*er*) ecotype background with a wild type plant in the Columbia (Col) background. Plants from the F_1 generation were allowed to self-fertilize and the resulting seeds were routinely stratified and sown as described above. Tissue was collected from rosette leaves of all plants approximately four to five weeks after germination and stored at -70°C. DNA was isolated from tissue harvested from all the plants which were scored *tt9* (based on lighter seed

color). The tissue was ground in a cetyltrimethylammonium bromide (c-tab) buffer [2% (w/v) (c-tab), 1.4M NaCl, 0.2% (v/v) 2-mercaptoethanol, 20mM EDTA, pH 8.0 and 100mM Tris, pH 8.0] and incubated at 65°C for 30 minutes, followed by extraction with chloroform: isoamyl alcohol (24:1), and precipitation of nucleic acids with isopropanol. Precipitated DNA was washed with 70 percent ethanol and re-suspended in Tris-EDTA buffer (10mM Tris, pH 8.0, 1mM EDTA, pH 8.0) to stabilize the DNA for long term storage, generally at -20°C (Liu et al., 1999).

An alternative DNA isolation procedure was employed to facilitate rapid analysis of all F_2 progeny (phenotypically wild type and *tt9*). In this case, DNA was isolated from fresh leaf tissue by grinding in 0.5M NaOH (approximately 1μl per 1mg leaf tissue), boiling for 30 seconds and neutralizing the base by adding Tris-EDTA buffer (200mM Tris, pH 8.0, 1.2mM EDTA, pH 8.0) in a 10:1 ratio of buffer to NaOH.

PCR-based molecular markers were used initially to score mapping lines at locations flanking the region containing *TT9* (8.62Mbp or 10.52Mbp and 11.1Mbp on chromosome 3) and subsequently to test lines in which at least one recombination event was observed between the flanking markers at relevant internal sites (10.58Mb, 10.64Mb, 10.67Mb, 10.72Mb, 10.78Mb). The PCR was carried out using primers designed to amplify regions containing polymorphic loci (Appendix, Table A-1). All reactions involving simple sequence length polymorphism (SSLP)-based markers were carried out with an initial denaturing step at 94°C for 30 seconds followed by 35 cycles of 30 seconds at 94°C, 30 seconds at the appropriate annealing temperature (Appendix, Table A-2) and 60 seconds per 1000bp product length at 72°C. The conditions for derived cleaved amplified polymorphic sequence (dCAPS)-based markers differed only in that the first ten cycles were carried out at a lower annealing temperature (2-4°C) to account for the mismatched nucleotide(s) in one of the primers. The products of the PCR were separated

according to size by electrophoresis in 3-3.5% (w/v) agarose/TBE (89mM boric acid, 89mM Tris, and 2mM EDTA, pH 8.0) gels stained with ethidium bromide (0.5μg/ml) and scored against Col, L*er* and Het controls. In the case of the dCAPS markers, prior to separation by electrophoresis, PCR products were also digested with the appropriate restriction enzyme (Appendix, Table A-2) using 1-5 units enzyme per 20μl PCR reaction with additional enzyme buffer at a 1X concentration (Appendix, Table A-3) in a 40μl total volume and incubating at 37°C for a minimum of 4 hours.

RNA expression analysis of candidate genes

RNA was extracted from 30-60mg of silique tissue – harvested under liquid nitrogen from L*er* or *tt9* plants grown under identical conditions and stored at -70°C – using the SV Total RNA Isolation System (Promega, #Z3101) according to the manufacturer's recommendations, and quantified spectrophotometrically. RNA expression was assayed for nine genes located within the region containing *tt9* as well as for *Chalcone Synthase* (*CHS*), *BANYULS* and *Transparent Testa 10* (*TT10*) using Real Time PCR. L*er* and *tt9* cDNA was synthesized using a Bio-Rad iScript™ cDNA synthesis kit (Bio-Rad #170-8890) which includes oligo (dT) and random hexamer primers. The reactions were carried out according to the manufacturer's recommendations using 500ng L*er* or *tt9* RNA as template. Real-time PCR was carried out in a Bio-Rad MiniOpticon Thermal Cycler using iQ™ SYBR® Green Supermix (Bio-Rad #170-8880) according to the manufacturer's recommendations. The reactions were assembled in replicates of three using gene-specific primers (Appendix, Table A-4). The RT-PCR program consisted of an initial denaturing step at 94°C for 3 minutes followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 56°C and 1 minute 15 seconds at 72°C. The resulting data were analyzed

with the Opticon Monitor 3 Software to determine the minimum number of cycles required to break background fluorescence (threshold cycle). A melting curve was taken for each of the reactions to determine whether the reactions were specific. The specificity of the reactions was also confirmed by separating the products by electrophoresis in 1% agarose/TBE gels stained with ethidium bromide.

T-DNA tagged lines

T-DNA insertion lines for the genes within the mapped interval were found using the Salk Institute Genomic Analysis Laboratory (SIGnAL) T-DNA Express database. Seeds for 13 T-DNA tagged lines harboring insertions in one of 10 candidate genes within the mapping interval were obtained from the *Arabidopsis* Biological Resource Center (ABRC, Columbus, OH). Approximately 20 seeds from each line were cultivated as described above. Plants were scored on the basis of seed-color by comparison with wild-type and *tt9*.

Cloning and *in planta* transformation

A clone of At3g28340 was generated by PCR using AccuPrime™ *Pfx* DNA polymerase (Invitrogen #12344-024) and primers designed to amplify a section including roughly 1000bp of promoter and 500bp downstream of the translational stop codon while also incorporating *HindIII* recognition sequences at each end of the amplicon (Forward: GCAGAAGCTTTATGTAGAGA TGCACATGG, Reverse: GCAGGATCCAGCTCCAAATATCGTTGCA). The PCR reaction was assembled (10μM For primer, 10μM reverse primer, 1x *Pfx* reaction buffer, 250ng template DNA, 2.5 units *Pfx* polymerase) and run using a program consisting of an initial denaturing step at 95°C for 2 minutes, followed by 15 cycles of 95°C for 15 seconds, 54°C for 30 seconds and

68°C for 4 minutes, 25 cycles of 95°C for 15 seconds and 68°C for 4 minutes and a final extension step at 68°C for 7 minutes. The PCR product was purified using a QIAquick PCR Purification kit (Qiagen, #28104) and digested in a total volume of 50µl with 5 units *Hind*III (Promega) in 1x Buffer E (Promega). The construct was ligated into the binary vector pGreenII0229 which carries a gluphosinate resistance marker for plant selection. The pGreenII0229-At3g28340 construct, once generated, was introduced into *Agrobacterium tumefaciens* strain GV3101 by freeze/thaw transformation in chemically competent cells (Pikaard, 2005).

For complementation experiments using *in planta* transformation *transparent testa* 9 seeds were sown, without prior stratification, three per 2"x 2" container in moistened potting mix (Scotts Metro Mix 360) and covered with window screen to facilitate dipping (described below). The seeds were allowed to germinate in a growth chamber under the conditions described above. After approximately one week, the photoperiod was changed to 16 hour light/8 hour dark. *In planta* transformation was performed by inverting whole plants (all above-ground tissues) into a 5% sucrose solution containing GV3101-At3g28340 (OD_{600} between 0.8 and 1.5), and 0.5ml/L of the surfactant Silwet L-77 (Vac `n` Stuff®, Lehle Seeds). Dipped plants were placed in a sealed tray and stored out of direct light overnight then returned to the growth chamber under conditions described above. The transformation procedure was repeated with the same plants one week after the initial dip (Clough and Bent, 1998).

V. Results

Mapping of the *TT9* Locus

Transparent testa 9 was among 10 distinct mutants isolated in a screen of chemically mutagenized or irradiated plants displaying diminished seed-coat pigmentation (Koornneef, 1990). It was mapped previously to chromosome three near *GL1* by linkage analysis with a *GL1* marker line (Shirley et al., 1995). Genetic mapping was carried out in this lab in two phases. The locus was first narrowed to a region of approximately 580kb (between 10.52Mb and 11.1Mb) by analyzing approximately 150 *tt9* lines from among more than 500 F_2 plants resulting from a cross between a mutant *tt9* plant in the L*er* background and a wild-type plant in the Col background. Once it was determined that markers at 10.52Mb and 11.1Mb flanked *tt9*, rapid screening of more than 700 additional lines, both wild-type and *tt9* was carried out by first genotyping plants at 10.52 and 11.1 and identifying lines with a single recombination event between these two markers. Given the estimated recombination rate between these two markers from the previous lot of lines analyzed (~4.4%), it was estimated that under these conditions, a double recombinant (e.g. L*er* at 10.52 and 11.1 but heterozygous (Het) through a discrete interval between these two markers) would only be missed approximately 0.2 percent of the time. Therefore, out of our approximately 700 lines, perhaps only 1-2 double recombinants might have been misidentified as lines with no recombination and discarded.

Lines that were identified as having a single recombination event between 10.52 and 11.1 were grown to maturity and scored wild-type or *tt9* based on seed-color phenotype. In some cases wild-type plants were carried out to the F_3 generation to determine whether they segregated for the *tt9* phenotype, which would indicate that the F_2 parent had been Het at the *TT9* locus. Those which failed to segregate were scored as having an F_2 parent that was Col at *TT9*. All of

the recombinant lines were further analyzed at adjacent markers within the mapped region located at 10.64Mb and 10.67Mb. A single line (line 1642) was identified which indicated that 10.67 was the right border of the region containing *tt9*, however given that this particular line in fact represented a double recombinant, and given the possibility of error in identifying a single mutant, further analysis was performed with the other recombinant lines using markers at 10.58, 10.72 and 10.78. It should be pointed out, however, that analysis of a pool of tissue collected from F_3 plants originating from line 1642 confirmed the genotype of this line for all of the markers tested, and all of the F_3 plants were *tt9* with regard to seed-color phenotype, indicating that the F_2 had been correctly scored as *tt9*.

	10.52	**10.58**	**10.64**	**10.67**	**10.72**	**10.78**	**11.1**
Recombinants between marker and *tt9*	13	2	0	1	2	7	41
Total plants analyzed	797	794	797	797	785	787	797
Approximate map distance	0.816	0.126	<0.063*	0.063	0.127	0.445	2.572

Table 2: Summary of Mapping Data. The total number of plants analyzed is based on the number of plants that were either physically scored at the given marker, or for which the genotype could be reasonably deduced based on the genotype at adjacent markers. The variation in the number of plants analyzed is due to some plants dying without setting seed or before subsequent analysis could be undertaken. *Map distance calculated for one recombinant chromosome.

Based on the results of the mapping data obtained, the *TT9* locus has been placed between 10.58Mb and 10.72Mb on the third chromosome at distances of 0.126 map units and 0.127 map units respectively (Table 2). The interval defined by these two markers contains 30 putative genes. Although there is evidence that 10.67Mb may in fact be the right border, linked to *tt9* by 0.063 map units, subsequent analysis of candidate genes was based on the broader map for reasons outlined above regarding the single line defining the right border. Although there were only two lines defining each the 10.58 and 10.72 borders, the fact that there was more than one independent line to confirm this interval provides greater confidence in the legitimacy of these results. Furthermore, the plant lines which were informative in defining the mapping

interval were re-scored using independently isolated tissue samples, which confirmed the previous results.

In Silico Wild-type RNA Expression Analysis

Wild-type RNA expression levels for the 30 genes found within the mapped interval containing *TT9* were assessed *in silico* by searching the publicly available data in the Genevestigator microarray database (Zimmermann et al, 2004). Given that *tt9* plants are affected at the level of seed coat pigmentation, it would be expected that the gene responsible would be expressed in seeds and/or siliques. Based on this inference, microarray data was examined for each of the genes to identify those which were expressed in the specified tissues. The expression levels are summarized below (Table 3). There were several predicted genes within the interval which were not found within the Genevestigator database, and therefore it was not possible to obtain expression data for these genes. The data obtained in this way were used as a guide toward further analysis of candidate genes such that efforts could be focused on evaluating those genes whose expression pattern suggested a plausible role in seed coat pigmentation. All genes which were expressed above background levels in either siliques or seeds are indicated in boldface type below (Table 3). Because it has been observed with respect to the normalized data presented within the Genevestigator database that genes whose signal intensity was below 200 for a given experiment were not expressed above background "noise", a signal intensity of 200 was recommend as the threshold for determining whether a gene is expressed in a given experiment.

Gene	Description	Siliques	Seed
At3g28315	Retrotransposon	Not found	Not found
At3g28320	**Expressed protein**	**93**	**242**
At3g28330	**F-box family protein**	**240**	**239**
At3g28340	**Galactinol synthase**	**600**	**200**
At3g28345	**P-glycoprotein-related**	**48**	**400**
At3g28350	Hypothetical protein	19	35
At3g28360	**P-glycoprotein-related**	**850**	**200**
At3g28370	Hypothetical protein	Not found	Not found
At3g28380	P-glycoprotein-related	<100	<100
At3g28390	P-glycoprotein-related	<100	<100
At3g28400	hAT-like transposase	Not found	Not found
At3g28410	F-box family protein	17	22
At3g28415	P-glycoprotein-related	Not found	Not found
At3g28420	**Expressed protein**	**227**	**84**
At3g28430	**Expressed protein**	**217**	**300**
At3g28450	**Leucine-rich repeat, tm protein kinase**	**800**	**800**
At3g28460	**Expressed protein**	**700**	**700**
At3g28470	MYB family TF	Not expressed	Not expressed
At3g28480	**Oxidoreductase**	**1100**	**1100**
At3g28490	Oxidoreductase	Not expressed	Not expressed
At3g28500	60S acidic ribosomal protein	N/A	N/A
At3g28510	Expressed protein	Not expressed	Not expressed
At3g28520	AAA-type ATPase	55	34
At3g28530	**Expressed protein**	**280**	**270**
At3g28540	**AAA-type ATPase**	**580**	**39**
At3g28550	Proline rich extension-like family	Not expressed	Not expressed
At3g28560	Expressed protein	Not expressed	Not expressed
At3g28570	AAA-type ATPase	Not found	Not found
At3g28580	AAA-type ATPase	61	29
At3g28590	Expressed protein	178	98

Table 3: *In silico* Gene Expression Analysis. The values listed represent the signal intensity obtained for a given gene in the tissues indicated. They represent the mean values obtained for all arrays in the database. Genetic loci indicated in bold are expressed above background in either seeds or siliques or both.

Real-Time PCR Analysis of Mutant versus Wild-Type Gene Expression

Mutant and wild-type expression of nine candidate genes within the mapped interval was assayed using Real-Time Reverse Transcription (RT)-PCR. Actin-2 was initially chosen as a control since this gene is fairly ubiquitously expressed in plants, however difficulty encountered with the gene specific primers used resulting in a failure to amplify Actin-2 led us to instead use chalcone synthase (CHS) as expression of this gene has been shown to be unaffected in *tt9* (Shirley et al., 1995). Silique tissue was chosen for analysis for reasons outlined above concerning the putative involvement of *tt9* within the plant. It was expected that a disruption in the nucleotide sequence of a gene may impact the expression level of that gene within the plant. Several plausible mechanisms exist through which this might occur. If the mutation occurred in a promoter region, for example, it may interfere with appropriate signaling molecules or transcription factors, thereby decreasing the efficiency with which the gene is transcribed and ultimately leading to decreased levels of the transcript. Mutations affecting the stability of the RNA, including but not limited to those occurring in the 3` untranslated region (UTR), may also result in fewer transcripts detected since the RNA may be more readily degraded.

Among the nine candidate genes selected for analysis, none exhibited differences in transcript levels between mutant and wild-type (Figure 13). At3g28480 appeared in some assays to be up-regulated in *tt9* as compared to wild-type, however these results were not reproducible and therefore do not suggest a significant change in expression for this gene. Melting curves for each of these reactions, as well as gel electrophoresis of the resulting products confirmed the specificity of the reactions (data not shown). While a result that indicated a decreased transcript level for a particular gene in the mutant plant as compared to wild-type might have provided strong evidence for the correlation of that gene with the mutant, results indicating little or no

difference between mutant and wild-type transcript levels, as we observed, may not definitively rule out such candidates. It is conceivable, for example, that a single nucleotide deletion may result in a frame shift resulting in a transcript that encodes either a truncated or otherwise significantly altered protein and yet not sufficiently disrupt the stability of the mRNA to result in reduced transcript levels. In this case the mutation still results in a non-functional protein leading to the particular phenotype observed, however this may not be discernible at the RNA level.

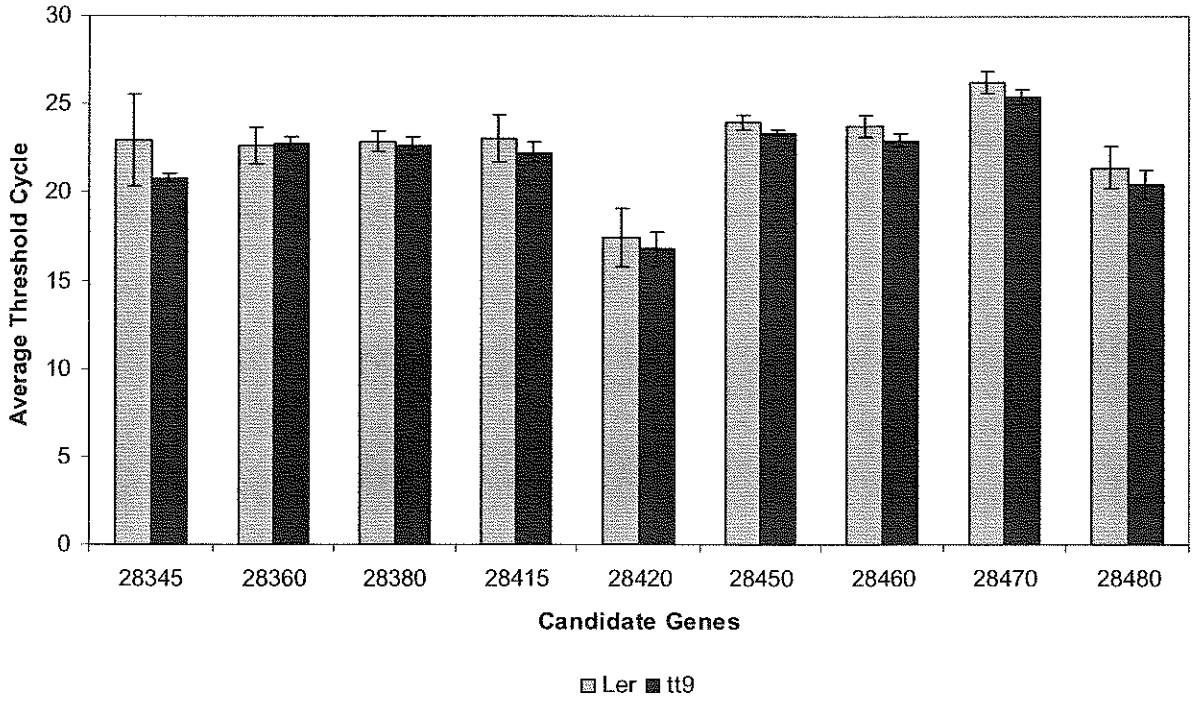


Figure 13: Real-Time PCR Data for Nine Candidate Genes. The graph illustrates the results obtained following real-time PCR analysis of nine genes within the mapped interval. Threshold cycles were averaged over three replicate reactions for each gene assayed. The results indicate that none of the nine genes tested show an altered transcript accumulation. Error bars indicate +/- 1 standard deviation from the mean.

BANYULS (*BAN*), the gene encoding anthocyanin reductase in *Arabidopsis* was also assayed since several regulatory mutants have been shown to produce fewer or no transcripts for this gene (Nesi et al., 2000, 2001; Pourcel et al., 2005; Shirley et al., 1995; Xie et al., 2003). Expression of *TT10,* a recently characterized gene which was shown to encode a laccase-like

polyphenol oxidase, was also assayed in *tt9*. *TT10* acts late in flavonoid biosynthesis and does not appear to be regulated by the same factors coordinating *BANYULS* expression (Pourcel et al., 2005). Given that *tt9* also appears to be affected late in the pathway, it was inferred that a change in *TT10* transcript level in our mutant may suggest a regulatory role for *tt9* in the expression of *TT10*. Real-time PCR data indicated that *TT10* appeared to be expressed at equal levels in the mutant and wild-type, while *BAN* appeared to be down-regulated in *tt9* between 4- and 16-fold (Figure 14). Melting curves as well as gel electrophoresis indicated that all of the reactions were specific (Figure 15). *BAN* expression was assayed four separate times and with three independent RNA samples and in each case there was a minimum of a two-cycle lag between the threshold cycle for L*er* and that for *tt9* (Figure 16). This may suggest a role for *TT9* in the regulation of *BAN*.

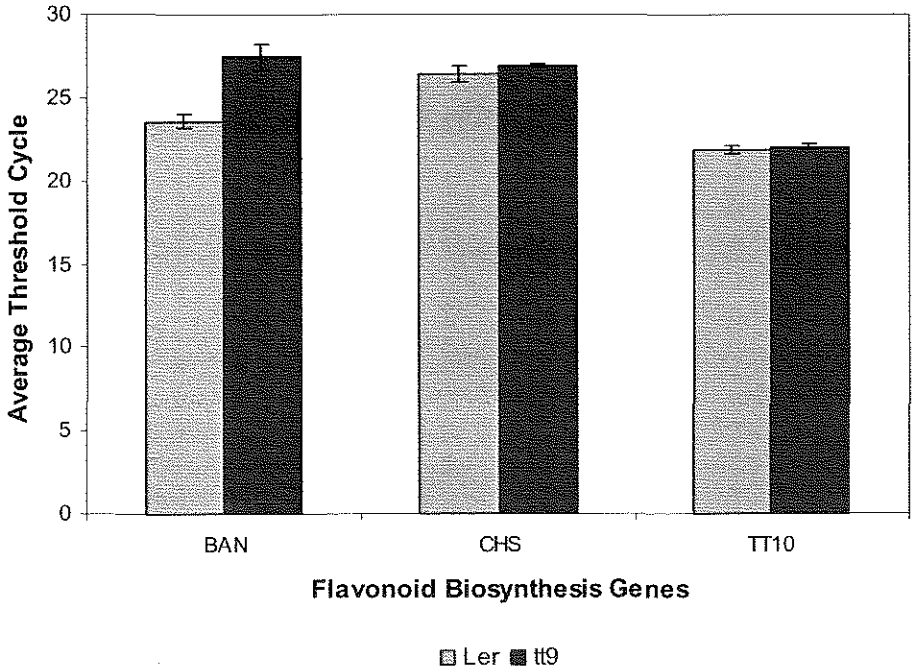


Figure 14: Real-Time Data for *BAN*, CHS and *TT10*. The graph indicates the average threshold cycle for replicates of three identical reactions for the genes assayed. Based on the greater threshold cycle required for *BAN*, this gene appears to be slightly down-regulated in the mutant plant. Error bars indicate +/- 1 standard deviation from the mean.

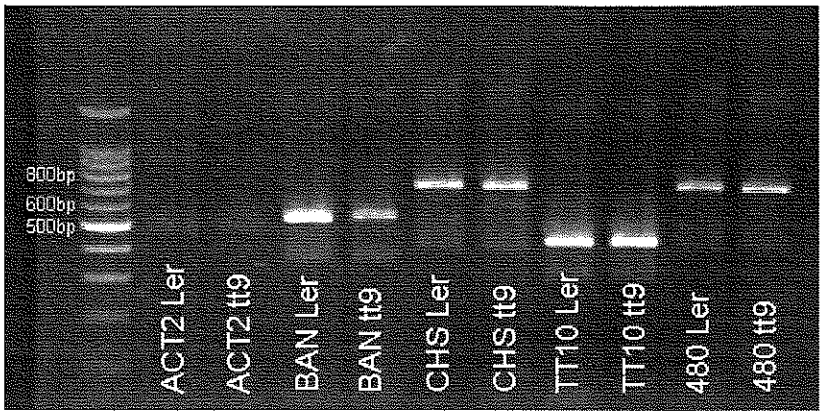


Figure 15: RT-PCR Results for Actin-2, *BANYULS*, CHS and TT10. Actin-2 was initially included as a positive control, however because the Actin-2 gene failed to amplify, CHS was instead used as the positive control. The gel indicates that single products of the appropriate sizes were amplified. *BANYULS* expression appears to be down-regulated in *tt9* as compared to wild-type. Real-time data based on the minimum threshold cycle required to rise above background fluorescence indicated that *BAN* expression was between 4- and 16-fold reduced in *tt9*.

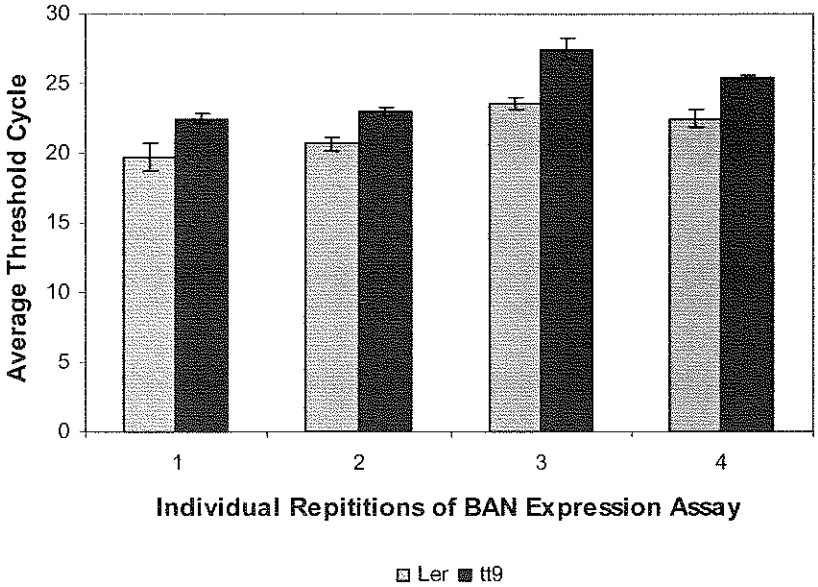


Figure 16: Real-Time PCR Data for Four Separate *BAN* Assays. The graph indicates the results obtained in four separate real-time PCR assays. Each bar represents an average of three replicate reactions for the experiment listed. Error bars indicate +/- 1 standard deviation from the mean.

T-DNA tagged lines

Thirteen T-DNA tagged lines from either the Salk Institute for Genomic Research or Syngenta Arabidopsis Insertion Library (SAIL) collections with insertions in ten candidate genes within the region of interest were obtained from the *Arabidopsis* Biological Resource Center (ABRC, Columbus, OH). The tagged lines were generated by *Agrobacterium*-mediated transformation of Col plants and the insertion sites were determined either by a sequence-based high-throughput insertion site recovery system or by thermal asymmetric interlaced PCR (TAIL-PCR), and Basic Local Alignment Search Tool (BLAST) comparison of product sequences with the published Col sequence (Alonso et al., 2003; Sessions et al., 2002). It was expected that any lines homozygous for the T-DNA insertion would be knock-outs with regard to the gene in question. Therefore, if any such lines exhibited the *tt9* phenotype, it could be inferred that the insertion line was allelic to *tt9* and the precise gene involved would be known. Subsequent cloning and complementation with the gene in question could then be performed to confirm allelism.

A minimum of ten individual plants were scored for each T-DNA insertion line, of which all displayed wild-type seed color (Table 4). The seed stocks obtained from the ABRC represented T3 or T4 generation transformants which could have been homozygous or hemizygous for the insertion, or wild-type (i.e. no insertion). Because the seed was pooled for distribution, it was expected the seed stocks would be homozygous, hemizygous or wild-type in a theoretical ratio of approximately 1:2:1. If these ratios hold true, and if a Poisson distribution is assumed, a minimum of twelve individual T3 or T4 plants must be analyzed in order to ensure that at least one homozygous line would be found within the analyzed lines with a probability greater than 0.95. Based on these findings it can be asserted with reasonable confidence that the

genes for which T-DNA tagged lines were analyzed may be ruled out as candidates for *tt9* with the exception of At3g28340, for which only ten individual plants from the corresponding Salk line were analyzed.

Gene	Insertion Line	Location	Plants scored
At3g28340	SALK_072808	1000-Promoter	10 (WT)
At3g28345	SAIL_177_A01	Exon	29 (WT)
	SAIL_1187_C04	Exon	28 (WT)
At3g28360	SALK_148699	Exon	21 (WT)
	SALK_006491	Exon	35 (WT)
At3g28370	SALK_127144	Exon	18 (WT)
At3g28420	SALK_064362	1000-Promoter	20 (WT)
At3g28430	SAIL_289_G06	Exon	16 (WT)
At3g28450	SALK_111477	Exon	16 (WT)
At3g28490	SALK_067682	Exon	20 (WT)
At3g28570	SALK_007203	Exon	20 (WT)
	SALK_076105	Exon	22 (WT)
At3g28580	SALK_031517	Exon	16 (WT)

Table 4: Summary of T-DNA Tagged Lines Analyzed to Date. All T-DNA tagged lines were obtained from the *Arabidopsis* Biological Resource Center (ABRC, Columbus, OH). Salk lines belong to the SIGnAL Salk insertion collection and were generated by vacuum infiltration of Col plants using *Agrobacterium* containing the pROK2 vector (Alonso et al., 2003). SAIL lines belong to the SAIL insertion collection and were generated by vacuum infiltration of Col plants using *Agrobacterium* containing either pCSA110 or pDAP101 (Sessions et al., 2002). "1000-Promoter" indicates that the T-DNA is found in the promoter region within 1000bp of the translational start codon.

Cloning and Complementation

Six genes located between 10.58Mbp and 10.67Mbp – At3g28340, At3g28345, At3g28360, At3g28390, At3g28420 and At3g28430 – were initially selected for a focused cloning approach aimed at complementing the *tt9* phenotype. The goal of this approach was to select several candidate genes which appeared to be fairly strong candidates and to generate wild-type constructs of these genes for *in planta* transformation. This was performed with the expectation that if the gene in question was the same gene disrupted in *tt9,* insertion of a wild-type copy would complement the *tt9* mutation, thus allowing the mutant plant to produce seeds with wild-type tannin accumulation. The genes selected for complementation were chosen based

on map position as well as on wild-type expression levels above background in siliques and/or seeds as reported in Genevestigator or as obtained in our lab (data not shown) (Alonso et al., 2003; Zimmermann et al., 2004). Gene specific primers were designed to amplify the coding region of each of the genes in addition to approximately 1000bp of promoter and 500bp of the 3`UTR. Of the six genes attempted, four were able to be amplified to generate a product of the appropriate size; however only one – At3g28340 – was subsequently successfully cloned (Figure 17). In every case except At3g28340, multiple, non-specific products were generated in the PCR reactions frequently to the exclusion of the target product. It was inferred that the difficulty encountered in cloning these genes may have been in part due to primers binding to highly conserved sequences in multiple locations and thus producing non-specific products. T-DNA tagged lines with exon insertions were later scored for most of the genes which we were unable to clone, and all of these lines produced seeds with wild-type coloration. Based on these findings, it was decided that it was no longer imperative to pursue the remaining genes for cloning and complementation.

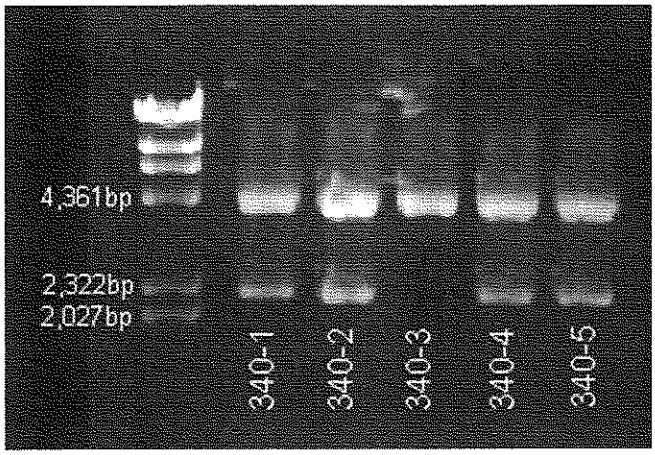


Figure 17: Agarose Gel Separation of Digested Vector Containing At3g28340 Clone. DNA was isolated from cells derived from single colonies growing on the At3g28340 transformation plates. The presence of two bands in the gel in the case of 340-1, 340-2, 340-4 and 340-5 indicates the presence of the At3g28340 insert in the pGreenII0229 vector.

At3g28340 was cloned into the binary vector pGreenII0229 and co-transformed into *Agrobacterium* strain GV3101 with pSoup, a plasmid carrying the *vir* genes required for successful plant transformation, as well as the *RepA* gene which acts *in trans* on the pSa *Ori* (origin of replication) sequence in pGreen. Co-transformation with pSoup is necessary in order for the pGreen vector to replicate in *Agrobacterium*. Mutant *tt9* plants were transformed according to the Floral Dip method of Clough and Bent (1998). The overall health of the *tt9* plants used was not optimal for this procedure; however the timing of the experiment was such that this could not be readily avoided. Due to this factor, very few seeds were produced and collected from the dipped plants and given the low reported transformation efficiencies associated with this method, it appears unlikely that any of the fewer than 100 seeds collected will represent transformants (Clough and Bent, 1998). For this reason, the transformation was repeated with 10 additional plants, from which seed has now been collected. pGreenII0229 carries a gluphosinate resistance gene as a plant selectable marker which allows transformants to be selected by spraying seedlings with Basta® and retaining those which are resistant to this treatment. Selection of transformants has not yet been carried out as the required herbicide is currently unavailable in this area of the United States.

VI. Discussion

With the sequencing of the *Arabidopsis* genome in 2000, there has been a fervent drive to move beyond the genome sequence to a thorough understanding of the functions of the more than 25,000 putative genes. Project 2010, for example, is an initiative aimed precisely at achieving this goal – i.e. at identifying a function for every *Arabidopsis* gene by the year 2010. This has led to the development of multiple tools for studying gene function in *Arabidopsis*, providing investigators with significant resources for carrying out work in this area.

The flavonoid biosynthetic pathway is one which has been extensively studied in *Arabidopsis* and one which has not been a stranger to the initiatives described above. In recent years several new structural and regulatory factors have been uncovered, particularly through investigation of insertion-mutants with altered seed-color phenotypes. Most of the *transparent testa* mutants first described by Marteen Koornneef (1987) have now been characterized, providing insight into the precise transcriptional regulation of certain elements of the pathway, as well as further insight into late steps in the pathway following the branch point to proanthocyanidin synthesis. Despite our current understanding, there remain ambiguous points for which further characterization of mutants in this pathway may potentially offer significant insight. Furthermore, while *Arabidopsis* is not itself a significant target for biotechnological manipulation, the insights gained through investigation in this relatively simple model provide important insights for other species.

Through map-based cloning, the *transparent testa 9* (*tt9*) locus has been mapped to a 140kb region on chromosome three, near the *GL1* locus. Representing significant progress in the effort to identify the gene responsible for *tt9*, this narrow map location was achieved in part using a rapid and efficient method for screening mapping lines. The nature of genetic mapping as

carried out in this lab is such that as the mapping interval becomes narrower, it becomes increasingly difficult to identify lines that may be relevant for further narrowing the interval. This is due to the fact that these methods rely upon recombination events occurring between a given marker and the mutant locus, and such events become rarer as the interval separating the marker from the locus of interest becomes narrower. For this reason, the development of cost-effective, high-throughput protocols allowing for the analysis of hundreds of mapping population plants in a short period of time becomes increasingly important as the project progresses. Although not truly high-throughput, the method developed and used in the later stages of the mapping initiative in this lab provided a significant advantage over previous methods in terms of efficiency, and proved instrumental in allowing for the identification of plant lines which define our current mapping interval.

With the region containing *tt9* narrowed to 140kb, an interval defined and confirmed by only a handful of recombinant lines, it became clear that it would be necessary to investigate the 30 predicted genes contained within this interval. Toward this end, several methods were employed including assessment of wild-type expression levels in siliques and seeds, comparative RNA expression assays between mutant and wild-type, sequencing of predicted coding regions, analysis of T-DNA tagged lines and generation of wild-type gene constructs for complementation. Each of these methods harbors certain benefits and drawbacks, but ultimately the approach taken was meant to maximize the number of genes which could be evaluated in at least one of the ways described above, with the exception of the *in silico* RNA expression analysis, which was assessed for nearly all of the genes.

Roughly half of the 30 putative genes can be tentatively ruled out based on data obtained from mutant versus wild-type RNA expression assays and/or T-DNA tagged lines (summarized

in Appendix, Tables A-5 and A-6). Real-time RNA expression data for nine genes within the interval revealed no notable differences between mutant and wild-type expression. Sequencing of eight of these genes (data not shown) revealed no differences between the mutant and wild-type sequences in all but one case. Sequence data from At3g28460 revealed a single nucleotide change however this mutation occurred in the third nucleotide in a codon, and would not result in a change in the encoded amino acid. Sequencing only involved the coding regions of the genes of interest. However, while it may be possible that a mutation in a promoter region or un-translated region (UTR) may result in an altered transcript, in a non-functional protein and ultimately in phenotypic alterations, it would be expected that such a mutation would affect the transcript level for that gene in the mutant as compared to wild-type. As stated above, no changes in RNA expression levels were observed.

An additional five genes may be ruled out based on results from analysis of T-DNA tagged lines which revealed that all lines analyzed produced seeds with wild-type pigmentation. It should be pointed out, however, that none of the T-DNA tagged lines were confirmed homozygous by molecular analysis. Although it might have been expected based on the assumed ratio of wild-type to hemizygous to homozygous lines that at least one of the lines scored would have been homozygous (with probability greater than 0.95 in most cases) the expected ratios assume that all plants produce an equal number of seeds and that all seeds are equally viable. This however, is probably not entirely accurate particularly in light of the apparent germination defect in *tt9*. Furthermore, as indicated above (Table 4), not all of the lines harbored insertions within exons of the predicted coding region. The extent to which insertions in promoter regions, UTRs or introns might affect the phenotype of the resulting plants is not known, although there

have been reported cases of insertion lines such as these which had sufficient interference with the native gene to cause phenotypic alterations (B. Winkel-Shirley, personal communication).

In order to gain greater confidence in the results from the T-DNA tagged lines, molecular analysis of individual insertion lines should be carried out in order to identify homozygotes, which may then be grown to maturity and scored on the basis of seed color. Future work in this lab will be aimed in this direction, both in confirming the results obtained to date by re-evaluating these lines, as well as by analyzing all the remaining genes within our mapping interval for which insertion lines are available. Homozygous lines will be identified by performing PCR using a left border (LB)-specific primer and a gene specific primer, as well as with two gene specific primers such that a homozygous line may be distinguished from a hemizygous or wild-type line based on the success of the PCR amplification and the size of the product generated (reviewed in Østergaard et al., 2004). If a homozygous line goes on to exhibit a wild-type phenotype with regard to seed color it will then be determined with greater confidence that the gene in question is not responsible for *tt9*.

Typically employed as the final step in confirming that a candidate gene is responsible for the mutant being investigated, cloning and complementation were pursued in a focused manner in order to provide definitive evidence as to whether the gene in question was responsible for *tt9*. One clone has been generated to date with which mutant plants were transformed. The putative transformants are pending selection. Future efforts with regard to cloning will be focused on evaluating genes which are not amenable to other methods, particularly those for which T-DNA tagged lines are not available.

Considerable progress has been made in the effort to isolate the gene responsible for *tt9*. Although the gene remains elusive to date, efforts continue in the evaluation of candidate genes

and the strategies as outlined above should allow for definitive analysis of all remaining genes within the mapped interval. It is expected that by isolating and characterizing the *tt9* locus, the mechanisms involved in flavonoid biosynthesis may be further refined and understood, potentially lending to efforts involved in the manipulation of this pathway. It has been shown that *tt9* seedlings produce near-normal anthocyanin pigments suggesting that *tt9* acts late in the biosynthetic pathway (Shirley et al., 1995). This inference is further supported by data which suggest that *tt12*, which encodes a multi-drug and toxic compound extrusion (MATE) family transporter supposed to be involved in proanthocyanidin sequestration, is epistatic to *tt9* (Debeaujon et al. 2001). Given these findings, it is expected that characterizing *tt9* may lead to greater depth in understanding of late events in the flavonoid, and specifically the PA pathway; events which to date remain relatively ambiguous. Furthermore, given that *tt9* appears to have a negative effect on germination, elucidating its specific function may lead to a better understanding of how flavonoids are involved in this physiological process.

VII. References

1. Abrahams S, Tanner GJ, Larkin PJ, Ashton AR (2002) Identification and biochemical characterization of mutants in the proanthocyanidin pathway in *Arabidopsis*. Plant Physiology **130**: 561-576

2. Abrahams S, Lee E, Walker AR, Tanner GJ, Larkin PJ, Ashton AR (2003) The *Arabidopsis TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. The Plant Journal **35**: 624-636

3. Albert S, Delseny M, Devic M (1997) BANYULS, a novel negative regulator of flavonoid biosynthesis in the *Arabidopsis* seed coat. The Plant Journal **11**: 289-299

4. Alonso JM, Stepanova AN, Leisse T, Kim C, Chen H, Shinn P, Stevenson D, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers C, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmid I, Weigel D, Brogden D, Zeko A, Crosby W, Berry C, Ecker J (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301**: 653-657

5. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. The Plant Journal **39**: 366-380

6. Baxter IR, Young JC, Armstrong G, Foster N, Bogenschutz N, Cordova T, Peer WA, Hazen SP, Murphy AS, Harper JF (2005) A plasma membrane H+-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences **102**: 2649-2654

7. Bechtold N, Jaudeau B, Jolivet S, Maba B, Vezon D, Voisin R, Pelletier G (2000) The maternal chromosome set is the target of the T-DNA in the *in Planta* transformation of *Arabidopsis thaliana*. Genetics **155**: 1875-1887

8. Bieza K, Lois R (2001) An *Arabidopsis* mutant tolerant to lethal ultraviolet-B levels shows constitutively elevated accumulation of flavonoids and other phenolics. Plant Physiology **126**: 1105-1115

9. Bravo L (1998) Polyphenols: chemistry, dietary sources, metabolism, and nutritional significance. Nutrition Reviews **56**: 317-333

10. Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proceedings of the National Academy of the Sciences **85**: 6856-6860

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

12. Cos P, De Bruyne T, Hermans N, Apers S, Vanden Berghe D, Vlietinck AJ (2003) Proanthocyanidins in health care: Current and new trends. Current Medicinal Chemistry **10**: 1345-1359

13. de la Riva GA, González-Cabrera J, Vázquez-Padrón R, Ayra-Pardo C (1998) *Agrobacterium tumefaciens*: a natural tool for plant transformation. Electronic Journal of Biotechnology **1**

14. Debeaujon I, Leon-Kloosterziel KM, Koornneef M (2000) Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. Plant Physiology **122**: 403-414

15. Debeaujon I, Peeters AJM, Leon-Kloosterziel KM, Koornneef M (2001) The *TRANSPARENT TESTA12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. Plant Cell **13**: 853-872

16. Desfeux C, Clough SJ, Bent AF (2000) Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation be the *Arabidopsis* floral-dip method. Plant Physiology **123**: 895-904

17. Devic M, Guilleminot J, Debeaujon I, Bechtold N, Bensaude E, Koorneef M, Pelletier G, Delseny M (1999) The *BANYULS* gene encodes a DFR-like protein and is a marker of early seed coat development. The Plant Journal **19**: 387-398

18. Dixon RA, Xie D-Y, Sharma SB (2005) Proanthocyanidins - a final frontier in flavonoid research? New Phytologist **165**: 9-28

19. Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? Journal of Experimental Botany **55**: 1445-1454

20. Hellens R, Mullineaux P, Klee H (2000) A guide to *Agrobacterium* binary Ti vectors. Trends in Plant Science **5**: 446-451

21. Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiology **129**: 440-450

22. Johnson CS, Kolevski B, Smyth DR (2002) *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. Plant Cell **14**: 1359-1375

23. Kitamura S, Shikazano N, Tanaka A (2004) *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. The Plant Journal **37**: 104-114

24. Konieczny A, Ausubel FM (1993) A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. The Plant Journal **4**: 403-410

25. Koornneef M (1981) The complex syndrome of *TTG* mutants. Arabidopsis Information Service **18**: 45-51

26. Koornneef M, Luiten W, de Vlaming P, Schram AW (1982) A gene controlling flavonoid 3`-hydroxylation in *Arabidopsis*. Arabidopsis Information Service **19**: 113-115

27. Koornneef M, Hanhart CJ (1983) Linkage marker stocks of *Arabidopsis thaliana*. Arabidopsis Information Service **20**: 89-92

28. Koornneef M (1990) Mutations affecting the testa colour in *Arabidopsis*. Arabidopsis Information Service **27**: 1-4

29. Kreuzaler F, Ragg H, Fautz E, Kuhn DN, Hahlbrock K (1983) UV-induction of chalcone synthase mRNA in cell suspension cultures of *Petroselinum hortense*. Proceedings of the National Academy of the Sciences **80**: 2591-2593

30. Leon-Kloosterziel KM, Keijzer CJ, Koornneef M (1994) A seed shape mutant of *Arabidopsis* that Is affected in integument development. Plant Cell **6**: 385-392

31. Li J, Ou-Lee T-M, Raba R, Amundson RG, Last RL (1993) *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. Plant Cell **5**: 171-179

32. Liu D, Zhang S, Fauquet C, Crawford NM (1999) The *Arabidopsis* transposon TagI is active in rice, undergoing germinal transposition and restricted, late somatic excision. Molecular Genetics and Genomics **262**: 413-420

33. Lukowitz W, Gillmor CS, Scheible W-R (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiology **123**: 795-805

34. Meyers BC, Tej SS, Vu TH, Haudenschild CD, Agrawal V, Edberg SB, Ghazal H, Decola S (2004) The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. Genome Research **14**: 1641-1653

35. Mosier JS, Nonogaki H (2003) Integrative seed biology. Department of Horticulture at Oregon State University. www.science.oregonstate.edu/isb/arabidopsis.html

36. Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. The Plant Journal **14**: 387-392

37. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* Genes in *Arabidopsis* Siliques. Plant Cell **12**: 1863-1878

38. Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L (2001) The *Arabidopsis TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell **13**: 2099-2114

39. Nesi N, Debeaujon I, Jond C, Stewart AJ, Jenkins GI, Caboche M, Lepiniec L (2002) The *TRANSPARENT TESTA16* locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell **14**: 2463-2479

40. Nichenametla SN, Taruscio TG, Barney D, Exon JH (2006) A review of the effects and mechanisms of polyphenolics in cancer. Critical Reviews in Food Science and Nutrition **46**: 161-183

41. Østergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana*. The Plant Journal **39**: 682-696

42. Pelletier MK, Murrell JR, Shirley BW (1997) Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis* (Further evidence for differential regulation of "early" and "late" genes). Plant Physiology **113**: 1437-1445

43. Pelletier MK, Shirley BW (1996) Analysis of flavanone 3-hydroxylase in *Arabidopsis* seedlings (Coordinate regulation with chalcone synthase and chalcone isomerase). Plant Physiology **111**: 339-345

44. Pelletier MK (1997) Molecular and biochemical genetics of 2-oxogluterate-dependent dioxygenases required for flavonoid biosynthesis in *Arabidopsis thaliana*. Virginia Polytechnic Institute and State University

45. Pikaard C (2005) *Agrobacterium* growth and transformation. Washington University in St Louis. http://www.biology.wustl.edu/pikaard/

46. Pourcel L, Routaboul J-M, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005) *Transparent Testa10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. Plant Cell **17**: 2966-2980

47. Rein D, Lotito S, Holt RR, Keen CL, Schmitz HH, Fraga CG (2000) Epicatechin in human plasma: In vivo determination and effect of chocolate consumption on plasma oxidation status. The Journal of Nutrition **130**: 2109S-2114S

48. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P. The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. Nucleic Acids Research **31**, 224-228. 2003.

49. Sagasser M, Lu G-H, Hahlbrock K, Weisshaar B (2002) *A. thaliana TRANSPARENT TESTA 1* is involved in seed coat development and defines the WIP subfamily of plant zinc finger proteins. Genes & Development **16**: 138-149

50. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, Clarke JD, Cotton D, Bullis D, Snell J, Miguel T, Hutchison D, Kimmerly B, Mitzel T, Katagiri F, Glazebrook J, Law M, Goff SA (2002) A high-throughput *Arabidopsis* reverse genetics system. Plant Cell **14**: 2985-2994

51. Sharma S, Dixon RA (2005) Metabolic engineering of proanthocyanidins by ectopic expression of transcription factors in *Arabidopsos thaliana*. The Plant Journal **44**: 62-75

52. Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: Analysis of two *Arabidopsis transparent testa* mutations. Plant Cell **4**: 333-347

53. Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koorneef M, Ausubel FM, Goodman HM (1995) Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. The Plant Journal **8**: 659-671

54. Singer T, Burke E (2003) High-throughput TAIL-PCR as a tool to identify DNA flanking insertions. Methods in Molecular Biology **236**: 241-272

55. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**: 796-815

56. Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, Blundell TL, Esch JJ, Marks MD, Gray JC (1999) The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. Plant Cell **11**: 1337-1350

57. Winkel-Shirley B (2001) It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. Plant Physiology **127**: 1399-1404

58. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology **126**: 485-493

59. Xie D-Y, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science **299**: 396-399

60. Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiology **136**: 2621-2632

VIII. Appendix
Primer Sequences and Reaction Conditions

Primer name	Type	Forward Primer Sequence	Reverse Primer Sequence	Product sizes	
				Col	*Ler*
MKP8.62	SSLP	GATACACTCCTTTGTACGC	CAACATTGGCGGCATTGC	234	222
10.12NdeI	dCAPS	AATTTTATTTATTCCTGACATA	CTACTTACTATTCAACGTC	179	199
10.52 EcoRV	dCAPS	ACCGACTCCATCATGTT	TTAATTTTTAATAACCAGATAT	181	198
10.58Seq	Seq/SSLP*	TGCAAACTGTGGATGATGC	TTTGAACCAAGCACTAATGC	528	572
10.64 DdeI	dCAPS	AAGGTTCTCTCCTCTTG	GTAAAGTGAATAAACACTTA	221	204
10.67 DpnI	dCAPS	AGCTATGGGAGCTTCGTGAT	AGAATTGGAGGTCGACATCC	220	201
10.72Seq	Seq	TTTGATACCTAGATTATGACG	GATCAATCCGAACACATGC	439	439
10.78 DdeI	dCAPS	ATTCACTGTTTATGTCGC	TTCGATTTCTCATATTCTAA	225	209
MKP11.1	SSLP	CTTCCAAACATCAAAGCCG	TGCAGCAGGATAGGTTGG	247	184

Table A-1: Mapping Primer Sequences. All primers are listed from the 5` end. The underlined nucleotides in the dCAPS primer sequences indicate mismatched nucleotides designed to incorporate a restriction endonuclease recognition sequence into the amplicon derived from either the L*er* or the Col allele, depending on the particular primer set. The dCAPS primer product sizes indicate the sizes expected following digestion with the appropriate endonuclease (indicated in the primer name). SSLP = Simple Sequence Length Polymorphism; dCAPS = derived Cleave Amplified Polymorphic Sequence; Seq = Sequencing. *The 10.58Seq marker was originally defined as a sequence-based marker; however, a large insertion/deletion was detected which was not published in the Monsanto Co. Polymorphism Collection, which allowed this marker to be used as an SSLP marker.

Primer Name	T_{A1}	# cycles	T_{A2}	# cycles	Restriction Enzyme (Allele cut)	Restriction Enzyme Buffer
MKP8.62	56	35	N/A	N/A	N/A	N/A
10.12NdeI	50	10	52	25	*Nde*I (Col)	D
10.52 EcoRV	48	10	50	25	*EcoR*V (Col)	D
10.58Seq	56	35	N/A	N/A	N/A	N/A
10.64 DdeI	48	10	50	25	*Dde*I (Ler)	D
10.67 DpnI	56	10	60	25	*Mbo*I (Ler)	C
10.72Seq	56	40	N/A	N/A	N/A	N/A
10.78 DdeI	48	10	50	25	*Dde*I (Ler)	D
MKP11.1	56	35	N/A	N/A	N/A	N/A

Table A-2: Reaction Conditions for Mapping Primers. All of the reactions were carried out with an initial denaturing step at 94°C for 2 minutes, the indicated number of cycles of 94°C for 30 seconds, the indicated annealing temperature for 30 seconds and an extension step at 72°C for 30 seconds. In the case of the dCAPS markers, the first set of ten cycles was then immediately followed up with 25 cycles using the second annealing temperature. T_A = annealing temperature. Restriction enzymes and buffers (Promega).

Buffer	pH (at 37°C)	Tris-HCl (mM)	$MgCl_2$ (mM)	NaCl (mM)	KCl (mM)	DTT (mM)
C	7.9	10	10	50	-	1
D	7.9	6	6	150	-	1

Table A-3: Buffer Composition for Promega Restriction Enzyme Buffers.

Gene Name	Primer 1 (For)	Primer 2 (Rev)	Product Size (bp)	
			cDNA	Genomic
AT3G28345	GTATGGTTGCATAAGCGC	TAAACGTCTCGAACAAAGC	**797**	**870**
AT3G28360	CGGATGGTTATGAACTATG	TCGGTGATTTGATTAACGC	**731**	**883**
AT3G28380	TGCATCACATATGGAGGC	TTGGTGACTTGATTATCGC	**640**	**816**
AT3G28390	GGATGGTTATGAACCATG	CACAATAGTTGTACGACC	**797**	**966**
AT3G28415	TTGAAGGCTCAGTGAAGC	TGTACGACCAATAGTGGC	**850**	**1003**
At3g28420	ATTTCGTATAAGAATCGATCG	AAACGAAGCCAAAGAACGAA	**790**	**790**
At3g28450	TTGACGAGGATTTCGATGC	AAGTAGTACAAGATGATATAG	**582**	**582**
At3g28460	TCCTGGAACGTGCTGATA	AGCTCTTCAGTCACAGTCT	**358**	**617**
AT3G28470	CCTTGTTGTGACAAGTCC	TGACATGTTCCCGAGGC	**830**	**1020**

Table A-4: RT-PCR Primers. All primers are listed from the 5' end. The RT-PCR program consisted of an initial denaturing step at 94°C for 3 minutes followed by 40 cycles of 30 seconds at 94°C, 30 seconds at 56°C and 1 minute 15 seconds at 72°C. The fluorescence signal was measured after the each annealing step and after each extension step.

Summary of Candidate Gene Analysis

Gene	Description	WT RNA Expression		Sequencing	RT-PCR	T-DNA	Clone
		Siliques	Seed				
At3g28315	Retrotransposon	Not Found	Not Found				
At3g28320	Expressed protein	93	242			*	
At3g28330	F-box family protein	240	239			*	
At3g28340	Galactinol synthase	600	200			Yes*	In progress
At3g28345	P-glycoprotein-related	48	400	Yes	Yes	Yes*	
At3g28350	Hypothetical protein	19	35			*	
At3g28360	P-glycoprotein-related	850	200	Yes	Yes	Yes*	
At3g28370	Hypothetical protein	Not Found	Not Found			Yes*	
At3g28380	P-glycoprotein-related	<100	<100		Yes	*	
At3g28390	P-glycoprotein-related	<100	<100			*	
At3g28400	hAT-like transposase	Not Found	Not Found			*	
At3g28410	F-box family protein	17	22			*	
At3g28415	P-glycoprotein-related	Not Found	Not Found	Yes	Yes	*	
At3g28420	Expressed protein	227	84	Yes	Yes	Yes*	
At3g28430	Expressed protein	217	300			Yes*	
At3g28450	Leucine-rich repeat, transmembrane protein kinase	>800	>800	Yes	Yes	Yes*	

Table A-5: Candidate Gene Summary (10.58-10.67): The candidate genes listed above represent the predicted genes found between the 10.58 and 10.67 molecular markers (10.58Mb and 10.67Mb on chromosome 3). WT (wild type) RNA expression represents data obtained from an *in silico* search for WT expression patterns. Signals above 200 indicate expression above background levels. The sequencing column indicates those genes for which the coding region has been sequenced from a mutant plant and was found not to differ from wild type. The RT-PCR column indicates genes for which mutant versus wild-type RNA expression analysis was performed using Real-Time RT-PCR. In all cases indicated above, no differences in RNA transcript level were observed between mutant and wild-type. The T-DNA column indicates whether T-DNA tagged lines have been analyzed for the genes indicated. All those for which analysis has been completed produced seed with wild-type coloration. Currently, many of these lines as well as others (indicated by *) are being assessed in order to confirm that the individual plants scored are homozygous for the T-DNA insertion. The clone column indicates those genes for which genomic clones have been generated for complementation experiments. In the case of At3g28340 mutant plants have been transformed with the At3g28340 T-DNA construct and are awaiting further analysis.

Gene	Description	WT RNA Expression		Sequenced	RT-PCR	T-DNA	Clone
		Siliques	Seed				
At3g28460	Expressed protein	>700	>700	Yes	Yes	*	In progress
At3g28470	MYB family TF	Not Expressed	Not Expressed	Yes	Yes		
At3g28480	Oxidoreductase	>1100	>1100	Yes	Yes	*	In progress
At3g28490	Oxidoreductase	Not Expressed	Not Expressed	Yes		Yes*	
At3g28500	60S acidic ribosomal protein	N/A	N/A				
At3g28510	Expressed protein	Not Expressed	Not Expressed			*	
At3g28520	AAA-type ATPase	55	34			*	
At3g28530	Expressed protein	280	270				
At3g28540	AAA-type ATPase	580	39				
At3g28550	Proline rich extension-like family	Not Expressed	Not Expressed				
At3g28560	Expressed protein	Not Expressed	Not Expressed				
At3g28570	AAA-type ATPase	Not Found	Not Found			Yes*	
At3g28580	AAA-type ATPase	61	29			Yes*	
At3g28590	Expressed protein	178	98				

Table A-6: Candidate Gene Summary (10.67-10.72): The candidate genes listed above represent the predicted genes found between the 10.67 and 10.72 molecular markers (10.67Mb and 10.72Mb on chromosome 3). See Table A-5 above for description of specific fields. With regard to cloning, in the case of At3g28460 and At3g28480 wild-type genomic clones have been generated but have not yet been used to transform mutant plants.