Genetic studies of phenotypic variants in the woodland strawberry, *(Fragaria vesca)*.

Sarah Hudson Holt

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Horticulture

Richard E. Veilleux
Yinghui Dan
Vladimir Shulaev
James Tokuhisa
Brenda S. J. Winkel

September 2, 2011
Blacksburg, VA

Keywords: Rosaceae, T-DNA, Somaclonal variation, *SEPALLATA, F3H, EMP*
Abstract

Sarah Hudson Holt

The diploid woodland strawberry (*Fragaria vesca*) is a rapidly developing translational model for members of the family Rosaceae and other plants. This thesis represents some of the first forward genetics studies evaluating putative T-DNA insertion mutants in *F. vesca*. The observed phenotypes include alterations to floral development, anthocyanin pigmentation and leaf structure.

The floral development mutant named *green petal (gp)* was not associated with the T-DNA insertions present. Based on similar phenotypes induced by mutation of transcription factors involved in floral development of *Arabidopsis thaliana*, we used a BLAST search of the *F. vesca* genome hybrid gene models to identify 30 candidate genes that may have caused the *gp* phenotype. Expression analysis of these genes revealed that it was due to a 37 bp deletion in a *SEPALLATA3*-like E-Class MADS box transcription factor. This mutation altered organ structure in the three inner whorls of the flower, affecting fertility and fruit development. The deletion was demonstrated to segregate with the mutant phenotype in a segregating population of 92 individuals, 22 of which had green petals.

The anthocyanin biosynthesis mutant named *white runner (wr)* lacked red pigmentation in the stems and runners. The T-DNA insertion in this line was located in a highly repetitive LTR retrotransposon region, which complicated analysis. Segregation analysis of the *wr* lines revealed that the phenotype was unassociated with the T-DNA insertion as well. We used a targeted expression analysis of three critical structural genes in the flavonoid biosynthesis pathway that revealed a 20 bp deletion in the gene encoding flavanone 3-hydroxylase, an enzyme necessary for the production of flavonols, anthocyanins and proanthocyanidins. In an F$_2$ segregating population, this deletion co-segregated with the phenotype.

The third mutant line presented here displayed a *curly leaf (cl)* phenotype and was found to harbor a T-DNA insertion in a gene encoding a putative erythroblast macrophage attacher protein (EMP). Sequence and protein domain analysis indicated that FvEMP was related to the mammalian EMP protein that functions in cytoskeletal dynamics and red blood cell enucleation. Complementation analysis confirmed that introduction of the wild type *FvEMP* gene into the *cl*
mutant plants restored wild type leaf phenotype. Further morphological analysis revealed additional pleiotropic effects of the mutation, including abnormalities in seed set and germination, pollen tube growth, adhesion of the abaxial epidermal layer to the mesophyll layer and reduced petiolule length. These phenotypes are consistent with actin binding and microtubule associated protein mutants in other plant species.

Insertional mutagenesis is a critical molecular tool for model crop development. These studies highlight the precautions that must be taken when evaluating insertional mutants. These mutants are excellent tools for studying their respective disrupted gene function. The in depth molecular analysis of the mutants presented in this work was only possible because of the availability of the *Fragaria vesca* genome which was used extensively to identify T-DNA insertion sites and recover candidate gene sequences for expression analysis.
Table of Contents

Short and Sweet. An introduction to strawberry.........................................................1

References ......................................................................................................................3

Chapter 1. The Cautionary Tale of T-DNA Insertional Mutagenesis............................4

Introduction ..................................................................................................................4

Foreign DNA mutations ..............................................................................................6

Indels ............................................................................................................................9

Cytogenetic mutations ..............................................................................................11

References ..................................................................................................................13

Chapter 2. Flower and Fruit Developmental Defects Caused by a Spontaneous DNA
Deletion in a SEPALLATA3-like Gene in Fragaria vesca ...........................................18

Abstract ....................................................................................................................18

Introduction ...............................................................................................................18

Methods .....................................................................................................................22

Results .......................................................................................................................25

Discussion ...............................................................................................................29

References .................................................................................................................47

Chapter 3. Molecular Characterization of an Anthocyanin Deficient Mutant in the
Woodland Strawberry, Fragaria vesca. ......................................................................51

Abstract ....................................................................................................................51

Introduction ...............................................................................................................52

Methods .....................................................................................................................56

Results .......................................................................................................................60

Discussion ...............................................................................................................64

References .................................................................................................................79

Chapter 4. Characterizing a Curly Leaf Mutant in Fragaria vesca ...............................84

Abstract ....................................................................................................................84

Introduction ...............................................................................................................84
Methods........................................................................................................................................87
Results...........................................................................................................................................90
Discussion .....................................................................................................................................94
References ...................................................................................................................................112

Chapter 5 . Conclusions and Future Directions.................................................................115
List of Tables
Table 2.1 Primers used for Multiplex PCR, CAPS deletion analysis and creation of the complementation construct...........................................................................................................34
Table 2.2 Segregation data of T-DNA insertion 1, T-DNA insertion 2 and the phenotype. ..........35
Table 2.3 Genes coding for MADS-Box TF candidates, the originating species and the sequence used to identify homologous genes in F. vesca. .................................................................36
Table 2.4 Primers designed for 30 F. vesca hybrid gene model MADS box candidate genes ......38
Table 3.1. Primers used for RT-PCR and PCR amplifications ..................................................68
Table 4.1 Primer sequences used for genomic PCR, hi-TAIL-PCR, RT-PCR and zygosity PCR to characterize insertional mutant line Fv10-C04-136 and its complementation transformations. ...........................................................................................................99
List of Figures

Figure 2.1 Proposed MADS model of floral development for *Fragaria*.................................40
Figure 2.2 Phenotypic traits observed in the *gp* mutant.........................................................41
Figure 2.3 Molecular analysis of the T-DNA insertion sites..........................................................42
Figure 2.4 T-DNA insertion sites and zygosity analysis ...............................................................43
Figure 2.5 RT-PCR of hybrid gene model candidate genes............................................................44
Figure 2.6 Analysis of gene07201 ................................................................................................45
Figure 2.7 Amino acid alignment (ClustalW) of putative SEP3 proteins from different species...46
Figure 3.1 Diagram of the flavonoid biosynthetic pathway for the main branches addressed here for strawberry ..........................................................69
Figure 3.2 Phenotypic analysis of the *white runner* (*wr*) mutant .............................................70
Figure 3.3 Q1 scan .......................................................................................................................71
Figure 3.4 Metabolite analysis of FV10-C04-354-B17-10 red runner and white runner plant petioles.................................................................................................................72
Figure 3.5 T-DNA insertion site ....................................................................................................73
Figure 3.6 Molecular analysis of candidate structural genes and sequence analysis of *FvF3H* (gene14611) .........................................................................................................................74
Figure 3.7 Amino acid alignment of putative F3H protein sequences (ClustalW) from *N. tabacum, P. hybrida, I. purpurea, F. xananassa, F. vesca, A. thaliana,* and *D. caryophyllus.* 75
Figure 3.8 CAPS analysis of the *FvF3H* (gene14611) on individual FV10-C04-354 B17-10 and B17-12 T3 sibling plants..............................................................................................................76
Figure 3.9 RT-PCR analysis of candidate structural genes and TFs in red and white fruit.........77
Figure 3.10 Developing callus from the *AtF3H* complementation transformation in strawberry....78
Figure 4.1 Photographs of the *curly leaf* phenotype .....................................................................100
Figure 4.2 Molecular analysis of *curly leaf* plants and T-DNA insertion site .............................101
Figure 4.3 Amino acid sequence analysis of putative EMP ............................................................102
Figure 4.4 NCBI BLAST Tree View phylogenetic tree. ..................................................................103
Figure 4.5 Molecular analysis of the C0 #38 FV10-C04-136 complementation transformants.....104
Figure 4.6 RT-PCR of #38 B45c C1 sibling plants.........................................................................105
Figure 4.7 Representative plants harboring a WT copy of recombinant *FvEMP* from the RT-PCR results in Figure 4.5.................................................................106

Figure 4.8 Seed set and germination analysis ..............................................................................107

Figure 4.9 Pollen tube growth ......................................................................................................108

Figure 4.10 Petiolule length analysis .............................................................................................109

Figure 4.11 Epidermal Peels .........................................................................................................110

Figure 4.12 Hand drawn epidermal pave cells from Figure 4.11....................................................111
Acknowledgements

I would like to thank Dr. Richard E Veilleux and Dr. Vladimir Shulaev for the opportunity to work in their labs on the strawberry project at Virginia Tech and Dr. Yinghui Dan her support on the project at the Institute for Advanced Learning and Research in Danville, VA. I greatly appreciate all of the ideas, guidance and assistance that these three mentors have given me. I would also like to thank my other committee members Dr. James Tokuhisa and Dr. Brenda Winkel for their ideas, guidance and reassurance during dark moments of the project. I would like to thank an unofficial member of my committee, Dr. Barry Flinn, for all of the ideas, financial support and mentoring you have given. I am thankful to all of the past and present members of the Veilleux Lab, especially Dr. Juan Ruiz-Rojas, Kerri Mills, Kendal Upham, Robert Lindsey and Nan Lu; the Shulaev Lab, especially Dr. Teruko Oosumi, Dr. Joel Shuman, and Dr. Diego Cortes; and all of the members of the ISRR lab, for all of the assistance, advice, friendship, and privilege to work with them. I would like to thank the many undergraduates that that have assisted me over the years. I would like to thank the group at Plant and Food New Zealand, for teaching me many skills that I implemented in this project. I would like to thank Joyce Shelton, Maura Wood and Donna Long in the Horticulture Department and Nancy Eanes, Diana Carter and Pam Patterson at IALR for all of the paperwork, room reservations, and mailed packages. I must thank all of the members of the IT department at IALR, without whom the many long distance committee meetings would not have been possible. I am incredibly thankful to my fiancé Aaron Baxter for all the support and understanding needed over the distance. Finally, none of this would have ever been possible without the love, support and encouragement of my family, Richard, Marian and Aisling Holt.
Short and Sweet. An introduction to strawberry

Research in economically-important rosaceous fruit crops such as apple, peach, pear, cherry and almond is hindered by constraints such as plant size, age to reproductive maturity, self-incompatibility, logistics of handling insertional mutant populations and propagation limitations. Members of the genus Fragaria have many favorable attributes that make the strawberry an attractive translational model for many other species (Folta and Dhingra 2006; Shulaev et al. 2008). The cultivated strawberry Fragaria × ananassa is a highly valued crop species in the US and worldwide. This small fruit has a delicious flavor and an abundance of nutritional benefits, including high flavonoid and vitamin C content (Hannum 2004). The cultivated strawberry was developed about 250 years ago in a French garden, with Fragaria chiloensis as the maternal parent and F. virginiana as the paternal parent (Darrow 1966). This hybrid was named F. × ananassa after the aroma that was reminiscent of pineapple (genus Ananas) (Darrow 1966; Hancock 1999). Formal breeding programs began in England in 1817, and the strawberry that we know and love today still resembles these original cultivars (Hancock 1999). Genetic studies of F. × ananassa are complicated by its octoploid allopolyploid genetic structure (Rousseau-Gueutin et al. 2008).

Many of the favorable attributes found in the cultivated strawberry are also present in the genetically simpler, diploid species, thus presenting excellent candidate model systems. The diploid strawberry species, Fragaria vesca L. ‘Hawaii-4’ (PI 551572), is comparable to Arabidopsis thaliana in that it maintains a small footprint using minimal growing space, has abundant seed set, has a small genome size (240 Mb, 1.7 × A. thaliana), is easily transformed and regenerated from tissue culture, has a short life cycle going from seed to seed in 12-16 weeks, and has an abundance of genetic resources, including a fully sequenced genome (Oosumi et al. 2006; Shulaev et al. 2010). Additionally, F. vesca L. Hawaii-4 and other Fragaria species have attributes not found in A. thaliana including clonal propagation, perenniality, fleshy fruits that facilitate more extensive secondary metabolite studies, and direct agricultural relevance. Studies employing forward and reverse genetics are made possible with efficient transformation and plant regeneration protocols (Folta et al. 2006; Oosumi et al. 2006). Insertional mutant populations have been developed and are being characterized for flanking sequences, molecular markers, and phenotypes (Oosumi et al. 2010; Ruiz-Rojas et al. 2010). Transient expression
assays in fruit and protoplasts have been developed and can be used for transformation construct validation (Hoffmann et al. 2006; Nyman and Wallin 1992).

The work presented in the following chapters represents some of the first studies evaluating *F. vesca* putative T-DNA insertional mutants that show a visible abnormal phenotype. The mutations present in these lines affect floral development, flavonoid accumulation in vegetative tissues, and leaf morphology. These studies demonstrate the caution that must be taken when evaluating putative T-DNA insertional mutants that have been generated by *Agrobacterium tumefaciens* infection and transformation involving intensive *in vitro* culture. The goals of these studies were to evaluate the T-DNA insertion present in each line: this included characterization of the insertion site and segregation analysis of the T-DNA with respect to the phenotype. Intensive phenotypic analysis was employed to understand the extent of the phenotype and determine the most likely candidate gene pathway affected. Simplistic ideas concerning the nature of insertional mutants, i.e., that each independent transformation harbors a single T-DNA copy neatly inserted into a gene with intact right and left borders and that this insertion is responsible for the phenotype observed, had to be abandoned in exposing the genetic realities made possible through the application of molecular and incipient genomic tools. Candidate gene analysis (Pflieger et al. 2001) was used to determine the source of some mutant phenotypes. Finally, verification of the genetic mutation causing the phenotype was completed using segregation analysis and complementation transformation. These studies demonstrate the complexity and utility of phenotypic variants in *Fragaria vesca* as well as the potential of the recently completed strawberry genome sequence (Shulaev et al. 2010) to address genetic conundrums.
References

Darrow GM (1966) The strawberry: history, breeding and physiology


Chapter 1. The Cautionary Tale of T-DNA Insertional Mutagenesis

Introduction

As food security becomes an increasingly important issue worldwide, plant biotechnology research has been challenged to develop more productive, nutritious, and resilient crops to serve the growing population’s needs. Each plant family has traits that can aid in this development, such as alkaloid metabolites in Solanaceae, stress tolerances in members of Poaceae, nitrogen fixing capability in Fabaceae, glucosinolates in Brassicaceae and extensive flavonoid metabolites in Rosaceae. In order to understand the genetic mechanisms that control these traits, molecular tools are used to manipulate “simple” models and gather information that is then translated to crops with more complex genomes. This often requires the use of forward and reverse genetics, using plant transformation as a means to understand gene expression. This avenue must be met with caution, as disruption of a living system can have unpredictable consequences.

Plant tissue culture is a critical tool in the genetic transformation of many economically relevant crops. It is well known that plant tissue forced to grow in vitro, in an undifferentiated callus state before regeneration, has a high occurrence of uncontrolled and spontaneous genetic mutation, causing variation that is known as somaclonal variation (Larkin and Scowcroft 1981). Plant tissue culture accelerates a normally slow natural process, potentially creating many novel variants with heritable changes in phenotype. There are many factors within tissue culture manipulations that can act as mutagens resulting in this clonal variation. Species, cultivar, age of the tissue culture donor plant, tissue type used as explant, growth regulators and their combinations, antibiotics, salt concentration, UV exposure, and length of time in culture have all been implicated as causative agents of somaclonal variation (Bairu et al. 2011; Jain 2001; Joyce et al. 2003; Veilleux 1998).

The types of genetic mutations that arise due to these factors are nearly as numerous (Lee and Phillips 1988). One intensively-studied source of variation is the activation of transposable elements in vitro. Tissue culture conditions in rice, maize, citrus, rye, Arabidopsis and other species have resulted in the activation of transposons and retrotransposons causing somaclonal variation (Barret et al. 2006; Courtial et al. 2001; De Felice et al. 2009; Grandbastien 1998; Hirochika 1993; Hirochika et al. 1996; Huang et al. 2009; Linacero and Vázquez 1993). Other
types of mutations that are caused by tissue culture are Indels (Insertions/deletions), chromosome rearrangement or translocations, SNPs (Single Nucleotide Polymorphisms), polyploidy, aneuploidy and changes in methylation status (Jain 2001; Kaeppler et al. 2000; Lee and Phillips 1988; Veilleux 1998). Genome-wide mutation has been detected in some plant species as a result of regeneration (Jiang et al. 2011). It is assumed that somaclonal mutations occur in many plant species, but many go undetected and have no impact on the fitness of the plant. This is because these changes happen in non-coding regions outside of gene boundaries, within intronic regions that do not affect RNA processing or in heterochromatic regions. They may be inherited, but do not impair proper gene function. Other mutations are not so innocuous. Mutations in coding regions can have deleterious effects on plant growth and organ function (Bairu et al. 2011). Various tools that are used to identify mutations genome-wide are amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPDs), transposon display, simple sequence repeat (SSR) markers and now even whole-genome sequencing (Bairu et al. 2011; Veilleux 1998). Many of these tools are unable to detect mutations that occur across the genome, due to the fact that SNPs and small indels are difficult to observe. In all cases verification of specific mutations requires DNA sequencing and comparison of sequence from the donor plant tissue with its clonal or sexual progeny.

For the biotechnological manipulation of many crops, *in vitro* culture and regeneration are necessary steps for genetic modification, i.e., integration of foreign DNA. The tools used for genetic modification include mechanical means of foreign DNA integration, such as microinjection, chemically-mediated transfection of protoplasts, gene bombardment and electroporation; and biological means of gene integration, such as viral mediated transformation and *Agrobacterium* mediated transformation. This review will focus on *Agrobacterium* transformation, and the deleterious genetic aberrations detected in transgenic plants generated through *in vitro* and *in planta* methods.

*Agrobacterium*-mediated transformation is a process that uses the Gram-negative bacterium to infect plant tissue, utilizing its horizontal gene transfer capability to generate mutants with T-DNA, or Transfer DNA, insertions (Stachel and Zambryski 1989; Zambryski et al. 1989; Zupan et al. 2000). The carefully-engineered T-DNA can knock out or knock down endogenous gene function, serve as a trap for endogenous genes or promoters, act as a marker or
assist in marker development for mapping studies, and even introduce visual marker fusions to aid in understanding gene expression. The simple infection process of plant tissue by Agrobacterium is mutagenic, even in populations of T-DNA mutants that were generated without in vitro regeneration (Castle et al. 1993; Negruk et al. 1996). Due to the overlap of mutation types, abnormalities found in transgenic plants that have been subjected to both in vitro culture and Agrobacterium infection are difficult to resolve.

**Foreign DNA mutations**

The purpose of Agrobacterium-mediated insertional mutagenesis is insertion of the T-DNA at one genomic site, with known, defined borders. This is a deliberate mutational agent; however unintentional consequences of this process may include the insertion of multiple T-DNAs, backbone vector sequence and extraneous bacterial chromosomal DNA in a myriad of permutations. These types of mutations happen irrespective of the method of infection and plant regeneration. There are many steps throughout T-DNA integration that can lead to integration of abnormal T-DNA.

Modern engineered Agrobacterium transformation strains harbor a tumor-inducing (Ti) plasmid that contains the necessary virulence (vir) genes required for proper transformation and the T-DNA located on a second, much smaller binary vector (Hood et al. 1993). Wounded plant cells produce signaling molecules that upon interaction with the Agrobacterium cell activate VirA, a membrane-bound sensor, which in turn activates VirG, a transcriptional activator, though a phospho-relay system (Winans 1991). VirG activates transcription of the other vir genes, virB, virC, virD, virE and virF. These proteins are necessary for the T-DNA processing and transport in both the bacterial and plant cell, as well as aid in nuclear localization and integration into the plant genome (Howard et al. 1992; Tinland et al. 1992). The use of Agrobacterium vir knockouts and plant cell vir over-expression lines has illuminated how each of these proteins contributes to this process. The T-DNA is defined by two 24 bp imperfect repeat border regions that are recognized by the VirD1/D2 complex. These proteins nick the bottom strand of the T-DNA within the repeats and a single VirD2 protein binds to the 5’ end at the right border (Herrera-Estrella et al. 1988; Pansegrau et al. 1993; Ward and Barnes 1988) protecting this end from degradation (Dürrenberger et al. 1989) and serving as a guide into the plant cell. Another VirD2 molecule binds to the 5’ end of the non-T-DNA strand at the left border (Herrera-Estrella
Rolling circle replication originating from the 3’ nick at the right border displaces the VirD-bound single-stranded T-DNA and fills in the gap left by the T-DNA until it reaches the VirD2-bound to the 5’ end at the non-T-DNA border. Then the VirD2-bound single-stranded T-DNA molecule is transported to the host cell though a channel formed by the VirB1-11 and VirD4 proteins (Christie et al. 1989; Okamoto et al. 1991). These channels also transport other Vir proteins into the host cell, specifically VirE and VirF (Vergunst et al. 2005). Upon entry into the host cell the single stranded T-DNA is coated with VirE2 molecules to prevent degradation and assist in nuclear localization (Christie et al. 1988; Citovsky et al. 1992). It is believed that plant cell factors interact with the T-DNA complex, assisting in the transport into the nucleus and guidance to the sites of integration (Howard et al. 1992; Tinland et al. 1992). Once the T-DNA complex is inside the host cell nucleus the next steps leading to and including integration into the host genome are not completely known. There is evidence that VirF recruits host cell proteins forming a complex that activates the proteasomal degradation of the T-DNA protein coat (Tzfira et al. 2004b). There is evidence that this is coupled with second strand synthesis leading to the formation of a double stranded T-DNA molecule before integration (Tzfira et al. 2004a). The current theory on T-DNA integration is that the double stranded molecule inserts into the host genome via non-homologous recombination at double stranded breaks (Lacroix et al. 2006; Tzfira et al. 2004a). These are the processes believed to happen during a single canonical T-DNA insertion event.

There are two steps in this process that can result in many of the observed abnormal T-DNA integration patterns. First, defects in the VirD removal of the T-DNA from the binary vector in the Agrobacterium cell can lead to the integration of vector backbone sequence in the plant genome. If the VirD complex does not nick the left border sequence and bind to the 5’ end of the non-T-DNA strand, the displacement (and rolling circle replication) would not end at the left border. This can result in read through of a portion or the entire binary vector from left border. Often the read through terminates at some point within the vector backbone sequence. Occasionally replication displaces the entire vector through the right border and through the T-DNA a second time. These insertions are two T-DNAs connected by the vector backbone sequence. Such insertion types can complicate the analysis of the insertion site and lead to additional phenotypes. Kononov et al. (1997) used a GUS selectable marker located on the backbone of the binary vector employed for T-DNA transformation in tobacco and found that
75% of analyzed lines contained integrated vector backbone sequence, in some cases linked to
the T-DNA and sometimes the backbone integration was independent of the T-DNA insertion.
Screening methods detected that six of the plants may have had the entire binary vector
integrated into the plant genome. A study looking at the possible complex patterns of T-DNA
integration in rice found 18 different combinations of T-DNA integration patterns, with only
48% of the analyzed population having canonical single insertions (Yin and Wang 2000). This
study demonstrated the necessity of thorough T-DNA insertion site characterization with
multiple primer sets situated within and outside of the T-DNA insertion. Three of the analyzed
independent transformants were found to contain single copy long T-DNAs which were greater-
than-unit-length binary plasmid. These insertions were two T-DNAs connected by the vector
backbone sequence, as described above. The T-DNA insertion site in tobacco and carrot were
found to contain integrated backbone vector sequence in 44% of the independent transformants,
with no significant difference found between the two species (Pukhnacheva et al. 2005). In the
event that there is an error with the right border sequence or VirD2 processing, it is also possible
for the T-DNA strand displacement to start at the 5’ end of the non-T-DNA strand at the left
border. This can result in a T-DNA formed out of the vector backbone, with replication
potentially reading through and skipping the right border into the T-DNA and terminating at the
left border. This type of abnormal T-DNA consists of the entire binary vector with one copy of
the T-DNA.

The second step of the Agrobacterium transformation process that can lead to abnormal
T-DNA insertions is during the T-DNA integration into the host cell genome. Ideally, only one
T-DNA is integrated at a site at one time, however there is evidence that tandem duplications of
T-DNA insertions are the result of multiple T-DNAs integrating at the same site. Tandem T-
DNAs have been found head-to-head, head-to-tail, and tail-to-tail, with even more complex
patterns observed if any of the T-DNAs are truncated. Co-integration of different sequence T-
DNAs from separate bacterial cells can result in tandem integration patterns, with variable
orientations and alterations of the T-DNA structure (De Neve et al. 1997). Although it can be
difficult to resolve the mechanism by which an abnormal T-DNA insertion took place, often
more difficult is the process of determining the amount of sequence that has been inserted. The
duplicate sequence found in these abnormal insertions complicates analysis using molecular
tools.
In addition to the expected T-DNA insertion and consequential insertion of vector backbone sequences, integration of Agrobacterium chromosomal DNA (AchrDNA) has been detected in four Arabidopsis T-DNA insertion FST (flanking sequences tag) databases, FLAG, SAIL, GABI-Kat, and SIGnAL (Ulke et al. 2008). All four of these populations were generated using the *in planta* floral dip method, each using different combinations of binary T-DNA vectors and Agrobacterium strains. This study only analyzed the publicly available FST sequences, and the authors suggest that integration of AchrDNA is only associated with T-DNA insertion and not transferred independently. They postulate that the T-DNA occasionally integrates into the Agrobacterium chromosome and mobilizes from that site upon plant transformation when the integrity of one of the T-DNA border sequences has been maintained. Southern blot analysis using probes designed to the AchrDNA confirmed linkage of the T-DNA to the AchrDNA in three GABI-Kat T₃ lines (Ulker et al. 2008). The authors also reported evidence of AchrDNA transfer to rice insertional mutant lines as well. In the *A. thaliana salade* mutant, the T-DNA insertion was not solely responsible for the dwarf phenotype, but was complicated by a deletion of at least 1.5 kb at the T-DNA integration site as well as the integration of a 6.7 kb bacterial transposon Tn5393 (Zhao et al. 2009). The authors found that the bacterial transposon was present in the Agrobacterium strain LBA4404 used to generate the insertional mutant population in *A. thaliana*, and was most likely the result of transposition into the T-DNA prior to plant infection. These unexpected insertion types can complicate the process of analyzing insertional mutant lines in all species. The abnormal integration patterns can also cause the formation of fusion proteins that can behave unexpectedly once inside the plant cell. The risk of unintentional horizontal gene transfer of bacterial genes into plants with germinal transmission to progeny must be considered.

**Indels**

Indels are genetic polymorphisms caused by insertions or deletions of native DNA. These nucleotide changes can be large, spanning hundreds of kb, or small microindels, consisting of the removal or addition of a single nucleotide. One of the best-studied sources of indel polymorphism is transposable element associated indels. Transposable elements are mobile DNA elements that are divided into two classes based of the mode of movement (Wicker et al. 2007). Class 1 elements, or retrotransposons, move by a “copy and paste” mechanism that replicates itself as an RNA intermediate and inserts a DNA copy in a new place.
Retrotransposon movement results in insertion polymorphisms, although the integration process may also delete portions of the insertion site. This class of mobile elements has been implicated as the causative factor of genome expansion in *Zea mays* (SanMiguel et al. 1996). Class 2 elements, or DNA transposons, move by a “cut and paste” mechanism. DNA transposons can yield both insertion polymorphisms, through reinsertion into a new location and the incomplete excision of the transposon from the donor site, and deletion polymorphisms from the incomplete excision of the transposon removing nucleotides from the original insertion site (Kidwell and Lisch 1997). There are many families of transposable elements within both classes, defined by many characteristics including structure, length, species specificity and movement trigger (Feschotte et al. 2002). In the *Ds* based *A. thaliana* T-DNA insertional mutant, *halfman*, 150 kb of plant genomic DNA were deleted at the site of *Ds* insertion (Oh et al. 2003). Introduction of the tobacco retrotransposon *Tnt1* into *A. thaliana*, by root explant transformation (an *in vitro* based protocol), resulted in somatic transposition with independent lines carrying between 0-26 copies of the transposon (Courtil et al. 2001). In parallel, the *Tnt1* T-DNA introduced through whole plant infiltration or floral dip, an *in planta* method, had no somatic or germinal transposition evaluated through the T\textsubscript{3} progeny (Courtil et al. 2001).

Alternative causes of indels in transgenic lines may be incomplete, aborted T-DNA integration. Two T-DNA mutants, *ant1* and *ant2*, showing defects in ovule development in *A. thaliana*, were determined unlinked to the T-DNA and rather due to a 22 bp deletion (*ant1*) and a G-to-A transition (*ant2*) in *AINTEGUMENTA*, a floral morphology gene known to be involved in female gametophyte development (Klucher et al. 1996). In the epicuticular wax deficient *A. thaliana* mutant, *cer*, two phenotype mutants generated through seed transformation were found to be unlinked to the T-DNA insertion (Negruk et al. 1996). One of these mutants was found to harbor a 17 bp deletion in the second exon of the gene, and another contained a 2 bp substitution followed by a 2 bp insertion, both lines resulting in frame shift mutations causing residue disruption and premature termination of the amino acid sequence. The authors proposed that the causes of these mutations were unsuccessful T-DNA integration. Since these lines were generated using *in planta* methods of transformation, bypassing the *in vitro* regeneration step, the mutation mechanism is most likely due to the *Agrobacterium* infection or T-DNA integration process, but may be impossible to determine exactly what occurred. Finding these microindels
and small <50 bp indels can be a difficult process, since these mutations escape most polymorphism detection tools.

**Cytogenetic mutations**

Cytogenetic mutations are changes to chromosome structure caused by translocations, inversions, deletions and aneuploidy. When chromosome imbalance occurs, visible and/or inducible phenotypes may result. *Agrobacterium* mediated transformation has been implicated in causing extensive cytogenetic aberrations. Originally presenting as a T-DNA insertion segregating with the mutant phenotype, the *A. thaliana* line, ACL4, was found to contain an array of chromosomal changes (Nacry et al. 1998). This line bore a mutation named *ton1*, characterized by severe dwarfism; extensive molecular analysis showed one tandem head-to-head T-DNA insertion with the left border mapping to chromosome 2 and the right border mapping to chromosome 3, a truncated second T-DNA insertion with the right border mapping to chromosome 2 and the left border mapping to chromosome 3, a reciprocal translocation between chromosomes 2 and 3, a 40-cM inversion on chromosome 2, and a deletion on chromosome 3. The process of T-DNA integration induced extreme genome rearrangement (Nacry et al. 1998).

Many cytogenetic abnormalities have been attributed to the insertion of multiple T-DNAs. The *mgo2* mutant in *A. thaliana* has a 26 cM inversion on chromosome 1 that is flanked by two T-DNAs (Laufs et al. 1999). In a study of *A. thaliana* embryonic mutants generated through seed transformation (Feldmann and Marks 1987), it was estimated that 20% of the T-DNA tagged mutants contained chromosomal translocations, with many containing complex T-DNA loci (Castle et al. 1993). In a screen for seedling lethal mutants using T-DNA and *Ds* insertional mutant populations generated by *in planta* methods, only 29% and 33% were found to co-segregate with the insertion, respectively (Budziszewski et al. 2001). The untagged mutants were not characterized and the authors proposed that the seedling lethal mutations were most likely caused by incomplete T-DNA insertions or single nucleotide changes; however cytogenetic changes cannot be excluded as the cause. Screens for embryo and seedling lethal mutants may be enriched with gross cytogenetic abnormalities. Cytogenetic changes are also detected in insertional mutant lines with canonical integrations. In the *A. thaliana* line, BNP-23, a simple, single T-DNA insertion caused the reciprocal translocation of chromosomes 1 and 2,
resulting in gene disruption and severe phenotypic mutations (Lafleuriel et al. 2004). In the *A. thaliana pgd1* mutant, a single T-DNA insertion resulted in the deletion of 269 kb, deleting 65 genes in this line (Ding et al. 2007). The *A. thaliana hot* mutant has a 75.8 kb deletion at the T-DNA integration site at the bottom of chromosome 1 (Kaya et al. 2000). Tax and Vernon (2001) described two T-DNA associated interchromosomal rearrangements found in the course of mutant evaluation in *A. thaliana*. A comprehensive characterization of 64 of the *in planta* generated SALK T-DNA mutant lines revealed that at least 19% had chromosome translocations (Clark and Krysan 2010). These are only a few of the cytogenetic anomalies that have been described and reported in *Arabidopsis*. Also, care must be taken when crossing seemingly canonical T-DNA insertional mutant lines for the development of multiple mutants. Curtis et al. (2009) only discovered a reciprocal chromosome translocation associated with a T-DNA integration upon the crossing of the mutant with another insertional mutant line.

Recognizing that unintended genetic aberrations may be caused by *Agrobacterium* infection and transformation, as well as *in vitro* regeneration is important when utilizing T-DNA insertional mutants for forward and reverse genetic studies. The following chapters evaluate three *Fragaria vesca* mutants that had been generated by *Agrobacterium*-mediated transformation and *in vitro* regeneration and that expressed visible changes in phenotype. The lessons learned throughout the process of evaluating these phenotypic variants reiterate the caution that must be exercised when utilizing these populations.
References


Clark KA, Krysan PJ (2010) Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines. Plant J. 64: 990-1001


Herrera-Estrella A, Chen ZM, Vanmontagu M, Wang K (1988) VirD proteins of Agrobacterium tumafaciens are required for the formation of a covalent DNA protein complex at the 5' terminus of the t-strand molecules. EMBO J. 7: 4055-4062

Hirochika H (1993) Activation of tobacco retrotransposons during tissue culture. EMBO J. 12: 2521-2528


Pukhnacheva NV, Novoselia TV, Zotkevich EA, Deineko EV (2005) [Insertion of vector sequences in the genome of transgenic plants]. Genetika 41: 1203-1209


Warld ER, Barnes WM (1988) VirD2 protein of Agrobacterium tumefaciens very tightly linked to the 5' end of the T-strand DNA Science 242: 927-930


Chapter 2. Flower and Fruit Developmental Defects Caused by a Spontaneous DNA Deletion in a SEPALLATA3-like Gene in *Fragaria vesca*

Abstract

Floral morphogenesis is critical in the plant life cycle. Genetic control of floral patterning depends upon gene interactions, orchestrated by the MADS box transcription factors (TFs). These TFs are divided into five classes (ABCDE) corresponding with the ABCDE model of floral morphology. These genes have overlapping function across the floral whorls and are necessary for proper organ development. Understanding how each TF impacts floral morphology has been greatly aided by the use of knockout mutants, either caused by random genetic mutation or intentional T-DNA insertional mutagenesis.

*Fragaria vesca* L. has been rapidly gaining prominence as a model for Rosaceae and other fruit crops. This diploid strawberry possesses simple flowers and fleshy fruits that provide an excellent opportunity for studying MADS box TFs. Phenotypic screening of a T1 pCambia insertional mutant population revealed a floral morphology mutation in *F. vesca* named *green petal* (*gp*). The phenotypic effects of the *gp* mutation include sepaloid petal morphology, pollen abortion, indehiscent anthers, spontaneous conversion of anthers to sepals, presence of trichomes on pistils, green carpels and ectopic fruit formation, thus affecting all of the three inner whorls. T1 Segregation analysis through T-DNA Zygosity PCR showed that neither T-DNA insertion associated with the *gp* phenotype. Based on the *gp* phenotype, we selected 30 MADS box TFs from strawberry as potential candidate genes responsible for this phenotype. This analysis led to the discovery of a 37 bp deletion of the C-terminal domain in a putative *SEPALLATA3* E-class MADS TF. Segregation analysis confirmed association of the deletion with the *gp* phenotype. The wide ranging effects of the mutation suggest that this gene’s function may affect many downstream processes, offering new insight into flower development.

Introduction

Floral morphology and development are critical in the plant life cycle. Genetic control of floral patterning depends upon gene interactions, orchestrated by the MIKC-type MADS box transcription factors (TFs). The MADS box TFs are found among plants, yeast and mammals, named for MCM1 in *Saccharomyces cerevisiae*, AGAMOUS in *Arabidopsis*, DEFICIENS in

In angiosperm floral development these TFs are divided into five classes (ABCDE) corresponding with the ABCDE model of floral morphology (Figure 2.1) (Theissen 2001). These genes have overlapping function across the floral whorls and are necessary for proper organ development. For example, A and E class genes specify sepal identity, whereas, A, B and E are necessary for proper petal development. Continuing to the inner whorls B, C, and E class genes are necessary for stamen development, while C and E determine carpel identity and C, D, and E determine ovule identity.

These genes have been characterized in many flowering species including, but not limited to, Arabidopsis thaliana, Antirrhinum majus, Petunia hybrida, and Gerbera hybrida. In Arabidopsis, the A class genes are APETALA1 (AP1) and APETALA2 (AP2, not a MADS TF), the B class genes are APETALA3 (AP3) and PISTILLATA (PI); C function is controlled by AGAMOUS (AG); and SEEDSTICK (STK), SHATTERPROOF (SHP1 and SHP2) contribute to D-function (Mandel et al. 1992; Yanofsky et al. 1990). The E function, which is necessary in all whorls, is controlled by the SEPELLATA (SEP) genes, of which there are four in Arabidopsis and they seem to display some functional redundancy. Many of these whorl regulators also act outside of the floral whorls and interact with MADS box TFs FRUITFULL (FUL), CAULIFLOWER (CAL) and LEAFY (LFY, not a MADS TF) to specify floral meristem identity.

The expression of the A, B, C, D and E TFs is spatially restricted by activation and repression mechanisms and ectopic expression results in additional floral defects. Understanding how each TF impacts floral morphology has also been greatly aided by the use of knockout mutants, either caused by random genetic mutation or intentional T-DNA insertional mutagenesis. Depending on the gene, mutations can carry varying levels of severity, often depending on the protein domain to which the mutation is targeted. The MIKC-type MADS TFs comprise four domains: the MADS domain composed of 56 conserved amino acids involved in DNA binding; the Intervening domain, a less conserved region and the K-box, a domain resembling the coiled-coil domain of Keratin, both involved in protein dimerization; and the C-terminal domain, implicated in transcriptional activation. These TFs are tetramers in vivo and bind to simple motifs termed CArG boxes (for ‘CC-ArC-GG’) in the upstream regulatory
sequences of the activation targets. How the MADS multimer recognizes the proper cis element sequence is a complex question.

The use of knockout mutants and over-expression lines has revealed that many of the MADS box TFs have multiple functions of gene activation and repression depending on their spatial and temporal expression. Mutations in AP1 result in the conversion of sepals to leaf-like outer organs, the development of secondary flowers at the base of the leaf-like sepals and failure of the petal whorl to develop (Irish and Sussex 1990). Early characterization of mutations in AP2 showed temperature sensitive defects in the first two outer whorls, with no disruption of the reproductive organs (Bowman et al. 1989; Jofuku et al. 1994). Floral architecture is severely disrupted in ap1 ap2 double mutants, with floral meristems continually converted to inflorescence meristems (Irish and Sussex 1990). This phenotype is similar to ap1 cal double mutants (Kempin et al. 1995). This pattern of disrupted development confirms that multiple MADS floral factors are required for the transition from inflorescence meristem to a floral meristem and the specification of the first two floral whorls (Hempel et al. 1997; Mandel et al. 1992; Yant et al. 2010). The reduced phenotypic effect of single mutations also indicates that many of these genes are functionally redundant and can compensate for the loss of each other (Ferrándiz et al. 2000; Kempin et al. 1995). In AP3, allele ap3-1 contains a single nucleotide polymorphism that converts a lysine to methionine at the carboxyl end of the K box (Jack et al. 1992). This weak allele results in a temperature sensitive defect; at 25°C second whorl petals are converted to sepals and third whorl stamens are converted to carpeloid stamens, whereas at 15°C second whorl petals are converted to sepaloid petals and the third whorl stamens are normal (Bowman et al. 1989). Alleles ap3-3, ap3-4, and ap3-5 all have nucleotide changes that result in premature stop codons that severely truncate the proteins, causing complete conversion of petals to sepals and stamens to carpels (Jack et al. 1992). Severe mutations in the AP3 sister gene, PISTILLATA (PI), result in flowers that have sepals in the petal whorl and carpels in the stamen whorl (Bowman et al. 1989; 1991). Similarly, pi-5, a weak, temperature sensitive PI mutation due to the conversion of a conserved glutamic acid to a lysine in the K domain only affects the petal whorl (Yang et al. 2003). Transient expression assays using AP3 and PI::GUS fusions have shown that the co-expression of both genes is required for nuclear localization (McGonigle et al. 1996) and other floral-specific factors initiate early AP3 and PI expression, with AP3 and PI responsible for autoregulation of late expression (Honma and Goto 2000; Lamb et al. 2002).
Constitutive expression of PI and AP3 in a wild-type background overrides other floral-factor regulatory mechanisms and results in flowers with two outer whorls of petals, and two inner whorls of stamens, with cauline leaves developing petaloid-like characteristics in the leaf margins (Krizek and Meyerowitz 1996). Mutations in AG result in a flower within a flower, the stamens being replaced by petals and the fourth whorl replaced by whorls of petals (Bowman et al. 1989). This phenotype is the result of the non-functionality of AG and its inability to repress expression of the MADS box TFs directing petal identity in the inner whorls.

Resolving the function of each of the E-class MADS box genes, SEPALLATA1-4, in *A. thaliana* has been a difficult task due to the subtle or normal phenotypes displayed by single mutants, indicating functional redundancy among the four genes (Ditta et al. 2004; Ma et al. 1991; Mandel and Yanofsky 1998; Pelaz et al. 2000). However, RNA localization studies have indicated expression of the SEP genes occurs early in flower development and throughout the floral whorls, indicating a requirement for this gene for specification of meristem and organ identity (Ditta et al. 2004; Ma et al. 1991; Mandel and Yanofsky 1998). Additionally, mutant phenotypes in putative SEP orthologues in other species provide evidence of the importance of the SEP genes in floral development (Ampomah-Dwamena et al. 2002; Angenent et al. 1992; Ferrario et al. 2003; Matsubara et al. 2008; Pnueli et al. 1994; Uimari et al. 2004; Vrebalov et al. 2002). New model plants with greater complexity in floral architecture and fruit development will aid in the understanding of how the interactions of these genes direct this critical stage of plant development.

*Fragaria vesca* L. is a developing model for Rosaceae and other fruit crops. This diploid strawberry has simple flowers and fleshy fruits that provide an excellent opportunity for studying MADS box TFs. Phenotypic screening of a T1 pCAMBIA insertional mutant population, which was generated from an *Agrobacterium*-mediated transformation, revealed a floral morphology mutation in *F. vesca* named green petal (gp) (Ruiz-Rojas et al. 2010). This floral homeotic mutation displays extensive changes in floral organ structure, with the mutation affecting whorls 2, 3 and 4. Additionally this mutant exhibits defects in fruit and seed development. Methods presented in this study are: 1) a phenotypic characterization of the gp mutant, 2) a molecular characterization of the T-DNA insertions present in this line, and 3) gene expression analysis of candidate MADS box TFs to determine the cause of the gp mutant.
Methods

Plant ID and growth conditions. T₁ seeds (n=92) of insertional mutant FV10-CO2-721 GenBank Accession # HM012640) were obtained from the F. vesca L. ‘Hawaii 4’ pCAMBIA (AF234298, Cambia, Canberra, Australia) insertional mutant seed repository at Virginia Polytechnic Institute and State University (Veilleux Lab, Virginia Tech, Blacksburg, VA, USA). Germination was induced in B-5 liquid germination medium under sterile conditions (Lindsey 2010). Upon germination seedlings were transferred to 72 cell flats in Pro-Mix, grown in the Virginia Tech greenhouses. Seedlings were transferred to the Institute for Sustainable and Renewable Resources (ISRR) greenhouse and planted into 8” pots in Miracle-Gro (The Scott’s Company LLC, Marysville, OH, USA) potting medium and grown under greenhouse conditions. Previous analysis showed that this line segregated for the gp phenotype (Ruiz-Rojas et al. 2010).

Molecular analysis. Genomic DNA was extracted from each of the T₁ plants as in Oosumi et al.(2006) by Kendall Upham. Plants were screened for the T-DNA insertion by Multiplex PCR by Kendal Upham using Immomix Red according to manufacturer’s instructions (Bioline USA inc., Tauton, MA, USA). Primers used for Multiplex PCR were designed from the plasmid sequence pCAMBIA 1304 (GenBank accession # AF234300.1) to amplify GFP (green fluorescent protein, expected band size: 179 bp), HYG (hygromycin, expected band size: 318 bp), and the region spanning the left border (expected band size: 550) and the right border (expected band size: 718 bp) of the T-DNA to check for backbone integration. As a DNA quality control primers were designed for F. vesca ANTHOCYANIN SYNTHASE (FvANS, Genbank Accession # AY695818.1) (expected band size: 422 bp). Primer sequences are listed in Table 2.1. High-efficiency Thermal Asymmetric Interlaced (hi-TAIL) PCR was used (Liu and Chen 2007) to amplify the F. vesca genomic DNA flanking the T-DNA insertions sites. Fragments were sequenced on a Beckman Coulter CEQ 8800 according to manufacturer’s instructions (Beckman Coulter Inc., Brea, CA, USA). Obtained sequences were annotated with SeqBuilder (DNASTAR, Inc., Lasergene 8, Madison, WI, USA). Flanking sequences were compared to the F. vesca genome scaffold database, using the BLASTn tool, to determine the T-DNA integration site (Shulaev et al. 2010).

Phenotypic characterization. Flowers of the wild type and gp mutant plants collected at anthesis, but before anther dehiscence were dissected under a microscope into individual whorls to document variation. Floral organs were observed under an Olympus SZX 12 light microscope
(Olympus America, Center Valley, PA, USA) and pictures taken using an Olympus DP 70 camera and Image Pro Software.

_Pollen viability staining._ Anthers were collected from flowers after anthesis, but before dehiscence from wild type and _gp_ plants. The pollen grains were fixed and stained according to Peterson et al. (2010). The pollen was viewed at 20× magnification using an Olympus BX 51 microscope, photographed with an Olympus DP 70 camera and Image Pro software (Media Cybernetics Inc, Bethesda, MD, USA).

_Candidate gene selection._ Previously characterized E-class and B-class MADS box TFs were chosen from _A. thaliana_ and other rosaceous species (Table 2.3) and compared to the _F. vesca_ genome hybrid gene model database (09/2010) using the BLAST tools. The top 40 hybrid gene model hits from each search were compared to isolate the top 30 hybrid gene model candidates overall. Each gene model was aligned to GenBank to verify completeness of the gene model and annotated as necessary. All sequences were documented using SeqBuilder. Primers were designed in the putative exon regions of each gene, to amplify as much as the putative transcript as possible. This doubled as an internal control of RNA integrity; since the primers were designed to span one or more introns, expected band size ensured the absence of contaminating genomic DNA. Each primer pair was designed to amplify at an annealing temperature of 55°C. Primers were obtained from Operon (Eurofins MWG Operon, Huntsville, AL, USA).

_RNA isolation and reverse transcriptase polymerase chain reaction (RT-PCR) analysis._ RNA extractions were performed on unopened flower bud tissue from wild type and _gp_ plants. The RNA extraction method was followed according to Moser et al. (2004), with omission of the proteinase K step. First strand cDNA synthesis was prepared using Invitrogen SuperScript III cDNA synthesis kit (Life Technologies, Carlsbad, CA, USA). RT-PCR was carried out in a 25 µl reaction, which used 0.25 pmol gene model specific primers, 0.16 mM dNTPs and 0.2 U Ex Taq polymerase (Takara Bio Inc, Japan). The thermocycler program for all primer pairs was as follows: one cycle (95°C for 3 min), 35 cycles (95°C 1 min, 55°C 1 min, 72°C 1 min), and one cycle (72°C 3 min). Actin primers were used as a positive control. All products were stained 1x with EZVision (Ameresco, Solon, OH, USA) and separated on a 1% agarose/ 1x TAE gel stained 0.5× with Sybersafe (Invitrogen). The remaining product for gene07201 was fractionated under the same stain conditions on a 2% agarose gel.
**Sequencing of gene07201.** PCR products of the wild type and gp mutant plants from the RT-PCR reaction of gene07201 were excised from the gel and purified using Qiagen’s Gel Extraction Kit (Qiagen Inc., Valencia, CA, USA). The products were sequenced using the Beckman Coulter CEQ 8800.

**Analysis of gene07201 sequence.** Sequences were aligned using the ClustalW 2.1 alignment program (Larkin et al. 2007). Restriction site analysis was performed in SeqBuilder. Primers flanking the deletion site were designed using the primer design feature within SeqBuilder. PCR was carried out in a 25 µl reaction, which used 0.25 pmol of gene model specific primers, 0.16 mM dNTPs and 0.2 U Ex Taq polymerase (Takara Bio Inc., Japan) and 100 ng of genomic DNA. The primers flanking the deletion site were named 07201delR and 07201delF, listed in Table 2.1. The thermal cycling program was as follows: one cycle (95°C for 3 min), 35 cycles (95°C 1 min, 55°C 1 min, 72°C 1 min), and one cycle (72°C 3 min). The product (10 µl) of was stained 0.5x with EzVision and were run on a 2% agarose/ 1x TAE gel stained 0.5x with SyberSafe. To the remaining 15 µl of PCR product 2 µl of NEB 10x Buffer 4, 2 µl of water and 1 µl of Nde1 (10u/µl) (New England Biolabs Inc. (NEB), Ipswich, MA, USA) were added. The samples were incubated at 37°C for 3 h. The entire digest of 20 µl was stained 0.5x with EzVision and run on a 2% agarose/ 1x TAE gel stained 0.5x with SyberSafe.

**Design and assembly of a complementation construct.** The complementation transformation construct was assembled from the FvSEP3 (gene07201) cDNA transcript and the native FvSEP3 promoter. The primers for the gene07201 transcript were designed in the 5’ region upstream of the translation start site containing a BsrGI restriction site and in the 3’ region downstream of the stop codon. Primers for the cDNA transcript named 07201Seq1F and 07201cDNAR, listed in Table 2.1. The primers for the gene07201 promoter region were designed to amplify 1.6 kb upstream of the FvSEP3 start site. The 07201ProF primer was designed in Exon 1 of the coding region spanning the same BsrGI restriction site as in the cDNA transcript. Primers for the gene07201 promoter region were named 07201ProF and 07201Seq7R, listed in Table 2.1. The cDNA transcript was amplified from WT flower bud cDNA and the promoter from WT genomic DNA with Bio-Rad’s iPROOF High-Fidelity Master Mix according to instructions (Bio-Rad Laboratories Inc., Hercules, CA, USA). The PCR product (10 µl) was stained 1x with EzVision and run on a 1% agarose/1x TAE gel to verify amplification and band size. The remaining
product was purified with Qiagen’s PCR Purification Kit according to manufacturer’s instructions. The purified products were digested with the BsrG1 restriction enzyme (NEB), stained 1x with EzVision and run on a 1% low melting point (LMP) agarose/1x TAE gel stained 1x with SyberSafe. The bands were excised from the gel and melted at 70°C. The cDNA and promoter were ligated at 16°C overnight in LMP agarose using the T4 Rapid Ligase Kit (Promega Corporation, Madison, WI, USA). The vector pCAMBIA 1304 was digested with EcoI/CR1 (Promega), stained with EzVision and run on a 1% LMP agarose/1x TAE gel stained 1x with SyberSafe. The linearized vector band was excised from the gel, heated to 70°C and 2.5 µl was added to the cDNA-promoter ligation for an additional overnight ligation at 16°C. The ligation reaction was transformed into E. coli strain DH5α, and plated for blue-white selection. We screened white colonies for insertion by colony PCR using primer combinations, which were: 1) 07201cDNAR and 07201ProF, 2) 07201Seq1F and 07201cDNAR and 3) 07201ProF and 07201Seq7R. Plasmid DNA from two positive colonies was sent to Quintara BioSciences (Quintara BioSciences, Berkeley, CA, USA) for sequencing to verify that no mutations had been introduced during the cloning process. These colonies were transformed into A. tumafaciens strain EHA105 using Bio-Rad’s Gene Pulser XCell Electroporation system according to manufacturer’s instructions.

**Complementation transformation.** The above FvSEP3 complementation construct was transformed into FV10-C02-721-A4 gp petiole tissue as in Oosumi et al. (2006) and Zhao et al. (2004).

**Results**

**Green Petal (gp) segregation.** The segregating population of 92 T₁ plants of F. vesca mutant line FV10-C02-721 was scored for the gp phenotype: 70 individuals had wild type flowers and 22 presented with the gp phenotype, following a 3:1 segregation ratio ($\chi^2 = 0.59$, $p > 0.1$). This segregation ratio suggested that a single gene mutation was responsible for the gp phenotype.

**Insertion analysis.** Multiplex PCR analysis resulted in 85 plants that were positive and seven plants that were negative for the T-DNA, following a 15:1 segregation ratio for two T-DNA insertions ($\chi^2 = 0.59$, $p > 0.1$) (Figure 2.3A). HiTAIL PCR yielded three recoverable bands (Figure 2.3B), which when sequenced confirmed the presence of two unique insertion sites. The location of insertion 1 was scaffold 0513105, disrupting the last exon of hybrid gene model
gene30919, a putative cytoplasmic dynein light chain (Figure 2.4A). The location of insertion 2 was scaffold 0513098, with the T-DNA in an intron of hybrid gene model gene32482, a putative unidentified plant protein (Figure 2.4C). Zygosity PCR analysis of T-DNA insertion 1 yielded 29 plants that were homozygous negative (six of these were gp), 56 plants hemizygous for the insertion (13 of these were gp) and no plants homozygous for the insertion (Figure 2.4B; Table 2.2). This suggests that T-DNA insertion 1 was lethal in the homozygous state and segregated independently of the gp phenotype (assuming a 0:2:1 segregation, $\chi^2 = 0.88, p > 0.1$). Zygosity PCR analysis of T-DNA insertion 2 yielded 28 plants that were homozygous negative (six of these were gp), 38 plants were hemizygous (seven of these were gp) and 18 plants were homozygous positive (six of these were gp) (Figure 2.4D; Table 2.2). This segregation pattern followed a 1:2:1 ratio ($\chi^2 = 0.21, p > 0.1$), with the gp phenotype segregating independently of T-DNA insertion 2. The results of the Zygosity PCR analysis indicated that the gp phenotype was not associated with either T-DNA insertion and it was likely a somaclonal variant.

**Phenotype analysis.** The gp phenotype presented with abnormal development of the floral organs in whorls 2, 3 and 4, as well as abnormal fruit development and seed set (Figure 2.2).

*Petal morphology.* Petals of the WT plants are white in color, round with smooth or slightly scalloped edges (Figure 2.2B, C, left side of panel). There may be slight yellow or green pigmentation at the base of the petal. The petals on the gp mutant have many developmental defects (Figure 2.2B, C, right side of panel). The petals are green on the adaxial face and pink on the abaxial face. Both faces of the petal have trichomes and the petal margins are deeply serrated.

*Anther morphology.* Anthers of the WT plants are pale yellow and oval in shape, rounded at the top (Figure 2.2D, top of panel). There is a furrow along the margin of the anther, corresponding to the site of dehiscence. The anther is attached to a filament, making up the stamen. The anther-filament junction separates easily. Anthers on the gp mutant are initially yellow, but many develop sectors of green pigmentation and grow into sepal-like structures (Figure 2.2D, bottom of panel; Figure 2.2F). The shape of the gp anther is altered from the wild type shape, in that the anther is longer than wild type and more triangular. The distal end of the anther is acuminate, like a serration point of a sepal. The dehiscence furrow is present in the gp anther,
however dehiscence does not occur and the anther does not detach from the filament as in wild
type anthers.

Pollen viability. Modified Alexander’s stain was used on pollen to test for viability. In wild type
pollen, the cell walls appeared blue-green and the cytoplasm bright magenta when anthers had
been collected from flowers just after anthesis, but before dehiscence (Figure 2.2G). Pollen
collected from gp flowers after anthesis showed a much different staining pattern. The cell walls
stained pale blue-green; however the cytoplasm appeared to be detaching from the cell walls and
stained a much lighter pink (Figure 2.2H). This indicated the gp pollen grains aborted after
anthesis.

Carpel morphology. Carpels from wild type flowers have a pale yellow stigma, style and ovary
(Figure 2.2E, top of panel). The receptive stigma is rounded and without trichomes. Carpels on
the gp flowers are pale green in color (Figure 2.2E, bottom of panel). The stigma is pointed at
the ends and the carpels appear to have trichomes.

Seed inviability. Seeds from wild type fruits germinated and contained a pale white embryo
when dissected (data not shown). Seeds from gp fruits did not germinate and embryos were not
found on dissection (data not shown).

Ectopic fruit formation. Wild type fruits expanded and matured to a pale white/yellow fleshy
fruit, with viable brown seeds and a strong aroma (Figure 2.2I, left side of panel). The gp fruits
expanded to a pale green fruit that was harder than wild type and had no fruity aroma (Figure
2.2I, right side of panel).

Petal/ Fruit indehiscence. Once pollination occurred in wild type flowers the petals usually
senesced and detached from the flower. In the gp flower the petals did not senesce or detach
from the developing flower/fruit. Once a wild type fruit was fully mature it could be easily
detached from the hull (sepal whorl). The gp fruits did not detach from the hull.

Selection of candidate genes. Since the gp phenotype did not co-segregate with either T-DNA
insertion, we presumed it was somaclonal in nature and pursued its characterization using a
candidate gene screening approach to determine if alterations in MADS box gene expression in
the gp mutant may have been responsible for the phenotype. Previously characterized E-class
and B-class MADS box TF sequences from A. thaliana and other rosaceous species were chosen
from GenBank and used to search the *F. vesca* hybrid gene model database using the nucleotide BLAST algorithm (Table 2.3). The top hits from each search were analyzed and the 30 *F. vesca* hybrid gene models with the greatest similarity, based on E-values, to the *A. thaliana* and rosaceous TFs were chosen as candidate genes (Table 2.4).

**RT-PCR analysis.** RT-PCR experiments were carried out on cDNA derived from unopened flower buds from wild-type and *gp* plants. All 30 candidate genes were expressed in both the wild type and *gp* flower buds (Figure 2.5). A band shift was detected in the *gp* product for gene07201 for the *gp* mutant (Figure 2.5, indicated by box around gene 07201), a putative E-class *SEP3* MADS box TF. This band shift was further analyzed on a 2% agarose gel to determine if this was a gel artifact (Figure 2.6A). The change in band size became more obvious in the higher percentage gel. The gene07201 bands for wild type and *gp* were excised and purified for sequence analysis.

**Sequence and putative protein analysis of gene07201.** The putative mRNA sequence for gene07201 was obtained from the *F. vesca* genome database and compared to GenBank to determine predicted gene assignment. The BLASTx alignment indicated that the predicted sequence for gene07201 was orthologous to the E-class MADS box TF *SEPALLATA3*, but was incomplete in the region corresponding to the N-terminus, specifically the MADS box domain (data not shown). Analysis of nucleotide sequence of the upstream genomic region revealed that the first exon (the MADS box domain) was 2 kb upstream of the original predicted start site. A nucleotide sequencing error in the *F. vesca* genome obscured the real start site codon. Subsequent RT-PCR during generation of the complementation construct confirmed that the upstream region was connected to the original predicted mRNA transcript. Sequence analysis of the RT-PCR products for WT and *gp* for gene07201 showed that the *gp* transcript was missing 37 bp from the seventh exon (of eight), corresponding to the C-terminal end (Figure 2.6B). At the protein level the deletion would have disrupted the last 60 amino acids (Figure 2.6C; Figure 2.7, deletion start site denoted by arrow).

**Deletion genotyping results.** In order to determine whether the 37 bp deletion in the C-Terminal region of gene07201 in the *gp* mutant was associated with the *gp* phenotype we analyzed the population of 92 T1 FV10-CO2-721 plants segregating for the *gp* phenotype for the deletion. Primers were designed flanking the deletion site in the *gp* gene07201 to utilize a unique *Nde1*
restriction site present within the 37 bp deletion converting the amplified PCR product into a Cleaved Amplified Polymorphic Site (CAPS) marker (Figure 2.6 C, D and E). Of the 92 T₁ plants segregating for the gp phenotype, 21 were homozygous WT, 49 were hemizygous for the deletion and 22 were homozygous for the deletion, following a 1:2:1 segregation ratio ($\chi^2 = 0.81, p > 0.1$). All of the plants homozygous for the deletion were the gp phenotype. The gp phenotype was not present in the homozygous WT or hemizygous plants. This is strong evidence that the deletion caused the gp phenotype.

**Discussion**

The T-DNA insertional mutant, Fv10-C02-721 segregating for the gp phenotype, was found to segregate independently of both insertions. The presence of defective organs in all three inner whorls of the flower led to an expression analysis of candidate MADS box TFs. RT-PCR amplification of the candidate genes revealed a band size difference that resulted from a 37 bp deletion in *F. vesca* hybrid gene model gene07201, a SEPALLATA3-like gene in the gp mutant.

It is well known that *Agrobacterium* infection and transformation, as well as lengthy in vitro regeneration protocols can induce genetic mutations resulting in somaclonal variation (reviewed in Chapter 1, (Bairu et al. 2011; Larkin and Scowcroft 1981; Veilleux 1998)). Such mutations can complicate forward genetics studies that seek to determine the gene mutation in an insertional mutant with visible phenotype. Segregation analysis of the T-DNA mutant line is a critical step of the evaluation process to determine if the insertion and phenotype are linked. Upon determination of non-linkage, researchers are left with few options for identifying the genetic cause of the phenotype; these include construction and analysis of a mapping population with the mutant or candidate gene analysis. In some cases generating a mapping population is hindered by mutations that affect fertility and unknown cytogenetic aberrations caused by the *Agrobacterium* infection/transformation/regeneration processes. Either approach is time intensive and is not guaranteed success.

In the present study we used candidate gene analysis to identify a 37 bp deletion in *FvSEP3* as the cause of the gp phenotype. This analysis of 30 highly conserved candidate genes was only possible because of the availability of the *F. vesca* genome sequence that facilitated candidate gene identification and gene specific primer design (Shulaev et al. 2010). Determining
the cause of the deletion is impossible, but there are many scenarios. Small insertion/deletions (indels, < 50 bp) are difficult to find through most polymorphism detection methods simply because they go unnoticed in electrophoretic band sizes. SNPs and microindels (1 bp indels) are often attributed to mistakes during DNA replication, naturally occurring or stress induced. Comparisons of the genome sequence of five A. thaliana plants from 30 generations of single seed decent found that a small array of mutations occurred during the replication process (Ossowski et al. 2010). The authors attributed the mutations to mistakes in DNA replication due to UV stress and de-amination of methylated cytosines (Ossowski et al. 2010). In the F. vesca regeneration protocol used to generate the T-DNA insertional mutants, the prolonged time in tissue culture, exposure to UV light through GFP screening, exposure to growth regulators, and antibiotic selection have all been implicated as causative factors of somaclonal variation (Bairu et al. 2011; Veilleux 1998).

There are few phenotypic mutants that originated as T-DNA insertional mutants that have been associated with small (<50 bp) deletions. Two A. thaliana mutants, ant1 and cer, were not associated with their respective T-DNA insertions, but due to a 22 bp deletion and a 17 bp deletion, respectively (Kluch et al. 1996; Negruk et al. 1996). The proposed cause of these deletions was incomplete, aborted T-DNA insertion that removed nucleotides from the attempted site of integration.

The in vitro regeneration process is also a known inducer of transposable element movement in many plant species. Small indels may be generated by the insertion and then faulty excision of transposable elements resulting in deletion of nucleotides from the insertion site (Kidwell and Lisch 1997). For this theory to be possible, integration and excision would have to happen during callus growth prior to regeneration. Regardless of the mutation mechanism that transpired, the discovery of the FvSEP3 deletion in the gp mutant raises the question of the presence of additional mutations in this line. An answer would require extensive genome-wide polymorphism analysis or whole-genome sequencing.

The FvSEP3 gene is a member of the E-Class MADS box TF gene family, which are the SEPALLATA (SEP) genes, initially named in Arabidopsis thaliana. In A. thaliana there are four SEP genes (Ditta et al. 2004; Pelaz et al. 2000). Mutations in SEP1, SEP2, and SEP3 individually result in subtle phenotypes (Pelaz et al. 2000). The sep1 sep2 sep3 triple mutants
have normal sepals in the 1st whorl, conversion of petals and stamens to sepals in the 2nd and 3rd whorls, respectively, and a reiteration of sepals of the first three whorls within whorl 4. Scanning electron microscopy confirmed the conversion to ectopic sepaloid cells in the inner whorls and the presence of stomata, structures not normally found on petals, stamens or carpels. These triple flowers form endless, indeterminate whorls of sepals (Ditta et al. 2004). The forth SEPALLATA gene, SEP4, controls proper sepal development and sep1 sep2 sep3 sep4 quadruple mutants present like the triple mutants, except that the cell surface of the “sepals” have leaf-like epidermal cell structure and stellate trichomes, normally only seen on leaves (Ditta et al. 2004). These results indicate that the SEPALLATA genes are necessary for the proper formation of all the floral whorls and that they overlap in expression and may be functionally redundant in A. thaliana.

The lack of phenotypes in the single mutants and the extremity of the effects of triple and quadruple mutants make dissecting the individual function of each SEP gene difficult. In order to understand the function of each SEP gene, triple mutants containing one active allele for each were generated. sep1 sep2/(sep3+/+) mutants have wild type appearance except for a reduced number of stamens in the 3rd whorl; however they are still fertile (Ditta et al. 2004). The (sep1/+) sep2 sep3 and the sep1 (sep2+/+) sep3 mutants display normal flowers except for severe defects in ovule development that abolish fertility (Favaro et al. 2003). This indicates that of all the SEP genes, SEP3 may possess a more critical role in development. In the FvSEP3 gp mutant the mutation results in abnormal development in each of the three inner floral whorls. Each whorl develops into the correct organ, however the abnormalities prevent the organs from functioning properly. This indicates that FvSEP3 may not be complemented by the other putative FvSEP genes and has a greater role in floral development in the strawberry.

In 2001, 10 years after the original ABC model of floral morphology was presented, a new model was published, called the ‘floral quartet’ model of floral organ identity (Coen and Meyerowitz 1991; Theissen 2001). This model utilized the many previously published yeast-hybrid interaction studies to explain how different combinations of the ABCDE-class MADS box TFs interact at the molecular level to specify the identity of the organs within each floral whorl. As an example AP3 and PI, both of which are MADS box TFs, are essential to petal development, demonstrated by the phenotypes presented in single mutants (Bowman et al. 1989; 1991; Meyerowitz et al. 1991). Petal defects are also seen in ap1 mutants and 35s::SEP3
antisense transgenic lines (Mandel et al. 1992; Pelaz et al. 2001). Yeast two-hybrid assays and co-immunoprecipitation experiments have shown that AP3 and PI form heterodimers but were unable to activate transcription alone (Honma and Goto 2001). Yeast two-hybrid assays of AP3 and PI separately with AP1 or SEP3 did not support yeast growth, but when PI+AP3 were used together in a yeast three-hybrid assay, positive interactions occurred between SEP3 and AP1. Both AP1 and SEP3 contain transcriptional-activator domains, appear to be the only MADS box genes that interact with PI+AP3 that have transcriptional-activator domains and confer that ability to PI+AP3 when binding as tri/tetramers (Honma and Goto 2001). In wild-type plants these PI+AP3+SEP3+AP1 tetraramers are then able to regulate downstream targets resulting in proper petal formation. PI+AP3 does not interact with AG alone, but when SEP3 is introduced PI+AP3-SEP3+AG interactions take place. The “indeterminate whorls of sepals” phenotype in the sep1 sep2 sep3 triple was similar phenotype to that of the bc double mutants ap3 ag or pi ag (Bowman et al. 1991; Pelaz et al. 2000). However, expression of SEP1, SEP2 and SEP3 is not altered in the bc double mutants, nor is expression of AP3, PI or AG altered in the sep1 sep2 sep3 triple mutant, indicating that expression is not co-dependent. Studies of 35S::SEP3 sense lines have shown that AG and AP3 are ectopically induced in leaf tissue in strong over-expression lines (Castillejo et al. 2005). AP3, PI and AP1 are all required for proper development of petals and AP3, PI and AG are all required for proper stamen development, with SEPALLATA genes mediating the in planta interactions.

In our F. vesca gp mutant, the 37 bp nucleotide deletion was located in the second to last exon of FvSEP3, causing a frame-shift that changed the last 60 amino acids of the C-terminal domain (Figure 2.7, deletion start site denoted by arrow). Yeast hybrid studies using modified AtSEP3, showed that the transcriptional activity was abolished upon deletion of the C-terminal domain (Honma and Goto 2001). ChIP-SEQ (chromatin-immunoprecipitation sequencing) experiments screening for the DNA binding sites of SEP3 in wild type A. thaliana predicted that SEP3 has over 3,400 gene targets (Kaufmann et al. 2009). It would be expected that deletion/disruption of the transcriptional-activator C-terminal domain in the FvSEP3 protein would impact the expression of many genes, resulting in the extensive floral defects observed in the gp phenotype. The complementation analysis designed to restore wild type phenotype in the gp mutant is underway to definitively confirm that the deletion in FvSEP3 caused the gp phenotype.
In conclusion, the complexity of the strawberry flower and fruit development provides new insight into the function of the \textit{SEP3} gene and its role in development. This stable mutation will be useful in studying the \textit{SEP3} involvement with auxin response factors, aroma development, and flavonoid biosynthesis. This study is a good example of the utility of phenotypic mutants in \textit{F. vesca}, a cautionary tale of T-DNA insertional mutagenesis and the advantage of the \textit{F. vesca} genome.
**Table 2.1 Primers used for Multiplex PCR, CAPS deletion analysis and creation of the complementation construct.**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYGF</td>
<td>5’- TCACAGTTTGCCAGTGATAC -3’</td>
</tr>
<tr>
<td>HYGR</td>
<td>5’- ATCGTTATGTTCATCGGCAC -3’</td>
</tr>
<tr>
<td>mGFPF</td>
<td>5’- CCGCTGATCATTATCAACAAA -3’</td>
</tr>
<tr>
<td>mGFPR</td>
<td>5’- TGTTATAGTTCATCCATGCCA -3’</td>
</tr>
<tr>
<td>LBF</td>
<td>5’- AAGTTGTGGTTCAAGTTGTC -3’</td>
</tr>
<tr>
<td>LBR</td>
<td>5’- CCAGATAAGGGAATTAGGCT -3’</td>
</tr>
<tr>
<td>RBF</td>
<td>5’- TGTGATTAGAGTCCCGCAAT -3’</td>
</tr>
<tr>
<td>RBR</td>
<td>5’- GCAATGTCCAGTAGGTTCG -3’</td>
</tr>
<tr>
<td>FVANSF</td>
<td>5’- GGGAGGACTAAGTTTCCC -3’</td>
</tr>
<tr>
<td>FVANSR</td>
<td>5’- AGGATGAAGGTGAGTGCA -3’</td>
</tr>
<tr>
<td>07201delR</td>
<td>5’- GCATGCTGATCAGCTACAGA -3’</td>
</tr>
<tr>
<td>07201delF</td>
<td>5’- ATGGCAACCATTCCTCCATGCA -3’</td>
</tr>
<tr>
<td>07201Seq1F</td>
<td>5’- CCAAGCTCTGCAATGGACC -3’</td>
</tr>
<tr>
<td>07201cDNAR</td>
<td>5’- TTAGGTTAGGTTGGTTCATAGTCA -3’</td>
</tr>
<tr>
<td>07201ProF</td>
<td>5’- CTTTCTTCTCAATAGTCGTTCTTC -3’</td>
</tr>
<tr>
<td>07201Seq7R</td>
<td>5’- TCACACGTTAACATCCTTAGTCAAGAGG -3’</td>
</tr>
</tbody>
</table>

1 HYG: Hygromycin, mGFP: Green fluorescent protein, LB: Left border, RB: Right border, FVANS: *F. vesca ANTHOCYANIN SYNTHASE*
Table 2.2 Segregation data of T-DNA insertion 1, T-DNA insertion 2 and the phenotype.

<table>
<thead>
<tr>
<th>Zygosity</th>
<th>+/+²</th>
<th>+/-²</th>
<th>-/-²</th>
<th>Total</th>
<th>$\chi^2$ ³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insertion 1¹</td>
<td>0</td>
<td>56</td>
<td>29</td>
<td>85</td>
<td>NS²</td>
</tr>
<tr>
<td>wt/gp</td>
<td>0/0</td>
<td>43/13</td>
<td>23/6</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Insertion 2¹</td>
<td>19</td>
<td>38</td>
<td>28</td>
<td>85</td>
<td>NS³</td>
</tr>
<tr>
<td>wt/gp</td>
<td>13/6</td>
<td>31/7</td>
<td>22/6</td>
<td>85</td>
<td></td>
</tr>
</tbody>
</table>

¹ Numbers in rows for Insertion 1 and Insertion 2 correspond with the number of plants from the T₁ segregating population (n=85) with the genotype assigned at the top of the column.

² The + sign indicates allele without T-DNA insertion, the – indicates allele with T-DNA insertion.

³ Chi-squared analysis for each insertion was non-significant, NS² assuming 0:2:1 segregation and NS³ assuming 1:2:1 segregation.
Table 2.3 Genes coding for MADS-Box TF candidates, the originating species and the sequence used to identify homologous genes in *F. vesca*.

<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>GB Accession</th>
<th>Top 5 BLASTx <em>F. vesca</em> hits* (E-)</th>
</tr>
</thead>
</table>
| *A. thaliana*| **SEPALLATA3**| AT1G24260.2  | gene04229 (6e-77) gene07201 (1e-59) gene04563 (5e-53)
|              |              |              | gene19428 (4e-52) gene26118 (9e-47)                 |
| *A. thaliana*| **SEPALLATA4**| NM_126418.2  | gene04229 (2e-65) gene04563 (4e-58) gene26118 (1e-51)
|              |              |              | gene19428 (2e-47) gene26119 (5e-44)                 |
| *A. thaliana*| **SEPALLATA2**| NM_111098.3  | gene04229 (4e-89) gene04563 (4e-56) gene19428 (1e-53)
|              |              |              | gene26118 (5e-52) gene07201 (6e-45)                 |
| *A. thaliana*| **PISTALLATA**| NM_122031.3  | gene11267 (6e-43) gene11268 (6e-31) gene20015 (4e-28)
|              |              |              | gene13356 (9e-28) gene24494 (3e-27)                 |
| *M. xdomestica* | **JOINTLESS** | DQ402055.2  | gene20808 (6e-75) gene12120 (2e-42) gene03714 (7e-34)
|              |              |              | gene12119 (1e-28) gene30741 (3e-25)                  |
| *M. xdomestica* | **AP1-like**    | AY071921.1  | gene04562 (1e-97) gene26119 (2e-67) gene04228 (4e-62)
|              |              |              | gene04563 (4e-46) gene04229 (2e-42)                 |
| *M. xdomestica* | **MADS16**     | AB370212.1  | gene13354 (4e-41) gene25070 (1e-35) gene04563 (7e-34)
|              |              |              | gene24494 (9e-34) gene32354 (1e-32)                 |

<sup>a</sup> did not make the list of 30 candidate genes
<table>
<thead>
<tr>
<th>Species</th>
<th>Gene</th>
<th>GB Accession</th>
<th>Top 5 BLASTx F. vesca hits* (E-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M. xdomestica</td>
<td>AGAMOUS-like</td>
<td>AF401637.1</td>
<td>gene24494 (e-103) gene24852 (2e-78) gene25070 (8e-69) gene19428 (2e-40) gene26119 (6e-40)</td>
</tr>
<tr>
<td>P. pyrofolia</td>
<td>SEPALATA</td>
<td>AB571596.1</td>
<td>gene07201 (3e-83) gene04229 (2e-77) gene04563 (1e-52) gene19428 (9e-51) gene26119 (8e-46)</td>
</tr>
<tr>
<td>F. xananassa</td>
<td>AGAMOUS, STAG1</td>
<td>AF168468.1</td>
<td>gene24852 (e-130) gene24494 (6e-75) gene25070 (2e-68) gene26119 (2e-38) gene04229 (6e-38)</td>
</tr>
<tr>
<td>F. xananassa</td>
<td>RIN</td>
<td>AF484683.1</td>
<td>gene04229 (e-133) gene04563 (6e-52) gene19428 (6e-49) gene26118 (7e-48) gene26119 (2e-39)</td>
</tr>
<tr>
<td>F. chiloensis</td>
<td>MADS-box 2</td>
<td>GQ398010.1</td>
<td>gene07201 (e-101) gene04229 (2e-74) gene04563 (2e-52) gene19428 (5e-47) gene26118 (8e-45)</td>
</tr>
<tr>
<td>F. chiloensis</td>
<td>MADS-box 1</td>
<td>GQ398009.1</td>
<td>gene07201 (2e-96) gene04229 (1e-73) gene04563 (3e-52) gene19428 (1e-46) gene26118 (2e-43)</td>
</tr>
<tr>
<td>P. persica</td>
<td>PISTILLATA-like</td>
<td>EU005663.1</td>
<td>gene11267 (9e-65) gene11268 (6e-37) gene13356 (9e-31) gene04562 (3e-30) gene20015 (1e-29)</td>
</tr>
<tr>
<td>Gene #</td>
<td>Putative Class</td>
<td>Primer Name</td>
<td>Primer Sequence</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>gene26118</td>
<td>E</td>
<td>26118 F</td>
<td>ATGGGGAGAGGAGGAGGTTGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26118 R</td>
<td>AGCATCCACCCAGGAGAATGAT</td>
</tr>
<tr>
<td>gene04229</td>
<td>E</td>
<td>04229 F</td>
<td>ATGGGGAGGGGAAAGGAGTGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>04229 R</td>
<td>AGCATCCACCCAGGGAATGAAAC</td>
</tr>
<tr>
<td>gene07201</td>
<td>E</td>
<td>07201 F</td>
<td>ATGGCTCAAAAAACTGGAGAGAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>07201 R</td>
<td>ATGGCAAACCATCTCCATCCAT</td>
</tr>
<tr>
<td>gene04563</td>
<td>E</td>
<td>04563 F</td>
<td>ATGGGAAGAGGGAGGAGGTGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>04563 R</td>
<td>TTAGTTGTTGCAATTGAAACGCC</td>
</tr>
<tr>
<td>gene26119</td>
<td>A</td>
<td>26119 F</td>
<td>ATGGGGAGGGGGAAGGAGGTTCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26119 R</td>
<td>CTATTCCATCAGGGTGGGAGGCG</td>
</tr>
<tr>
<td>gene20808</td>
<td>E</td>
<td>20808 F</td>
<td>ATGGCTCGGGAGAAGATTCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20808 R</td>
<td>CCAACTTGAGAGACGTTGACTGA</td>
</tr>
<tr>
<td>gene04562</td>
<td>A</td>
<td>04562 F</td>
<td>ATGGGAAGGGGTAGGGTTCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>04562 R</td>
<td>TGAAGCAAAGCATCCAGGTG</td>
</tr>
<tr>
<td>gene19428</td>
<td>E</td>
<td>19428 F</td>
<td>ATGGGGAGGGGAGGAGGTGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19428 R</td>
<td>CATCCCTGGAGAATTTGCT</td>
</tr>
<tr>
<td>gene25070</td>
<td>B</td>
<td>25070 F</td>
<td>ATGGGGAGAGGAAAGATTTGAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25070 R</td>
<td>CTAAGGATATGAGAAATTACAC</td>
</tr>
<tr>
<td>gene24494</td>
<td>C</td>
<td>24494 F</td>
<td>ATGGGAGATCCCACAAACACTAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24494 R</td>
<td>TTACACAAGGTTGAGGATGCT</td>
</tr>
<tr>
<td>gene24852</td>
<td>D</td>
<td>24852 F</td>
<td>ATGGCCTATGAAAACAAAAACAAAAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24852 R</td>
<td>TTACACTAACTGGAAGGAAACTTGGT</td>
</tr>
<tr>
<td>gene13356</td>
<td>E</td>
<td>13356 F</td>
<td>ATGGGGAGGGGGAAGGAGTGGTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13356 R</td>
<td>CTGCGTAGTGCTTCTAATTGATCTAA</td>
</tr>
<tr>
<td>gene04228</td>
<td>A</td>
<td>04228 F</td>
<td>ATGGGAAGAGGGCAAGGTTTCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>04228 R</td>
<td>CCATCCATGTTGAGCACTGACT</td>
</tr>
<tr>
<td>gene26429</td>
<td>B</td>
<td>26429 F</td>
<td>ATGGGAAGGGGAAAGGATTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26429 R</td>
<td>GGTGCTTTGGCAATTATGAGAC</td>
</tr>
<tr>
<td>gene20134</td>
<td>-</td>
<td>20134 F</td>
<td>ATCGAGAACAGAAATAAACCGGCA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20134 R</td>
<td>TCAAGCTCAAGATGAGCTGTT</td>
</tr>
<tr>
<td>gene32354</td>
<td>E</td>
<td>32354 F</td>
<td>GGAAGACCTCAGATGAGCAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32354 R</td>
<td>CTTGGGACTTCTTTGCTTTGG</td>
</tr>
<tr>
<td>gene20015</td>
<td>A</td>
<td>20015 F</td>
<td>AGGATAGGAACAAAACCACAAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20015 R</td>
<td>GAAAGGCTTCTTTCCACCCAC</td>
</tr>
<tr>
<td>gene29503</td>
<td>C</td>
<td>29503 F</td>
<td>ATGGGTAGGGGGAAGGATTGAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29503 R</td>
<td>TCCCTTTCTGGAGCCTTCCCT</td>
</tr>
<tr>
<td>gene12120</td>
<td>C</td>
<td>12120 F</td>
<td>AAGCCCGCGAGGAAATAGATA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12120 R</td>
<td>AAGCTTGAGAGATAAGGTGTCGAG</td>
</tr>
<tr>
<td>Gene #</td>
<td>Putative Class</td>
<td>Primer Name</td>
<td>Primer sequence</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>gene27105</td>
<td>-</td>
<td>27105 F</td>
<td>ATGGCTCGTGGAAGGGTGTCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27105 R</td>
<td>GGGTATATGGGATATCAGGTCG</td>
</tr>
<tr>
<td>gene13354</td>
<td>E</td>
<td>13354 F</td>
<td>GAGATGAAAAGAAATTGAAAATGCGAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13354 R</td>
<td>CCGATGAAAGGAGCCAGTCTCC</td>
</tr>
<tr>
<td>gene15202</td>
<td>B</td>
<td>15202 F</td>
<td>ATGGCGAGAGGTAGATCCAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15202 R</td>
<td>GAACCTCTCTCTCTCCATCC</td>
</tr>
<tr>
<td>gene11267</td>
<td>B</td>
<td>11267 F</td>
<td>ATGGGAAGGAGATGGAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11267 R</td>
<td>AATAGGCTGGAGCTGAAGAATG</td>
</tr>
<tr>
<td>gene11268</td>
<td>B</td>
<td>11268 F</td>
<td>ATGGGGGAGGGAGATGGAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11268 R</td>
<td>CGTGGAGATCGCTGATTAGG</td>
</tr>
<tr>
<td>gene14896</td>
<td>-</td>
<td>14896 F</td>
<td>GGGTCGAGGGAAGGTTGAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14896 R</td>
<td>GACATAAAATGACCTCCACCGTGT</td>
</tr>
<tr>
<td>gene06301</td>
<td>-</td>
<td>06301 F</td>
<td>GGGTAGGGGAAAGTGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>06301 R</td>
<td>TGGTTGGCAGCCCAGCTTTTA</td>
</tr>
<tr>
<td>gene12119</td>
<td>-</td>
<td>12119 F</td>
<td>GCCGGAGAGTTATCGGTTCTTTG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12119 R</td>
<td>GCTCCCTTTTGGTCAAGTCCGG</td>
</tr>
<tr>
<td>gene05027</td>
<td>-</td>
<td>05027 F</td>
<td>TCAGTAGCAGGCAAGTCCCTTTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>05027 R</td>
<td>GAATCATTAGAATGGGGCGG</td>
</tr>
<tr>
<td>gene19425</td>
<td>C</td>
<td>19425 F</td>
<td>TGAGAGGCAAACCCAGATGGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19425 R</td>
<td>CTCCAGTCTCTCTCTGCGACGT</td>
</tr>
<tr>
<td>gene03714</td>
<td>-</td>
<td>03714 F</td>
<td>AAGCCCGCAGCAGAAGAGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03714 R</td>
<td>GCGCCAGAGTCATCATCTACAGAA</td>
</tr>
</tbody>
</table>
Figure 2.1 Proposed MADS model of floral development for *Fragaria*
Figure 2.2 Phenotypic traits observed in the gp mutant
A. WT flower (left) gp flower (right) at anthesis; B. Adaxial side of WT petal (left) and gp petal (right); C. Abaxial side of WT petal (left) and gp petal (right); D. WT stamens (top) and gp stamens (bottom); E. WT carpels (top) and gp carpels (bottom); F. gp anthers that converted to sepals; Pollen viability stain of WT pollen (G) and gp pollen (H); I. WT fruit (left) collected 32 days after anthesis and gp fruit (right) collected 40 days after anthesis. WT, wild type; gp, green petal.
Figure 2.3 Molecular analysis of the T-DNA insertion sites.
A. Multiplex PCR of FV10-C02-721 T₁ segregating plants (n=92). B. Hi-TAIL PCR of FV10-C02-721. LAD 1/3 and LAD 2/4 refer to the combinations of Long Arbitrary Degenerate primers used for each lane pair. Bands boxed in gray were sequenced. HYG: hygromycin, GFP, green fluorescent protein, LB: left border, RB: right border, FVANS: *F. vesca ANTHOCYANIN SYNTHASE.*
**Figure 2.4 T-DNA insertion sites and zygosity analysis**

A. GBrowse screenshot for the scaffold location of Insertion 1. The flanking sequence alignment is indicated by the gray bar. B. Representative Zygosity PCR for segregation of Insertion 1 in the T₁ FV10-C02-721 segregating population. C. GBrowse screenshot for the scaffold location of Insertion 2. The flanking sequence alignment is indicated by the gray bar. D. Representative Zygosity PCR for segregation of Insertion 2 in the T₁ FV10-C02-721 segregating population.
<table>
<thead>
<tr>
<th>Hybrid Gene ID</th>
<th>Class</th>
<th>WT GP</th>
<th>Hybrid Gene ID</th>
<th>Class</th>
<th>WT GP</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene26118</td>
<td>E</td>
<td></td>
<td>gene13356</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>gene04229</td>
<td>E</td>
<td></td>
<td>gene04228</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>gene07201</td>
<td>E</td>
<td></td>
<td>gene26429</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>gene04563</td>
<td>E</td>
<td></td>
<td>gene20134</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>gene26119</td>
<td>A</td>
<td></td>
<td>gene32354</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>gene20808</td>
<td>E</td>
<td></td>
<td>gene20015</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td>gene04562</td>
<td>A</td>
<td></td>
<td>gene29503</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>gene19428</td>
<td>E</td>
<td></td>
<td>gene12120</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>gene25070</td>
<td>B</td>
<td></td>
<td>gene27105</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>gene24420</td>
<td>C</td>
<td></td>
<td>gene13354</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>gene14896</td>
<td>-</td>
<td></td>
<td>gene11267</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>gene06301</td>
<td>-</td>
<td></td>
<td>gene05027</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>gene12119</td>
<td>-</td>
<td></td>
<td>gene19425</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>gene24852</td>
<td>D</td>
<td></td>
<td>gene15202*</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>gene03714</td>
<td>-</td>
<td></td>
<td>gene11268</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Actin</td>
<td>+</td>
<td></td>
<td>Actin</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.5 RT-PCR of hybrid gene model candidate genes**

This table shows the RT-PCR product amplification for wild type (WT) and *green petal* (GP) for each of the hybrid gene model candidates listed in Table 2.3. The class is the tentative ABCDE model class assigned based on comparison to the GenBank database, using the BLASTx tool. The – symbol was assigned if the class could not be determined. A box is placed around gene07201, to highlight the observed band shift between the WT and GP products.
Figure 2.6 Analysis of gene07201
A. RT-PCR products for amplification of gene07201 on a 2% agarose gel. B. Alignment of sequence obtained from bands in A, for the wild type (WT) and green petal (GP) samples displaying the 37 bp deletion. C. Gene diagram of FvSEP3 (gene07201), containing primer placement for deletion genotyping and complementation construct (I: Intervening domain, K: Keratin-like domain, C: C-terminal domain, arrows represent primers, Roman numeral represent exons). D. Representative gel showing gene07201 deletion PCR results and E. Representative gel showing NdeI digest of the PCR bands in panel D.
Figure 2.7 Amino acid alignment (ClustalW) of putative SEP3 proteins from different species.

Alignment of putative SEP3 amino acid sequences from *A. thaliana* (At1G24260.2), *F. chiloensis* (ADF49576.1) and *M. domestica* (ADL36740.1) with FvSEP3. The white letters shaded with black indicate residues identical to the consensus sequence. The gp deletion start site is indicated with the black arrow.
**References**


Ma H, Yanofsky MF, Meyerowitz EM (1991) AGL1-AGL6, an Arabidopsis gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. 5: 484-495


**Abstract**

Strawberry, genus *Fragaria*, is an attractive model to study flavonoid biosynthesis due to the accumulation of flavonoid pigments throughout the plant in different tissues and organs, i.e., fruits, achenes, petiole and stolon tissue, petals and leaf margins under different environmental conditions. The various spatial pattern of expression is an indication that different regulatory gene combinations control the expression of structural genes coding for enzymes in the flavonoid biosynthesis pathway in different tissues. *Fragaria vesca* L. Hawaii 4 is a diploid accession with white fruit, red petioles, red runners, brown achenes, perfect flowers and a day neutral flowering habit. In this study we describe the characterization of a putative *F. vesca* T-DNA insertional mutant, called white runner (*wr*) that lacks red pigment in the stems and runners. Nucleotide sequence of the T-DNA flanking region showed that the T-DNA had integrated into a highly repetitive Long Terminal Repeat (LTR) retrotransposon region lacking potential candidate genes in the adjacent region. Analysis of a T₂ population of *wr* revealed that the *wr* phenotype segregated as a single recessive mutation, but did not cosegregate with the T-DNA insertion, suggesting that the *wr* mutant was generated by heritable somaclonal variation. Our hypothesis was that if the *wr* phenotype is caused by a mutation in a transcription factor (TF) we would expect altered expression of multiple structural genes encoding key enzymes in the flavonoid pathway. Alternatively, if no changes in the gene expression of the structural genes were detected, then the mutation likely resides within a structural gene resulting in a modified enzyme. Three genes encoding critical enzymes in flavonoid metabolism, chalcone synthase (*FvCHS*), flavanone 3-hydroxylase (*FvF3H*), and flavonoid 3’-hydroxylase (*FvF3’H*), were selected for gene expression analysis in both wild type and *wr* tissue. Our gene expression and sequence analysis revealed a 20 bp deletion in the second exon of *FvF3H*, changing 182 amino acids in the conserved C-terminal region. Complementation analysis with *AtF3H* and *FvF3H* in the *wr* mutant was used to verify the cause of the *wr* phenotype.
Introduction

Flavonoids are an important class of plant-specific secondary metabolites that contribute to the rainbow of colors seen in fruits, flowers, seeds, stems and leaves. The extensively studied flavonoid biosynthesis pathway has illuminated the origin of many diverse compounds including both colorless flavonols and isoflavonoids, as well as colored pigments such as anthocyanins, phlobaphenes and proanthocyanidins. Flavonoid biosynthesis has been studied in plants as far back as 1664, when Robert Boyle described the acid/base effects on plant pigments (Winkel-Shirley 2001). Later, in the 19th century, Gregor Mendel utilized flower color to conduct his famous pea experiments that every biology undergraduate learns about today. In the 1940s and 1950s, Barbara McClintock utilized the variable coloration of maize kernels to demonstrate the properties of transposable elements as they moved into and out of flavonoid biosynthetic genes.

Flavonoids have a wide range of biological functions within plants. These include protection from UV light, recruitment of pollinators, male fertility and auxin transport (Brown et al. 2001; Harborne and Williams 2000; Mo et al. 1992; Winkel-Shirley 2002). These secondary metabolites, along with carotenoids, are responsible for the range of fall leaf colors as well as the red, blue and purple pigments in flower, stem and leaf tissues. Often located in the epidermal layers of leaves and tissue exposed to UV light, flavonoids protect cells from photo-oxidative damage (Beevers 2002; Winkel-Shirley 2002). Abnormal accumulation of red or purple pigments in tissues is often a sign of physiological stress.

In the autumn as the leaves senesce, plants actively transport nutrients from the leaves to storage tissues for use in the next season (Beevers 2002). High light intensity and cool temperatures may increase the production of anthocyanins, resulting in the red coloration in senescing leaf tissue (Matile 2000). Anthocyanins preferentially absorb blue-green light and their accumulation may improve the quality and quantity of light absorbed by chlorophyll (Beevers 2002; Merzlyak and Chivkunova 2000). As chlorophyll is broken down anthocyanin accumulation acts as a “pigment screen” shielding the photosynthetic machinery from photo-oxidative damage by light and reducing the oxidative stress caused by reactive oxygen species. This protective effect improves the nutrient recovery in the plant.

Insects detect different wavelengths than hummingbirds and flower color is critical for attracting the appropriate pollinator (Kevan et al. 1996). Changes in floral color in a
subpopulation of plants can lead to pollinator species shifts resulting in genetic divergence, sexual incompatibility and speciation (Bradshaw and Schemske 2003). Coloration in fruits attracts herbivores that facilitate seed dispersal and ultimately leads to success of reproduction. These functions of flavonoids are critical to the life cycle of flowering plants.

Several species, including maize and petunia, require flavonoids for proper pollen germination and tube growth. Flavonoids give pollen its yellow color and constitute as much as 4% of the dry weight. Mo et al. (1992) showed that transgenic petunias with white pollen were self-sterile. Chemical analysis confirmed the absence of kaempferol, the flavonol required for pollen germination in petunia. Tests were done with both mutant petunia and mutant maize, administering wild-type stigma extracts during pollen germination. Self-crosses in the presence of wild-type flavonol extracts produced > 90% viable seed, whereas self-crosses without wild-type flavonol extracts yielded <1% seed set (Mo et al. 1992). The similar reproductive defects in the two unrelated species indicate a generic need for flavonoids during pollen germination, though not all species have this requirement, as A. thaliana chalcone synthase transparent testa (tt4) mutants are unable to produce flavonols but are completely self-fertile (Burbulis et al. 1996; Ylstra et al. 1996).

Studies with A. thaliana and Medicago truncatula have shown that certain flavonoids are endogenous negative regulators of auxin transport (Brown et al. 2001; Wasson et al. 2006). A. thaliana seedlings grown in the presence of the flavonoid naringenin exhibited severely stunted root growth, or agravitropism resulting in curved root growth. The A. thaliana flavonoid-deficient mutants, tt4 (2YY6), lack anthocyanins and show growth characteristics expected of elevated auxin transport. These mutants showed reduced apical dominance, increased secondary inflorescence branching and increased lateral root branching (Brown et al. 2001). Transgenic M. truncatula, with silenced flavonoid synthesis through RNAi and Agrobacterium rhizogenes transformation, were unable to form nodules; the phenotype was rescued by the application of two flavonoids, naringenin and liquiritigenin, supporting the hypothesis that flavonoids are required for nodule formation (Wasson et al. 2006).

One of the strongest forces driving research on flavonoids is the potential impact to human health and nutrition. The pigments that color favorite fruits such as grapes, blueberries and apples also have bioactivities, e.g., anti-inflammatory, anti-oxidant, bactericidal and
fungicidal, providing protection against human diseases (Aron and Kennedy 2008; Hannum 2004; Kris-Etherton et al. 2002; Prior and Wu 2006). Famous examples of research include the benefits of resveratrol consumption, a polyphenol in red wine, thought to exert cardioprotective effects, the effects of flavonoid rich dark chocolate on cholesterol levels and the impact of pomegranate juice consumption on atherosclerosis (Aviram et al. 2000; Baba et al. 2007; Freedman et al. 2001; Pace-Asciak et al. 1995; Rein et al. 2000; Sanbongi et al. 1997; Wan et al. 2001). These are only a few of the many potential benefits flavonoids have on human health (Crozier et al. 2009; Kris-Etherton et al. 2002; Kris-Etherton and Keen 2002; Prior and Wu 2006). This has created an interest in biotechnology and the creation of high flavonoid ‘functional’ foods.

There are two genetic components of flavonoid biosynthesis: structural genes that encode enzymes that catalyze the reactions and regulatory genes that encode for transcription factors (TFs) that control expression of the structural genes. This pathway has been studied in detail in many species, including but not limited to A. thaliana, Antirrhinum majus, Petunia hybrida, Vitis vinifera, many Rosaceous species, Zea mays, Perilla frutescens, and Ipomoea nil. Many biotechnology tools and mutagenic anomalies have been used to understand the effects of genetic defects in the pathway. The tools include T-DNA insertional mutants, RNA interference, transposable element mutations, spontaneous DNA indels and over-expression mutants.

Much of the research has focused on the MYB, bHLH and WD40 TFs, responsible for regulating the expression of the structural genes. Like the MADS box TFs, those controlling flavonoid biosynthesis are tightly controlled spatially and temporally, with tissue specificity and expression in response to environmental cues, external stress and stimuli. The MYB, bHLH and WD40 TFs are involved in many cellular processes beyond regulating flavonoid biosynthetic gene transcription, including trichome and root hair formation, cell division, chromatin remodeling and signal transduction (Larkin et al. 1994; Morita et al. 2006; van Nocker and Ludwig 2003).

TF and structural gene knock-out mutants and over-expression lines have contributed to our understanding of the roles of these genes in the pigment accumulation in specific organs of numerous plant species. In Japanese morning glory (I. nil), a 2 bp deletion in InMYB1 yields plants with white flowers, red stems and colored seeds, whereas a 7 bp insertion in InWDR1
Yields plants with white flowers, green stems, ivory seeds and defects in seed trichome formation (Morita et al. 2006). Allelic variation in MYB transcription factors has been implicated in causing the variation in pigmentation in the peel and flesh of apple fruit (Allan et al. 2008; Espley et al. 2007; Kim et al. 2005). Constitutive expression of the MYB transcription factor Production of Anthocyanin Pigment 1 (PAP1) causes upregulation of many genes in the flavonoid pathway, yielding red pigmentation in the leaves of A. thaliana (Rowan et al. 2009). Over-expression of P. hybrida chalcone isomerase in tomato resulted in significantly elevated flavonols in the fruit peels (Muir et al. 2001). Elevated levels of maize dihydroflavonol 4-reductase (DFR) in P. hybrida reduced its substrate specificity, converting dihydrokaempflerol into pelargonidin based anthocyanins resulting in red flowers (Meyer et al. 1987). A tri-transgene experiment in tobacco, suppressing flavonol synthase (FLS) and flavonoid 3’hydroxylase (F3’H), and over-expressing G. hybrida DFR simultaneously resulted in red flowers (Nakatsuka et al. 2007). An 800 bp deletion in DFR was responsible for the change from red to yellow color in onion skin (Kim et al. 2005). Movement of transposable elements in the gene encoding chalcone synthase is responsible for unstable variegated flower pigmentation in snapdragon (Martin et al. 1985).

Fragaria is an attractive experimental system to study flavonoid biosynthesis due to the localization of flavonoid pigments throughout the plant in different tissue types, i.e., fruits (receptacles), achenes, petioles, runners, petals and leaf margins in response to environmental stress. The spatial patterns of accumulation may be an indication that alternative combinations of regulatory genes may operate in different tissues. Strawberry fruit and achene color is composed of pelargonidins, cyanidins and proanthocyanidins. Different combinations of derivatives of these compounds result in the shades of pink and red found in various cultivars and species of strawberry. There are also strawberries that have no pigment at all, referred to as white- or yellow-fruited strawberries. Fragaria vesca L. Hawaii 4 is a diploid accession with white fruit, red petioles, red runners, brown achenes, perfect flowers and a day-neutral flowering habit.

The following study presents a F. vesca mutation, called white runner (wr) that lacks red pigment in the stems and runners. The mutant was identified in the T1 generation after self-pollination of T0 insertional mutants that had been transformed with the binary plant transformation vector, pCAMBIA1304. The flanking region of the T-DNA insertion site was
analyzed at the sequence level and for association with the phenotype. Because of the phenotype we expected altered expression levels of multiple structural genes encoding enzymes in the flavonoid pathway, caused by a mutation in a regulatory MYB, bHLH or WD40 TF. Alternatively, if no changes in gene expression in the structural genes could be detected, then the mutation may reside within a structural gene resulting in a modified enzyme. To test this possibility, we selected three critical structural genes for expression analysis in both wild type (WT) and wr in order to identify the mutation responsible for the wr phenotype.

Methods

Plant material. WT F. vesca Hawaii 4 (PI551572) was transformed with pCAMBIA1304 through Agrobacterium mediated transformation and regenerated in vitro, described by Oosumi et al. (2006). Independent T-DNA insertional mutant lines were grown from seeds collected from T₀ plants and screened for phenotype and insertion site, described in Ruiz-Rojas et al. (2010). In T₁ insertional mutant line FV10-C04-354 there was a phenotype that segregated as a single recessive Mendelian trait characterized by an absence of red pigmentation in the petiole and runner tissue, and it was therefore named white runner.

Molecular analysis of T-DNA insertion. Hi-TAIL PCR was used to amplify the F. vesca genomic sequence flanking the T-DNA insertional site, as described in Liu and Chen (2007). The PCR amplification products were sequenced at the Tuft’s CORE sequencing facility (Tuft’s University CORE Facility, Boston, MA). Flanking DNA sequences were aligned to the F. vesca scaffold database using the BLASTn tool, (July 2011; Shulaev et al. 2010).

Plant growth. T₂ seed of FV10-C04-354 GFP-negative lines B17-10 and B17-12 were germinated in MS medium, screened under an Olympus SZX 51 microscope with an SZX-mGFP filter to verify absence of GFP expression, then seedlings were planted in hydrated and autoclaved Jiffy-7 peat plugs in Magenta boxes. Plants were grown at room temperature, approximately 22°C, on a rack with four fluorescent bulbs, under a 16 h photoperiod. The wr phenotype was easily scored in 3-week-old plants.

Histological staining. Red Runners and white runners from FV10-C04-354-B17-10 plants, were de-colored overnight in absolute ethanol with gentle shaking. Runner tissue was stained with 1.25 % Vanillin-HCL for 5 min or 0.5% p-dimethylaminocinnamaldehyde (DMACA) (Acros Organics, Thermo Fisher Scientific) for 3 h and destained with 70% ethanol. Staining was
documented using an Olympus Rebel XTi SLR camera. T2 FV10-C04-354-B17-10 and B17-12 seeds were washed with water, dried on filter papers and stained for 4 days in 2% DMACA. Seeds were then washed five times with 70% ethanol and dried on filter paper. Stained and unstained seeds were viewed using an Olympus SZX 12 microscope and photographed with an Olympus DP70 camera and Image Pro Plus Software.

**Metabolite analysis.** Petiole tissue from two FV10-C04-136-B17-10 red runner plants and two B17-10 white runner plants was collected for metabolite analysis. Approximately 60 mg of tissue from each plant was collected in methanol-rinsed, 2 ml tubes containing two 2.5 mm stainless steel beads and snap frozen on liquid nitrogen. Tissue was ground in a paint shaker for three cycles at 30 sec each, with liquid nitrogen poured over the tubes between cycles. Ten µl of extraction buffer (99% methanol/ 1% acetic acid) were added for every mg of tissue. Tubes were vortexed to distribute buffer and kept on ice for the remainder of the extraction period. Samples were centrifuged at 15,000 rpm at 0°C for 15 min and the supernatant was transferred to a clean methanol-rinsed tube. Samples were centrifuged again at the same conditions and supernatant was transferred a second time. Extracts were dried in a vacuum dryer and resuspended with solvent B (10 mM ammonium acetate, 1% acetic acid in methanol) at a rate of 5 µl of solvent B per mg original weight. Extracts were stored at -80°C until analysis. Analysis was performed on an Agilent 1100 LC binary pump coupled to 3200 QTrap Linear Ion Trap Quadrupole Mass Spectrometer (ABSciex Instruments, Foster City, CA, USA).

Chromatographic separation was performed on a Thermo Hypersil Beta Basic 18 column (100 mm x 1 mm, 5 µm particle size). The mobile phases were (A) 10 mM ammonium acetate, 1% acetic acid in water, and (B) 10 mM ammonium acetate, 1% acetic acid in methanol. The LC gradient used was 0-5 min, 2% B; 5-19 min, 2-38% B; 19-24 min, 38-100% B; 24-31 min, 100% B; 31-34 min, 100-2% B; 34-50 min, 2% B. The parameters used for the Q1 MS scans were 4.7 kV for the ion spray voltage, 63 V for the declustering potential, 6.3 V for the entrance potential, with the interface heater set to 120°C and the curtain and nebulizer gases set to 20 and 25 psi, respectively. Compounds were analyzed in the positive ion mode and scanned from m/z of 200 to 800 amu. Data were analyzed with Analyst software (ABSciex).

**Selection of structural genes.** Three critical structural genes were chosen based on their position in the flavonoid pathway. Chalcone synthase, flavanone 3-hydroxylase, and flavonoid 3’-
hydroxylase are required for the synthesis of cyanidins, pelargonidins and proanthocyanidins (Figure 3.1). *A. thaliana* mRNA sequence for chalcone synthase (*CHS*, AT5G13930), flavanone 3-hydroxylase (*F3H*, AT3G51240), flavonoid 3’-hydroxylase (*F3’H*, AT5G07990), and *transparent testa glabra 1* (*TTG1*, AT5G24520) were obtained from The Arabidopsis Information Resource (TAIR) and used to identify homologs in the *F. vesca* hybrid gene model database using the BLAST tool. Nucleotide sequence from *F. vesca* gene models: gene26825 (*FvCHS*), gene14611 (*FvF3H*), gene25801 (*FvF3’H*) and gene12450 (*FvTTG1*) were translated into all three frames and compared to the GenBank database using the BLASTx tool to check for completeness and annotated for coding regions in SeqBuilder (DNASTAR Inc., Lasergene 8, Madison, WI, USA). Primers for RT-PCR were designed using SeqBuilder.

**Molecular analysis.** RNA from FV10-C04-354-B17-10 (WT red runner) and B17-12 (*wr*) tissue, *F. vesca* PI551572 de-seeded white-fruit tissue and *F. vesca* PI602578 de-seeded red fruit tissue was extracted according to Moser et al. (2004), with one change: omittance of the proteinase K step. First strand cDNA was synthesized using Superscript III First-strand Synthesis System (Invitrogen), according to instructions. *ExTaq* DNA polymerase (Takara, Japan) was used for RT-PCR amplification of the structural genes. Thermal cycling conditions used for all primer sets as follows: 94°C for 3 min, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, followed by 3 min at 72°C. Products stained 1x with EzVision (Amresco) were run on a 1% agarose/ 1xTAE gel stained 0.5x with SyberSafe (Invitrogen) and imaged using an Alpha Imager 3400 (Alpha Innotech, Protein Simple, Santa Clara, CA, USA) and FluorChem SP (Alpha Innotech) software. PCR fragments for gene14611 were sent to Quintara Biosciences (Berkeley, CA) for sequencing. Sequences were aligned using ClustalW (Goujon et al. 2010; Larkin et al. 2007) and nucleotide and predicted protein sequences were annotated SeqBuilder.

**Genotyping for gene14611 deletion.** Genomic DNA was isolated from young leaf buds and runner tissues using methods described previously (Oosumi et al. 2006). Primers flanking the deletion site in *FvF3H* (gene14611) were designed using SeqBuilder, to utilize a unique Neos restriction site within the deletion for PCR product digest. *ExTaq* DNA polymerase was used for amplification of the deletion site and NEB restriction enzyme NcoI was used to genotype for the deletion. The primers used for amplification of the deletion region were named 14611_delF and 14611_delR, yielding an expected product size of 388 bp (Table 3.1). Thermal cycling
conditions used were as follows: 94°C for 3 min, followed by 35 cycles of (94°C for 1 min, 55°C for 1 min and 72°C for 1 min), followed by 3 min at 72°C. PCR Products stained to a 1x concentration with ExVision (Amresco) were run on a 2% agarose/1x TAE gel stained to a 1x concentration with SyberSafe (Invitrogen). NcoI restriction digests were carried out at 37°C for 3 h and the entire digest volume was stained to a 1x concentration with EzVision and run on a 2% agarose/ 1x TAE gel stained to a 1x concentration with SyberSafe (Invitrogen). If homozygous for the deletion, the PCR product did not contain the NcoI site, and the band size remained 388 bp. If hemizygous for the deletion, the expected band sizes of the digested product were 164 bp and 225 bp, yielding three bands, two from the WT after cleavage and the intact band, from the deletion line. If homozygous WT both PCR alleles would be cleaved, yielding only two band sizes of 164 bp and 225 bp.

**Protein sequence analysis of FvF3H.** Protein sequences for *A. thaliana* F3H (NP_190692.1), *Fragaria x ananassa* F3H (AAU04792.1), *Ipomea purpurea* F3H (AAB41102.1), *Nicotiana tabacum* F3H (BAF96938.1), *Dianthus caryophyllus* F3H (CAA49839.1) and *Petunia hybrida* F3H (Q07353.1) from GenBank were compared to the protein sequence of *Fragaria vesca* F3H (gene14611) and the *F. vesca* F3H with deletion by aligning with ClustalW. The protein sequence alignment was annotated to highlight the conserved residues.

**Assembly of F3H complementation construct.** RT-PCR primers were designed in the upstream 5’ and 3’ UTR regions of *FvF3H* (gene14611) to amplify the entire cDNA transcript using Sequence Builder. The product was amplified using BioRad’s iPROOF High-Fidelity master mix *Pfu* polymerase according to instructions. The binary vector AKK 1434 (kindly provided by Christopher Taylor, Donald Danforth Science Center, St. Louis, MO) containing the Strawberry Vein Banding Virus promoter (SVBV) ligated to the coding region of GUS was digested with *EcoRV* and *Eco*I CRI to release the GUS cassette. The digested vector was run on a 1% low melting point agarose (LMP)/ 1x TAE gel. The ~4000 bp vector fragment was excised from the gel and incubated at 70°C for 10 min. The *FvF3H* RT-PCR product was blunt-end ligated into the modified AKK 1434 vector in LMP agarose with Promega T4 Rapid Ligation reagents. The ligation reaction was used for transformation of *E.coli* strain DH5α and colonies were screened for the insertion and correct orientation of the SVBV promoter with the *FvF3H* cDNA by colony PCR with the primers SVBV Pro and gene14611R (Table 3.1). The SVBV: *FvF3H* cassette was
PCR amplified using primers SVBV Pro and SVBV Term with iPROOF High-Fidelity master mix (Bio-Rad) according to manufacturer’s instructions and fractionated on a 1% LMP agarose/1x TAE gel. The 2600 bp fragment was excised and the gel slice was heated to 70°C to melt the agarose in preparation for ligation. The binary transformation vector, pCAMBIA1304, was digested with EcoRI and fractionated on a 1% LMP agarose/1x TAE gel. The SVBV:FvF3H cassette in LMP agarose was blunt-end ligated into pCAMBIA1304 in LMP agarose using T4 Rapid Ligation reagents (Promega). The ligation reaction was used for transformation of E. coli strain DH5α and antibiotic resistant colonies were screened for presence of the gene insertion by colony PCR. The correctly assembled complementation construct was introduced into A. tumefaciens strain EHA105 using electroporation. A pWWTY2 vector containing AtF3H coding region construct from A. thaliana (AT3G51240) in GV3101(pMP90) was kindly provided by Sherry Hildreth (Winkel Lab, Virginia Tech, Blacksburg VA).

**Complementation analysis.** The above FvF3H and AtF3H complementation constructs were used for transformation of FV10-C04-354-B17-12 wr tissue as described by Oosumi et al. (2006) and (Zhao et al. 2004).

**Results**

**Identification of the white runner phenotype and molecular analysis of the T-DNA.** The WT F. vesca accession used in this study has white fruits, brown achenes, and petioles and runners with red pigmentation. T₁ segregation analysis of a F. vesca pCAMBIA1304 insertional mutant population revealed a mutation, in line FV10-C04-354, affecting secondary metabolite accumulation, characterized by an absence of red pigmentation in the petioles and runners. Analysis of multiple Hi-TAIL PCR fragments showed that each one was the same sequence, indicating that FV10-C04-354 contains a single T-DNA insertion (data not shown). A nucleotide sequence alignment to the F. vesca scaffold database returned significant homology to five scaffolds (Figure 3.5.B). Closer inspection of the insertion site in scaffold 513177 revealed that the insertion site is located in a putative LTR retrotransposon region (Figure 3.5.A). The repetitive, homologous sequence between the five scaffold hits prevented primer design for segregation analysis. The lack of candidate genes that could be responsible for the wr phenotype in the region flanking the insertion site indicated that the T-DNA insertion might be unrelated to the phenotype. The T-DNA insertion may have mobilized a retrotransposon
resulting in the wr phenotype, or somaclonal mutation as a result of the Agrobacterium infection and in vitro regeneration may be the cause.

**Segregation analysis of GFP negative FV10-C04-354.** T2 seedlings from FV10-CO4-354 that were collected as GFP negative T-DNA insertional segregates showed the wr phenotype. This absence of linkage indicates that the T-DNA and the phenotype are unrelated. Of the T2 segregating population of line FV10-CO4-354-B17-10, 12 plants had the red runner phenotype and 4 plants had the white runner phenotype, a 3:1 segregation ratio suggesting the wr phenotype was due to a single recessive mutation ($\chi^2 = 0, P = 1$). All 15 plants of T2 line FV10-CO4-354-B17-12 exhibited the wr phenotype, indicating that the trait was homozygous in this line. These two populations were used for genetic analysis of the trait.

**Histological staining of runners and seeds.** Vanillin-HCl and DMACA were used an indicators of flavan-3-ol localization in red runners vs. white runners and seeds in the two lines, FV10-C04-354-B17-10 and B17-12. Seeds from the lines FV10-CO4-354-B17-10 and FV10-CO4-354-B17-12 both showed brown pigmentation, indicating that the homozygous wr line B17-12 still produced proanthocyanidins despite the lack of pigment in the stems and runners (Figure 3.2.G, I). Both lines showed no difference in DMACA staining (Figure 3.2.H, J). Vanillin-HCl staining of the FV10-C04-354-B17-10 red runner showed vivid pink staining throughout the runner, petiole, leaf veins and tips of the leaf (Figure 3.2.C). Vanillin-HCl staining of the FV10-C04-354-B17-10 wr sibling only showed pink staining at the base of the crown where the root primordia emerge (Figure 3.2.D). DMACA staining of the red runner showed deep blue stain in the tips of the leaves and the sheath covering of the crown, the runner node and the developing runner tip (Figure 3.2.E). Blue staining was also observed at the base of the trichomes throughout the runner and petiole tissue. DMACA staining of the wr showed paler blue staining in the same regions as the red runner, which was nearly absent at the base of the trichomes (Figure 3.2.F). This result further indicates that the wr plants were able to produce proanthocyanidins.

**Metabolite analysis.** Metabolite analysis of petiole segments from two red FV10-C04-354-B17-10 red runner and white runner plants (Figure 3.4) showed in the Q1 scan trace an absence of major peaks in the white runner samples (Figure 3.3). Two peaks were present in the white runner samples that were not present in the red runner samples, possibly indicating accumulation
of precursor compounds, or new compounds caused by the redirection of the pathway (Figure 3.3). There were reduced levels of five phenolics in the white runner petioles, consistent with the following compounds: A. \( m/z \) 579, pelargonidin 3-rutinoside; B. \( m/z \) 625, unknown; C. \( m/z \) 463, quercetin 3-glucoside; D. \( m/z \) 449, cyanidin 3-glucoside; E. \( m/z \) 611, quercetin 3-rutinoside (Figure 3.4). Tandem mass spectrometry analysis will have to be performed to make more confident identification of the compounds. For all of the compounds the \( wr \) samples show lower levels of absolute abundance when compared to the red runner sample B1710-11. There is variability between the two red runner samples, which will have to be resolved by performing additional biological and technical replicates to determine the extent of the difference. These data do show that there are reduced levels of compounds in two branches of the flavonoid pathway (Figure 3.4): flavonols (quercetins) and anthocyanins (pelagonidins and cyanidins) in the \( wr \) mutants.

**Selection of key structural flavonoid biosynthesis genes and expression analysis.** The flavonoid pathway is controlled by structural genes encoding enzymes in the pathway and TFs that regulate structural gene expression. It was hypothesized that if the mutation causing the \( wr \) phenotype resided in a TF then expression of one or more structural genes would be affected, relative to WT. Alternatively, if the expression of the structural genes was unchanged then the mutation may lie within a structural gene. The structural genes that were selected were chalcone synthase (FvCHS), flavanone 3-hydroxylase (FvF3H), and flavonoid 3’-hydroxylase (FvF3’H). *A. thaliana* mRNA sequence for these three genes was obtained from TAIR and compared to the *F. vesca* hybrid gene model database with the BLASTn tool to retrieve the putative *F. vesca* mRNA sequences. These three structural genes were selected because of their critical roles in flavonoid biosynthesis. *TTG1*, a gene coding for a WD40 repeat TF known to control anthocyanin accumulation in leaves and stems, as well as root hair and trichome development in *A. thaliana* (Larkin et al. 1994), was selected for expression analysis in the *F. vesca* red runner and \( wr \) as a control. Trichome development was unimpaired in the \( wr \) mutants, so changes in expression of FvTTG1 were not expected. RT-PCR amplification of the candidate structural genes and FvTTG1 showed no difference in expression between white and red runnered plants. This suggests that the mutation causing the \( wr \) phenotype may reside within a structural gene in the flavonoid pathway rather than a gene coding for a TF. A slight band shift was detected in the PCR product of FvF3H (gene14611) between the WT and \( wr \), which sequence analysis revealed
as a 20 bp deletion in the middle of the second exon of *FvF3H* in the *wr* (Figure 3.6). We then used CAPS genotype analysis on populations FV10-CO4-354-B17-10 and FV10-CO4-354-B17-12 with primers designed to flank the deletion site followed by Nco1 digestion of the PCR products exploiting a unique restriction site within the deletion, so that only the WT allele products would be cleaved. The deletion co-segregated perfectly with the *wr* phenotype (Figure 3.8).

**Impact of deletion on FvF3H.** The 20 bp deletion in *FvF3H* results in a frameshift beginning at encoded amino acid position L\textsuperscript{174}, with the protein sequence change beginning at V\textsuperscript{176}, and ending with a premature stop codon at amino acid position 275 (Figure 3.7). ClustalW alignment of the seven F3H proteins from GenBank with WT *FvF3H* and the predicted protein in the *wr* line, shows a high degree of conservation, especially in the region downstream of the deletion site. The deletion most likely results in a nonfunctional F3H protein. The amino acid sequence of FvF3H was aligned back to the *F. vesca* hybrid gene model database using the BLASTp tool to determine if there are other copies of F3H present. This alignment returned many putative F3H sequences, though none of the sequences grouped with FvF3H in a phylogenetic analysis and all had less than 40% identity with FvF3H (data not shown).

**Analysis of the candidate gene expression in fruits.** Given that the white fruit phenotype may be due to a mutation associated with FvF3H, we conducted gene expression analysis on red fruit from *F. vesca* PI602578 and white fruit from PI551572 (Figure 3.9.A). The expression of the three structural candidate genes in white fruit was much lower relative to expression in the red fruit (Figure 3.9B). Because all three structural genes appear to be under-expressed in fruit tissue in the white-fruited accession, this indicates that the mutation causing the white fruit phenotype in *F. vesca* Hawaii 4 (PI551572) may not be located in *FvF3H*, but is more likely to be in a gene coding for a regulatory TF.

**Complementation analysis.** In order to confirm that the deletion in exon 2 of *FvF3H* caused the white runner phenotype, we assembled a complementation construct using the cDNA transcript of *FvF3H* from red-runnering strawberry tissue isolated from FV10-CO4-354-B17-10 with a constitutive promoter, the *Strawberry Vein Banding Virus* (SVBV2) promoter. This promoter-cDNA cassette was cloned into binary vector pCAMBIA1304. Additionally, *AtF3H* driven by a CaMV35s promoter was also transformed into *wr* tissue to determine if the *A. thaliana* gene
could substitute for FvF3H. Two A. tumefaciens transformation methods were used to transform runner segments of FV10-C04-354-B17-12 wr plants. Preliminary observations are the callus is regenerating on selection medium, and spots of red pigmentation are forming in the developing callus (Figure 3.10A, B). A few young shoots show darkened pigmentation in the leaf tissue (Figure 3.10C, D). The transformed regenerates will be analyzed for phenotype and expression of the rescue F3H.

**Discussion**

As more genetic resources for members of the Rosaceae family become available, we further our understanding of how the intricate networks of genes control many secondary metabolic pathways. Species within Rosaceae vary for flavonoid pigmentation in leaves, stems, fruits and seeds. Species within *Fragaria* are particularly well suited for genetic studies, due to their small size and short generation time, facilitating rapid mapping population development, amenability to *Agrobacterium* transformation and *in vitro* regeneration. The availability of a public genome sequence for *F. vesca* facilitates the rapid cloning and characterization of structural genes and TFs of the flavonoid pathway from *F. vesca* and other *Fragaria* species. As insertional mutant populations for *F. vesca* become more widely available, flavonoid biosynthesis studies will be aided by phenotypic variants.

In this study we have isolated a white runner phenotypic variant that was originally identified in a pCAMBIA 1304 insertional mutant population developed in a white-fruit accession of *F. vesca*. WT *F. vesca* Hawaii 4 has red pigmentation in the stems and runner tissue. Segregation analysis revealed that the wr phenotype was not associated with the T-DNA insertion site. The primary goal of this study was to identify the mutation leading to the wr phenotype.

Expression analysis of three key flavonoid genes led to the detection of a 20 bp deletion in exon 2 of FvF3H. In *F. vesca* FvF3H is predicted to comprise 364 amino acids, with the deletion occurring after position 176, the C-terminal region containing known iron and oxoglutarate binding sites. Besides causing the removal of six amino acids, the deletion caused a frame shift, changing the encoded amino acids up to a premature stop codon at position 275. This may result in a completely nonfunctional enzyme. This hypothesis is supported by C-terminal deletion studies of the orthologous F3H in *P. hybrida*; in these studies, loss of 29 C-
terminal amino acids resulted in significantly reduced enzyme activity (Lukacin et al. 2000b; Wellmann et al. 2004). Our \textit{wr} strawberry would be expected to have an even greater C-terminal deletion of FvF3H. Additionally, there are known to be strictly conserved residues in the C-terminal domain (H\textsuperscript{277}, H\textsuperscript{219}, R\textsuperscript{287}, D\textsuperscript{221}, numbers refer to FvF3H sequence) that, upon site directed mutagenesis in \textit{P. hybrida} F3H converting the WT amino acid to Q, N, Q and Q, respectively, resulted in reduction of catalytic activity to 0%, 0.5%, 0.1% and 0.5%, respectively, relative to wild type (Lukacin and Britsch 1997). Replacement of S\textsuperscript{289} by threonine, alanine and valine yielded reductions to 20%, 8% and 1%, respectively, relative to wild type (Lukacin et al. 2000a). The FvF3H deletion completely eliminates these conserved residues further supporting the hypothesis that \textit{wr} FvF3H is nonfunctional.

Since the WT \textit{F. vesca} Hawaii 4 does not have pigmentation in the fruits the impact on fruit due to the FvF3H deletion could not be assessed; however achenes produced on the white fruits of both the WT and \textit{wr} were brown indicating the accumulation of proanthocyanidins. Vanillin-HCl and DMACA staining of the runner tissue indicated that reduced amounts of proanthocyanidins were formed, relative to wild type (Figure 3.2.C-F). The unstained and DMACA stained achenes of B17-10 and B17-12 showed no difference also indicating the presence of proanthocyanidins. Due to the truncation of the \textit{wr} FvF3H the production of proanthocyanidins is most likely not due to a ‘leaky’ mutation. A BLASTp alignment of the amino acid sequence of FvF3H yielded additional putative F3H copies in \textit{F. vesca}; however identity and phylogenetic analysis cannot confirm the gene assignment of these sequences. There may be rare copies of F3H, whose expression is tightly regulated and only functions in specific tissues.

The F3H knockout mutants in Arabidopsis provide an alternative hypothesis to explain the proanthocyanidin accumulation in the strawberry \textit{wr} mutant. The \textit{transparent testa 6 (tt6)} locus is the homologous F3H locus in \textit{A. thaliana} and multiple \textit{tt6} mutants have been characterized (Shirley et al. 1995; Wisman et al. 1998). All of the \textit{tt6} mutants have pale brown seeds, indicating the production of proanthocyanidins, despite mutations disrupting AtF3H (Owens et al. 2008b). F3H is a 2-oxoglutarate dependent (2-OG) dioxygenase, using naringenin to produce dihydrokaempferol, the precursor substrate for flavonols, anthocyanins and proanthocyanidins. Enzymes in the 2-OG dioxygenase superfamily are found in eukaryotes and microorganisms (Martens et al. 2010). These enzymes are involved in many processes; all
catalyze reactions where two molecules of oxygen are incorporated into the substrates (Aravind and Koonin 2001; Decarolis and Deluca 1994). Five enzymes in the flavonoid pathway are members of this superfamily: flavanone 3-hydroxylase (F3H), flavonol synthase (FLS), anthocyanidin synthase (ANS), flavone synthase I (FNSI), and flavonol 6-hydroxylase (F6H) (Martens et al. 2010; Owens et al. 2008b). Owens et al. (2008) suggested that another 2-oxoglutarate dioxygenase, FLS or ANS, in the flavonoid pathway can partially complement the missing F3H activity in the *A. thaliana* tt6 mutants, resulting in pale brown seeds. This enzyme substitution has been validated *in vitro*, but not *in planta* (Owens et al. 2008a; Turnbull et al. 2004; Welford et al. 2001). F3H mutants in soybean and carnation also showed phenotypes indicating ‘leaky’ expression, which may have alternative explanations due to F3H complementation by other flavonoid 2-oxoglutarate dioxygenases (Dedio et al. 1995; Zabala and Vodkin 2005). Metabolic phenotypes in F3H mutants in *A. thaliana* and carnation also indicate that nonfunctional F3H may cause a diversion of substrate, shunting of precursor molecules into other side pathways (Owens et al. 2008b; Zuker et al. 2002). Strong antisense F3H transgenics in carnation resulted in flowers with significantly elevated fragrance, demonstrating redirection of the pathway (Zuker et al. 2002). Redirection of the pathway may also be indicated in the *F. vesca* FvF3H mutants from preliminary metabolite analysis showing new peaks in the white runner samples. Though F3H has been identified and characterized in many species, there have not been many F3H mutants identified. This may be due to rare copies of F3H that are uncharacterized or other 2-OG dioxygenases are compensating for the loss of F3H, however effectively, and masking the phenotypic identification of F3H mutants.

Previous work mapping the white-fruit locus in *F. vesca* ‘Yellow Wonder’ indicated that this locus was situated at the end of linkage group 1, tightly linked to F3H. However, it now seems unlikely that the white-fruit trait in *F. vesca* Hawaii 4 is caused by a mutation in FvF3H, due to the presence of red pigment in the petioles and runners. It is possible that there are cis-elements, affecting TF complex binding to the promoter region causing the white-fruit phenotype. If this were the case then it would be expected that only the expression level of FvF3H would be affected among the three candidate structural genes. Gene expression analysis was therefore carried out via RT-PCR on the three candidate genes: FvCHS, FvF3’H and FvF3H in de-seeded red fruit and white fruit. Expression analysis of the white fruit tissue indicated that there was reduced expression of all three structural genes, compared to the red fruit and the
positive control. This suggests that the white-fruit phenotype was more likely caused by mutation in a regulatory gene, a TF specifically expressed in fruit tissue affecting the expression of multiple structural gene targets. There are many white-fruiting cultivars in both the diploid and octaploid strawberry, with variation of pigmentation observed between the runner/petiole, seed, and fruit (personal observations of the USDA Clonal Germplasm Repository Records for Fragaria, http://www.ars.usda.gov/Main/docs.htm?docid=11324). Dissecting the mechanisms that underlie variation in fruit color would help to understand how the network of regulatory TFs and structural genes encoding enzymes interact in Rosaceae.

In summary, a T-DNA insertional mutant named white runner, characterized by the absence of red pigment in the petiole and runner tissue was determined to not be associated with the T-DNA but due to a 20 bp deletion in the gene encoding F3H. Histological staining and metabolite analysis indicated reduced levels of compounds associated with three branches of flavonoid metabolism: flavonols, proanthocyanidins, and anthocyanins; however the residual accumulation would indicate that perhaps an uncharacterized tissue-specific $F3H$ gene is functional and expressed or a similar enzyme in the pathway is compensating for the loss of $F3H$. This is consistent with findings in $A. thaliana$. Preliminary results from complementation analysis indicate the strawberry $F3H$ gene and a verified $A. thaliana F3H$ gene could restore WT phenotype in the wr mutant. Preliminary observations indicate that constitutive expression may result in increased pigmentation in developing leaf tissue. These experiments are underway and will be evaluated shortly.
<table>
<thead>
<tr>
<th>Gene</th>
<th>ID</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enzymes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chalcone Synthase</td>
<td>gene26825 F</td>
<td>5’-AGGGTCCGGCCACTGTCTTG-3’</td>
</tr>
<tr>
<td></td>
<td>gene26825 R</td>
<td>5’-GATGACTGTGAAGCACGACGCTTCT-3’</td>
</tr>
<tr>
<td>Flavonoid 3’-hydroxylase</td>
<td>gene25801 F</td>
<td>5’-TCACCCTTTCTCGGCCGCTG-3’</td>
</tr>
<tr>
<td></td>
<td>gene25801 R</td>
<td>5’-GTGAGCCCAAATGCTTCATCCAT-3’</td>
</tr>
<tr>
<td>Flavanone 3-hydroxylase</td>
<td>gene14611 F</td>
<td>5’-GCCTCTACTCTACTCTGGACGCT-3’</td>
</tr>
<tr>
<td></td>
<td>gene14611 R</td>
<td>5’-TCTGCTCCAATGCGCTTCC-3’</td>
</tr>
<tr>
<td></td>
<td>14611_outerF</td>
<td>5’-CCATAACAAACATGCATAAAGGTACTC-3’</td>
</tr>
<tr>
<td></td>
<td>14611_outerR</td>
<td>5’-CCAATAACTCATCTCTTCTCTTCCTTCATC-3’</td>
</tr>
<tr>
<td></td>
<td>14611_delF</td>
<td>5’-GTCCTCCTCCACAGGTGACGCT-3’</td>
</tr>
<tr>
<td></td>
<td>14611_delR</td>
<td>5’-TGCAGGATTGCGCGAGATT-3’</td>
</tr>
<tr>
<td><strong>Transcription Factors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparent Testa Glabra 1</td>
<td>gene12450 F</td>
<td>5’-AATCCGTCCTACGACTCC-3’</td>
</tr>
<tr>
<td></td>
<td>gene12450 R</td>
<td>5’-CTTCAAGAGCTGACTCTTGAGGA-3’</td>
</tr>
<tr>
<td>Actin</td>
<td>ActinF</td>
<td>5’-GGTCTCGGACATTGCTTGCT-3’</td>
</tr>
<tr>
<td></td>
<td>ActinR</td>
<td>5’-AGGCCCGGGTTTGTGACG-3’</td>
</tr>
<tr>
<td>SVBV Promoter</td>
<td>SVBV Pro</td>
<td>5’-TCGAGGTCGCGTCGATAAAGCT-3’</td>
</tr>
<tr>
<td>SVBV Terminator</td>
<td>SVBV Term</td>
<td>5’-GGCGCGGCTTAACTTAAAGAGATCC-3’</td>
</tr>
</tbody>
</table>

aPrimers used for RT-PCR amplifications
bPrimers used for complementation construct
cPrimers used for PCR genotyping for 14611 deletion
dRT-PCR control primers
Figure 3.1 Diagram of the flavonoid biosynthetic pathway for the main branches addressed here for strawberry.
The candidate structural genes chosen for this study were CHS, F3H, and F3’H (indicated by circles); Abbreviations: CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; F3’H, flavonoid 3’-hydroxylase; DFR, dihydroflavonol reductase; ANS, anthocyanin synthase; ANR, anthocyanidin reductase; LAR, leucoanthocyanidin reductase.
Figure 3.2 Phenotypic analysis of the white runner (wr) mutant.
A. A FV10-C04-354-B17-10 T₃ red runner sibling; B. A FV10-C04-354-B17-12 T₃ wr sibling; Vanillin-HCl histological staining of the red runner (C) and wr (E), the circle indicates the location of pink pigment in E.; DMACA histological staining of the red runner (D) and wr (F), the black arrows indicate the sites of blue staining.; Washed, unstained T₃ achenes from FV10-C04-354 B17-10 (G) and B17-12 (I); DMACA staining of achenes from B17-10 (H) and B17-12 (J).
Figure 3.3  Q1 scan
Q1 scan from 14 to 22 min for A. red runner plant B1710-1, B. red runner plant B1710-11, C. white runner plant B1710-8 and D. white runner plant B1710-15. The x-axis represents the m/z, mass-to-charge ratio, with the mass in amu (atomic mass units) and the y-axis is the peak intensity, cps (counts per second).
Figure 3.4 Metabolite analysis of FV10-C04-354-B17-10 red runner and white runner plant petioles.
Each plot is the absolute peak area for a m/z for two red runner B17-10 samples and two white runner B17-10 samples. A. m/z 579; B. m/z 625; C. m/z 463; D. m/z 499; E. m/z 611; F. tentative identities of each of the m/z values in graphs A-E.
Figure 3.5 T-DNA insertion site
A. GBrowse screenshot of T-DNA site in scaffold 513177. The gray bar indicates the flanking region obtained from the T-DNA right border. The insert appears to be located in a putative long terminal repeat (LTR) retrotransposon region, indicated by the horizontal yellow bar; B. A table documenting the other BLASTn hits for the T-DNA flanking region alignment. The right border flanking region also showed significant similarity to four other scaffold regions, demonstrated by the E-values.
Figure 3.6 Molecular analysis of candidate structural genes and sequence analysis of *FvF3H* (gene14611).

A. RT-PCR amplification of the three candidate structural genes (*F3H, F3'H*, and *CHS*) and TF (*TTG1*). Actin was used as a control. The upper gel panel is the PCR products run on a 1% gel and the lower gel panel are the same PCR products for Actin and *F3H* run on a 2% gel.; B. Sequence alignment of red runner *FvF3H* (RRR) and white runner *FvF3H* (WRR) displaying the 20 bp deletion missing in the WRR sequence. Deletion is highlighted in gray.; C. Diagram of *FvF3H* gene structure with the location of the 20 bp deletion indicated by a black arrow. The primers designed for CAPS deletion genotyping and complementation construct development are indicated by horizontal arrows.; RT-PCR, reverse transcriptase-PCR; *F3H*, flavanone 3-hydroxylase; *F3'H*, flavonoid 3’-hydroxylase; *CHS*, chalcone synthase; *TTG1*, transparent test glabra 1; CAPS, cleaved amplified polymorphic sequence. Ladder: 1 kb+ (Invitrogen Corporation, Carlsbad, CA)
Figure 3.7 Amino acid alignment of putative F3H protein sequences (ClustalW) from *N. tabacum*, *P. hybrida*, *I. purpurea*, *F. xananassa*, *F. vesca*, *A. thaliana*, and *D. caryophyllus*. The conserved residues are highlighted in black. The location of the deletion start is indicated by the black triangle. The locations of conserved residues in the C-terminal region are indicated by black arrows. The GenBank accession numbers for all of the putative protein sequences are located in the Methods section.
Figure 3.8 CAPS analysis of the FvF3H (gene14611) on individual FV10-C04-354 B17-10 and B17-12 T3 sibling plants.

Plants with a red runner phenotype are indicated with red lane text and those with the white runner phenotype are indicated with white text. The top gel panel is the PCR amplification using primers 14611delF and 14611delR. The expected product size is 388 bp. The lower gel panel is the Nco1 digest of the above PCR product. If the product has the deletion the Nco1 restriction site would be absent and the product would not digest, remaining 388 bp. If the product does not have the deletion, the expected band sizes would be 164 and 225 bp. Only the lanes indicating that the plant possesses the wr phenotype do not digest either allele product. The ladder is the GeneRuler 1kb+ ladder (Fermentas Inc., Glen Burnie, MD).
Figure 3.9 RT-PCR analysis of candidate structural genes and TFs in red and white fruit. A. Images of the de-seeded white fruit (WF, PI551572) (top panel) and red fruit (RF, PI602578) (bottom panel) used for RNA extraction. Bar = 5 mm.; RT-PCR amplification of candidate structural genes CHS, F3H and F3’H, and Actin as a control. Ladder: GeneRuler 1 kb+.
Figure 3.10 Developing callus from the *AtF3H* complementation transformation in strawberry.
A, B. Callus with spots of red pigment, C, D. Young leaf primordium with darkened pigmentation.
References


Chapter 4. Characterizing a Curly Leaf Mutant in *Fragaria vesca*

**Abstract**
An insertional mutant in *Fragaria vesca*, identified by a curly leaf phenotype, was found to harbor a T-DNA in a gene of unknown function in plants. The deduced amino acid sequence of the coding region of the WT cDNA indicated that this protein was an orthologue of the erythroblast macrophage attacher protein (EMP) in mammalian cells. This protein mediates the interaction between erythroblasts and macrophages facilitating erythroblast maturation, but is ubiquitous in all human cell types functioning in actin binding and cell division. Three functional domains have been described for EMP; the LisH domain, the CTLH domain and the CRA domain, all of which are implicated in microtubule dynamics. In *F. vesca* complementation transformations were performed to restore wild type phenotype. Through the course of the evaluation additional phenotypic effects were observed including abnormalities in seed set and germination, pollen tube growth, and petiolule length. These phenotypes are consistent with developmental abnormalities observed in mutants harboring mutations in actin binding and microtubule associated genes in plants. This study confirmed that genetic complementation restores the leaf phenotype in the cl mutant background. The possibility of FvEMP function in cytoskeletal dynamics is proposed.

**Introduction**
Genes that affect plant cell structure and morphology can have a wide array of developmental defects at the cellular and whole organ level when disrupted. The networks of mechanisms that control this process are extensive and require the coordinated interaction of many cellular components. The components in these interactions are often conserved across many species including plants and animals, with the cell-scaffolding-ubiquitous molecules, such as actin and tubulin, which form microfilaments (MF) and microtubules (MT), respectively, directing the show. Despite the obvious constraints on whole plant mobility, plant cells are dynamic ever rearranging, changing structures. In plant cells, there are two types of growth processes: tip growth, meaning new wall material is deposited at one site such that the cell expands unidirectionally, and diffuse growth, meaning that new wall material is deposited across the cell surface and the cell expands anisotropically. During tip growth in germinating pollen the deposition of actin at one site on the cell wall begins to form bundled filaments that secretory
vesicles travel along to the cell wall. The fusion of those vesicles at the plasma membrane adds new wall and membrane material creating a nascent pollen tube. The continual actin cable nucleation and vesicle fusion to the wall drive pollen tube growth, while MTs, located distally, function in callose and cellulose deposition, maintaining structure and nuclear movement. In epidermal pavement cells actin filaments and MTs work together in multi-directional cell expansion. MTs form bands across the cell, thickening the walls and pulling them in, while actin MFs deposit at the thinner cell wall sites creating bulges or lobes pushing the cell walls out, resulting in puzzle piece-like shapes, characteristic of the mature epidermal layer.

Both MFs and MTs are dynamic polymers, undergoing nucleation, elongation and maturation, or depolymerization; in microtubules this is called dynamic instability. This constant rearrangement and de novo synthesis enable the structure, division, growth, elongation and shape of the many different cell types and processes in cells. MTs have functions in cell division, determine the direction of cell expansion in diffuse growing cells, direct the alignment of cellulose microfibrils in the cell wall, and serve as a guide in the movement of secretory Golgi vesicles. MFs are involved in similar processes and larger actin-protein complexes called cables are known to function in organelle movement and cytoplasmic streaming. Due to the extensive range of functions there are several possible mutant phenotypes that result from disruption of MF and MT arrangement.

There are many Microtubule Associated Proteins (MAPs) and Actin Binding Proteins (ABP) that are necessary for MF and MT formation, arrangement and function. These proteins are responsible for the nucleation, severing, rearrangement, cross-linking, capping, stabilization and organization of MFs and MTs. Discovery of these proteins has been aided by the use of knockout and knockdown mutants as well as new visualization tools that can detect differences in MF and MT organization. The MAP mutants spirall (sprl) and tortifolia (torl) display helical organ torsion or twisting of leaves and stems, indicating that these proteins are required for proper MT orientation (Buschmann et al. 2004; Furutani et al. 2000). Mutations in TONNEAU2, a putative type 2A protein phosphatase, result in severe defects in cortical MT organization producing thick dwarf plants, with an array of developmental abnormalities and the inability to form a preprophase band (Camilleri et al. 2002). The ton1 mutant is the only other mutant known to be defective in preprophase band formation, and the TON1 proteins are conserved among plants and have domains that are highly similar to those found in human
centrosomal proteins (Azimzadeh et al. 2008). VILLIN, an ABP, is involved in actin nucleation, bundling, capping and severing actin filaments. Defects in VLN5, an *A. thaliana* VILLIN protein, result in defective pollen tube growth (Zhang et al. 2010). Mutations in CAP1, a cyclase associated protein that binds to actin, result in mutants with severely reduced stature and rosette diameter, defects in root length and inflorescence height, impaired pollen germination and defects in trichome structure (Deeks et al. 2007). By transforming WT and *cap1* *A. thaliana* with a CaMV35S:GFP:FABD2 construct, which is a second actin binding protein fused with GFP, Deeks et al. (2007) observed that the *cap1*:GFP:FABD2 actin bundles in the root hairs were short and aggregated, as opposed to the long streaming bundles seen in WT::35s:FABD2 root hairs. *A. thaliana* FIMBRIN5 (*fim5*) loss-of-function mutants display curled pollen tubes with reduction in pollen germination, delayed penetration of pollen tubes through the stigma to reach the ovule *in vivo* (Wu et al. 2010). FIMBRIN gene family members are actin binding proteins involved in actin bundling (Kovar et al. 2000). Fluorescent Alexa-488 phallodin staining of wild type and *fim5* mutant pollen grains and developing pollen tubes show that the actin bundles are disorganized and chunky in the mutant compared with the wild-type (Wu et al. 2010). Some actin binding proteins interact with MTs as well as MFs. In the *A. thaliana* FORMIN14 loss-of-function *afh14* mutants, there were defects in the arrangement of microtubule arrays during cytokinesis, leading to defects in microspore formation (Li et al. 2010). The Actin Related Protein, Arp2/3 complex is known to activate actin nucleation from the side of an already existing F-actin filament leading to branched actin structures. In the *A. thaliana* Arp2/3 complex mutants, *arp2-1*, *arp3-1*, and *arp5-1*, show defects in the organization of diffuse cortical F-actin resulting in defects in branched trichome development and changes to epidermal pave cell shape (Li et al. 2003). The above discussion represents only a few of the many MAPs and ABPs that affect MT and MF arrangement, organ and cell structure.

In mammalian cells an extracellular protein was identified that mediated the interaction between macrophages and erythroblasts, necessary for the successful enucleation and maturation of erythroblasts. Later studies have shown that this protein is ubiquitous across all cell types, binds to actin and is involved in cell division. This study presents a putative actin binding protein named Erythroblast Macrophage Attacher protein (EMP) identified in *Fragaria vesca*. To date there are no studies documenting the function of this protein in plants. A T-DNA insertional mutant line FV10-CO4-136 with a visible phenotype named *curly leaf* (*cl*) was found
to harbor its T-DNA insertion within this gene. Initially the goal of this study was to characterize a population of previously generated transgenic *F. vesca* plants to determine if the *FvEMP* gene could complement the *curly leaf* phenotype. Additional phenotypic effects of the insertion were discovered over the course of that analysis and in addition the predicted protein sequence of FvEMP was analyzed for domains that may indicate its function in the plant cell.

**Methods**

**Background work.** The *curly leaf* (*cl*) mutant was originally identified by Teruko Oosumi in a T<sub>1</sub> pCAMBIA 1304 insertional mutant screen. The predominant feature of this mutant was the curling of the leaf margin with a downward leaf drooping. Oosumi identified the mutation, used segregation analysis to verify that the T-DNA insertion segregated with the *cl* phenotype, used TAIL-PCR to amplify the flanking region for sequencing and cloned the wild type gene into pCAMBIA 2300 for complementation transformation of the *cl* mutant. In August of 2006 Aaron Baxter was in the process of completing the complementation transformation. I assisted him in the transformations and brought the C<sub>0</sub> FV10-C04-136 rescue lines out of culture. C<sub>0</sub> leaf tissue and C<sub>1</sub> seed were collected from these lines. At that time there were FV10-C04-136 T<sub>1</sub> *cl* plants in a growth chamber at the Virginia Bioinformatics Institute. T<sub>2</sub> seeds were collected from these plants.

**Nucleotide and protein sequence analysis.** The FV10-C04-136 T-DNA flanking sequence was retrieved from GenBank (DQ830764) and aligned to the *F. vesca* genome scaffold database (Shulaev et al. 2010). The *FvEMP* predicted mRNA sequences from the *ab initio* gene model gene23286 and hybrid gene model gene23350 were retrieved and compared to the deduced amino acid sequence from the GenBank database with the BLASTx tool to determine accuracy. Neither model appeared correct so the hybrid model gene 23350 was trimmed and annotated using SeqBuilder (DNAStar, Lasergene8) according to the GenBank BLAST alignment. The translated FvEMP amino acid (AA) sequence was aligned to EMP AA sequences from *A. thaliana* (AT3G55070, obtained from TAIR), *V. vinifera* (GenBank Accession (GB) XP_002281688.1), *P. trichocarpa* (GB XP_002311128.1), *R. communis* (GB XP_002525069.1), *Bombus terrestris* (GB XP_003397163.1), *Solenopsis invicta* (GB EFZ12028.1), *Danio rerio* (GB NP_955843.1), *Gallus gallus* (GB NP_001012622.1), *Homo sapiens* (GB NP_001017405.1), *Mus musculus* (GB NP_067475.1), *Xenopus tropicalis* (GB...
NP_001007963.1) using ClustalW in MegAlign (DNASTAR, Lasergene 8) (Goujon et al. 2010; Larkin et al. 2007). SMART was used for protein domain analysis of FvEMP (gene23286) (Letunic et al. 2009; Schultz et al. 1998). The predicted FvEMP protein was BLASTp against GenBank and the NCBI Tree View tool using the pairwise alignment was generated using the Neighbor Joining Method allowing 65% maximum sequence difference.

**Plant Growth.** T2 seeds collected from FV10-CO4-136 GFP positive plants with the cl phenotype were germinated on MS media and grown in greenhouse conditions. C1 seeds collected from FV10-CO4-136 complementation double mutants were germinated on MS media and grown in greenhouse conditions. Plants were grown in Miracle-Gro potting medium initially in 7.6 cm pots and transplanted to 2 L pots.

**DNA extraction and PCR.** DNA extractions were carried out as described in (Oosumi et al. 2006). DNA was extracted from C0 tissue from seven putative independent complementation transformants. These lines were given the IDs #38 A45, A43, C18, B45c, B9b, B18b, and C2b. These lines were screened by PCR: 1) for the presence of the original pCAMBIA1304 T-DNA insertion using primers designed for hygromycin (primer ID: HYG); 2) to verify that the original pCAMBIA1304 insertion was homozygous using primers designed in the F. vesca genomic sequence flanking the insertion site (primer ID: FV136); and 3) for presence of the second complementation insertion pCAMBIA2300 using primers designed for kanamycin (primer ID: NPTII). The FV136 forward and reverse primers were designed in exons V and VI (Figure 4.2D) of FvEMP and serve a dual function. This primer set will not amplify the region across the original T-DNA insertion site when homozygous and will amplify a smaller product from the genomic DNA of the complementation transformants because the WT cDNA transcript was used to create the complementation construct. All primer sequences are listed in Table 4.1. The thermal cycling program was as follows: one cycle (95°C for 3 min), 35 cycles (94°C 1 min, 55°C 1 min, 72°C 2 min), and one cycle (72°C 3 min). A modified hiTAIL PCR protocol was used to amplify the flanking region of the pCAMBIA2300 insertion to determine if the C0 transformation events were independent. Since the border regions of pCAMBIA1304 and pCAMBIA2300 are highly similar, the T-DNA specific primer for the pre-amplification reaction was designed in the NPTII gene sequence unique to pCAMBIA2300. The primary and secondary hiTAIL primers were designed from the left border of pCAMBIA2300. This
amplification scheme selectively amplified the flanking region only from the pCAMBIA2300 insertion. The primer sequences are listed in Table 4.1. The hiTAIL bands were purified using the Qiagen Gel Extraction kit (Qiagen, Carlsbad, CA) and sequenced with the Beckman Coulter CEQ 8800 according to manufacturer’s instructions. The obtained sequence was used to search the F. vesca genome scaffold database using the nucleotide BLAST algorithm (September 2010) to determine insertion site. DNA from leaf bud tissue was extracted from T_2 FV10-C04-136 cl lines, C_1 #38 B9b and B45c lines and F. vesca Hawaii 4 fourth generation inbred wild type. The T_2 FV10-C04-136 lines were screened by PCR using primers: HYG, NPTII, FV136 and Actin. The thermal cycling program was as follows: one cycle (95°C for 3 min), 35 cycles (94°C 1 min, 55°C 1 min, 72°C 2 min), and one cycle (72°C 3 min). Zygosity PCR primers were designed from flanking regions surrounding the complementation transformation insertion sites for #38 B9b and #38 B45c to identify homozygous transgenic pCAMBIA2300 insertions. Primers B9bF and R as well as B45cF and R are listed in Table 4.1. The thermal cycling program was as follows: one cycle (94°C for 3 min), 35 cycles (94°C 1 min, 55°C 1 min, 72°C 1 min), and one cycle (72°C 3 min). For all reactions, we used the Takara Ex Taq polymerase kit (Takara Bio Inc, Japan).

**RNA extraction and RT-PCR.** RNA extractions were carried out according to (Moser et al. 2004), with omittance of the proteinase K step. RNA was extracted from leaf bud tissue from T_2 lines FV10-C04-136: 16T_1 F1T_2, 19T_1 D1T_2, 19T_1 B1T_2 and C_1 line #38 B45c and wild type. First strand cDNA synthesis was prepared using SuperScript III First Strand Synthesis System (Invitrogen Corporation, Carlsbad, CA). RT-PCR was carried out in a 25 µl reaction, using 0.25 pmol gene model specific primers, 0.16 mM dNTPs and 0.2 U Ex Taq polymerase (Takara, Japan). The BioRad MyCycler thermocycler program for all primer pairs was as follows: one cycle (95°C for 3 min), 35 cycles (94°C 1 min, 55°C 1 min, 72°C 2 min), and one cycle (72°C 3 min). All primers are listed in Table 4.1.

**Phenotype analysis**

**Pollen germination.** Anthers were collected from open flowers the morning before expected dehiscence. Flowers were kept in a Petri plate on the greenhouse bench until dehiscence occurred. Liquid pollen germination medium consisted of 9.1% sucrose and 0.084% boric acid. The hanging drop method was used for germination: one drop of medium was placed on a clean
microscope slide using a Pasteur pipette, a dehiscent anther was placed in the drop using forceps under a dissecting microscope, then tapped gently to release the pollen grains into the medium, the microscope slide was inverted quickly and placed on a 60 mm Petri dish nested inside a 150 mm Petri dish. The smaller Petri dish was slowly filled with water, enough to cover the bottom of the dish, but not so much that the surface of the water touched the hanging drop. This humidity chamber was covered with the lid of the 150 mm dish and gently placed in a sandwich-sized Ziploc bag. The chambers were left at room temperature overnight in the dark. The germinated pollen tubes were viewed under an Olympus SZX 12 microscope with a 1.6× objective and photographed with an Olympus DP70 camera after 21 h.

Seed set. Fruits were collected, dried and seeds were manually extracted, then counted. Data were analyzed using a One-Way Analysis of Variance (ANOVA) and a Tukey Test with unequal sample sizes.

Seed germination. Seed germination was performed as in (Lindsey 2010). Seeds from nine C₁ #38 B9b siblings and ten C₁ #38 B45c siblings were counted into 50 seed aliquots with three replications for each genotype. Seeds were sterilized and placed in Gamborg’s B-5 liquid germination medium. The bottles were placed on a shaker at 100 rpm, at room temperature under four fluorescent bulbs. Germinated seeds were removed under sterile conditions and counted every 7 days for 4 weeks. Data were analyzed using a One-way ANOVA and a Tukey test for equal sample sizes.

Petiolule length. Petiolule length was measured in the primary, center leaflet of fully expanded trifoliate leaves. The distance measured was from the base of the petiolule where it joined the other petiolules, to the site where leaf tissue started to grow outward from the midvein. Data were analyzed using a One-way ANOVA and a Tukey test for equal sample sizes.

Epidermal peels. Abaxial epidermal peels were taken from the leaves of 3 month indoor grown plants. Epidermal peels were fixed overnight in methanol and mounted in water. Peels were viewed at 20× magnification under a light microscope (Olympus BX51).

Results

Characterization of the T₂ FV10-C04-136 T-DNA mutants. The T-DNA insertional mutant line FV10-C04-136 was originally selected because of unusual leaf morphology, including
downward leaf cupping and a rolling under of the leaf margin, resulting in a downward leaf direction (Figure 4.1B, D). The T-DNA was found to be located in the intron between exons five and six of a putative Erythroblast Macrophage Attacher Protein (FvEMP) (Figure 4.2D).

Genomic PCR of T2 FV10-C04-136 leaf bud tissue with primer sets: Actin to demonstrate that the DNA was sufficient for control amplification, HYG to demonstrate the presence of the T-DNA and primer set FV136 (primers that flank the T-DNA insertion site in exons five and six) to confirm that the T-DNA was homozygous in these lines because only WT alleles would be expected to amplify under the PCR conditions used (Figure 4.2A). RT-PCR of T2 leaf bud tissue with primer set FV136 showed low levels of expression of FvEMP, indicating that the T-DNA affected RNA processing, leading to leaky expression of the gene (Figure 4.2B).

Protein sequence and structure analysis. Comparison of the three amino acid reading frames of the putative RNA sequence of FvEMP with GenBank indicated that this gene was highly conserved across many plant and animal species. ClustalW protein alignment of EMP predicted amino acid sequences from F. vesca, V. vinifera, R. communis, P. trichocarpa, A. thaliana, B. terrestris, S. invicta, D. rerio, M. musculus, H. sapiens, X. tropicalis, and G. gallus (Figure 4.3 A) showed conservation throughout the sequence. SMART protein domain analysis (Letunic et al. 2009; Schultz et al. 1998) of FvEMP confidently predicted three functional domains: LisH (Lis1 Homologous, E-value: 2.84e-02), CTLH (C-terminal to the LisH, E-value: 6.71e-13) and CRA (C-terminal of the Ran Binding Protein, E-value: 1.34e-17) (Figure 4.3 B). The amino acid residues corresponding to the three functional domains are placed within the alignment in Figure 4.3 A. Phylogenetic analysis in Figure 4.4 shows the conservation of the EMP protein within the animals and the grouping of the plants, animals and insects observed in the alignment in Figure 4.3 A.

Characterization of the FV10-C04-136 complementation transformants. Seven FV10-C04-136 C0 complementation transformants were analyzed for presence of the pCAMBIA2300 T-DNA insertion. This T-DNA contains the gene encoding kanamycin (NPTII) resistance and constitutively expressed FvEMP. PCR using primers designed for NPTII and FvEMP (FV136) showed that all seven lines were positive for NPTII and for FvEMP (Figure 4.5A). Hi-TAIL PCR was used to amplify the F. vesca genomic DNA flanking the pCAMBIA2300 complementation T-DNA insertion in the seven C0 #38 FV10-C04-136 complementation transformants (Figure 4.5B). Sequence analysis of the boxed bands (Figure 4.5B) showed that
the complementation T-DNA of #38 B9b, B45c, C18, C2b, and B48b were all situated in different genomic positions, indicating independent insertion events (data not shown). Complementation lines #38 A45 and A43 were duplicate double insertion events (data not shown). Lines #38 B9b and B45c were chosen for further analysis. The results of a nuclotide BLAST comparison against the *F. vesca* genome indicated that the complementation T-DNAs were inserted in scaffolds 513018 and 513098, respectively, both within 3’UTR regions of a putative ProFAR isomerase-like gene and a predicted F-box/Leucine Rich Repeat gene (Figure 4.5C, D).

Of 33 C1 plants derived after self-pollination of C0#38 B45c FV10-C04-136, zygosity PCR of the complementation T-DNA indicated that seven plants were homozygous wild type, 21 plants were hemizygous for the T-DNA and five plants were homozygous for the complementation T-DNA, following a 1:2:1 segregation ratio ($\chi^2 = 0.25$ p < 0.05). Only the homozygous wild type (negative for the complementation T-DNA) expressed the cl phenotype; all of the plants containing the complementation T-DNA displayed wild type leaf phenotype.

RT-PCR was completed on leaf bud cDNA for 15 #38 B45c C1 lines, six with the cl phenotype and nine with wild type phenotype (Figure 4.6). All of the cl samples displayed the reduced leaky expression seen in the T2 FV10-C04-136 RT-PCR. All of the wild type lines displayed greater expression levels than cl phenotypes, consistent with untransformed wild type (Figure 4.6). Representative plants corresponding to the #38 B45c lines in the RT-PCR results are shown in Figure 4.7.

Twenty C1 #38 B9b FV10-C04-136 complementation sibling plants were tested using zygosity PCR for the pCAMBIA 2300 complementation insertion. In this population only one plant displayed the cl phenotype and was homozygous wild type (negative) for the complementation insertion (data not shown). Of the remaining 19 C1 #38 B9b plants, 12 were hemizygous and seven were homozygous for the complementation insertion, following a 1:2:1 segregation ratio ($\chi^2 = 0.11$, p < 0.05). All of these plants displayed a wild type leaf phenotype (data not shown). The PCR and phenotype results for #38 B45c and B9b FV10-C04-136 complementation lines showed that introducing the wild type *FvEMP* into the cl mutant restored wild type leaf phenotype.
**Phenotypic analysis.** More comprehensive phenotypic analysis resulted in the discovery of additional phenotypic expression in the cl plants. Seed set analysis revealed that the cl #38 C₁ sibling plants had reduced seed set numbers (analyzed by One-Way ANOVA and significance determined by Tukey’s Rank Test) compared to their wild type #38 C₁ siblings (Figure 4.8A, C). Seed germination experiments (analyzed by One-Way ANOVA and significance determined by Tukey’s Rank Test) also showed that in both C₁ lines the cl siblings had reduced germination when compared to the wild type siblings (Figure 4.8B, D). The reduced seed germination in the cl lines would affect the seed set data, and for the lines where there is no significant difference between cl and WT in the current data set (example: Figure 4.8A. cl line B9b B3 vs. WT lines B9b E3, C3, and B2) significance may occur if the seed germination reduction is taken into account. This requires the question of where does the seed germination abnormality exist? Are the seeds non-viable to begin with and there is error in the judgment of which seeds are viable or is there an additional developmental abnormality in germination? Reduced seed set in the cl plants led to an investigation of pollen tube growth. In vitro pollen germination showed that pollen from cl plants (Figure 4.9B, D, F) appeared to have reduced pollen germination in comparison to wild type plants (Figure 4.9A, C, E). C₁ plants #38 B9b D3 and #38 B45c F₁ displayed wild type leaf phenotype and normal pollen tube growth. Further observation of the plants with the cl phenotype revealed that there was a reduction in petiolule length (Figure 4.10C, D). Measurements of petiolule length of C₁ #38 B9b and B45c sibling plants (n=5) were analyzed by One-way ANOVA and significance was determined using Tukey’s rank Test. The results showed that the petiolule lengths in the plants with cl phenotype were significantly reduced in comparison to wild type sibling plants (Figure 4.10A, B).

The additional phenotypic abnormalities observed in the cl mutants indicate that the effects of the T-DNA insertion into FvEMP are much more extensive than originally observed. These phenotypic effects are consistent with those observed in actin binding (ABP) and microtubule associated protein (MAP) mutants, which exhibit abnormal leaf shape, pollen tube growth deficiency, reduction in organ lengths, and reduced fertility (Camilleri et al. 2002; Furutani et al. 2000; Li et al. 2010; Zhang et al. 2010). Cytoskeletal mutants also display changes in abaxial epidermal pave cell shape and size. In order to investigate if there was a pave cell phenotype in the cl plants, we removed epidermal peels from the abaxial sides of leaves of 3-month indoor-grown plants (Figure 4.11). During this process we observed that the cl plants had less adhesion
between the epidermal pave cell layer and the parenchyma cells underneath, than the WT and complementation WT plants. The epidermal layer in the cl plants would move as though detached when prodded with forceps. This yielded cleaner peels with clearly defined cell borders and less background noise from additional layers of cells (Figure 4.11). Identically sized circles were placed on each of the photographs (Figure 4.11) and ten cells within or on this boundary were hand drawn to determine if there were any obvious differences in shape or approximate size (Figure 4.12). No obvious differences were observed between the cl, WT or complementation WT peels; however the changes are often subtle and require measurements of cell area, distance across the narrow regions and distance from valley to peak of the lobes.

**Discussion**

T-DNA insertional mutant line FV10-CO4-136 named *curly leaf* (*cl*) exhibited an array of developmental abnormalities including altered leaf morphology, reduced pollen germination, seed/fruit set, seed germination, petiolule length and altered cell morphology. These pleiotropic effects are consistent with abnormal phenotypes observed in actin binding and microtubule associated protein mutations. However, in plant species, the function of a macrophage erythroblast attacher protein, named for its importance in animal systems, remains unknown. The T-DNA insertion disrupted a putative macrophage erythroblast attacher protein (FvEMP), orthologue of which can found in many plant and animal species. SMART, the Simple Modular Architecture Research Tool, confidently predicted three protein domains: LisH, a Lis1 Homologous domain, CTLH, C-terminal to the LisH domain, and CRA, a CT11-RanBPM Ran binding domain of the microtubule organizing center. Since there are no studies elucidating the function of this gene, or any related gene with the same combination of protein domains, in any plant species, the individual functions of each domain present in characterized mammalian and plant proteins will be discussed here.

*LisH domain* The LisH domain was originally identified in the *LIS1* gene mutation, resulting in the human disease Miller-Dieker lissencephaly, or smooth brain, which results in severe neuronal migration defects (Reiner et al. 1993). LIS1 has been shown to associate with cytoplasmic microtubules and mutations within the LisH domain of this protein, resulting in reduced protein half-life, inhibition of dimer formation and defects in actin regulation (Gerlitz et al. 2005). Treacher Collins syndrome and oral-facial-digital-type 1 are two other human diseases
associated with mutations in LisH domains (Emes and Ponting 2001). The cause of these diseases has also been attributed to defects in cell migration. One hundred eukaryotic proteins had been identified as of 2001 that contained a LisH domain. This domain is an α-helical motif that is a protein-binding domain involved in microtubule dynamics. Characterized proteins in A. thaliana containing a LisH domain include LEUNIG (also called ROTUNDA) and TONNEAU. LEUNIG (LUG) acts as a transcriptional repressor of AGAMOUS, a MADS box TF, restricting expression to floral whorls 3 and 4 (Conner and Liu 2000). Yeast two-hybrid assays using truncated forms of LUG showed that the LUFS domain (composed of the LisH domain and a Pfam domain SSDP) was necessary for protein interaction, further indication that the LisH domain is a protein-protein interaction domain (Sridhar et al. 2004). Mutants for this gene have also shown variation in leaf morphology, specifically increased cell size, alterations in pave cell structure and abnormal leaf shape (Cnops et al. 2004). TONNEAU1A and TONNEAU1B are similar tandem positioned genes that contain a LisH domain and are involved in the cortical microtubule organization of the preprophase band (PPB) (Azimzadeh et al. 2008; Camilleri et al. 2002). Mutations in these two genes result in plants unable to form a PPB, and have severe growth effects and microtubule organization defects.

**CTLH domain** The C-Terminus to the LisH (CTLH) motif is often found downstream of the LisH domain, but the two domains occur separately in some proteins, warranting their individual identities. This is a predicted α-helical sequence thought to also be involved in protein-protein interactions. SMART protein domain analysis has identified 22 Arabidopsis proteins containing a CTLH domain, of which 20 also contain a LisH domain. Other domains associated with this complex are RING finger, WD40 repeat, Pfam::SPRY, armadillo repeat and the CRA domain (Kobayashi et al. 2007).

**CRA domain** The Interpro abstract describes the CRA domain as the CT11-RanBPM domain, with RanBPM representing the Ran Binding Protein in the Microtubule-Organizing center. RanBPM is involved with microtubule organization at the centrosome (Nakamura et al. 1998). This domain is found in Ran-binding proteins, with the CRA domain implicated in protein-protein interaction by yeast two-hybrid studies with the fragile X mental retardation protein (FMRP) in humans (Menon et al. 2004). Kobayashi et al. (2007) investigated the multi-protein complex reported to be associated with RanBPM and found that protein components of the
complex contained combinations of the LisH, CTLH and CRA domains, including p48EMLP, a splicing variant of the erythroblast macrophage attacher protein. The 20S protein complex was named the CTLH complex, but the function remains unclear. The authors proposed that based on the involved functional domains the complex may function in cell migration, microtubule dynamics, cell adhesion and/or nucleokinesis.

In order to understand the possible role of the Fragaria vesca erythroblast macrophage attacher protein (FvEMP) in the woodland strawberry, we will examine the function of EMP in mammalian cells.

Mammalian EMP, erythroblast macrophage protein, was first discovered in human erythroblasts and macrophages, which together form erythroblastic islands in human bone marrow during erythropoiesis. Hanspal and Hanspal (1994) used a cell attachment assay to identify proteins and protein complexes that erythroblasts and macrophages attach to, isolated from macrophage membranes. The identified protein was found on the surface of both macrophages and erythroblasts. In vitro cultures of erythroblasts without macrophages were unable to enucleate and fully mature. In addition the use of heparin, a protein that binds EMP making it unavailable for the E-M interaction, leads to the decline of the E-M cultures and the number of mature enucleated erythroid cells (Hanspal and Hanspal 1994). In a later study Hanspal et al. (1998) cloned and characterized the human EMP cDNA. Through the use of N and C-terminal fusion constructs, in erythroblasts and macrophages, they found that EMP consists of a small N-terminal extracellular domain, a transmembrane domain and a large cytoplasmic domain (Hanspal et al. 1998). Using Northern blot analysis of human heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas they found EMP transcripts in nonerythroid cells and tissues. As verification they repeated this analysis on pure human cell lines, to account for the possibility of peripheral blood contamination. This would suggest a function of EMP beyond erythroblast enucleation. Sequence analysis and transfection assays have indicated that the N-terminus of EMP is extracellular in both macrophages and erythroblasts possibly facilitating the interactions between these two cell types. Though transcripts are present in HeLa cell lines, co-culture studies have shown that erythroblasts do not adhere to this cell line. Transfection of HeLa cells with an overexpression construct of the full length Emp cDNA resulted in erythroblast-HeLa cell attachment. This result indicates that alternative processing and fate of the EMP protein is highly cell specific. The ubiquitous
expression across tissue types is also observed in RNASeq data included among the genomic resources for potato based on the \textit{S. tuberosum} Group Phureja DM genome (Potato Genome Sequencing Consortium 2011) and \textit{A. thaliana} microarray data (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi using the \textit{AtEMP} TAIR ID AT3G55070).

Bala et al. (2006) ask the question of the function of EMP in non-erythroid cells. Using Western blot analysis with an anti-EMP probe, they detected EMP protein in human macrophage cell lines, human erythroid cell lines, as well as human cervix epithelioid carcinoma cells (HeLa) (Bala et al. 2006). In order to study the nuclear and cytoplasmic localization of the EMP protein, they used human embryonic kidney (HEK) cells, which contain extremely low levels of EMP, to allow transfection of epitope tagged EMP. They found that EMP localizes to the nuclear matrix and co-localizes with actin by western blot analysis in HEK (non-erythroid/macroage) cells. They further confirmed the co-localization of EMP and actin using a GST-pull-down assay. In order to understand the localization of EMP during cell division they immunolabeled synchronized HEK cells fixed at different stages of mitosis. EMP showed extensive changes in localization exhibiting chromatin localization during prophase, moving towards the spindle poles during metaphase, then localized with the condensed chromosomes during anaphase, moving out towards the contractile ring in telophase. This extensive rearrangement suggests a critical role in cell cycle and division.

Soni et al. (2006) found that mice null for EMP were unable to form mature enucleated erythrocytes and died perinatally. Further analysis showed that EMP null fetal liver macrophages were unable to mature properly leading to the inability of erythroblast-macrophage island formation, a necessary step in the erythroblast enucleation process. This study further confirmed the co-localization of actin and EMP and also showed that actin filament organization was altered in EMP-null erythroblasts and macrophages. In wild type erythroblasts F-actin/EMP localization was found in both the plasma membrane and throughout the cytoplasm, whereas in mutant erythroblasts actin was only found along the plasma membrane (Soni et al. 2006).

A strawberry plant has neither erythroblasts nor macrophages, yet our study indicates that the orthologous protein has dramatic effects on various aspects of plant development. Considering the functional evidence for involvement of this protein in cell division and evidence for its role in actin binding, we can deduce that FvEMP is involved in cytoskeletal function.
There are many more experiments required to understand the function of FvEMP in the plant cell. The conservation of the functional domains and amino acid sequence indicate that this gene is an orthologue to the mammalian EMP; however how much of the mammalian function has been conserved has yet to be determined. This study sought to verify that reintroducing the WT FvEMP into the cl mutant restores WT phenotype. The analysis of two independent complementation transformants has shown that WT leaf phenotype was restored; however through the course of the work additional pleiotropic effects were discovered that had not originally been identified in the cl mutant. The phenotypes are consistent with those observed for other ABP and MAP mutants, correlating with the functional domain analysis that indicates that FvEMP most likely plays a role in cytoskeletal dynamics. Future experiments to illuminate the function of FvEMP could include immunoprecipitation assays to determine FvEMP’s binding partners, transformation of WT, cl, and complementation WT with ABP and MAP fluorescent protein fusions to observe in vivo the arrangement of microfilaments and microtubules, evaluation of A. thaliana AtEMP mutants and confirmation that all of the pleitropic effects segregate with the T-DNA insertion and are rescued by the complementation transformation. This gene presents an exciting opportunity to understand how the function of a protein conserved between plants and animals has diverged over time.
Table 4.1 Primer sequences used for genomic PCR, hi-TAIL-PCR, RT-PCR and zygosity PCR to characterize insertional mutant line Fv10-C04-136 and its complementation transformations.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>FV136F&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>5’-CGCTTGAGCCTCTGCTGAAT-3’</td>
</tr>
<tr>
<td>FV136R&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>5’-GGCAGAAGTGTGGGATTTC-3’</td>
</tr>
<tr>
<td>HYGF&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5’-CTTGGCCCTCGGACGAGTG-3’</td>
</tr>
<tr>
<td>HYGR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5’-CTCACCACGCAGTCTGGTCG-3’</td>
</tr>
<tr>
<td>NPTIIF&lt;sup&gt;a,b,c&lt;/sup&gt;</td>
<td>5’-TCAGAGAAACTCGTCAAGAAGGC-3’</td>
</tr>
<tr>
<td>NPTIIR&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>5’-ATGGGGATTGAACATGGATT-3’</td>
</tr>
<tr>
<td>ActinF&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>5’-GGTCTCGAACATTCTGGGTCAAT-3’</td>
</tr>
<tr>
<td>ActinR&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>5’-AGGCGGTTTGCTGGGAG-3’</td>
</tr>
<tr>
<td>2300_1a&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5’-CCAAAATCCAGTACTAAAATCCAGTCCC-3’</td>
</tr>
<tr>
<td>2300_2a&lt;sup&gt;c,d&lt;/sup&gt;</td>
<td>5’-CGTCCGCAATGGTATTCATAGGCTAG-3’</td>
</tr>
<tr>
<td>B9bF&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5’-TGCCATATGCTCAGGAATTTAGC-3’</td>
</tr>
<tr>
<td>B9bR&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5’-AGTCGACGTTTGATGGCTGGT-3’</td>
</tr>
<tr>
<td>B45cF&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5’-CGACATGTCATCTGGCTCTC-3’</td>
</tr>
<tr>
<td>B45cR&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5’-GCATGATCGTTTCCTGGCCTAG-3’</td>
</tr>
</tbody>
</table>

<sup>a</sup>Primers used for genomic DNA PCR
<sup>b</sup>Primers used for RT-PCR
<sup>c</sup>Primers used for HiTAIL PCR
<sup>d</sup>Primers used for zygosity PCR
Figure 4.1 Photographs of the *curly leaf* phenotype.
A. Wild type *F. vesca*, B. *cl* plant, C. close-up of wild type leaves, D. close-up of *cl* leaves
Figure 4.2 Molecular analysis of curly leaf plants and T-DNA insertion site
A. Genomic PCR for T2 FV10-C04-136-(16T1-A1, 16T1-F1, 19T1-B1, 19T1-D1) cl plants with primer sets for actin, FV136, HYG and NPTII; B. RT-PCR for T2 FV10-C04-136 leaf bud (LB) tissue with primer sets for actin and FV136. C. Diagram of the pCAMBIA 1304 T-DNA, HYG, gene encoding for hygromycin resistance; CaMV35s, cauliflower mosaic virus promoter; GFP::GUS, selectable marker fusion of green fluorescent protein and β-glucuronidase. D. Diagram of FvEMP gene structure with the T-DNA insertion site. The exons are designated by roman numerals, the position of the FV136 primers are shown on this diagram. Abbreviations: 2300, pCAMBIA 2300 plasmid DNA; 1304, pCAMBIA 1304 plasmid DNA; FV136, primers designed flanking the 1304 T-DNA insertion site in FvEMP; HYG, primers designed to amplify hygromycin from 1304; NPTII, primers designed to amplify kanamycin from 2300.
Figure 4.3 Amino acid sequence analysis of putative EMP
A. ClustalW alignment of predicted EMP amino acid sequences from F. vesca (FvEMP), A. thaliana (AT3G55070), V. vinifera (Vv_XP_002281688.1), P. trichocarpa (Pt_XP_002311128.pro), R. communis (Rc_XP_002323069.pro), B. terrestris (Bt_XP_003397163.pro), S. invicta (Si_EFZ12028.pro), D. rerio (Dr_NP_955843.pro), G. gallus (Gg_NP_001012622.pro), H. sapiens (Hs_NP_001017405.pro), M. musculus (Mm_NP_067475.pro), X. tropicalis (Xt_NP_001007963.pro), the blue highlights the consensus residues relative to H. sapiens, the black highlights the consensus residues relative to V. vinifera; ___ indicates LisH domain, ___ indicates CTLH domain, ___ indicates CRA domain, the vertical yellow line indicates the exon V/VI boundary (site of T-DNA insertion). B. SMART protein domain predictions for FvEMP
Figure 4.4 NCBI BLAST Tree View phylogenetic tree.
Tree was built using the pairwise BLASTp alignment with the Neighbor Joining method allowing for 65% maximum sequence difference.
Figure 4.5 Molecular analysis of the C₀ #38 FV10-C04-136 complementation transformants
A. PCR analysis of the C₀ #38 FV10-C04-136-A45, A43, C18, B45c, B9b, B18b, and C2b complementation transformants using primer sets actin, FV136 and NPTII; B. HiTAIL PCR of the C₀ #38 FV10-C04-136 complementation transformants, the bands boxed in white were sequenced, A45 and A43 were duplicate transformants; C. GBrowse screenshot of the #38 B9b pCambia2300 T-DNA insertion site; D. GBrowse screenshot of the #38 B45c pCambia2300 T-DNA insertion site.
Figure 4.6 RT-PCR of #38 B45c C₁ sibling plants.
Actin (A) and FV136 (B). Leaf phenotype is below the plant ID: WT, wild type; CL, curly leaf. Ladder: 1kb + (Invitrogen Corporation, Carlsbad, CA)
Figure 4.7 Representative plants harboring a WT copy of recombinant \textit{FvEMP} from the RT-PCR results in Figure 4.5. 
In each frame the plant on the left is wild type, the plant in the middle is the \#38 B45c C$_1$ sibling and the plant on the right is the T$_2$ \textit{cl} mutant. A. \#38 B45c \textit{cl} C$_1$; B. \#38 B45c H4 WT C$_1$; C. \#38 B45c B5 \textit{cl} C$_1$; D. \#38 B45c C6 \textit{cl} C$_1$; E. \#38 B45c D1 WT C$_1$; F. \#38 B45c A1 WT C$_1$. 
Figure 4.8 Seed set and germination analysis
A. #38 B9b C₁ seeds per fruit (n=5); B. #38 B9b C₁ seed germination data (n = 3); C. #38 B45c C₁ seeds per fruit (n=5); D. #38 B45c C₁ seed germination data (n = 3). Error bars represent mean values ± SE. Different letters indicate significant differences among the means as according to Tukey’s test at $P \leq 0.05$. 
**Figure 4.9 Pollen tube growth**
A. wild type; B. FV10-C04-136-19T1-D1T2 cl mutant; C. #38 B9b D3 wild type C1 sibling; D. #38 B9b B3 cl C1 sibling; E. #38 B45c F1 wild type C1 sibling; F. #38 B45c A5 cl C1 sibling.
Figure 4.10 Petiolule length analysis
A. Plot of petiolule length in C1 #38 B9b siblings; B. Plot of petiolule length in C1 #38 B45c siblings; C. Example of petiolule in C1 rescue; D. Example of petiolule in C1 cl non-rescue. Error bars represent mean values ± SE (n = 5). Different letters indicate significant differences among the means according to Tukey’s Test at $P \leq 0.05$. 
Figure 4.11 Epidermal Peels
A. Wild type (WT) abaxial epidermal peel; B. FV10-C04-136 T3 cl epidermal peel; C. #38 B45c C2 WT epidermal peel; D. #38 B9b C2 WT epidermal peel. The identically sized circles are the regions from which individual cells were hand drawn in Figure 4.12. Bars = 100 µM.
Figure 4.12 Hand drawn epidermal pave cells from Figure 4.11.
A. Epidermal pave cells from wild type; B. FV10-C04-136 T3 cl phenotype plant; C. #38 B45c C2 WT phenotype plant; D. #38 B9b C2 WT phenotype plant. Bars = 100 µM.
References


112


Chapter 5. Conclusions and Future Directions

T-DNA insertional mutagenesis is a useful tool for forward and reverse genetics studies and genetic marker development in many plant species. Generation of T-DNA tagged mutant lines requires various transformation techniques and methods of plant regeneration, dictated by the plant of interest. These methods can also lead to additional mutations that can complicate analysis of T-DNA tagged lines. This dissertation describes the analysis of three T-DNA insertional mutants in *Fragaria vesca* that exhibited visible changes to wild type phenotype, affecting floral development (*green petal*), flavonoid biosynthesis (*white runner*) and leaf morphology (*curly leaf*). The intention of this project was to explore the utility of T-DNA insertion mutants in *F. vesca*, analogous to what has been done in *A. thaliana*, through phenotype characterization, analysis of the T-DNA insertion site and complementation of the phenotype with a wild type copy of the disrupted gene. Segregation analysis of the T-DNA insertional mutant lines displaying the *green petal* and *white runner* phenotypes revealed that the T-DNA and phenotype segregated independently. Both of these mutant phenotypes were caused by small deletions in the coding region of a candidate gene. The exact causes of the deletions are unknown, but are likely caused by the *Agrobacterium* transformation (T-DNA integration) or stress induced mutations linked to the *in vitro* culture/regeneration process. Also unknown is the extent to which additional deletions may have been introduced in the genomes of these lines.

Improving the mutant generation process to minimize somaclonal mutations is difficult because there are many components of *in vitro* regeneration that can cause somaclonal variation. To improve the process there must also be methods in place to determine if changes to the system reduce the mutation rate. Small indels (< 50 bp) are difficult to detect in large genome-wide polymorphism screens due to the minimal changes in electrophoretic band sizes, although there are ongoing efforts to develop technologies to identify small differences in size and/or sequence between whole genomes including next generation sequencing.

If *in vitro* regeneration must be used, a number of steps can be taken that may reduce the occurrence of somaclonal mutations. Minimizing the time in culture would likely reduce the opportunity for mutation and there are many steps that can be taken to ensure rapid regeneration. These include using young explant tissue, minimal sterilization to prevent tissue damage or growing donor tissue aseptically before transformation, using multiple selectable markers to allow for reduced antibiotic concentrations, and minimizing U.V. exposure. Development of *in*
Planta Agrobacterium transformation methods in *F. vesca* would eliminate somaclonal mutations caused by *in vitro* culture and regeneration. Unfortunately, there are characteristics of *F. vesca* that prohibit the development of these methods. The low efficiency of *in planta* transformation methods would require large numbers of *F. vesca* plants (flower buds) and seeds, as well as extensive greenhouse space to generate sufficient transformants. For the efficient generation of large numbers of T-DNA insertional mutants the *in planta* transformation method would best be coupled with a transposon tagging system. A dual system would generate large numbers of transposon tagged lines that would not have the somaclonal mutations caused by *in vitro* regeneration. Unfortunately, this scheme would require two generations before seeds could be screened as independent transposants, equaling a year under optimal conditions. These plants would not be immune from the mutations caused by the Agrobacterium infection and T-DNA integration process (reviewed in Chapter 1). The importance of time versus possibility of somaclonal mutation must be weighed to determine the best transformation scheme.

Complementation analysis of the *green petal* and *white runner* mutants is currently underway to confirm that expression of a wild type copy of the mutated gene can restore wild type phenotype. Even though these mutants do not segregate with their respective T-DNA insertion, their distinctive phenotypes facilitated analysis that led to the discovery of the causative mutation. The effects that the *green petal* mutation has on fruit development may be helpful in understanding the genetic control of auxin, abscission and biosynthesis of aroma related compounds. The *white runner* mutation resides in a gene encoding flavanone 3-hydroxylase (F3H), a key enzyme in flavonoid biosynthesis. Future experiments include investigating why the *white runner* mutant produces brown seeds, an indication of proanthocyanidin production that also depends on a functional F3H. This may indicate the presence of additional copies of F3H that are highly regulated to specific tissues. Preliminary metabolite experiments also indicate that the F3H mutation pathway “block” may lead to production of unusual products from the redirection of the pathway. Since the mutations were discovered and markers developed the *green petal* and *white runner* phenotypes can be tracked in seed lines and maintained. These phenotypic variants can now go into the bank of available mutants for others to utilize. Though the path to this point did not go as originally planned, the same end goal was achieved.
The *curly leaf* phenotype co-segregated with the T-DNA insertion and the insertion site was identified as a putative erythroblast macrophage attacher-like (EMP) gene, based on the homology to a gene characterized in mammalian cells. Complementation of the mutant phenotype with a wild type copy of the gene resulted in plants with normal leaves. This mutant is a good example of the utility of T-DNA insertional mutants in *F. vesca*. During the complementation analysis, additional phenotypes were identified that segregated with the *curly leaf* phenotype, including deficiencies in pollen germination, seed set, seed germination, petiolule length and epidermal layer adhesion. These phenotypes, including the *curly leaf*, are consistent with phenotypes observed in actin binding and microtubule associated protein mutants. Future experiments could be focused on visualizing microfilament and microtubule organization in the wild type, *curly leaf* and complemented mutants to determine if changes exist. This visualization could be accomplished using fluorescent stains that target actin and tubulin, or an immunohistochemical antibody targeting the EMP protein directly. Using stable transformation of the three genotypes with fluorescent protein tagged actin binding or microtubule associated genes, real-time microfilament and microtubule dynamics could be observed *in vivo* in rapidly changing cell types such as pollen tubes and root hairs. Co-immunoprecipitation techniques would be used to determine the *F. vesca* EMP protein complex partners. These experiments would be duplicated with the *A. thaliana* EMP mutants to utilize the speed and ease of the *A. thaliana* transformation system and compare EMP function and phenotype. These experiments would aid in understanding what the function of a protein, highly similar to one in mammals that is critical to red blood cell maturation, is in plants.

In conclusion, this project is not a testament to the functionality of the *F. vesca* T-DNA insertional mutant population for forward genetics studies, as had been hoped. Additional forward genetics studies like the ones presented here are necessary to determine how frequently somaclonal mutations are occurring and causing observed phenotypes. This study does demonstrate the valuable information that *F. vesca* can contribute as a developed model and with the critical importance of a publicly available, annotated genome sequence for defining the molecular basis for essential traits. *F. vesca* is still early in its development as an experimental model. As interest in human health and functional foods increases, strawberry offers many attractive qualities for manipulation in a lab setting that ensure its success.