Winter 2010

The genetics of sex determination in Fragaria virginiana: A candidate gene approach

Benjamin Bryant Orcheski
University of New Hampshire, Durham

THE GENETICS OF SEX DETERMINATION IN *FRAGARIA VIRGINIANA*: A CANDIDATE GENE APPROACH

BY

BENJAMIN BRYANT ORCHESKI

B.S., University of New Hampshire, 2007

THESIS

Submitted to the University of New Hampshire
in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
in
Genetics

December, 2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.

Thesis Director, Thomas M. Davis

Professor of Plant Biology and Genetics

Estelle M. Hrabak

Associate Professor of Molecular Biology and Genetics

James E. Pollard

Associate Professor of Plant Biology

12-13-10

Date

We choose to go to the moon. We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win, and the others, too.

- **JFK. Moon Speech - Rice University, 1962**

ACKNOWLEDGEMENTS

It's an honor just to be nominated. I suppose that thanks are in order to a whole bunch of people. This is arguably the most awkward portion of the whole thesis endeavor but in the end its better to have lots of people to thank than nobody at all. So let's start at the beginning (a very good place to start). I'll begin with my parents, Conrad and Leslie to whom I owe absolutely everything in my life. It is you who gave me love, taught me the value of a good education and a strong work ethic, and showed me that the world is truly a fascinating place. And perhaps above all you instilled in me a sense of burning curiosity, which is probably why I am where I am, doing what I'm doing, and not living in the basement (or the unfinished attic). I am more like you than you will ever know. And before I go any further, I also to express my appreciation to Bob Gabriel, the nicest human being I have ever had the pleasure to know.

I am also indebted to a number of people at UNH, especially to the members of the Davis Lab. You are some of the most fun and interesting people I have met and it has been an extreme pleasure to work with you. You have been like a second family to me over the years and I wish you the best life has to offer. Thanks are in order to Megan "Stewie" Thompson, an incredible friend and scientist whose "measure twice, cut once" brand of science has always been a delightful foil to my method of "shoot first ask questions later". Behind every great woman there is a great man. Having said that I want to acknowledge Jason "JayJay" Thompson a great friend, whose humor and wit makes him always enjoyable company.

I also have to acknowledge the advice of one Dr. Andy Laudano who suggested I find a lab to work in even before I began my stint at UNH. If not for that chance encounter things might have played out quite differently. Life is a funny thing.

To, Dr. Estelle Hrabak and Dr. Jim Pollard I want to extend my great appreciation. You have left your marks on me, not only as outstanding teachers but also as committee members, serving as a source of wisdom and guidance during this perilous endeavor.

Most of all I would like to thank the big man upstairs, Dr. Tom Davis who took a gamble more than seven years ago by hiring a stupid freshman with no formal scientific background. Then, miraculously three years ago, against all probable odds, he took that very same student on as a master's candidate. Go figure. So here's to you Tom. This degree is as much yours as it is mine. I would like to thank you for the encouragement you gave me and confidence you had in my abilities to take on the task at hand. And arguably most important, it was you who taught me more about being a scientist, than all the classes, labs, and tests over these past years. You let me be independent, curious, fiddle around, and to take my sweet time getting things done. You were there when I needed you and weren't when I didn't. Finally I'm indebted to your contagious enthusiasm and the thrill that you find in the pursuit of scientific knowledge, both of which have rubbed off on me. Thanks.

Well, that wasn't so bad. So here we go. Grab a glass of wine, find a comfy chair and enjoy. I can only hope you have as much fun reading this as I had during my years of working on it.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

THE GENETICS OF SEX DETERMINATION

IN *FRAGARIA VIRGINIANA*: A CANDIDATE GENE APPROACH

BY

BENJAMIN BRYANT ORCHESKI

University of New Hampshire, December, 2010

The evolutionary genetics of plant reproduction is a fascinating aspect of plant biology, involving an astounding diversity of reproductive strategies. While the majority of flowering plants are hermaphroditic (monoecy), many species have opted for an alternate mode, whereby male and female gametes originate on different plants (dioecy).

Wild strawberry, *Fragaria virginiana*, is a particularly interesting species, as its members can be female, male, or hermaphroditic. Its female flowers display arrested development of male reproductive organs (stamens), evidently conditioned by a single dominant allele of an unknown genetic locus. My aim was to isolate and characterize the gene(s) responsible for male sterility. A candidate gene approach was used to identify *F. virginiana* homologs of seven genes known to affect stamen development, representing three gene families. Their characterization included linkage analysis, results of which suggest particular ongoing focus on the *SUPERMAN*-like gene *FRASUP5*.

INTRODUCTION

Strawberry General Information

What could be more interesting than strawberries? Perhaps apples, but that remains to be seen. Strawberries belong to the genus *Fragaria* which is a member of the Rosaceae family. Members of this family have been cultivated and consumed since antiquity and include many ubiquitous fruits such as apples (*Malus*), pears (*Pyrus*), cherries, peaches, plums, apricots, and almonds (*Prunus*), blackberries and raspberries (*Rubus*) as well as many ornamentals like roses (*Rosa*), cinquefoil (*Potentilla*) and hawthorn (*Crataegus*). The genus *Fragaria* is exciting to study from both a genetic standpoint and because of its global economic importance. In 2004, world production of strawberries was estimated at 6.9 billion pounds, with the United States contributing 2.2 billion pounds (Rieger, 2007). The market value for strawberries produced in the U.S. was 1.9 billion dollars in 2008 (http://www.agmrc.org/commodities__ products/fruits/strawberries/ commodity_strawberry _profile.cfm, http://www.uga.edu /fruit/strawbry.html). The genus is interesting genetically for its range of ploidy levels, the species' ability to interspecifically hybridize, and the relatively small haploid genome size of 206Mb (Folta and Davis, 2006), which in recent years has made the diploid *Fragaria vesca* (2N =2X = 14) a model for Rosaceae functional genomics.

From a genetic standpoint, the genus *Fragaria* is appealing for study because it exemplifies the phenomenon of polyploidy. Unlike the vast majority of animals, polyploidy is not only tolerated but evolutionarily encouraged in plants. In fact

polyploidy appears to be an essential part of plant evolution and it is believed that most if not all plant families had one or more polyploidization event in their past (Blanc and Wolfe, 2004). *Fragaria* contains roughly twenty one species which range in ploidy from diploid to decaploid (Hummer and Hancock, 2009). In the genus, hybridization is thought to be an important vehicle for speciation and is a necessary precursor to allopolyploidy. This is evident in the genome organization of the three strawberry octoploid species, *F. ×ananassa*, *F. virginiana* and *F. chiloensis* (all 2N = 8X = 56).

From an economic standpoint, one species of *Fragaria* accounts for the great majority of income generated by the genus. The cultivated or dessert strawberry, *Fragaria ×ananassa* Duchesne is ubiquitous, grown around the globe from tropical climates to sub-arctic regions. This is a testament not only to the demand for the fruit but also the intensive breeding that has developed numerous cultivars suited to particular climates. From both a historical and evolutionary view *F. ×ananassa* is a recent species at slightly over 250 years old. It is an allopolyploid or alloautopolyploid (2N = 8X = 56), derived from an accidental cross between two wild octoploid species; *F. virginiana* (2N = 8X = 56) and *F. chiloensis* (2N = 8X = 56) (Rousseau-Gueutin, 2008). Both are new world species, *F. virginiana* being native to the eastern portion of North America while *F. chiloensis* is native to the western coasts of North and South America. During the colonization of the Americas the two geographically separated species were brought back to Europe as novelties. Grown side by side in botanical and novelty gardens in the early to mid 18th century (Staudt, 1962), the inevitable interspecific consummation, with a female clone of *F. chiloensis* serving as the seed parent, gave rise to a novel octoploid species; *Fragaria ×ananassa*. The economic success of the cultivated strawberry is due

to the melding of traits from its two parents. *F. chiloensis* has large fruit but lacks the characteristic sweetness and taste associated with the modern strawberry. On the other hand, *F. virginiana* has the distinctive sweetness and taste of the modern strawberry yet lacks the associated fruit size. The cross between these two wild species and subsequent breeding gave us the modern *F.* ×*ananassa*, a species greater than the sum of its parts with large sweet fruits and a distinctively strawberry taste.

Importance of Wild Strawberry Species

Although the varying levels of ploidy are interesting, they can serve as a barrier for strawberry improvement via traditional breeding. The difference in ploidy between the cultivated octoploid strawberry and its numerous wild diploid relatives means that desirable traits cannot be introgressed from the diploids to the octoploid without manipulation of the diploids. To create a fertile cross with *F.* ×*ananassa* and a diploid, the diploid must go through successive chromosome doubling to create an artificial octoploid (Bors and Sullivan, 2005). Otherwise, a cross of this nature would result in a sterile pentaploid (2N = 5X = 35). Therefore, wild relatives already at the octoploid level greatly increase the ease of genetic transfer between species.

As they are the direct progenitors of the cultivated strawberry, *F. virginiana* and *F. chiloensis* are prime candidates for the breeding of economically important traits into *F.* ×*ananassa* via introgression. A wide geographical distribution over the Americas for both wild relatives creates the possibility for a large number of desirable traits that could be introduced into the cultivated strawberry (Hancock and Luby, 1993). While these breeding techniques are being actively pursued, there are some limitations, namely sex. Cultivated varieties of *F.* ×*ananassa* are hermaphrodite, having both male and female sex

organs on the same flower. This is the most beneficial system for large scale cultivation as it maximizes the chances of pollination and thus fruit set. In the wild, things are not so cut and dried. The prevailing sexual strategy of *F. virginiana* and *F. chiloensis* is gynodioecy yet certain populations, may exhibit trioecy. Gynodioecy occurs when a species contains members that are both hermaphroditic and female. Trioecy is defined as females, males, and hermaphrodites occurring within the same species. The debate over which sexual strategy the species' use is complicated by the defining of males within a population. It has been reported in *F. virginiana* (Staudt, 1999) that hermaphrodite plants lose their ability to set fruit as they age which may cause females to be mistaken as males. Fruit set in hermaphroditic *F. chiloensis* is variable which may also lead to misidentified males. Hancock and Bringhurst (1979) contend that *F. chiloensis* is completely dioecious. However, it has been our experience with greenhouse populations of *F. chiloensis* that self-fertile hermaphrodites do exist in this species. The overall evolutionary advantage of dioecy is to facilitate out-crossing which promotes genetic diversity. Vegetative reproduction by runnering (stolons) in these species is also common and may be necessary if only a single sex colonizes a new area.

All known species of polyploid strawberries are known to display some form of sex differentiation, either dioecy or some intermediate form (heteroecy). Sex differentiation is even seen in the *F.* ×*ananassa* subspecies *cuneifolia* although this subspecies is not cultivated (Staudt, 2009). Given the fact that there are varying types of sex expression throughout the different levels of ploidy, it is likely that sex differentiation evolved multiple times in the genus (Staudt, 2009). It is also of interest to note that all the

diploid species are monoecious, although Ahmadi and Bringhurst (1991) report that dioecy has been seen in *F. vesca* subsp. *bracteata*.

The differing sex morphs of *F. virginiana* and *F. chiloensis* present a problem for the introgression of economically important traits into *F.* ×*ananassa*. As mentioned above, female and male strawberries are not desirable from an agricultural standpoint. It is always possible to use wild hermaphrodite relatives for introgression; however this may not always be a satisfactory option, as when a desirable trait is closely linked to sex. Therefore it is of practical as well as basic scientific importance to understand the genetic basis of sex determination in *Fragaria*. The discovery of markers or actual genes associated with the male or female trait will be extremely important to strawberry breeding, particularly in the area of Marker Assisted Selection (MAS). MAS uses molecular markers linked to desirable (or undesirable) traits to screen the progeny of a cross when they are young. Offspring harboring alleles linked to desirable traits are kept while those with alleles linked to undesirable traits are discarded. For example, with this technique wild female relatives can be used in traditional breeding. If the beneficial trait is not linked to sex, marker(s) for sex can be used to screen the F1 for females to discard. If the trait is linked to sex, marker(s) for sex can be used to screen the F1 for recombinants.

Sex Determination in Octoploids

From a transmission genetics standpoint, the current model of sex determination in the octoploid species *F. virginiana* and *F. chiloensis* presented by Ahmadi and Bringhurst (1991) is that three alleles at a single locus determine sex in these organisms. Under this model, femaleness (F) is dominant to hermaphrodite (H) which is in turn

dominant to maleness (M). Hermaphrodites may be homozygous (H/H) or heterozygous (H/M) while males are always homozygous (M/M). Female plants are always heterozygous at this locus (F/H or F/M), and segregate 1:1 female to hermaphrodite (or male) except when both parents contain only one M allele (F/M x H/M), which results in a 1:2:1 hermaphrodite to female to male ratio.

Recently, Spigler et al. (2008; 2010) proposed that sex determination in *F. virginiana* is not under the control of a single locus. Rather, these authors concluded that the male and female sex determining loci are distinct, yet tightly linked. Their mapping population was an F1 derived from a cross between a female and a hermaphrodite. They found that 2 out of 184 F1 offspring were neuter, lacking both male and female fertility (Spigler et al. 2008). This led them to the conclusion that male and female fertility are controlled by two closely linked loci with a low frequency of recombination, the two observed neuter types being identified as rare recombinants. In this present study, the genetic basis of female fertility/sterility was not investigated, and detailed phenotypic data on this trait were not collected, thereby precluding the identification of neuters. Rather, only the phenotypic dimorphism of male sterility/fertility and its genetic basis were investigated, and as such for this present study the locus in question will be designated the "sex determination locus" unless explicitly stated otherwise.

In developing their model, Spigler et al. (2008; 2010) were greatly aided by the discovery of a Simple Sequence Repeat (SSR) marker designated ARSFL7, which was found to be tightly linked to sex determination in *F. virginiana*. Thus it served as a critical tool for confirming the presence of neuters in their mapping population by detecting recombination between the ARSFL7 marker and the sex determination loci.

This marker was also used extensively throughout this study and will be discussed in detail below.

With the results laid forth by Ahmadi and Bringhurst (1991) and Spigler et al. (2008; 2010), it is very likely that the sex determining locus is found in only one subgenome. This conclusion is based on several observations: i) disomic inheritance of markers segregating with sex (the SSR ARSFL7) as discussed below, ii) females are always heterozygous and produce ~50% female (male sterile) progeny, and iii) male fertility is recessive to male sterility. If the locus for sex determination was in all eight copies of a particular chromosome, it would be very difficult to have a homozygous recessive plant in a highly out-breeding octoploid species.

Inheritance Patterns in Octoploid Strawberry

The genomic makeup of the octoploid species is extremely fascinating and has profound implications for the study of inheritance in these organisms. The pattern of inheritance is of particular importance when using a candidate gene approach since it gives insight into the number and nature (dominant/recessive) of genes involved in a trait. It has been well-established that the two wild North American octoploid species, *F. virginiana* and *F. chiloensis,* are the result of allopolyploidization. By extension, the fact that *F.* ×*ananassa* is derived from these two species means that it is also an allopolyploid (Hummer and Hancock, 2009). Two questions arise from this evolutionary scenario and are the topics of a heated debate within the strawberry community. What type of inheritance is currently exhibited in these octoploid species? What were (or are) the *Fragaria* species that contributed their genomes to the octoploid genome?

Three models of subgenomic composition and inheritance have been published on *Fragaria* ×*ananassa* and are all based on cytological evidence of chromosome pairing and segregation. However, as they make fertile offspring (i.e. *F.* ×*ananassa*) and share a direct common ancestor, it is reasonable to speculate that *F. virginiana* and *F. chiloensis* contain the same genome composition and mode of inheritance.

The first model presented by Fedorova (1946) contends that the genome of *F.* ×*ananassa* contains three subgenomes designated AAAABBCC. The three subgenomes represent three different genome contributors in the octoploids' evolutionary past. Under this model, there is mixed disomic and tetrasomic inheritance. The homologous chromosomes of the B and C subgenomes form bivalents (characteristic of disomic inheritance) during meiosis while the A subgenome has four copies and forms either bivalents, or quadrivalents that disjoin in an orderly II + II manner indicative of tetrasomic inheritance.

The second model of octoploid genome composition presented by Senanayake and Bringhurst (1967) is an updated version of the previous model. According to Senanayake and Bringhurst, *F.* ×*ananassa* has the genome composition AAA'A'BBBB. This is quite similar to the model presented by Fedorova, except that there are no longer three subgenomes equally distinct from each other. Under this model, the octoploid genome is composed of three subgenomes, but two of those subgenomes, A and A', are more closely related to each other than to the B subgenome. The Senanayake and Bringhurst model poses that homologous chromosomes of the A and A' subgenomes each form bivalents during meiosis while the B subgenome forms quadrivalents.

However, the A chromosomes may sometimes pair with their A' homoeologs in interspecific hybrids.

According to the third and most current model presented by Bringhurst (1990), *F.* ×*ananassa* is composed of four subgenomes; AAA'A'BBB'B'. While still distinct subgenomes, A and A' are evolutionary closer to each other than to any B subgenome and vice versa. Under this model, *F.* ×*ananassa* is fully diploidized (acts like a diploid), forming only bivalents during meiosis and thus follows a pattern of disomic inheritance. Chromosomes within the same subgenome are homologs and homologous chromosomes pair during meiosis to undergo recombination with each other. Corresponding chromosomes that are in different subgenomes are homoeologs. They share a high degree of synteny conservation but, although they may contain the same genes (or alleles), they do not pair during meiosis and thus do not recombine.

A fully diploidized octoploid genome has profound implications on genome organization and study of these species. The existence of four subgenomes adds to the challenge of linkage mapping in the octoploid species, as there should be 28 linkage groups (four subgenomes with seven chromosomes per genome) for an octoploid. In terms of organization, lineage-specific adaptations may be confined to only a single subgenome (from that lineage) as appears to be the case regarding sex determination in *F. virginiana.* At least as far as sex is concerned, it seems that only a single subgenome governs the determination of a plant's sex Furthermore recombination suppression may be occurring at the sex locus in order to keep specific genes associated with a "sex haplotype." This is based on evidence presented by Ahmadi and Bringhurst (1991) and

from association studies with the ARSFL7 marker segregating with sex (Spigler et al., 2008; 2010) in *F. virginiana* (discussed below).

Unfortunately, there has been a notable scarcity of research studying the genome organization of the wild octoploid species; *F. virginiana* and *F. chiloensis*. However one recent study by Ashley et al. (2003) has provided evidence that *F. virginiana* follows a disomic pattern of inheritance. This small scale study tracked the segregation of four Simple Sequence Repeat (SSR) markers in the F1 of two separate female by hermaphrodite crosses. The researchers found that the allele compositions in the F1 were exclusionary, such that if two alleles of a marker were present in one of the parents only one of these alleles would be found in the F1. This is characteristic of disomic inheritance where the alternate alleles would separate during meiosis. All four SSR markers were found to segregate in a disomic fashion, although one did deviate significantly from the expected Mendelian ratio. The obvious limitation to this study is the small sampling size of the markers. As only four SSR's were studied and no indication was given as to what linkage group (chromosome) they belonged to, it is possible that all the markers were on a single pair of homologous chromosomes within a subgenome that exists in only two copies. Other parts of the genome could still display tetrasomic inheritance. To say anything conclusive about inheritance in the whole *F. virginiana* genome, more segregation of SSR markers covering every linkage group will need to be studied. Despite the study's shortcomings, the evidence provided by Ashley et al. (2003) along with the association study of the sex-linked marker ARSFL7 (Spigler et al., 2008; 2010) (discussed below), shows there is a strong indication that at least a subset of the *F. virginiana* genome behaves in a disomic manner.

A Candidate Gene Approach

Since the early 1990's a vast amount of research has been performed in the area of flower development with numerous associated genes being isolated and characterized (Albani and Coupland, 2010; van Nocker and Ek-Ramos, 2007). These genes have been implicated in virtually every aspect of flower development from the inception of the floral meristem to the maturation of the seed. Compared to what is known about sex determination in other plant species and especially in model organisms such as *Arabidopsis*, *Petunia*, and *Antirrhinum*, relatively little is known about sex determination in the genus *Fragaria*. The most logical approach to studying sex determination in strawberry is to determine what genes affect sexual development in other well-studied organisms and then isolate and characterize their orthologs in *Fragaria*. This transferability of known gene function in one organism to study a phenomenon of unknown genetic basis in a different organism is known as a candidate gene approach. For this study a total of seven *Fragaria* genes belonging to three gene families were studied to determine if any were implicated in sex determination. From strawberry, five members of the *SUPERMAN*-like gene family of putative transcription factors, along with the *AGAMOUS* transcription factor, and the ethylene synthesis gene *ACS-7* were chosen as candidate genes. The following presents the reasoning behind why these particular genes were chosen for further study.

The ABC Floral Homeotic Genes

Introduction

Genes that specify the identity of organs in a flower make natural candidates for the determination of sex in *Fragaria virginiana*. In order to understand why these genes

are candidates, we must first understand the current model of flower development from a genetic standpoint. A typical perfect flower is composed of four rings of organs, known as whorls. The first two whorls do not play a direct role in plant reproduction. The first whorl is comprised of sepals, which are leaf-like structures that protect the unopened flower. Petals, which are used to attract pollinators, compose the second whorl. The third whorl contains stamens, the male reproductive structures which produce pollen. In the center of the flower is the fourth whorl, made up of carpels, the female reproductive structures that contain ovules. The basic genetic model of the flower development pathway is remarkable in its simplicity. However, as we shall see, the regulation of the "master switch" genes involved in flower development can at times be stunningly complex.

Our current understanding of the genes underlying floral development was mapped out in the late 1980's to early 1990's (Meyerowitz et al. 1991). This work occurred primarily in the Meyerowitz lab and culminated in the "ABC Model of Flower Development" (Figure 1). The ABC Model postulates that the four whorls of organs in a wild type flower are created by the differential expression patterns of three classes of genes, A, B, and C (Meyerowitz et al. 1991; Weigel and Meyerowitz, 1994). These genes were first discovered in the model plant *Arabidopsis thaliana*, and a total of five were characterized: *APETALA1* (*AP1*), *APETALA2* (*AP2*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*), and *AGAMOUS* (*AG*) (Parenicová 2003). *AP1* and *AP2* are A-class genes, *AP3* and *PI* are B-class genes, and *AG* is the C-class gene. Expression of A-class genes alone controls sepal development, while A-class and B-class proteins direct petal development,

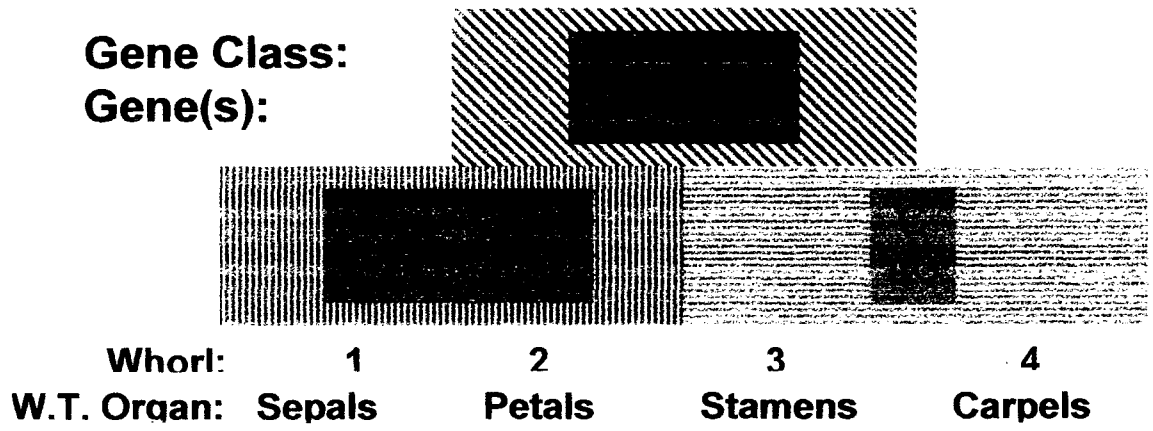


Figure 1. The ABC Model of Flower Development. The wild-type gene expression pattern in a perfect (hermaphroditic) flower based on the "ABC Model of Flower Development" (Meyerowitz et al. 1991). Whorl 1, composed of sepals requires the expression of the A-class genes, *AP1* and *AP2*. Whorl 2 composed of petals requires A-class as well as the B-class genes *AP3* and *PI*. Stamens comprise whorl 3 and rely on B-class expression as well as the C-class gene *AG*. *AG* acts alone in the whorl 4 to control carpel development.

B-class and C-class proteins control stamen development, and the C-class protein directs carpel development (Ng and Yanofsky, 2000).

These genes are known as Floral Homeotic Genes (FHG), analogous to the well-studied homeotic genes directing embryo development in *Drosophila*. When mutated (either by loss of function or ectopic expression) homeotic genes are able to direct the development of a morphologically normal organ in an incorrect position. As we shall see, the ABC homeotic genes are able to wreak havoc on normal flower morphology when mutated.

Since they direct stamen development, the B-class (*AP3* and *PI*) and the C-class (*AG*) genes are natural choices as candidate genes for sex determination in *Fragaria*. Indeed, some Arabidopsis mutants of these genes do appear to somewhat mimic the phenotype of female *Fragaria*, namely, undeveloped stamens with occasional petalloid stamens. Therefore the rationale for candidate genes will focus on the B and C-class genes rather than the A-class. However, it is important to know that none of the FHG classes act independently. There appears to be a complex network of synergistic and antagonistic relationships between the three classes. These relationships in turn govern the phenotypes observed in the various mutant (loss-of-function and ectopic expression) classes.

Before discussing the roles of the B- and C-class genes in flower development, some description of the gene family is warranted. The ABC genes are members of the MADS-box family (Weigel and Meyerowitz, 1994) with the exception of *AP2* which belongs to a separate *AP2*-like family (Weigel, 1995). All of the Floral Homeotic Genes

are transcription factors and thus have the capacity to bind DNA as either homodimers or heterodimers.

The MADS-box family appears to be conserved throughout higher eukaryotes, being found in animals and fungi in addition to plants. All members contain the highly conserved MADS-box which is the DNA binding domain of the protein. Every higher plant genome studied contains the MIKC-type MADS-box sub-family. MIKC-type MADS-box genes, which are plant specific, are characterized by four discernable domains. Starting from the N-terminus, the first domain is "M" for the MADS-box which is essential for DNA binding. Second is the "I" or intervening domain, which appears to control the specificity of DNA binding. Aside from the MADS-box, the "K" or keratin-like domain is the most conserved and mediates protein-protein interaction (dimerization). Last is the "C" or C-terminal domain which is the least conserved of the four domains. It is also the least understood but is thought to stabilize K-domain interactions and is also required by the B- and C-class genes for their function as transcriptional activators (Mizukami et al. 1996).

APETALA3* and *PISTILLATA

Let us begin our journey to the center of the flower with the two B-class genes, *APETALA3* (*AP3*) and *PISTILLATA* (*PI*). In the second whorl of wild type flowers, the B-class genes mediate petal development and in the third whorl they mediate stamen development. The two respective proteins work as a heterodimer in conjunction with the A-class and C-class proteins to activate the transcription of genes necessary to build petals and stamens respectively. As a consequence of this requirement to dimerize, the activity of either B-class protein is dependent upon the activity of the other. However, the

maintenance of expression of either B-class gene is independent of the other (Weigel and Meyerowitz, 1994).

Loss-of-function of either *AP3* or *PI* results in a loss of B-class activity. In Arabidopsis, this is characterized by a conversion of the second whorl petals to sepals and of third whorl stamens to carpels (Bowman et al. 1991a; Coen and Meyerowitz, 1991). Furthermore, in some severe loss-of-function *PI* alleles, organs are actually missing, as some organ primordia in the third whorl fail to develop (Bowman et al. 1991a). Since *AP3* and *PI* are not expressed in whorls one and four, development of these whorl organs remains phenotypically normal in these B-class mutants. From these data we can begin to see a picture whereby some allelic variant of a *Fragaria* B-class gene could provide the phenotype seen in female flowers. The female's vestigial stamens could be the result of a partial conversion of the stamens into carpelloid tissue. However, this scenario is not supported by the fact that in female flowers, second whorl petals are phenotypically normal and in no way resemble sepals. Work by Hill et al. (1998) has shown that the *APETALA3* gene has elements in its promoter that are specific to either second or third whorl expression. Based on the sequence similarity between *AP3* and *PI*, it is likely that *PISTILLATA* also contains whorl specific promoter elements. Therefore, one possible explanation for the phenotypically normal female petals is that the mutant allele simply lacks the elements to be expressed in the second whorl and so is only expressed in the third whorl. This explanation would only hold water if the B-class allele was a dominant negative, something that will be discussed below.

Over-expression studies performed by Krizek and Meyerowitz (1996) with both *APETALA3* and *PISTILLATA* showed drastic deviations from the wild type flower

morphology. This work built upon that done by Jack et al. (1994) who showed that the over-expression of the *AP3* gene under the constitutive Cauliflower Mosaic Virus 35S promoter converts carpels to stamens. Krizek and Meyerowitz (1996) generated transgenic *35S::PI* Arabidopsis plants that showed a partial conversion of first whorl sepals to petals. Transgenic plants had a first whorl that was a mosaic of sepal-petal tissue. When the two single mutant lines were crossed to create a *35S::AP3* and *35S::PI* double mutant, the transgenic plants showed a complete conversion of the first whorl sepals to petals and the fourth whorl carpels to stamens. Therefore it would appear that in the double mutants, the action of *AP3* enhances the conversion of sepals to petals observed in the *35S::PI* single mutant.

Neither of the over-expression phenotypes observed in the single or double mutant *AP3* and *PI* lines matches the characteristics of a female *Fragaria* flower. Female strawberry flowers retain their first whorl sepals and contain fully fertile fourth whorl carpels. Thus it would appear that the female trait in *Fragaria virginiana* is not due to over-expression of either of the B-class genes.

AGAMOUS

We shall conclude our journey to the center of the flower with a study of *AGAMOUS* (*AG*), the single C-class Floral Homeotic Genes. *AGAMOUS* was the first FHG to be cloned and sequenced and as such, has the distinction of being the best characterized of all the ABC genes (Yanofsky et al. 1990). *AGAMOUS* also has the distinction of playing three critical roles in proper flower development. First, *AG* expression is essential for development of the correct organs in flower whorls three and four. In the third whorl, *AG* works in concert with the two B-class genes (*AP3* and *PI*) to

promote the proper development of stamens, while in the fourth whorl *AG* acts alone to specify carpel development.

In *AG* loss-of-function mutants, the third whorl stamens are converted to petals. In the fourth whorl even more drastic homeotic conversions occur such that the fourth whorl develops as sepals rather than carpels. Inside these fourth whorl sepals, numerous extra floral whorls form such that "new" flowers develop inside the wild type flower. These observations bring us to the next two functions that *AG* orchestrates in flower development. *AG* appears not to be simply just an FHG but also a cadastral gene, controlling the expression boundaries of the A-class genes *AP1* and *AP2* (Irish and Sussex, 1990; Meyerowitz et al. 1991; Weigel and Meyerowitz, 1994). When normal *AG* function is removed from the flower, it allows ectopic A-class expression in the third and fourth whorl. This in turn converts the third whorl to petals (by the combined action of the A- and B-class genes) and the fourth whorl to sepals (by the sole action of the A-class genes). Thus *AG* is a negative regulator of A-class expression in the third and fourth whorls.

The *AGAMOUS* gene product also appears to be critical for floral meristem determinacy. This is based on the fact that in wild type Arabidopsis, a single floral meristem gives rise to a single flower. However, in *AG* loss-of-function mutants, multiple whorls and flowers develop interior to the fourth whorl of the wild type flower. Thus the AG protein directs the floral meristem to terminate after the spatial pattern of floral whorls is determined. *AGAMOUS* may also be a negative regulator of cell division, since in the *AG* mutants, the formation of extra whorls and organs (arising from organ

primordia) interior to the fourth whorl demands a great deal of cell proliferation (Meyerowitz et al., 1991).

None of the observed *AGAMOUS* loss-of-function phenotypes appears to explicitly mimic what is seen in female strawberry flowers. To begin with, female *Fragaria virginiana* have fully fertile carpels. There is no apparent conversion of carpels to sepals nor is there a second flower that develops interior to the carpels. With regards to the ectopic petal phenotype of Arabidopsis *AG* mutants, more similarities can be envisioned. It is not a stretch of the imagination to speculate that the vestigial stamens observed in female plants could be due to a partial conversion of stamens to petals, especially in light of the fact that females more often produced petalloid stamens than hermaphrodites (personal observation).

Fortunately, the two observed phenotypes in an *AG* loss-of-function mutant may not have to occur together. Work performed by Deyholos and Sieburth (2000) demonstrated that the large (3.8kb) second intron of the *AGAMOUS* gene controls whorl-specific expression. They were able to generate beta-glucuronidase (*GUS*) reporter gene constructs from various fragments of the *AG* second intron. Many of these constructs showed differential GUS staining such that certain constructs were expressed predominantly in the third whorl while others were expressed predominantly in the fourth whorl. From this work, it appears that *AG* contains whorl specific regulatory elements. This leads to the possibility of *AG* loss-of-function mutants that are only specific to a particular whorl as a consequence of mutations in either of the whorl specific regulatory sequences. One could now imagine a situation in strawberry whereby an *AG* allele lacks the elements for whorl three expression (resulting in ectopic petals) but retains regulatory

elements for normal fourth whorl expression. In addition, the *Fragaria AG* ortholog also contains the large (~4kb) second intron (presented herein). Just like the mutant B-class genes discussed previously, this scenario only works if the mutant *AG* allele is a dominant negative, something to be discussed below.

Over-expression of the *AGAMOUS* gene does not appear to mimic the phenotype of female *Fragaria virginiana* flowers. In 1992, Mizukami and Ma transformed Arabidopsis so it would over express an *AG* cDNA using the 35S promoter. They found that in the first whorl, sepals are converted to carpels and second whorl petals are converted to stamens. Whorls three and four remain phenotypically normal in transgenic plants. These results in no way resemble female strawberry flowers, where the only abnormality is the loss of stamens. All other whorls in these flowers remain wild type.

The *AGAMOUS* gene lends itself to be a candidate gene because it has been isolated from strawberry (*Fragaria ×ananassa*) and its expression has been well characterized. The putative strawberry ortholog of *AG*, designated *STAG1* (GenBank accession AF168468) was isolated from strawberry as a cDNA (Aharoni et al. 1999). Later on Rosin et al. (2003) performed an in-depth expression study of the *STAG1* gene, including Northern blot, in situ hybridization, and *GUS* reporter fusion constructs. To perform a Northern blot, RNA isolated from various organs was probed with a labeled *STAG1* probe. The probe only hybridized to transcript present in stamen and carpel tissues, while no signal was detected in RNA isolated from petals, leaves, or roots.

Strawberry fruit development is quite complex as fruit is made of many tissue types. To discern tissue specific expression pattern of *STAG1* in developing fruit, in situ hybridization was performed with developing green fruit using a second labeled probe.

No probe was found to hybridize to the receptacle or "false fruit." However, *STAG1* expression was found in the endothelium surrounding the embryo and the vascular strands which connect the developing achenes (true fruit) to the receptacle.

Finally the authors used a *GUS* promoter fusion construct to study *STAG1* expression. This was done by fusing a 3.6kb fragment of the *STAG1* promoter to the *GUS* gene and then transforming the construct into strawberry. Expression of the construct was flower-specific, as there was no GUS staining in leaves, roots or petioles. The authors found *GUS* expression in the stamens and the base of the receptacle and petals. Stamen expression was expected, however expression at the base of the petals was not expected, as *AG* is not normally expressed in the second whorl. In addition, no *STAG1* signal was detected from isolated petal RNA. The authors speculate that since *AG* is normally not expressed in the petal whorl, the observed *GUS* expression may be due to a spill over of the strong GUS staining from the stamens. Finally, GUS staining was observed in all stages of the developing achenes. Both the stamen and achene GUS staining is similar to the results of *AG* in situ hybridization observed in the late stages of the Arabidopsis flower which shows signal in the stamens and endothelium surrounding the ovary (Bowman et al. 1991b). Taken together, the expression data support the conclusion that the *STAG1* gene is the ortholog of the Arabidopsis *AGAMOUS* gene.

The expression of ABC genes in dioecious species is also of great interest to the study of sex determination in flowering plants. In addition, the differential expression of FHG's in one species may give insight to the cause of dioecy (or gynodioecy for that matter) in other species. In 2005 Sather et al. published their findings on the expression of the *AG* ortholog of spinach (*Spinacia oleracea*), a dioecious plant. There appears to be

a strong correlation in spinach between the expression pattern of the *SpAG* gene and the sex of the flower. To begin, based on Northern blot analysis, the *SpAG* gene is exclusively expressed in floral tissues. Early in flower development, in both male and female flowers, *SpAG* is expressed throughout the entire floral meristem. Later on, when the first whorl differentiates into sepal primordia, *SpAG* is only expressed in the primordia of the reproductive whorl (either stamen or carpel) of the respective sex. The sexes can be easily determined after the establishment of first whorl sepal primordia because females have two sepals while males have four. In male flowers, the fourth whorl is completely missing, comprising simply a flat layer of cells with no *SpAG* expression. In the females, the third whorl is a ring of cells with no *SpAG* expression.

The authors conclude that *SpAG* expression is important in establishing the sexual dimorphism in spinach. However, differential *SpAG* expression does not appear to be the direct cause of different sexes. This is because the number of sepal primordia is different in males and females and the primordia arise before *SpAG* is differentially expressed in the two flower types. What is remarkable about the *SpAG* gene in dioecious spinach is that it represents a natural case of whorl specific *AGAMOUS* expression. Thus it appears that in at least one dioecious species, the same gene can be regulated such that it can direct the development of different sex organs in different flowers. Given that *Fragaria virginiana* is gynodioecious it is conceivable that the strawberry *AG* ortholog could have differing expression patterns in the two sexes.

When all is said and done, the B- and C-class Floral Homeotic Genes do not shape up to be the best candidates for the observed phenotype of female strawberry flowers. In regards to the Floral Homeotic Genes, in Arabidopsis, most naturally

occurring alleles or those derived through mutagenesis are recessive. This makes them fine candidates for a diploid organism but *Fragaria virginiana* is an octoploid, containing eight sets of chromosomes. If all four homeologous chromosome pairs contain copies of any given FHG, then it is highly unlikely that eight mutated copies of a gene could exist in a plant. There is certainly no guarantee that all eight chromosome pairs would have a particular FHG. However, all octoploid species are of allopolyploid origin, and there is strong phylogenetic evidence that, diploids were the main (if not exclusive) genome contributors (Folta and Davis, 2006). As all extant diploid strawberry species are fully fertile (with the exception of one subspecies, *F. vesca* ssp. *bracteata*), it is likely that they each contributed a fully functional set of FHG's to *F. virginiana*.

Furthermore, it has been well established that femaleness is a dominant trait, and so by their very nature females must be heterozygous for the sex determining gene (Ahmadi and Bringhurst, 1991). This is because females cannot be crossed with females (or selfed) therefore pollen can only come from a male or hermaphrodite flower. Ahmadi and Bringhurst (1991) also demonstrated that, when selfed, hermaphrodites can only produce hermaphrodites (if they are H/H) or hermaphrodites and males (if they are H/M). Selfing a hermaphrodite cannot generate female progeny, therefore femaleness is a simple dominant trait.

Only through promoter fusions to create ectopic or over-expression lines can a Floral Homeotic Gene product become phenotypically dominant (Krizek and Meyerowitz, 1996; Mizukami and Ma, 1992). However, dominant negative alleles of both B- and C-class FHG's can occur. A dominant negative mutation occurs when a mutation in the coding region of a gene eliminates some but not all aspects of the protein's function

(Mizukami et al. 1996). An apt example would be DNA binding proteins that must form higher order complexes to bind DNA. It may be possible to have a mutant protein that can still dimerize (interact with the wild type protein) but that cannot bind DNA. Thus in a heterozygote, the mutant protein eliminates function and as such is dominantly negative.

Two examples of dominant negative B- and C-class mutants exist. The first report that a B-class gene could produce a dominant negative mutation was by Krizek and Meyerowitz (1996) who found that two alleles of the *PISTILLATA* gene, *pi-1* and *pi-3* acted in this manner. The *pi-1* allele contains a premature stop codon after the MADS-box domain while the *pi-3* allele contains a non-synonymous mutation in the MADS-box domain. It is known that both of these alleles are dominant negative because ectopic expression of the wild type *PI* gene in Arabidopsis cannot rescue the mutant phenotype like it can for other mutant *PI* alleles (Krizek and Meyerowitz, 1996). The authors speculate that even with an introduced wild type copy of the *PI* gene, these alleles generate enough mutant protein to interact with the wild type AP3 protein and inhibit the proper function of the AP3/PI heterodimer.

In the same year, Mizukami et al. (1996) found that certain deletion mutants of the *AGAMOUS* gene in Arabidopsis could act as dominant negatives. The researchers generated a series of deletion mutants by truncating the *AG* coding region to only include certain domains (see above). These constructs were driven by the 35S promoter and were transformed into wild type plants (containing the wild type *AG* gene). Some constructs were able to produce the *AG* over-expression phenotype (see above) while other transformants resembled wild type, indicating that the transgene had no function. Interestingly, the researchers found that the *AG* construct lacking codons for the C-

terminal end of the protein displayed an *AG* loss-of-function phenotype, despite the fact that plants already contained a wild type *AG* gene. They speculated that the mutant construct was able to dimerize with the native AG protein but that this dimer has no AG function, hence the loss-of-function phenotype. Therefore an *AG* allele that does not code for the C-terminal end of the protein acts in a dominant negative manner.

Unfortunately, dominant negative alleles suffer from mutations in the coding sequence rather than in the promoter. Therefore although it may not function properly, the expression pattern of the mutant allele should be identical to wild type (unless possibly the product regulates its own expression). This system would not work in the female flowers of *F. virginiana*. This is because a dominant negative FHG mutant would have a phenotype in the two out of four whorls it is expressed in. In female strawberry flowers only whorl three shows a mutant phenotype, the vestigial stamens. It would certainly be possible to have a mutant allele that has both expression and coding region mutations. However this would require two different mutations to occur within the gene, making it less likely to occur naturally.

Although some of the mutant phenotypes associated with the ABC Floral Homeotic Genes appear to mimic the phenotype observed in female *Fragaria virginiana*, many do not. Separable whorl-specific function of the ABC genes has been demonstrated, but the fact that most mutant FHG alleles are recessive (while femaleness is dominant) and that *F. virginiana* is octoploid, does not suit the conclusion that one of the B- or C-class Floral Homeotic Genes is responsible for sex determination in strawberry. As we shall see next, another class of genes appears to be much better candidates for the sexual dimorphism in *Fragaria*.

SUPERMAN and *RABBIT EARS*: The Gene Police

Introduction

Who will police the police? It is well established that the Floral Homeotic Genes are the master switches in deciding what organ will be found in a particular whorl. But what factors decide when and where the Floral Homeotic Genes will be expressed? It has been known for some time that certain FHG pairs act in an antagonistic fashion towards each other. For instance, the *AP2* gene expressed in whorls one and two prevents the expression of *AG* in these two whorls. Conversely the *AG* gene expressed in whorls three and four prevents the expression of *AP2* in these two whorls. However, other Floral Homeotic Gene pairs do not appear to have this sort of regulation. The spatial expression of both B-class genes (*AP3* and *PI*) and the C-class gene *AG* in the flower appears to be under the control of what are known as cadastral genes. Cadastral genes do not specify organ type like the Floral Homeotic Genes (Bowman et al., 1992). Rather, they control the expression boundary of FHG's, relegating them to only certain whorls. By delineating and maintaining expression boundaries, cadastral genes also play an important role in determining overall organ number. Given their important role in regulating FHG expression, cadastral genes are in all reality the gene police, and will be discussed below.

C2H2 Structure

SUPERMAN (*SUP*) and *RABBIT EARS* (*RBE*), two important cadastral genes, code for members of a much larger family of proteins known as C2H2 zinc-finger proteins (ZFP's). These family members are mostly nucleic acid binding proteins, able to bind DNA and RNA through their zinc-finger domain, although some members have protein binding capabilities (Englbrecht et al., 2004). The zinc-finger tetrahedrally

coordinates a single zinc atom through two cysteine (C) and two histidine (H) residues (hence the C2H2). The consensus sequence for the C2H2 domain is (F/Y)-X-C-X2–5-C-X3-(F/Y)-X5-ψ-X2-H-X3–5-H, ("X" representing any amino acid and ψ is a hydrophobic residue) with the amino acids forming a compact B-hairpin and A-helix structure to coordinate the zinc atom (Wolfe et al. 2000; Takatsuji and Matsumoto, 1996). The number of zinc-fingers in a protein can vary from one to five and can be arranged tandemly or dispersed along the length of the protein. The well-established DNA binding properties of certain members of this family, including *SUPERMAN* and *RABBIT EARS*, implicate them as transcription factors.

A comprehensive bioinformatic study of the C2H2 zinc-finger family has been undertaken by Englbrecht et al. (2004). They found that the Arabidopsis genome contains 176 C2H2 zinc-finger coding genes, few of which are conserved in other classes of eukaryotes. The authors found that there are significant differences in the structure of plant and animal C2H2 zinc-finger proteins. While most animal ZFP's have their zinc-fingers arranged tandemly, the majority of plant ZFP's with multiple fingers have them dispersed throughout the protein. Therefore it appears that the radiation of zinc-finger genes throughout the plant genome may have an important role in the evolution of plants. The method of classification used by Englbrecht et al. (2004) is based on the arrangement of zinc-fingers in the ZFP's. The three main types of ZFP's are set A, set B, and set C which is further divided into three subsets: C1, C2, and C3. Set A comprises ZFP's that contain a single tandem array of zinc-fingers. Set B contains proteins with more than one tandem array of zinc-fingers. Set C members either contain a single zinc-finger or multiple dispersed fingers. The three different subsets are distinguished by the number of

amino acids between the two histidine residues in the zinc-finger domain a follows: C1 H-X3-H, C2 H-X4-H, and C3 H-X5-H. Subset C1, of which *SUPERMAN* and *RABBIT EARS* are members, is the largest group with 64 members in the Arabidopsis genome. Subset C1 is also defined by a mostly invariant peptide motif QALGGH within the zinc-finger. The zinc-finger domains of the Arabidopsis SUPERMAN and RABBIT EARS proteins, highlighting the coordinating cysteine and histidine residues are: SUP, **C**SF**C**KREFRSAQALGG**H**MNV**H** and RBE, **C**SF**C**GREFKSAQALGG**H**MNV**H**.

The *SUPERMAN* Sub-Family

The *SUPERMAN* and *RABBIT EARS* genes belong to the *SUPERMAN*-like *RBE-SUP* sub-family of the larger C2H2 zinc-finger family (Takeda et al., 2004). All SUPERMAN-like proteins contain two defining characteristics: a single zinc-finger domain with the invariant QALGGH motif and a highly conserved leucine-rich DLELRL motif at the carboxy-terminal end of the protein (Hiratsu et al. 2004). These proteins are all putative transcription factors with a DNA binding domain (zinc-finger) and a transcriptional repression domain (leucine rich motif). To date, eight *SUPERMAN*-like genes have been reported in the Arabidopsis genome (Takeda et al., 2004).

Bowman et al. (1992) stress that *SUPERMAN* (and by extension *RABBIT EARS*) are not Floral Homeotic Genes like the ABC class MADS-box genes. They make the distinction that homeotic genes specify the identity of floral organs and so mutations in Floral Homeotic Genes transform one floral organ or part to another. Rather, *SUPERMAN* and *RABBIT EARS* are cadastral genes, controlling the expression boundaries of FHG's in the various flower whorls.

Dathan et al. (2002) undertook a comprehensive study to elucidate the DNA binding properties of the SUPERMAN protein. The authors wished to determine the essential amino acids involved in both DNA binding and DNA sequence specificity as well as the target sequence that the SUP protein recognizes. The DNA binding capabilities were assessed by generating recombinant SUP proteins expressed in *E. coli* cells. Purified recombinant protein was then combined with a labeled 20 bp oligonucleotide, Ep1ss (5′-AGGTTTGAC**AGT**GTCACTTT-3′), which is derived from a DNA consensus sequence that members of the C2H2 zinc-finger family from petunia are known to bind (Takatsuji et al., 1994). DNA-protein complexes could then be determined by polyacrylamide gel electrophoresis (PAGE). The results of the PAGE assay demonstrated that wild type SUP protein could strongly bind the Ep1ss oligonucleotide.

To determine the essential area(s) needed for DNA binding, a series of deletions in the *SUP* coding sequence were performed to see if the recombinant proteins could bind DNA as well as the wild type protein (Dathan et al. 2002). It was determined that deletions from amino acids 1 to 15 and 78 to 204 had no effect on the DNA binding capabilities of SUP. However deletions in the region of 15 to 78, which contains the single zinc-finger, either severely impaired or abolished the DNA binding ability of SUP. Therefore, the zinc-finger and the flanking amino acids were essential for the DNA binding properties of SUP.

Having established the critical region for DNA binding, Dathan et al. (2002) used a series of substitutions in the zinc-finger and surrounding region to determine which amino acids are critical for binding. They determined that amino acid substitutions in the zinc-finger region severely reduced the binding capacity of SUP and, within the

QALGGH motif, substitutions completely abolished DNA binding capabilities. Furthermore, the basic (arginine rich) areas flanking the zinc-finger increased affinity for DNA binding. The basic amino acids are important for the stability of the DNA-protein interaction since SUP cannot bind DNA as strongly when these amino acids are substituted with non-basic residues.

Earlier work with the DNA binding capacity of C2H2 zinc-finger proteins in petunia demonstrated that these proteins bind a core sequence of AGT within the Ep1ss oligonucleotide (5′-AGGTTTGAC**AGT**GTCACTTT-3′), (Takatsuji and Matsumoto, 1996). In order to determine if SUP also requires this core sequence for DNA binding, an alternate EP1ss labeled oligonucleotide was made with the AGT sequence changed to GAC (5′-AGGTTTGAC**GAC**GTCACTTT-3′). The ability of the SUP protein to bind the altered oligonucleotide was severely inhibited, indicating that SUP also requires the AGT core sequence for DNA-protein interaction (Dathan et al., 2002).

Although the DNA binding capability of the zinc-finger is an essential feature of SUPERMAN as a transcription factor, it appears that the leucine rich DLELRL motif comprises the "business end" of the protein. The first published report of DLELRL motif function was by Hiratsu et al. (2002), who studied the phenotypes of plants lacking the last 30 amino acids of the SUPERMAN protein. They described a series of experiments which indicated that these amino acids were important for proper flower development in Arabidopsis. First they generated a construct using the 35S promoter of the Cauliflower Mosaic Virus to constitutively express the Arabidopsis *SUP* gene lacking codons for the 30 most carboxy-terminal amino acids, designated *35S::SUPΔRD*. This construct was transformed into wild type plants. They found that the transgenic plants mimicked the

phenotype of *SUP* loss-of-function plants, namely, an increase in the number of stamens and a loss of female fertility due to the formation of stamen/carpel mosaic organs. They also compared these results to Arabidopsis over-expressing wild type *SUP*, which are dwarf, due to a lack of cell division and expansion. Plants expressing the *35S::SUPΔRD* were of normal size indicating that SUP without the 30 carboxy-terminal amino acids has lost its ability to repress cell proliferation (see below). Together, these results implicate SUP as a transcription factor with transcriptional repression abilities.

In a follow up study to their previous work, Hirastu et al. (2004) determined that the DLELRL motif was the only set of amino acid residues required for transcriptional repression by SUP. The researchers fused this motif to the carboxy-terminus of a transcription factor that binds to the *GAL4* binding site, a promoter element that induces transcription in yeast. This construct was transformed into Arabidopsis already containing a construct composed of the luciferase reporter gene driven by the *35S* promoter and five copies of the *GAL4* binding site. Transgenic plants transformed with the transcription factor lacking the DLELRL motif produced an ample amount of luciferase while plants with the DLELRL motif containing transcription factor produced five times less luciferase. These results demonstrate that, when bound to DNA, transcription factors containing the DLELRL motif, such as in SUP, are able to actively repress transcription. Lastly, the authors demonstrated through a series of substitution mutations that the repression activity of the DLELRL motif is inhibited if any of the leucine residues are replaced.

Mutant Phenotypes of *SUPERMAN* Over-Expression, Ectopic Expression, and Loss-of-Function

A variety of studies have been undertaken to study *SUP* phenotypes derived from over-expression, ectopic expression, and loss-of-function (Kater et al., 2000; Yun et al., 2002; Sakai et al., 1995; Bereterbide et al., 2001). Most of the pioneering work on the gene was done with the model organism *Arabidopsis thaliana*, as *SUP* was first identified and cloned from this organism (Sakai et al., 1995). Later on, a variety of work was performed with *Petunia hybrida*, a model organism for studying flower development and *Nicotiana tabacum*, used for its ease of genetic manipulation (Kater et al., 2000; Nakagawa et al., 2004). The first published report of *SUP* was by Schultz et al. (1991), who named the gene *Flo10* (*FLORAL MUTANT 10*). The gene was discovered through a mutant screen of seeds treated with ethyl methanesulfonate (EMS), a chemical that causes base substitutions. The outer three whorls (sepals, petals, and stamens) of the *flo10* mutant were all wild type in appearance, yet the fourth whorl contained supernumerary stamens and stamenoid carpels, reducing female fertility. Thus the mutant phenotype which is due to the expansion of the third whorl organ numbers was at the expense of the fourth whorl (carpel) organs. The *flo10* phenotype was found to be due to a single recessive allele (*flo10*) causing a loss-of-function of the gene product.

Further investigations into the supernumerary stamen phenotype by Bowman et al. (1992) showed that the increase in stamens was due to development of up to five rings of stamen primordia in the fourth whorl. Like Shultz et al. (1991), they also found that the number of stamens composing the third whorl was identical to wild type plants. Normally, the fourth whorl of Arabidopsis is composed of a single ring of organ primordia that develops into two carpels. In addition, researchers also found that in mutants, the number of extra stamens decreases up the inflorescence stalk (acropetally), from an average of

twenty on the lower flowers to an average of ten on the upper most flowers (Bowman et al., 1992).

Later, it was found that SUP plays an important role in female fertility. Gaiser et al. (1995) found through electron microscopy that *SUP* is also critical for the proper growth of ovule integument in Arabidopsis. In normal ovules, the integument grows asymmetrically such that more integument covers the ovule's abaxial side than the adaxial side. However, in *SUP* loss-of-function mutants, the integument grows equally on all sides of the ovule, resulting in severely reduced female fertility.

Nakagawa et al. (2004) reported the isolation and cloning of the petunia ortholog of *SUPERMAN* designated *PhSUP1*. By screening a transposon insertion library, researchers were able to find a plant containing a knockout of the *PhSUP1* gene. In concordance with findings by Schultz et al. (1991), the first three whorls of the *PhSUP1* loss-of-function line were phenotypically normal. The fourth whorl of wild type petunia contains two fused carpels and a single pistil, however *PhSUP1* loss-of-function mutants form extra carpels that are not fused and contain stamenoid tissue. In addition, one to three extra stamens grow out of the fourth whorl. The growth of extra stamens interior to the third whorl is consistent with the Arabidopsis *SUP* loss-of-function phenotype, indicating that another role of *SUP* is to control the number of stamens in a flower.

One recent publication by Kazama et al. (2009) has characterized the putative *SUPERMAN* ortholog from *Silene latifolia*, designated *SISUP*. Like the *SUPERMAN* genes from Arabidopsis and petunia, *SISUP* is a single copy autosomal gene. *S. latifolia* is the plant model organism for studying dioecy and the evolution of plant sex chromosomes. Sex determination in *S. latifolia* is based on a heteromorphic sex

chromosome system, whereby female plants are XX while male plants are XY. In order to characterize the orthology of the *SISUP* gene, a genomic fragment containing the *SISUP* gene was introduced into Arabidopsis *SUP* loss-of-function plants. Kazama et al. (2009) found that *SISUP* was able to complement the defective Arabidopsis *SUP* gene, such that the transgenic plants had a number of stamens comparable to the wild type (~5 per flower in the transgenic compared to six in the wild type) and also displayed normal carpel morphology.

Over-expression of *SISUP* in Arabidopsis using a *35S::SISUP* construct caused aberrations in petal and stamen development similar to results obtained by Yun et al. (2002) who ectopically expressed *SUP* under the control of the *AP1* promoter in Arabidopsis (see below). Transgenic plants with the *35S::SISUP* construct produced stamens that were fewer in number and smaller than wild type. Petals were also smaller than wild type but the transgenics maintained the same organ number and position as wild type. These results indicate that, like Arabidopsis *SUP*, *SISUP* is involved in the development of stamens and is a negative regulator of cell proliferation. Hiratsu et al. (2002) generated *35S::SUP* plants which, in addition to floral defects, were also dwarf. In contrast, the *35S::SISUP* Arabidopsis plants generated by Kazama et al. (2009) were of normal size. They postulated that this difference is due to the heterologous system, whereby the SISUP protein may not contain elements that can repress cell proliferation in vegetative tissues. Kazama et al. (2009) found that the *SISUP* gene plays a critical role in the development of female flowers in *Silene latifolia*. To determine if there was a discrepancy in expression of the putative *SUP* ortholog between male and female flowers, in situ hybridization and RT-PCR were performed. In situ hybridization of a *SISUP*

antisense probe was able to detect strong expression in female flowers but no signal was generated in the male flowers. Similarly, RT-PCR using *SISUP*-specific primers was able to amplify *SISUP* cDNA from female flower RNA extracts but not from male flowers. Taken together, these results indicate that the expression of *SISUP* is sex specific and strongly correlated with female development.

To obtain a more coherent picture of *SISUP* expression, in situ hybridization with the *SISUP* antisense probe at various stages of flower development in female flowers was performed. *Silene latifolia* flower development is divided into 12 stages, analogous to the 12 stages of Arabidopsis flower development (Smyth et al., 1990). From stages one through eight, there is no apparent difference in the development between male and female flowers in *S. latifolia*. However, after stage eight, stamen development in females is arrested, with only infertile, vestigial stamens remaining in these flowers. As with *SUP* in Arabidopsis flowers, *SISUP* expression is not detected until late in stage three where it is found in the region of cells that develop into whorls two and three. After stage eight and continuing to stage twelve, *SISUP* expression is limited to the base of the stamen primordia. The reduction in the spatial expression of *SISUP* at this period of flower development strongly correlates with the divergent development of male and female flowers after stage eight. Later in flower development at stage twelve, *SISUP* expression is found in the developing ovules. This result is analogous to the situation in both Arabidopsis and petunia where late *SUP* and *PhSUP1* expression is restricted to the ovule integuments (Gaiser et al., 1995; Nakagawa et al., 2004).

The study of sex expression in *Silene latifolia* is greatly aided by the use of a smut fungus infection system (Alexander, 1990). When infected by the smut fungus

Microbotryum violaceum, females develop normal fertile stamens and so become hermaphrodites. In combination with in situ hybridization, this system was used to study the expression pattern of *SISUP* in females in the process of reverting to hermaphrodites. Kazama et al. (2009) found that, compared to uninfected females, female plants infected with the smut fungus showed a significant decrease in *SISUP* expression over the course of floral development. This further indicates there is a strong correlation between *SISUP* expression and the suppression of stamen development. Interestingly, they found that no *SISUP* expression was detected in male flowers infected with the smut fungus. Therefore it would seem that, although *SISUP* may halt stamen development, the factor that suppresses *SISUP* expression appears to be some unknown sex determination gene(s) present on the Y chromosome (which females lack). The smut fungus somehow mimics this action of the sex determining factor(s) which in turn diminishes the expression of *SISUP*, leading to stamen development.

It appears that the overall effect of *SISUP* in sex determination is to act as a negative regulator of cell proliferation in developing stamens, thus arresting their development, resulting in female flowers (Kazama et al. 2009). If *SISUP* is truly the ortholog of *SUP*, this has important implications for the role of *SUP* in stamen development. The fact that *SISUP* is not expressed in male flowers despite the fact that they produce fertile stamens may mean that *SUP* is not critical for stamen development. Based on previous work, the critical role for *SUP* appears to the maintenance of the boundary between whorls three (stamens) and four (carpels) (Bowman et al., 1992). As there is no need to maintain this boundary in a male plant, it makes sense that no *SISUP* expression is needed in these flowers.

Phylogenetic analysis based on alignments of the *SISUP* gene with *SUP*, *PhSUP1* and *RBE* place *SISUP* in a clade with *SUP* and *PhSUP1* to the exclusion of *RABBIT EARS* (Kazama et al. 2009). This is further evidence that *SISUP* is indeed an ortholog of *SUPERMAN*. However, certain elements of the *SISUP* expression pattern mimic those of *RABBIT EARS* expression reported by Takeda et al. (2004). *RBE* is expressed in the area of whorls two and three, while it is well established that the *SUPERMAN* genes of Arabidopsis and petunia are expressed in the region between whorls three and four. Since *SUP* is not required for petal development in Arabidopsis, yet *SISUP* expression was observed in the vicinity of petal primordia it may mean that *SISUP* may play a dual role in *Silene*, acting as both the *SUPERMAN* and *RABBIT EARS* genes. Unfortunately, Kazama et al. (2009) did not determine whether *SISUP* could complement the Arabidopsis *RBE* loss-of-function phenotype.

Methylation at the *SUP* locus has also been shown to have a significant impact on the number of stamens produced in the Arabidopsis flower. Jacobsen and Meyerowitz (1997), Huang and Ma (1997) and Rohde et al. (1999) have shown that a variety of epigenetic *SUPERMAN* alleles known as "epi-mutants" exist. These alleles are characterized by hypermethylation of cytosine residues in the transcribed and coding regions of the gene. The hypermethylation is correlated with reduced *SUP* expression, such that the epi-mutants mimic the phenotype of *SUP* loss-of-function mutants, namely an increase in the number of stamens. In addition, the degree of methylation is highly correlated to the severity of the phenotype such that the more methylation an allele has, the more stamens will be produced.

A variety of work has been performed to elucidate the phenotype caused by over-expression of the *SUPERMAN* gene via constructs generated by attaching the Arabidopsis *SUP* cDNA to the Cauliflower Mosaic Virus 35S promoter (*35S::SUP*) (Bereterbide et al., 2001; 2002). The 35S promoter causes constitutive expression, meaning a large amount of gene product will be produced in all cells of the plant. Bereterbide et al. (2001) initiated the first study of *SUP* over-expression by transformation of tobacco with an Arabidopsis *SUP* cDNA driven by the 35S promoter. They found that transgenic plants were much smaller than their wild type counter parts. More importantly, flower morphology showed extreme defects in the transgenics. Transgenic flowers were significantly smaller than wild type flowers due to a reduction in the size of all floral organs, although number and positions of organs were maintained. In the third whorl, stamen filaments were reduced in length, and pollen fertility was only 30-45% of wild type.

Bereterbide et al. (2002) then used the *35S::SUP* construct to transform a cytoplasmic male sterile (CMS) tobacco line. The CMS line was the same size as wild type tobacco, but it showed severe flower aberrations such that the third whorl stamens and fourth whorl carpels were fused resulting in male sterility. As in their previous study (above), transgenic plants were dwarf, yet resulted in male fertile plants because the stamens were no longer fused to the carpels. This appears to be in contrast to their previous work (2001) that showed a reduction in male fertility. However there were some similarities, such that flowers were much smaller in the transgenic lines, and although the stamens were now free from the carpels, they were reduced in size.

As well as determining the loss-of-function phenotype, Nakagawa et al. (2004) studied the phenotype caused by *PhSUP1* over-expression. Up to this point, all the research on *SUP* over-expression was done using heterologous systems, i.e. the *SUP* gene from Arabidopsis was used in another species to elucidate its function. By introducing a *35S::PhSUP1* construct into wild type petunia, researchers found that transgenic plants exhibited a dwarf phenotype and a severe reduction (60-70%) in the size of all floral organs. Unfortunately, no report was given on male or female fertility in the transgenic petunia.

Loss-of-function and over-expression studies have not been the only two ways to characterize potential phenotypes of the *SUPERMAN* gene. Much work has been done on the ectopic expression of *SUP* in order to find smaller scale changes in phenotypes. Ectopic expression involves the fusion of a *SUP* cDNA to a promoter that drives gene expression only in a subset of cells or organs. In this way transformed plants will express *SUP* both in its wild type capacity and wherever the heterologous promoter drives expression. The choice of promoter is very important, and since wild type *SUP* is expressed exclusively in flowers, the promoters of other flower specific genes have been used to study *SUP* ectopic expression.

The first study of *SUP* ectopic expression was reported by Kater et al. (2000). They used the promoter of the *FLORAL BINDING PROTEIN 1* (*FBP1*) gene, the petunia ortholog of Arabidopsis *PISTILLATA* (*PI*) to drive the expression of Arabidopsis *SUP* cDNA. The ultimate goal was to transform wild type petunia and tobacco with the *FBP1::SUP* construct and assess the resulting phenotypes. *FBP1* is one of the B-class Floral Homeotic Genes, being expressed exclusively in the second and third whorls of the

flower. In accordance with the expression pattern of the *FBP1* promoter, only organs in the second and third whorls were affected in transgenic plants. In petunia, petals were much smaller than in wild type and they did not unfold properly, while in tobacco, petal tubes and limbs were much reduced in size. Researchers determined that the cell number in petals remained the same in both species; however elongation of the petal cells was severely affected. Stamens were also affected by the introduction of the construct. While the number of stamens remained the same in the transgenics, the filaments were much thinner and quite reduced in size. Like the petals, researchers found that the stamen aberrations were due to a reduction in cell expansion of the filaments since the transgenic filaments had the same cell number as wild type. Unfortunately, no mention was given to the male fertility of transgenics in either species.

The next report of ectopic *SUP* expression was from Yun et al. (2002) who used the promoter of Arabidopsis *APETALA1* (*AP1*) to drive the expression of Arabidopsis *SUP* cDNA. The *AP1* gene is one of the A-class Floral Homeotic Genes and is expressed in whorls one and two (sepals and petals) of developing flowers. However *AP1* is also expressed throughout the entire floral meristem early in flower development. Researchers found that introduction of the *AP1*::*SUP* construct into Arabidopsis caused ectopic expression of the *SUP* gene in all whorls of the flower and caused severe organ aberrations. In whorls two and three, there was a severe reduction in the number and size of organs, along with organ conversion in the second whorl, such that the petals were converted to sepals. Third whorl stamens were essentially absent except for occasional filamentous structures. In the fourth whorl, the number of carpels ranged from four to five, a marked increase from the wild type number of two. The increase in fourth whorl

carpels is believed to be at the expense of third whorl organs, and is in stark contrast to the *SUP* loss-of-function mutants in which the number of third whorl stamens increases at the expense of fourth whorl carpels.

The Genetic Regulation and Role of *SUPERMAN* in Wild Type Flower Development

Characterization of the SUPERMAN protein and its many mutants paved the way for a model of *SUP* regulation and the gene's role in normal flower development. Given the large body of work describing the putative function based on structural motifs of the SUPERMAN protein, it is clear that the protein is a transcription factor with the ability to repress transcription (Hiratsu, 2002; 2004). In addition to the protein structure data, the work done with over-expression, ectopic expression, and loss-of-function of *SUP* demonstrates that the gene has a profound impact on cell proliferation.

Loss-of-function of the *SUP* gene results in the formation of extra whorls, interior to whorl three that develop into stamens (Schultz et al., 1991; Bowman et al., 1992), and causes uncontrolled cell proliferation of the ovary integument in Arabidopsis (Gaiser et al., 1995). Conversely, over-expression and ectopic expression of the *SUP* gene result in a dwarf phenotype that lacks proper cell differentiation and smaller flowers with reduced or nonexistent stamens. From these data we see that the function of SUP is to be a negative regulator of cell expansion, division, and differentiation and a corresponding positive impact on these processes is seen when gene function is removed.

The actual role of *SUP* in the developing flower is clear. A variety of research has concluded that the function of the SUP protein is to prevent the expression of B-class Floral Homeotic Genes in the fourth whorl (Shultz, 1991; Bowman et al., 1992; Sakai et al., 1995). All three publications describe the *SUP* loss-of-function phenotype, namely

the development of extra stamens at the expense of fourth whorl carpels, as caused by an expansion of the expression of B-class genes (*AP3* and *PI*) into the fourth whorl. The fourth whorl normally expresses only the C-class gene *AG*, which stimulates the development of carpels, but when B- and C-class gene expression occurs in the same whorl, stamen growth is promoted. Study of the ectopic expression of *SUP* has demonstrated that increased *SUP* expression reduces the expression of B-class genes. Yun et al. (2002) found that plants expressing an *AP1::SUP* construct had severely reduced third whorl organs and enlarged fourth whorl carpels. They believed that this phenotype may be due to a reduction in B-class gene expression due to ectopic *SUP* expression. Indeed, they were able to establish that the mutant phenotype was caused by a reduction of *AP3* expression in the flower. Furthermore, they found that compared to wild type *AP3* expression, the area of whorl 3 *AP3* expression in the *AP1::SUP* mutant was constricted and that timing of *AP3* expression in flower development was delayed.

Clearly, the function of *SUP* is to create a boundary between the third and fourth whorls to restrict B-class expression. However, the mechanism by which it does this is a point of contention and has been addressed by two hypotheses. The first model proposes that either directly or indirectly, *SUP* initiates a signal at the third and fourth whorl boundary that prevents the expansion of B-class homeotic gene expression into the fourth whorl (Bowman et al., 1992; Sakai et al., 1995). It is unclear whether *SUP* prevents the expression of *AP3*, *PI* or both. As both B-class genes are needed for proper stamen development, either *AP3* or *PI* could be restricted from the fourth whorl to prevent stamen development. The second model proposes that SUP controls the boundary between whorls three and four by controlling cell proliferation between these two whorls.

According to Meyerowitz (1997) the *SUP* loss-of-function mutant phenotype is explained by continued cell proliferation in the third whorl cells, accompanied by a lack of cell division and expansion of fourth whorl cells. It is well established that SUP is a negative regulator of cell division and expansion. Sakai et al. (2000) propose that the increase in stamens in *SUP* loss-of-function mutants is not due to the fourth whorl carpels being converted to stamens but rather, the size of the third whorl (and hence the number of stamens) is expanded at the expense of the fourth whorl due to increased third whorl cell division and expansion. Therefore SUP does not prevent the expression of B-class genes in the fourth whorl but rather prevents the third whorl from expanding into and consuming the fourth whorl by controlling cell proliferation.

The action of *SUP* in maintaining appropriate levels of cell division and expansion is readily observed in work done by Bereterbide et al. (2002) in which the fertility of a cytoplasmic male sterile tobacco line was restored by over-expression of Arabidopsis *SUP*. The CMS was caused by uncontrolled growth of the third and fourth whorls such that the carpels and stamens became fused, forming male sterile stamen/carpel mosaic organs. By adding back copious amounts of SUP protein, cell proliferation was diminished and the stamens were freed from the carpel and produced functional pollen.

Over-expression of *SUP* suppresses cell expansion and division, and therefore it is easy to see *SUP* creating a boundary between whorls three and four through cell proliferation rather than controlling gene expression. It is also of interest to note that in Arabidopsis *SUP* loss-of-function mutants, *AP3* expression is not found in the center of the fourth whorl (Bowman et al., 1992). This result seems to support the latter model, as

it indicates that the very center of the flower retains its wild type fourth whorl identity. If the former model were true, it would be expected that *AP3* would be expressed in all tissue interior to the third whorl.

Regulation by *SUPERMAN* and its Mode of Action

Although *SUPERMAN* is by far the most well studied C2H2 zinc-finger gene in plants, very little is actually known about gene targets of the protein in terms of its downstream regulation. To date, only three papers have reported finding pathways that the SUP protein may regulate. The research of the first report was performed in transgenic tobacco over-expressing the Arabidopsis *SUPERMAN* gene (Bereterbide et al. 2001). Researchers found that plants expressing *SUP* under the 35S promoter (*35S::SUP*) showed a severe dwarfing phenotype due to a combinatorial lack of cell division, elongation, and differentiation. This phenotype mimicked that of mutants deficient for, or insensitive to, gibberellic acid (GA), a plant hormone that promotes growth and cell elongation. To determine if the dwarfing was due to altered GA synthesis, the transcript levels of GA-20 Oxidase were measured in *35S::SUP* plants. GA-20 Oxidase is a key enzyme in the pathway to synthesize GA1, a major plant gibberellin. They found that the transcript levels of GA-20 Oxidase were severely reduced in the *SUP* mutants. In addition, the application of exogenous GA3 was able to increase the cell size of dwarf plants 3-4 times, partially ameliorating the dwarfing. These results demonstrate that the SUP protein plays a key role in the regulation of GA synthesis.

A year later, the same team of researchers demonstrated that *SUP* plays a key role in cytoplasmic male sterility (CMS) of tobacco plants (Bereterbide et al. 2002). CMS plants, which result from incompatibility between the nuclear and mitochondrial genomes,

do not produce functional pollen and display fusion of the third and fourth whorls, which are normally free (not fused). Researchers transformed a tobacco CMS line with an Arabidopsis *35S::SUP* construct and found that transgenic plants were dwarf, as in their previous study (above). However, transgenic plants produced 20% more flowers than the CMS line and the flowers released functional pollen on stamens that were no longer fused to the carpel. These results demonstrate that *SUP* may play a critical role in signaling between the nuclear and mitochondrial genomes to maintain male fertility.

Although clearly the *SUPERMAN* gene plays a critical role in gibberellin synthesis and male sterility, it is still not clear whether SUP affects these (and possibly other) functions directly or whether it is simply the initiator of regulation. As a known transcription factor, it may likely control the expression of many genes in numerous pathways.

The most recent work into elucidating pathways that *SUP* might control was performed by Nibau et al. (2010). By over-expressing Arabidopsis *SUP* in tobacco under the 35S promoter, they were able to find strong evidence that *SUP* functions at least in part by controlling the production or signaling of both auxins and cytokinins. These two classes of molecules are important plant hormones and are involved in numerous processes including the control of cell division. Transgenic plants fell into three phenotypic classes, each of which was highly correlated with the level of *SUP* expression. Plants wild type in appearance had low levels of expression while high *SUP* expression caused a severe dwarfing phenotype which impaired proper development. Plants that were characterized by their dwarfing bushy habit as well as extensive root growth and curled leaves had an intermediate level of expression.

Nibau et al. (2010) found that with intermediate increases in ectopic *SUP* expression, plants showed a propensity for increased cell division in certain tissues. These plants had increased numbers of spongy mesophyll cell layers and xylem cell files. However, the proliferative effects of *SUP* expression appeared in only certain cell types as these plants were still dwarf and displayed various floral defects. However, after a certain level of *SUP* expression the positive effects on cell division are negated. Transformed lines with the highest levels of *SUP* transcript were severely dwarfed and did not survive into the adult phase or outside of tissue culture conditions.

The effects of *SUP* on cell division prompted the researchers to study how *SUP* expression might affect hormonal signals in the plants. This was done by transforming transgenic plants with the intermediate phenotype with two reporter gene constructs. The first introduced construct was *DR5*::*GUS*. The *DR5* promoter induces expression in the presence of auxin. It was found that in the doubly transgenic plants that *DR5* activated transcription of *GUS* in these plants and the amount of *GUS* activity increased with the severity of the *SUP* phenotype. No *GUS* expression was observed in wild type plants transformed with *DR5*::*GUS*. In their next experiment, transgenic plants with the intermediate *SUP* phenotype were transformed with the *ARR5*::*GUS* construct. The *ARR5* promoter induces expression in the presence of cytokinin. The doubly transgenic plants had *GUS* expression while wild type transformed with *ARR5*::*GUS* did not. These results indicate that *SUP* expression either increases endogenous levels of auxin and cytokinin or increases the sensitivity of these hormone signals.

To further understand the effects that *SUP* expression had on controlling cell division, the authors transformed the *35S*::*SUP* plants with a *CYCB1*::*GUS* reporter

construct. The *CYCB1* promoter is only activated by cell division. Researchers noted that the doubly transformed plants showed *GUS* expression and that increasing amounts of expression were correlated with the severity of the SUP phenotype. No *GUS* expression was observed in wild type plants transformed with *CYCB1*::*GUS*.

Finally, the authors addressed the affects of *SUP* over-expression on flower morphology. The observed floral phenotypes of the *35S*::*SUP* intermediate plants were similar to those seen by Bereterbide et al. (2001). In addition, Nibau et al. (2010) observed that stamens frequently had stigmatic tissue on anthers and flowers often developed more than the two wild type carpels.

Taken together, the evidence presented by Nibau et al. (2010) suggests that *SUP* plays a key role auxin and cytokinin signaling in tobacco. It is clear that *SUP* regulates either the levels of cytokinin and auxin or the sensitivity to which the plant perceives these hormone signals. They also found a strong correlation between this hormone signaling and cell proliferation. It appears that increased cytokinin and auxin signal inhibits cell division in most tissues as evidenced by the dwarf habits of the intermediate and extreme *SUP* over-expression phenotypes, and inhibition of stamen development. However, increased *SUP* expression appeared to stimulate cell proliferation in some tissues and cell types such as mesophyll cells, xylem cells, and the increased number of carpels in intermediate plants. The authors make the argument that the differing sensitivities of various cell types to the levels of SUP may control the boundary between whorls three and four by controlling differential cell division in the area between these two whorls.

Where and When is *SUPERMAN* Expressed?

Based on in situ hybridization of an antisense *SUP* probe, Sakai et al. (1995) were able to discern the expression pattern of *SUP* in wild type Arabidopsis plants. The earliest *SUP* expression is found at the boundary between whorls three and four. This occurs right after the earliest expression of the B-class genes in the whorl three organ primordia corresponding to between stage three and four of Arabidopsis flower development (Smyth et al. 1990). At stage eight of flower development, *SUP* transcripts are found at the base of the stamens and by stage nine SUP is detectable in the immature ovules. After stage eight, *SUP* expression in stamens and between the third and fourth whorl quickly diminishes to an unobservable level.

Nakagawa et al. (2004) studied the expression patterns of the petunia ortholog of *SUP*, *PhSUP*, using both in situ hybridization of an antisense *SUP* probe, and a *PhSUP*::*GUS* reporter gene. In situ hybridization showed *PhSUP* expression in the developing stamens and ovaries yet no mention of expression between whorls three and four was given. *GUS* expression was observed in developing ovaries and in the boundary between whorls three and four, consistent with the finding of Sakai et al. (1995).

Nakagawa et al. (2004) also found *PhSUP* expression in the boundary between whorls two and three in developing petunia organ primordia based on in situ hybridization. These results are in disagreement to reported *SUP* expression data by Sakai et al. (1995) which finds *SUP* only expressed at the boundary between whorls three and four, and later on in developing stamens and ovaries. A possible explanation for this is that the *SUP* antisense probe was hybridizing to *RBE* transcripts in the second and third whorl boundary due to their high degree of sequence homology. The authors may have been unfamiliar with the *RBE* gene as its sequence and expression were only published

several months prior their publication. However, Kazama et al. (2009) found the putative *Silene latifolia SUPERMAN* ortholog, *SISUP*, to be expressed in the region of whorls two and three. Therefore it may be that in some species, in addition to expression in the boundary between whorls three and four, *SUP* is also expressed in the boundary between whorls two and three.

Regulation of *SUPERMAN* Expression

A multitude of evidence has shown that the *SUPERMAN* gene plays a critical role in the proper development of angiosperm flowers. It is also clear that unregulated SUP expression has drastic consequences on both flower morphology and fertility. A variety of research has been conducted to determine which genes regulate the proper expression of SUP in a developing flower. The most common method used to study the regulation of SUP expression is to measure the level of expression both spatially and temporally in a variety of genetic backgrounds. These genetic backgrounds are mutants of other important floral genes such as the Floral Homeotic Genes whose expression has been altered.

Although *SUP* appears to be a negative regulator of *AP3*, it appears that *AP3* is a positive regulator of both spatial and temporal *SUP* expression. In Arabidopsis, *SUP* RNA is only detected in the developing flower after the initial expression of *AP3*. While this may simply be a coincidence, it is known that *SUP* expression between the third and fourth whorls in an *AP3* loss-of-function background is significantly reduced, but not eliminated (Sakai et al. 1995; 2000). The fact that the earliest *SUP* expression is not completely eliminated in these backgrounds indicates that some other factor is responsible for inducing the initial *SUP* expression. However, Sakai et al. (2000) found

that in either *AP3* or *PI* loss-of-function mutants, late *SUP* expression is still maintained at normal levels in the developing ovules, indicating that this facet of *SUP* expression is not dependent on B-class genes.

Furthermore, positive regulation of *SUP* is also seen when *AP3* is constitutively expressed by the 35S promoter. *35S::AP3* plants form extra whorls of stamens interior to the normal third whorl stamens. Ectopic *SUP* expression is clearly observed, as *SUP* RNA is detectable in these new stamen primordia (Sakai et al., 1995).

Additional evidence that B-class genes (*AP3* and *PI*) are required for normal *SUP* expression was elucidated by Bowman et al. (1992). A series of Arabidopsis loss-of-function double mutants with *SUP*, including *AP3* and *PI* was created. In *AP3* and *PI* loss-of-function single mutants, loss of B-class gene function causes the conversion of third whorl stamens to carpels. The researchers found that both the *AP3/SUP* and *PI/SUP* double mutants showed the same phenotype as the *AP3* and *PI* single mutants, indicating that *SUP* is epistatic to B-class gene function and that B-class genes are needed for proper SUP expression.

Sakai et al. (2000) make the claim that *AP3* and *PI* are not epistatic to *SUP* as previously reported. They found that in *AP3/SUP* and *PI/SUP* double mutants, bulges of tissue form around the gynoecium that resemble extra immature stamen whorls which are not seen in the *AP3* or *PI* single mutants. This indicates that the negative affects of cell proliferation by *SUP* are still present in *AP3* or *PI* single mutants. To state is another way, the positive effects of cell proliferation by *SUP* loss-of-function remain in *AP3/SUP* and *PI/SUP* double mutants.

Findings by Bowman et al. (1992) show that removal of *AGAMOUS* (*AG*) and *SUP* expression has a largely additive effect. *AG* is a negative regulator of cell growth in the developing floral meristem and fourth whorl such that its loss-of-function leads to a larger floral meristem with indeterminate growth (increased number of whorls). *SUPERMAN* is also a known negative regulator of growth such that its loss-of-function leads to the formation of extra stamenoid whorls at the expense of the fourth whorl. Bowman et al. (1992) found that Arabidopsis *AG/SUP* loss-of-function double mutants had an even larger floral meristem which led to an increase in indeterminate growth. Furthermore, in the *AG/SUP* double mutant, ectopic sepals, which are normally found in the *AG* loss-of-function single mutant, are replaced with petals. This is explained by the *SUP* loss-of-function aspect of the double mutant. Since *SUP* expression is missing, B-class expression pervades in the flower which in combination with pervasive A-class gene expression (a consequence of *AG* loss-of-function) converts the organ primordia to petals.

Sakai et al. (2000) demonstrated that *AG* is indeed a positive regulator of *SUP* expression. They found that *AG* loss-of-function in Arabidopsis has no effects on the initial expression of *SUP* but that expression quickly diminishes after stage five of flower development when whorl three and four primordia begin to separate. This demonstrates that *AG* is needed for the continued maintenance of *SUP* expression throughout floral development. Furthermore, they found that even with this diminished window of *SUP* expression, *AP3* expression was still restricted from the fourth whorl demonstrating that the continued maintenance of *SUP* expression is not required define the boundaries of the third and fourth whorls.

It also appears that the A-class Floral Homeotic Gene *APETALA2* (*AP2*) plays a role in the regulation of *SUPERMAN* expression (Bowman et al., 1992). Arabidopsis *AP2* loss-of-function mutants display a severe reduction in the number of third whorl stamens and stamens are completely missing in some allelic variants. It has been well established that the loss-of-function of *SUPERMAN* causes a great increase in the number of stamens interior to the third whorl (10-20 stamens); therefore it was of interest to Bowman et al. (1992) to assess the phenotype of plants harboring *AP2/SUP* loss of function. They found that in the double mutants, flowers had an intermediate number of stamens between the two single mutant phenotypes. The novel phenotype was also associated with the severity of the loss-of-function allele of the *AP2* gene. In *AP2-2/SUP* double mutants (*AP2-2* is a strong loss-of-function allele), an average of three stamens were observed and these were often stamen/carpel mosaic organs. In the *AP2-1/SUP* double mutants (*AP2-1* is a less severe loss-of-function allele), four to nine morphologically normal stamens are observed. These results demonstrate that *SUP* and *AP2* have an antagonistic relationship and that the genes may regulate one another to control proper flower development.

RABBIT EARS

Compared to *SUPERMAN*, relatively little is known about *RABBIT EARS* (*RBE*), the second whorl counterpart of *SUPERMAN*. Like *SUP*, *RBE* is a transcription factor that acts as a cadastral gene, defining the boundary between whorls two and three in Arabidopsis flowers (Takeda et al., 2004). *RBE* is expressed early in flower development in the petal primordia of the second whorl and appears to be essential for proper petal development, as plants homozygous for moderate loss-of-function alleles show

conversion of petals to stamen filaments. In severe loss-of-function alleles, no second whorl organ primordia develop.

Krizek et al. (2006) reported the molecular mechanism of *RBE* in proper flower development. They found that the role of *RBE* is to prevent the expression of *AGAMOUS* (*AG*) in the second whorl. *AGAMOUS*, a known suppressor of growth in the second whorl (Bowman et al., 1991) is expressed in the whorl 2 of *RBE* loss-of-function flowers, resulting in a suppression of second whorl growth. This mechanism is further evidenced by the fact that *RBE/AG* double loss-of-function mutants produce normal petals in the second whorl. Therefore it would appear that unlike the *SUPERMAN* gene, *RABBIT EARS* is a positive regulator of growth in the flower.

Takeda et al. (2004) generated lines of *35S::RBE* Arabidopsis and found that transgenic plants had no mutant phenotype. Since SUP and RBE both act to restrict Floral Homeotic Genes from particular whorls (albeit by different methods) it would be expected that ectopic or over-expression of these genes would have similar effects. When *SUP* (a negative regulator of B-class genes) is ectopically expressed (Yun et al. 2002), B-class gene expression is removed from the third whorl, stopping proper stamen development. Under this model it might be expected that ectopic or over-expression of *RBE* (a negative regulator of the C-class gene) would remove *AG* expression from the third whorl, also stopping proper stamen development. Under this model, the A-class genes, which are negatively regulated by *AG* would move into the third whorl, replacing wild type stamens with petals (Figure 2).

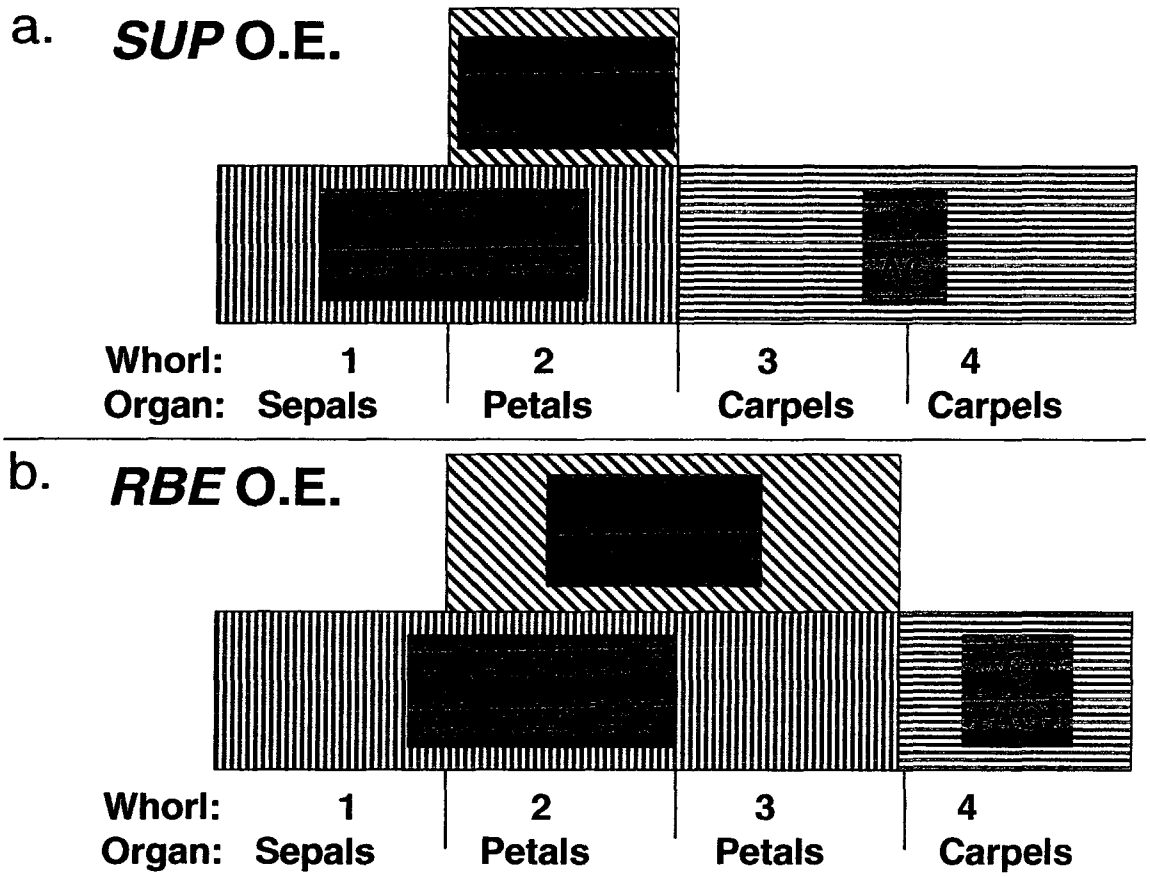


Figure 2. Phenotypic consequences of *SUP* and *RBE* over-expression (O.E.). a. With extra SUPERMAN gene product, AP3 and PI are removed from Whorl 3. This results in fewer stamens, undeveloped stamens or stamen/carpel mosaic organs in Whorl 3. **b.** With extra RABBIT EARS gene product, AG is removed from Whorl 3. This allows AP1 and AP2 to occupy Whorl 3, converting the organs to petals.

Conclusions

As evidenced by the diverse and important roles that *SUPERMAN* and *RABBIT EARS* play, it is clear that the genetic control of flower development is not as simple as the ABC model would lead one to believe. Flower development requires precise spatial and temporal control of the ABC genes. One of the most important aspects of this fine level of control is to make sure that expression of particular FHG's is confined to certain whorls. Creating FHG expression boundaries between whorls is essential to proper flower development and is the job of cadastral genes, *SUPERMAN* and *RABBIT EARS*, to create and maintain these barriers.

Asserting the hypothesis that femaleness in strawberry is due to a dominant allele that causes a gain of function, the logical conclusion is to search for candidate genes that promote female flower development (at the expense male fertility) through a gain of function rather than a loss of function. A cadastral gene, specifically the *SUPERMAN* gene, makes an excellent candidate for sex determination in *Fragaria virginiana*. The first and most obvious reason for *SUPERMAN* being a stellar candidate is that gain of function mutations (both ectopic and over-expression) appear to mimic the phenotype of female *F. virginiana* flowers. The most prominent distinguishing feature of a female *F. virginiana* flower is the presence of vestigial stamens. Numerous reports have demonstrated that a severe reduction in the size of stamens occurs when the *SUPERMAN* gene is ectopically or over-expressed (Kater et al., 2000; Yun et al., 2002; Bereterbide et al., 2001). Second, it is known that while the size of stamens in females is quite reduced, the number and position of vestigial stamens remains the same as in hermaphrodite plants (see description of strawberry flower phenotype below). In the case of *SUPERMAN*

expression, it appears that under some circumstances of over-expression of *SUP*, stamen size is reduced but the overall number and spatial organization of these organs remains wild type (Bereterbide et al., 2001). Finally, in strawberry, female flowers are noticeably smaller than their hermaphrodite counterparts. This phenotype is observed in petunia flowers which over express the *PhSUP1* gene. Flowers on these plants have an average reduction in size of 60-70% (Nakagawa et al., 2004). Taken together, the negative affects that *SUPERMAN* (and possibly other *SUPERMAN*-like genes) have on cell proliferation, which are in turn manifested by reduced organ size makes them intriguing candidates for the cause of sexual dimorphism in *Fragaria virginiana.*

Ethylene and Sex Determination: The *ACS-7* Gene

So far the candidate genes discussed have been transcription factors. However, these types of genes by no means have exclusive control over flower development. It has been known for some time that the small gaseous molecule ethylene is also a potent phytohormone. Although it is present in small quantities as an atmospheric air pollutant, plants produce their own endogenous ethylene through enzyme-catalyzed biosynthesis. During the 1950's it became clear that one role of ethylene as a plant hormone was to promote female flower development in cucumber. Application of exogenous ethylene was able to suppress the development of stamens and so promote female development. Therefore, the enzymes involved in both the regulation (transcription factors) and biosynthesis (enzymes) of ethylene play a direct role in the suppression of stamen development.

The genus *Cucumis* (melon and cucumber) represents a staggering variety of sex morphs, with four reproductive strategies represented in the various species. Within the

genus, species may be monoecious (separate male and female flowers on the same plant), andromonecious (male and hermaphrodites on the same plant), gynoecious (only female flowers) and hermaphrodite (perfect flowers). Female flowers develop stamen primordia but their development is arrested early in flower development. All four sex types can exist in *C. melo* (melon), with andromonecy under the control of the "A" locus.

The "A" locus of *C. melo* has been cloned and characterized (Boualem et al., 2008). Based on sequence alignment, it was determined that this locus contained an *ACS* gene with a high degree of homology to the Arabidopsis *ACS-7* gene. Thus the *Cucumis melo* gene was designated *CmACS-7*. *ACS* codes for the enzyme 1-aminocyclopropane-1-carboxylic acid synthase which catalyzes the penultimate and rate limiting step in ethylene biosynthesis.

Sequencing of *CmACS-7* alleles in both monoecious and andromonecious lines revealed a polymorphism in the coding region that constituted a nonsynonymous mutation. Monoecious lines contained an allele with alanine at position 57 while the andromonecious lines were homozygous for the allele with a valine at position 57. The alanine is conserved across seed plants and is known to be part of the enzyme's active site. To determine the enzyme's functional capability, the enzyme activity of both isoforms was assayed using the in vivo substrate concentration of *S*-Adenosyl methionine (SAM). It was found that the A57 allele had 50% maximal activity while the V57 allele had no detectable activity. Therefore, in vivo, the active A57 allele is dominant to the inactive V57 allele.

The *CmACS-7* activity assay provided direct evidence that differing ethylene production may affect sex. To further support this conclusion, *the CmACS-7* genes from

496 *C. melo* accessions of differing sex type were sequenced. It was found that all monoecious and gynoecious plants (without stamens) (n = 149) contained the dominant A57 allele while the andromonoecious and hermaphrodite plants (with stamens) (n = 347) were homozygous for the recessive V57 allele. The model suggested by this data is that plants able to make female flowers (monoecious and gynoecious) have the active A57 CmACS-7 allele while all plants that cannot make female flowers (andromonoecious or hermaphrodite) lack the active allele. From the work presented by Boualem et al. (2008) it is clear that in *Cucumis melo*, ethylene is either required for, or the direct cause of suppression of stamen development.

The fact that the *ACS-7* gene is highly correlated with the suppression of stamen development in melon makes it an intriguing candidate gene for sex determination in *Fragaria virginiana,* especially as it appears to prevent stamen development in a dominant fashion. Based on sequence homology to the *CmACS-7* and *AtACS-7* genes, the putative *Fragaria ACS-7* gene has been isolated and characterized, as described in this study. Although not investigated during this present study, it will be of interest in future work to determine the *ACS-7* allelic composition at the equivalent "57" amino acid position in different sex morphs of *Fragaria virginiana.*

CHAPTER 1.

METHODS AND MATERIALS USED THROUGHOUT THIS STUDY

Abstract

A variety of methods and materials, including strawberry species and their respective populations, were used extensively throughout this study. Therefore, a thorough discussion on this subject is warranted. The two strawberry species used in this research were the octoploid *Fragaria virginiana* and the diploid *Fragaria vesca*. Both species were an invaluable asset to this study, serving as a source of genetic material for gene isolation and characterization. The techniques and protocols herein were used throughout this study and will be presented in the General Methods section below. More specific methods sections in subsequent chapters will often refer to this General Methods section for information on a procedure. In addition, electronic resources, such as easily searchable genome databases that were used in this study, are also described.

Introduction

Plant Species and Populations

Fragaria virginiana. *Fragaria virginiana* is characterized as a gynodioecious species with plants producing either hermaphrodite flowers with fertile stamens and carpels or female flowers with only fertile carpels. It has been proposed that males are also present within this species. However this statement is quite subjective as males simply are defined as having lower female fertility than hermaphroditic plants (Ashman, 1999). It is

of importance to note that females do actually produce stamens in the same form and number as hermaphrodites. However, these stamens are vestigial and do not produce pollen. Female flowers are also noticeably smaller in size (personal observation). Therefore it would appear that there is some genetic factor underpinning the development of stamens in females that arrests their complete development. While arrested development of stamens is the most prominent feature of female flowers, it does not mean that the female flowers are actually missing stamens. Ashman (1999) reports that both female and hermaphrodite flowers have an average of 20-30 stamens.

One other characteristic of female flowers is their propensity to develop petalloid stamens in the third whorl. These organs appear to be stamens that have become flat and white at the top of the organ, a position normally occupied by the vestigial anthers. Hermaphroditic flowers also produce petalloid stamens; however it is a much rarer occurrence.

To study the genetic basis of sex determination in *F. virginiana*, a cross was performed between a female and a hermaphrodite accession. The resulting F1 population comprised both females and hermaphrodites, and thus was segregating for sex. A comparison of the flower morphology between the female and hermaphrodite members of the L1L2 segregating population (discussed below) can be seen in Figure 3.

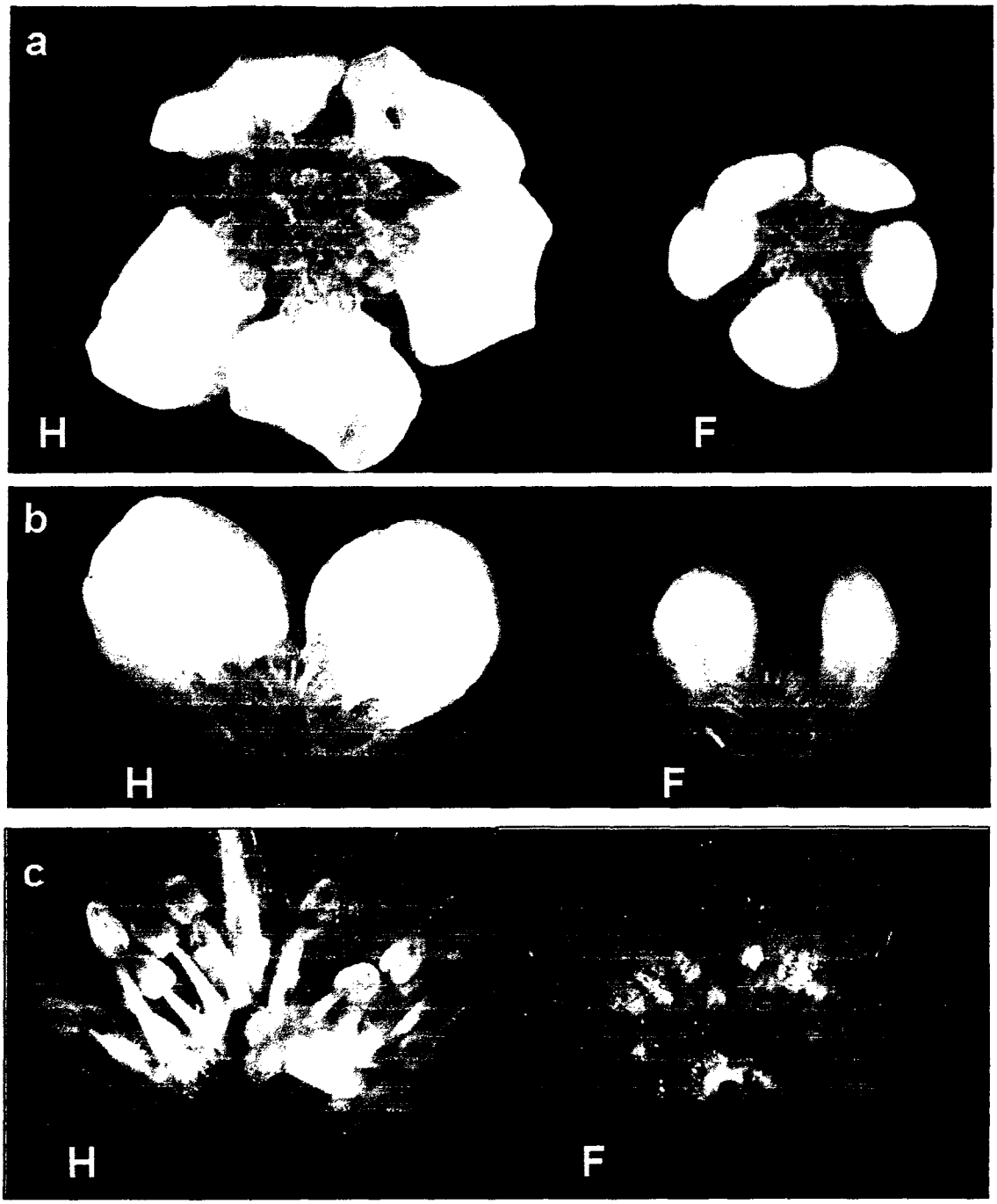


Figure 3. Morphology of hermaphrodite and female *F. virginiana* flowers. **a.** Top view of fully opened hermaphrodite (H) and female (F) flowers. **b.** Lateral view of the flowers as dissected through the center. **c.** Mature hermaphrodite and female flowers with receptacle removed. Note the difference between H and F in the size of the flowers (**a** and **b**), and of the stamens (**c**).

Fragaria vesca. *F. vesca* was used throughout this study to map and isolate candidate genes for subsequent transformation studies. *F. vesca* (2N =2X = 14), known colloquially as the Alpine or woodland strawberry, is the most geographically diverse of any strawberry

species, growing naturally in Asia, Europe, North America, and South America. Due to its extensive range, many subspecies of *F. vesca* exist, as do many varieties which have been bred for food or ornamental purposes.

F. vesca has become the cornerstone species for functional genetics and genomics in the strawberry community, as the species lends itself to this task for a number of reasons. First and arguably most important is that *F. vesca* is a diploid, making it genetically much less complex to work with as compared to the polyploid species. Secondly, it is self fertile which allows the development of inbred homozygous lines for mapping and transformation. Finally, the Alpine strawberry is easily propagated both vegetatively and by tissue culture. This makes it amenable to genetic transformation and an efficient *Agrobacterium*-mediated transformation system has been established using *F. vesca* variety Hawaii (Oosumi et al., 2006). 'Hawaii' was chosen as the variety for transformation due to its high rate of callus formation and regeneration.

In this study, 'Hawaii' was used to isolate the candidate genes for genetic transformation into Arabidopsis. 'Hawaii' was chosen for two reasons. First, the candidate genes will eventually be transformed back into this variety to deduce their function in vivo. Second, it is this variety that was chosen to be sequenced by the Virginia Bioinformatics Institute (discussed blow).

The mapping portion of this study also utilized *F. vesca*. The mapping population used in this study designated "YP" is an F2 derived from a cross between two accessions of *F. vesca*: 'Yellow Wonder' (YW) and 'Pawtuckaway' (PAWT). 'Yellow Wonder' is a variety of Alpine strawberry, *F. vesca* subsp. *vesca* semperflorens that has long been bred and cultivated throughout Europe. It is unique compared to the wild *F. vesca* in that it carries three easily recognizable recessive traits. First and most obvious is that, unlike most cultivated strawberries, the fruit of 'Yellow Wonder' is yellow. This is due to a mutation in the anthocyanin biosynthetic pathway, which is responsible for the familiar red color of strawberry fruit. The second trait that distinguishes 'Yellow Wonder' is that it does not produce runners. The third distinguishing feature is that this variety is ever bearing (day neutral), meaning that its ability to flower is not dependent upon day length. 'Pawtuckaway' is a wild accession of *F. vesca* subsp. *americana* collected at Pawtuckaway State Park, Nottingham, New Hampshire. 'Pawtuckaway' is typical of wild members of the North American subspecies as it produces bright red fruit, can reproduce vegetatively through vigorous runners, and has a facultative short day flowering cycle.

The main purpose of the YP F2 is to serve as a mapping population for various markers and genes of interest. The parents of this cross were chosen to maximize the amount of heterozygosity in the F1. 'Yellow Wonder' and 'Pawtuckaway' are ideal for mapping because both have accumulated many polymorphic regions within their genomes due to their widely disparate genetic backgrounds. 'Yellow Wonder', having been cultivated for a long period of time, is highly inbred (homozygous), and can only be propagated by seed (or crown divisions). On the other hand, 'Pawtuckaway' is a member of a highly outbreeding population and is presumed to be very heterozygous. Under these

circumstances, both parents have accumulated their own distinct set of polymorphisms that can be easily distinguished for mapping. The F2 YP population consists of 113 plants; however, only 72 are routinely used for mapping studies, providing a recombination resolution of 0.7cM.

An intraspecific cross between two distantly related accessions was seen as the best way to construct a mapping population. While interspecific crosses between diploid *Fragaria* species are possible, the use of these crosses for mapping may pose hidden challenges. Two different species of *Fragaria* are separated by much more evolutionary distance than two subspecies of the same species. Two different species may have acquired a variety of chromosome aberrations such as inversions, deletions, rearrangements, and translocations. This would create a loss of synteny along chromosomes which would hamper the ability to map certain areas of the genome. A population such as YP still has a wide array of genetic variation for mapping but also maintains chromosomal synteny.

Sequence Acquisition and Molecular Methods

The acquisition of novel gene sequence is often greatly aided by the use of previously isolated sequence, and this study is no exception in that regard. Given the overwhelming amount of nucleotide and protein sequence available in the age of genomics, scientists have come to rely heavily on well organized and easily searchable databases. In this study, two databases were readily taken advantage of: the GenBank database maintained by the National Center for Biotechnology Information (NCBI) and the *F. vesca* genome database maintained by Virginia Bioinformatics Institute (VBI).

Both of these sites have BLAST algorithms, allowing sequence homology to be readily assessed between any given gene and the database (Altschul et al., 1990).

While databases such as those maintained by NCBI and VBI allow heterologous sequences to be easily found, in order to actually isolate sequences of interest from *F. virginiana* and *F. vesca*, a number of molecular methods must be employed. Throughout this study are presented a variety of techniques tailored to a certain need. However, nearly all of the basic isolation and characterization of candidate genes and marker sequences required two widely used molecular methods: PCR and cloning. Presented below are the general methods used extensively in this study for the isolation and cloning of DNA sequence, and subsequent chapters will often make reference to these.

Materials and Methods

Databases

Two online databases served as invaluable resources during this study. The first is the National Center for Biotechnology Information (NCBI) GenBank database. This public database allows access to hundreds of thousands of deposited and annotated sequences. Furthermore, BLAST searches using both nucleotide and protein sequence can be performed against the database. This resource was used extensively to find heterologous sequence for primer design to isolate candidate genes. In addition, a total of five *Fragaria* candidate gene sequences isolated during this study have been deposited in this database.

The second database used in this study is maintained by the Virginia Bioinformatics Institute and houses the *F. vesca* variety Hawaii genome. This database

contains roughly 95% of the *F. vesca* genome in the form of contigs ranging in size from hundreds of thousands to several million base pairs (Shulaev, unpublished). The database also utilizes a BLAST algorithm that can align nucleotide (but not protein) sequence against the contig sequence.

General Methods

Standard PCR Protocol Using Invitrogen HiFi *Taq*: Invitrogen AccuPrime high fidelity (HiFi) *Taq* DNA polymerase (Invitrogen, Carlsbad, CA) was exclusively used for applications requiring extremely accurate DNA replication such as sequencing, restriction enzyme digestion of PCR products for mapping, and amplification of genes for transformation (discussed below). AccuPrime HiFi *Taq* is a proofreading polymerase with 3′ to 5′ exonuclease activity, offering a nine fold increase in fidelity over standard *Taq* polymerases. In addition to its proofreading abilities, HiFi *Taq* is also a hot-start polymerase due to the anti-*Taq* antibody in the buffer which prevents the polymerase from working until it is heated to 94°C.

Reaction Parameters: 2.5 µl of 10X buffer, 0.5 µl each of forward and reverse primers (20 µM stock), 0.1 µl (0.5 U) HiFi *Taq*, and 5-100 ng of genomic or plasmid template were combined for each reaction. Reactions were brought to 25 µl with deionized sterile water.

Thermocycler Parameters to generate products for sequencing and mapping: The thermocycler profile used for AccuPrime HiFi *Taq* was: initial denaturation at 94°C for 1 min; then 30 cycles of 94°C for 30 sec, 52°C for 30 sec, 68°C for 30 seconds for every 500 bp to be amplified; followed by a final extension at 68°C for 10 minutes. All primers, unless otherwise noted, were designed for optimal annealing at 52°C.

Standard PCR Protocol Using Lucigen EconoTaq: Lucigen EconoTaq (Lucigen, Middleton, WI) is a standard *Taq* DNA polymerase that lacks proofreading ability. As it lacks proofreading ability, EconoTaq was only used for amplifications that do not require accurate DNA synthesis such as colony screening and AFLP markers, etc.

Reaction Parameters: 2.5 µl of 10X buffer, 1 µl of 10 mM dNTP mix, 0.5 µl each of forward and reverse primers (20 µM stock), 0.2 µl (1U) EconoTaq, and 5-100 ng of genomic or plasmid template were combined for each reaction. Reactions were brought to 25 µl with sterile deionized water.

Thermocycler Parameters: The thermocycler profile used for Lucigen EconoTaq was: initial denaturation at 94°C for 1 min; then 30 cycles of 94°C for 30 sec, 50-60°C for 30 sec, 72°C for 30 seconds for every 500 bp to be amplified; followed by a final extension at 72°C for 10 minutes.

Standard Protocol for Cloning and Sequencing of PCR Products Using Invitrogen TOPO TA Cloning Kit: The Invitrogen TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) was used exclusively for the cloning of PCR products for sequencing.

Ligation: 1 µl of fresh PCR product amplified with Invitrogen AccuPrime high fidelity *Taq* was added to a reaction of 3.8 µl sterile water, 1 µl supplied salt solution and 0.2 µl (2 ng) of TOPO vector. Reactions were incubated at room temperature in the dark for 15 minutes.

Transformation: 2 µl of the ligation reaction was added to 25 µl of TOP10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA). The cells were incubated for 10 minutes on ice then transferred to a water bath at 42°C for 30 seconds and returned to ice. 125 µl of kit-supplied SOC medium was added to each tube of cells and the cells were

shaken at 250 RPM at 37°C for an hour. Cells were plated on LB agar (10 g/L tryptone, 10 g/L NaCl, 5 g/L hydrolyzed yeast, 15 g/L agar, pH 7.0) containing 50 or 100 μg/mL ampicillin and grown at 37°C for 12-16 hours.

Colony Screen: Individual clones were screened for inserts via PCR. Using half reactions of the standard Lucigen EconoTaq PCR, isolated colonies were picked from the plate and used directly as template for amplification. The thermocycler profile used for colony PCR was as follows: initial denaturation at 94°C for 1 min; then 30 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 30 seconds for every 500 bp to be amplified; followed by a final extension at 72°C for 10 minutes. In all colony screens either the kit-supplied T3/T7 or M13F/M13R vector primers (all at 20 μM) were used to amplify inserts.

Plasmid Isolation: Individual colonies containing the desired insert were grown in 10 mL Falcon tubes (Falcon Plastics, Brookings, SD) containing 2 mL of LB liquid media containing 100 μg/mL ampicillin. Tubes were shaken at 250 RPM at 37°C for 14-16 hours. The 2 ml culture was centrifuged at 10,000 RPM to pellet the cells. Plasmids were isolated from the cells using the Promega Wizard *Plus* SV Minipreps (Promega, Madison, WI) DNA Purification System per the manufacturer's instructions. Isolated plasmids were resuspended in 100 μl sterile deionized water and quantified on the NanoDrop (Thermo Scientific) spectrophotometer.

Plasmid Sequencing: 250-300 ng of isolated plasmid was added to a pre-sequencing (template and primer only) cocktail containing the plasmid, 1 μl (20 μM stock) of either M13F or M13R primer and sterile water to a 6 μl volume. The pre-sequencing cocktail was then used to perform a sequencing reaction, with the products separated on an Applied Biosystems 3130 Genetic Analyzer.

Results

Fragaria virginiana

To study the genetic basis of sex in *Fragaria*, an intraspecific cross was performed between two *F. virginiana* subsp. *virginiana* (2N=8X=56) with differing sex morphs. The female parent was designated L1 while the male parent was a hermaphrodite designated L2. Both parents were collected from the wild in Lincoln, New Hampshire by T.M. Davis. The F1 population derived from this cross was designated L1L2 and segregates for sex. Of the 57 members of the F1, 25 are female and 32 are hermaphrodite. Both the parents and the F1 progeny were maintained at the University of New Hampshire greenhouse facility.

CHAPTER 2.

ISOLATION AND CHARACTERIZATION OF CANDIDATE GENES

Abstract

A genetic study based on a candidate gene approach requires the isolation and characterization of orthologs from the species being studied. Unfortunately at the inception of this project, only the strawberry ortholog of *AGAMOUS* was publicly available. This meant that in order to progress any further, the *ACS-7* and *SUPERMAN*-like gene family members would first have to be isolated from a *Fragaria* species. To add further difficulty to the prospect of isolating these orthologous genes, the sequencing of the *F. vesca* genome was still some years away when this project was begun. In order to tackle this seemingly monumental task, a variety of available methods were employed to isolate these genes from strawberry. In addition, a novel method for paralog mining and sequence walking known as TVL-PCR was developed for this purpose. Despite the success of these gene isolation methods, the *F. vesca* genome sequence was still required to capture some members of the *SUPERMAN*-like gene family and the complete coding sequence of the strawberry *ACS-7* gene. A total of five *Fragaria SUPERMAN*-like genes have been isolated and characterized from *Fragaria virginiana*. As of now, the *F. vesca* genome is over 95% sequenced and has proved to be an invaluable resource not be taken for granted. Described below is the story of how the seven candidate genes were isolated from the strawberry genome and the subsequent characterization of these genes.

Introduction

Isolation and Characterization of *AGAMOUS* and *ACS-7*

Two papers have been published on the *F.* ×*ananassa* ortholog of the *AGAMOUS* gene (Aharoni et al., 1999; Rosin et al., 2003). This gene designated *STAG1* is available as a full length cDNA in GenBank as accession AF168468. This should make isolation of the full length (exon and intron) *AGAMOUS* ortholog from other *Fragaria* species quite easy. However, since the *STAG1* is a cDNA, the exact position of any introns are unknown. This means the intron size and positions must be either inferred from alignment with heterologous genomic sequence or by sequencing of a genomic *Fragaria AGAMOUS* clone.

A key word search on the NCBI GenBank Nucleotide database yielded no returns for any members of the *Fragaria ACS* gene family. However, this simply meant that no strawberry *ACS* genes were annotated. At least one member of the gene family may have been deposited as a "sequence dump" whereby large amounts of sequence, usually cDNAs are submitted to the database without a great deal of annotation in regards to homology or putative function. BLAST searches using *ACS* orthologs from other species yielded invaluable information about the *Fragaria ACS-7*, as discussed below.

Isolation and Characterization of *Fragaria SUPERMAN*-like genes

No such luck was had when searching for *Fragaria SUPERMAN*-like sequences on the GenBank database. This is not surprising since the gene sequences were most likely to be found as un-annotated cDNAs similar to the *Fragaria ACS* gene fragments. Given that *SUPERMAN*-like genes appear to be expressed at low levels, at least in floral tissues, they should be found very rarely in a cDNA library. As no relevant strawberry

sequence could be obtained from GenBank, a variety of PCR based strategies were employed to isolate *Fragaria SUPERMAN*-like genes. Later on however, another database, the Virginia Bioinformatics Institute's *F. vesca* reference genome database would yield invaluable information on this family of genes in strawberry.

The initial isolation of the *SUPERMAN* gene family in *Fragaria* relied on acquisition of heterologous sequences, i.e. *SUPERMAN* and *SUPERMAN*-like gene sequences from other organisms. This effort was broken down into two phases: isolation of strawberry *SUPERMAN*-like gene fragments, followed by isolation of the gene's entire coding region. At the project outset, sequences from the *SUPERMAN*-like gene family were available only in Arabidopsis, tobacco, petunia, rice, cotton, and apple. As described herein this heterologous sequence information was exploited to design degenerate primers targeted to two conserved motifs, the zinc-finger and the transcriptional repression domain (described above). In all, degenerate primer PCR was used to isolate two strawberry *SUPERMAN*-like gene fragments from L2 designated *FRASUP4* and *FRASUP5*. The sequence could then be used to isolate flanking sequences of these two new genes via one of many chromosome walking methods.

Various PCR-based sequence walking techniques have been developed for acquiring previously unknown genomic sequence flanking a known site. Such approaches universally employ one or more primers targeted to "internal" sites – i.e., sites within the known genomic sequence segment. These internal primers may be sequence-specific or degenerate, and are designed on the basis of known sequence information. With the inverse PCR technique, two internal primers are paired in outward orientation to amplify from circularized genomic template molecules (Ochman et al., 1993). Alternately, an

internal genomic primer may be paired with one or more "external" primers. Such external primers may be targeted specifically to ligated (Smith, 1992; Novak and Novak, 1997; Yuanxin et al., 2003) or otherwise appended (Cormack and Somssich, 1997) linkers (a fragment of known DNA ligated to the genomic template) or sequence extensions. Alternately, the external primers may be intended to anneal non-specifically to flanking genomic sequences, as in Thermal Asymmetric Interlaced PCR (TAIL-PCR) (Liu et al., 1995). Initially, TAIL-PCR was used for the *SUPERMAN*-like chromosome walk as it had been previously used in our lab with success (Vining, 2007).

High (Mullins et al., 2001) and low (White and Chen, 2006) success rates have been reported in studies using TAIL-PCR to recover sequence from regions flanking T-DNA insertions. We utilized TAIL-PCR to successfully extend one strawberry *SUPERMAN*-like gene fragment designated *FRASUP4* from *Fragaria virginiana* accession L2 in the 5′ and 3′ directions, eventually acquiring complete coding and extensive upstream sequence. However, TAIL-PCR could not acquire any flanking sequence of the second *SUPERMAN*-like gene fragment and was unable to obtain sequence from any additional *SUPERMAN*-like gene family members in the genome. Therefore, the challenges encountered during our subsequent efforts to extend these initial sequence segments in both the 5′ and 3′ directions, and to isolate additional *SUPERMAN* paralogs prompted the development of the novel sequence walking technique described herein known as TOPO vector-ligation PCR (TVL-PCR) (Orcheski and Davis, 2010). As an alternative approach, the technique of TVL-PCR has been successfully applied as described in detail below.

For TVL-PCR, the pCR4-TOPO vector was employed as the linker. In brief, total genomic DNA from the species of interest is subjected to restriction digestion, followed by end-repair to produce a pool of fragments with single, 3′ adenosine overhangs at each end. These fragments are shotgun ligated to pCR4-TOPO vector molecules (Figure 4A), creating chimeric molecules containing known priming sites at both ends: the vector on one end and known genomic sequence on the other (Figure 4B). The ligation product pool is used as template for the initial round of TVL-PCR (Figure 4C), and the product is then used as template in the fully- or semi-nested second round of TVL-PCR (Figure 4D). TVL-PCR generates a product or products that span a segment of known sequence, an adjacent segment of unknown sequence, and a short vector segment.

With regards to the *SUPERMAN*-like family, TVL-PCR was used to isolate the remaining 5′ and 3′ coding sequence from *FRASUP5* gene fragment. In addition, use of degenerate primers enabled the isolation of a novel strawberry *SUPERMAN*-like gene designated *FRASUP3* from L2.

While the isolation of strawberry *SUPERMAN*-like genes was actively being pursued, the Virginia Bioinformatics Institute was making significant inroads in the sequencing of the *F. vesca* var. Hawaii genome. By the time the three *FRASUP* genes had been isolated, a large portion of the genome had been assembled into contigs of a few thousand base pairs. More importantly, a BLAST algorithm was established to align nucleotide sequence against the contigs. It was now possible to BLAST the three previously isolated strawberry *SUPERMAN*-like genes against this database to determine if any had escaped the amplification by degenerate PCR or TVL-PCR.

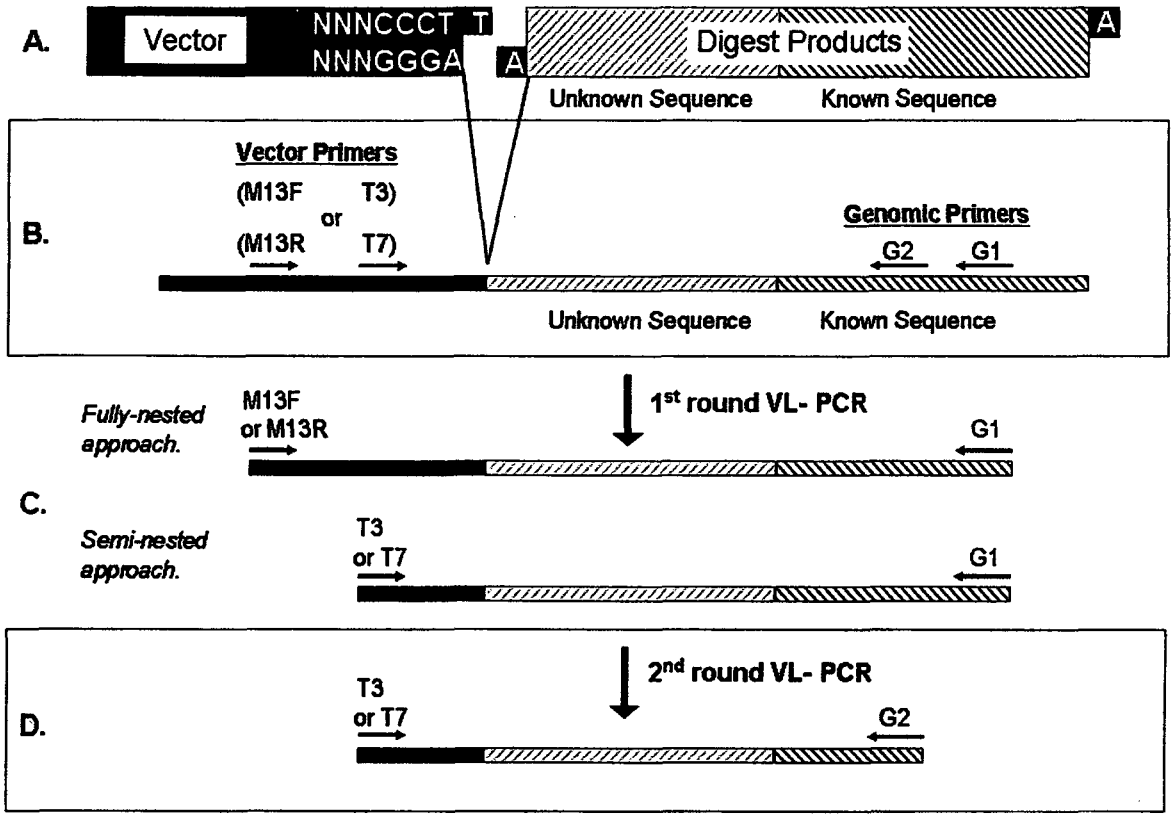


Figure 4. Overview of TVL-PCR. **A.** A pool of chimeric templates is generated using a ligation reaction that joins end-repaired genomic restriction fragments to TOPO vectors. Ligation is required only between one end of each genomic fragment and a vector molecule. **B.** A chimeric template molecule contains two priming sites (G1 and G2) within the genomic fragment and two appropriately oriented priming sites (M13F and T3, or M13R and T7) within the conjoined vector. **C.** In the first round of TVL-PCR, the G1 genomic primer is paired with a vector primer. In the fully nested approach an M13_ primer is used. Because the orientation of ligation is unknown, the G1 primer must be paired in separate reactions with M13F and with M13R. In the semi-nested approach, the G1 primer is paired with a T_ primer. **D.** In the second round of TVL-PCR, the G2 genomic primer is paired with either the T3 or the T7 primer, as appropriate. The T3 primer must be used if either the M13F or T3 primer was previously used in the first round of TVL-PCR, while the T7 primer must be used if either the M13R or T7 primer was previously used.

The isolation of the five *FRASUP* genes allowed their subsequent characterization. It was of great interest to this study to determine if all members contained the necessary sequence elements to belong to the *SUPERMAN* family. All *SUPERMAN*-like genes contain some variant of the highly conserved "QALGGH" motif of the zinc-finger and the "DLELRL" motif of the transcriptional repression domain. Therefore these two sequences are diagnostic for this family. Unfortunately there is little sequence conservation outside of these regions, making an assessment of orthology between the Arabidopsis *SUPERMAN* and *RABBIT EARS* genes and the *Fragaria SUPERMAN*-like genes quite difficult. In order to determine possible orthology, phylogenetic trees were created using the complete coding region of the genes. In total, alignments of both the nucleotide and protein sequences from the five *Fragaria SUPERMAN*-like genes, the Arabidopsis *SUPERMAN* and *RABBIT EARS* genes, and the *Petunia SUPERMAN* gene were used to generate a phylogeny.

Materials and Methods

Germplasm

All DNA used in the isolation of *Fragaria SUPERMAN*-like genes is genomic DNA isolated from *F. virginiana* accession L2. Total genomic DNA from *F. vesca* variety Yellow Wonder was used in the isolation of the *ACS-7* complete coding sequence. The protocol for DNA isolation was identical to that of Davis et al. (1995a) except that chloroform:octanol solution was not included in the microfuge tube to which the initial CTAB slurry was transferred.

Initial Sequence Acquisition and Primer Design

At the inception of this project, no *SUPERMAN*-like sequence information existed for *Fragaria* in publicly available databases. *SUPERMAN* and *SUPERMAN*-like nucleotide sequences from Arabidopsis and other species were obtained from NCBI GenBank via keyword and BLAST searches, and aligned using the Lasergene MegAlign program (DNAStar, Madison, WI). Two conserved sites, the zinc-finger and transcriptional repression domains, were identified as targets for designing degenerate PCR primers (Table 1) intended to amplify *SUPERMAN*-like sequence fragments from strawberry. The code for the degenerate bases can be found in Appendix B. Once an initial set of strawberry *SUPERMAN*-like sequence segments was obtained using this conventional, degenerate-primer PCR approach, the acquired strawberry sequences were added into the previous sequence alignment, providing the basis for identification of both conserved and non-conserved regions as target sites for genomic primers needed for TAIL-PCR and TVL-PCR.

Degenerate Primer PCR

A semi-nested degenerate PCR strategy was employed for the initial isolation of *SUPERMAN*-like genes. Two different primer combinations were tested to amplify the *SUPERMAN*-like genes. The first combination employed one round with just the QALGGH F1. Products from this reaction were then used as template in a second round using both LGGHMN F2 and LDLELR R1 primers. With the second combination, LDLELR R1 was used alone for the first round of PCR. The products were then used as template with just the QALGGH F1 primer. Template from this second reaction was used for the final round which employed both the LGGHMN F2 and LDLELR R1 primers.

Table 1. Initial gene isolation primers.

Primer Name	Primer Sequence		Melting (Tm)	Temp. (°C)
Degenerate PCR Primers		**Min.**	**Avg.**	**Max.**
QALGGH F1	5′-CARGCIYTIGGIGGICAY-3′	61.7°C	63.2°C	67.9°C
LGGHMN F2	5′-YTIGGIGGICAYATGAAY-3′	53.8°C	55.3°C	59.0°C
LDLELR R1	5′-ARICKIARYTCIARRTC-3′	44.3°C	50.5°C	59.8°C
Degenerate TAIL-PCR Primers				
AD1	5′-NGTC□ASWGANAWGAA-3′	42.0°C	45.6°C	48.5°C
AD2	5′-TGWGNAGSANCASAGA-3′	44.2°C	49.4°C	54.6°C
AD3	5′-AGWGNAGWANCAWAGG-3′	38.9°C	44.2°C	49.6°C
AD6	5′-WGTGNAGWANCANAGA-3′	37.9°C	44.6°C	50.9°C
Gene Specific TAIL-PCR Primers				
FRASUP4 5′ F1	5′-CAGAAGAGACAGAGCTAGACTC-3′		53.8°C	
FRASUP4 5′ F2	5′-ACTAGTGGTACTCCTAAACTAGC-3′		53.1°C	
FRASUP4 5′ F3	5′-CGCCTTAGAAATCATAGCCATG-3′		53.7°C	
FRASUP4 3′ F1	5′-GCCATAAGTCCACTTACCTCT-3′		53.9°C	
FRASUP4 3′ F2	5′-GCTGAGTTCATAGCAGAAGC-3′		53.5°C	
FRASUP4 3′ F3	5′-GTGCTTGCCTCTTCCC-3□		53.4°C	
FRASUP5 5′ F1	5′-GATTGACCTGGACGGAGG-3′		55.0°C	
FRASUP5 5′ F2	5′-GGCGATGAGAGAGGATTCG-3′		55.1°C	
FRASUP5 5′ F3	5′-GTGCCCAGAGTCGATGAG-3′		55.3°C	
FRASUP5 3′ F1	5′-CTCATCGACTCTGGGCAC-3′		55.3°C	
FRASUP5 3′ F2	5′-CGAATCCTCTCTCATCGC-3′		52.4°C	
FRASUP5 3′ F3	5′-CCTCCGTCCAGGTCAATC-3′		55.0°C	

The temperature profile used for all reactions of degenerate PCR was as follows: initial denaturation at 94°C for 1 min; 35 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 1 min; followed by a final extension at 72°C for 10 minutes. The 25 µl first-round PCR contained 100 ng genomic DNA as template, 0.4 µM of both the forward and reverse primer or 0.8 µM of a single primer, 2.5 mM $MgCl_2$, 1 U Taq DNA polymerase (Invitrogen), and 2.5 µl 10X PCR Buffer lacking Mg.

Second and third rounds of PCR were performed with 1 µl of previous round product as the template with the final primer concentrations remaining constant. The products of each round (10 µl loaded) were visualized on a 2% agarose 1X TBE gel run at 100 volts for 80 minutes. The final PCR products were cloned and sequenced using the TOPO TA cloning kit and colony screen protocols, as described in Chapter 1.

TAIL-PCR

Twelve sequence specific primers were designed to capture both the 5′ and 3′ ends of the two previously isolated strawberry *SUPERMAN*-like genes, *FRASUP4* and *FRASUP5* (Table 1). These primers were in turn paired with four degenerate primers to perform Thermal Asymmetric Interlaced PCR (TAIL-PCR, Liu et al., 1995; Liu and Whittier, 1995) using *F. virginiana* accession L2 as template. The thermal cycling parameters of the TAIL-PCR protocol can be found in Table 2. All twelve specific primers were tested with M13 vector primers using their respective *SUPERMAN*-like TOPO clone as template to determine that they would amplify.

TAIL-PCR was performed using Accuprime High Fidelity Taq DNA Polymerase (Invitrogen, Carlsbad, CA). For the first round of TAIL-PCR, a 25 µl reaction composed of 2.5 µl 10X Accuprime Buffer II, 3 µM degenerate primer, 0.4 µM first round specific

Table 2. Thermocycler reaction parameters used for TAIL-PCR (Liu et al., 1995).

Reaction	Thermal Paramaters	Cycles
1st Round	93°C, 1 min; 95°C, 1 min	1
	94°C, 30 sec; 62°C, 1 min; 72°C, 2.5 min	5
	94°C, 30 sec; 25°C, 3 min; Ramp 0.3°C/sec to 72°C; 72°C, 2.5 min	1
	94°C, 30 sec; 68°C, 1 min; 72°C, 2.5 min; 94°C, 30 sec; 68°C, 1 min; 72°C, 2.5 min; 94°C, 30 sec; 44°C, 1 min; 72°C, 2.5 min;	15
	72°C, 5min, hold @ 4°C	1
2nd Round	94°C, 30 sec; 64°C, 1 min; 72°C, 2.5 min; 94°C, 30 sec; 64°C, 1 min; 72°C, 2.5 min; 94°C, 30 sec; 44°C, 1 min; 72°C, 2.5 min;	12
	72°C, 5 min, hold @ 4°C	1
3rd Round	94°C, 30 sec; 44°C, 1 min; 72°C, 2.5 min;	40
	72°C, 5 min, hold @ 4°C	1

primer, 0.5 U Taq Polymerase, and 100 ng of template DNA was made for each combination of degenerate primer and first round specific primer. Products of the first round of TAIL-PCR were diluted 1:50 with deionized sterile water and 2.5 µl were used as template for the second round. The second round reactions were the same as the first except that 1.8 µM of the degenerate primer were used and 0.4 µM of the second round specific primer were used. Next, 2.5 µl of the second round products were used directly as template for the third round of TAIL-PCR. These reaction parameters were the same as the second round except that the third round specific primer was used in conjunction with the same degenerate primer.

After amplification, 10 µl of the third round TAIL-PCR products were then separated on a 2% agarose 1X TBE gel at 100 V for 80 minutes. Total PCR products from the following round three primer combinations were cloned using the TOPO TA cloning kit: FRASUP4-5′ F3/AD3, FRASUP4-3′ F3/AD2, FRASUP4-3′ F3/AD3, FRASUP4-3′ F3/AD6 and all combinations of FRASUP5-5′ F3 and FRASUP5-3′ F3.

A colony screen was performed and plasmids with inserts greater that 500 bp were isolated with the Wizard SV Plasmid Purification kit (Promega Madison, WI). Plasmids were sequenced bidirectionally using the M13 forward and reverse primer on an Applied Biosystems 3130 Genetic Analyzer. Sequence reads were assembled into a contig (length permitting) and aligned against the original *FRASUP* clone using the Lasergene Megalign software (DNASTAR, Madison, WI) to determine if any new gene sequence had been attained.

Template Preparation for TOPO Vector-Ligation PCR

Digestion of Strawberry Genomic DNA. For each reaction, 400 ng of genomic DNA was digested with 20 U of *Eco*RI, *Bam*HI, or *Hin*dIII (New England Biolabs, Ipswich, Massachusetts) in a 40 µl reaction. Enzymes used for TVL-PCR must produce recessed 3′ ends (5′ overhangs) or blunt ends. Digestions were performed overnight at 37°C. Digestion was verified by electrophoresis of 100 ng digested and undigested genomic DNA run side by side for 80 minutes at 100V on a 1% agarose 1X TBE gel.

Repair of Fragment Ends and Addition of 3′ A Overhang. End-repair was performed using 20 µl (200 ng) of each digested DNA sample with 1 µl of 10 mM dNTP mix, 1.3 µl sterile water, 2.5 µl of 10X EconoTaq buffer (Lucigen, Middleton, Wisconsin) and 0.2 µl (1 U) EconoTaq DNA polymerase. The reaction was incubated at 72°C for 30 minutes to fill in recessed 3′ ends of restriction enzyme cut sites, add a 3′ adenosine overhang and inactivate the restriction enzyme that was used to digest the genomic DNA.

Ligation. End-repaired DNA was then ligated to the pCR4-TOPO vector (Invitrogen) using 4.5 µl (36 ng) end-repaired DNA solution, by adding 0.5 µl (5 ng) of TOPO vector and 1 µl of supplied salt solution. The reaction was gently mixed and incubated in the dark at room temperature for one hour to ensure complete ligation.

TVL-PCR

The temperature profile used for both the first and second rounds of TVL-PCR was as follows: initial denaturation at 94°C for 1 min; 35 cycles of 94°C for 30 sec, 52°C for 30 sec, 68°C for 4 min; followed by a final extension at 68°C for 10 minutes. For each ligated template, two first-round amplification reactions were employed for the initial step of TVL-PCR, because the genomic fragment can ligate to the vector in alternate orientations (Figure 5). In one first-round TVL-PCR reaction (Figure 4C),

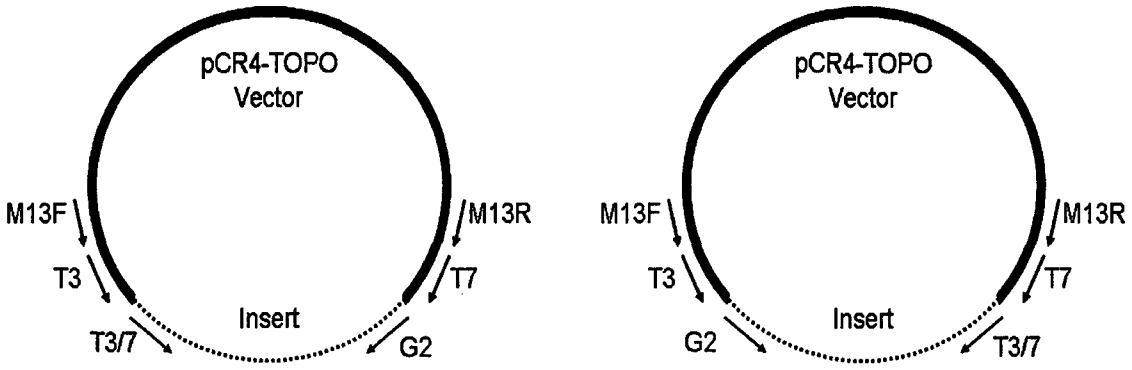


Figure 5. Recombinant vector with primer sites and alternate insert orientations. For sequencing, the products generated by TVL-PCR must be cloned into the TOPO vector. Two possible insert orientations are possible, as shown. M13F or M13R primers must be used for sequencing because they provide unique priming sites in the recombinant plasmid. The T3 or T7 primers are unsuitable for sequencing because they are represented in both the vector and the insert.

the distal genomic primer (G1) was paired with the M13F or T3 vector primer, while in the other the G1 primer was paired with the M13R or T7 vector primer. In the first round, use of a distal M13F or M13R vector primer allows for subsequent vector primer nesting, while use of a proximal T3 or T7 vector primer precludes subsequent vector primer nesting. For the second-round of TVL-PCR (Figure 4D), each first-round product was used as a template, and the nested genomic primer (G2) was paired with the appropriate T3 or T7 vector primer. The sequences of all the G1 and G2 primers used for TVL-PCR can be found in Table 3.

The 25 µl first-round TVL-PCR contained 3 µl ligation reaction template, 0.4 µM G1 genomic primer, 0.4 µM vector primer, 0.5 U AccuPrime *Taq* DNA Polymerase High Fidelity (Invitrogen), and 2.5 µl 10X AccuPrime PCR buffer II. A 10 µl portion of the first-round product was visualized on a 2% agarose 1X TBE gel run at 110 volts for 80 minutes. The second-round of TVL-PCR was performed with 1 µl of first-round product as the template, and 0.4 µM of both the nested G2 genomic primer and the appropriate vector primer. If the first-round vector primer was M13F or T3, then T3 was used as the second-round vector primer. If the first-round vector primer was M13R or T7, then T7 was used as the second-round vector primer.

Cloning and Sequencing of TVL-PCR Products

Total PCR products from each second round TVL-PCR were cloned using the Invitrogen TOPO TA Cloning kit for sequencing. Transformed cells were plated on LB agar plates containing 50 µg/ml ampicillin and were grown at 37°C overnight. Colony PCR was performed using the M13 primers provided with the TOPO cloning kit to confirm the insert size. Plasmids were isolated using the Promega Wizard *Plus* SV

Table 3. TVL-PCR gene isolation primers.

Primer Name	Primer Sequence	Melting Temp. (°C)		
Paralog Mining TVL-PCR Primers		**Min.**	**Avg.**	**Max.**
C2H2 F1	5′-ATGAATGTYCATAGRAGAGAYAG-3′	47.6°C	50.6°C	53.7°C
C2H2 F2	5′-YAGRAGAGAYAGRGCHAKRCT-3′	46.1°C	54.7°C	63.5°C
FRA SUP F1	5′-ATGAATGTYCACAGRAGAGAC-3′	50.1°C	52.1°C	54.1°C
FRA SUP F2	5′-GRAGAGACAGAGCTAGACTCA-3′	52.9°C	53.9°C	54.9°C
TVL-PCR Gene Specific Primers				
FRASUP3 5′ F1 (G1)	5′-AGGGTTAGGTTTAAGGTTGAG-3′		51.8°C	
FRASUP3 5′ F2 (G2)	5′-GTGAGGGTACGAAAGTAGG-3′		51.9°C	
FRASUP3 3′ F1 (G1)	5′-TCGCAAGTTGAACTATGTATCC-3′		52.6°C	
FRASUP3 3′ F2 (G2)	5′-CAGTTTGTTCAGGACTGAGT-3′		52.2°C	
ACS-7 5′ F1 (G1)	5′-CTCTGCAAGTCCCATCTG-3′		52.7 °C	
ACS-7 5′ F2 (G2)	5′-CATCGTAAGGGTTCTCATCG-3′		52.7 °C	
ACS-7 3′ F1 (G1)	5′-GAAAGTGAGAGGGGTCC-3′		51.9 °C	
ACS-7 3′ F2 (G2)	5′-CGAATCCATCCAATCCATTAGG-3′		53.3 °C	
TOPO Vector Primers				
M13F	5′-GTAAAACGACGGCCAG-3′		50.7°C	
M13R	5′-CAGGAAACAGCTATGAC-3′		47.0°C	
T3	5′-ATTAACCCTCACTAAAGGGA-3′		50.3°C	
T7	5′-TAATACGACTCACTATAGGG-3′		47.5°C	

Minipreps DNA Purification System. The plasmid inserts were sequenced bidirectionally on an Applied Biosystems 3130 Genetic Analyzer, using M13 forward and M13 reverse sequencing primers. The T3 and T7 sequencing primers cannot be used for this purpose, because their priming sites may be represented twice in the final TOPO clone: once in the cloning vector itself and once in the cloned product of second round TVL-PCR (Figure 5).

Fragaria SUPERMAN-like Alignments and Gene Phylogeny

After isolation of complete coding sequence from all five *FRASUP* genes, an alignment of the gene translations was performed using the Lasergene MegAlign program (DNAStar, Madison, WI). The alignment was done to determine if the proteins contained all of the necessary motifs to be considered *SUPERMAN*-like genes: a single zinc-finger and a transcriptional repression domain.

Nucleotide and protein alignments and their respective phylogenies were constructed using the software program MEGA (version 4.1). Nucleotide sequences of the eight genes from the start to stop codons were imported into the program. The sequences were translated and both the nucleotide and protein sequences were aligned using the ClustalW algorithm. After alignment, phylogenies of both nucleotide and protein sequences were constructed using both the Neighbor Joining (NJ) and Unweighted Pair Group Method with Arithmetic Mean (UPGMA) methods. Default program settings were used and all trees generated underwent 1000 bootstrap replicates.

Results

Isolation and Characterization of the *AGAMOUS* Gene from *Fragaria*

As confirmation of the gene identity, a BLAST search of the *STAG1* translation was performed against the Arabidopsis protein database. The best *Arabidopsis thaliana* hit was indeed the *AGAMOUS* gene (Locus: AT4G18960).

To determine the putative intron positions in the *STAG1* gene, the *STAG1* translation was aligned with the Arabidopsis AG protein. Intron positions in the Arabidopsis *AG* gene were designated as putative intron positions in the *STAG1* gene. A pair of primers (AG F: 5′-TGAAAACAAACCAAACACTGACCT-3′, AG R: 5′-TTCTTGTATCGTTCAATCGT-3′) was designed from the *STAG1* cDNA to span the large ~3kb intron found in the Arabidopsis *AG* gene. The product, amplified from *F. virginiana* L2, and roughly 3kb in size was cloned using the standard protocols for Invitrogen HiFi Taq and TOPO cloning. By sequencing a single clone bidirectionally using M13 vector primers, it was determined that the large *STAG1* intron had been amplified. This primer pair would be used later on to map the *Fragaria AGAMOUS* gene in the YP population (discussed below in Chapter 3).

Isolation and Characterization of the *ACS-7* Gene from *Fragaria*

A BLAST was performed with the homologous melon (*Cucumis*) *CmACS-7* gene (accession: EU791280) against the GenBank *Fragaria* EST-others and Nucleotide Collections using the tBLASTx algorithm. Two hits with low E-values were made to the Nucleotide Collection and three hits with low E-values were made to the EST-others Collection. The best BLAST hits from both databases were then used in a BLAST search against the *Arabidopsis thaliana* Nucleotide Collection database with the tBLASTx algorithm. The best *Fragaria* Nucleotide hit was to *AtACS-5* while the best *Fragaria* EST hit was to *AtACS-7*. Therefore, the *Fragaria* EST (Accession DY672246) hit was taken

to be the putative *Fragaria ACS-7* gene. This EST did not contain the complete coding sequence, and therefore a chromosome walk using TVL-PCR with *Fragaria ACS-7* gene specific primers (Table 1), was performed in order to capture the full coding sequence. Using *F. vesca* variety Yellow Wonder as template, TVL-PCR was able to capture the 3′ end of the gene with primers ACS-7 3′ F1, and ACS-7 3′ F2 but was unable to amplify the 5′ end with primers ACS-7 5′ F1, and ACS-7 5′ F2. However, eventually the *Fragaria ACS-7* gene fragment could be used in a BLAST against the Virginia Bioinformatics Institute's *F. vesca* reference genome database to obtain the putative full length sequence on a contig. This putative full length sequence was translated in all three reading frames and aligned with the *AtACS-7* and *CmACS-7* translations to find the putative start and stop codons. The putative *Fragaria* start codon was found easily due to the high degree of homology in this area between *Fragaria* and *Cucumis* (MAIEID for *Cucumis* and MAIEIE for *Fragaria*). The presumptive carboxy-terminal end of the *Fragaria ACS-7* gene appeared to have very little homology with the end of *CmACS-7*. However, there was significant homology between the *AtACS-7* and *Fragaria ACS-7* carboxy-terminal ends (FMDRRRRF* for Arabidopsis and FMEQRERRH* for *Fragaria*).

The *Fragaria* EST was the best BLAST hit when *CmACS-7* was used in a BLAST against the *Fragaria* database. However, this does not necessarily mean that the strawberry EST is the *ACS-7* ortholog. The *ACS* gene family is large with twelve members in the Arabidopsis genome (Yamagami et al., 2003). As there are so few *Fragaria ACS* gene sequences available on GenBank, the best *Fragaria* BLAST hit using the *CmACS-7* gene will not necessarily hit the *CmACS-7* gene first when it is used in a BLAST against the *Cucumis* database. To determine if the now-complete coding

sequence *Fragaria* EST was in fact the best match for the *CmACS-7* gene, the *CmACS-7* sequence was used in a BLAST against the VBI *F. vesca* genome database. The contig containing the best hit for *CmACS-7* was the same contig that the complete *Fragaria* EST was on. Based on these results it is probable that the *Fragaria* EST is the actual *ACS-7* gene.

Isolation and Characterization of the *FRASUP* Genes from *Fragaria*

Degenerate Primer PCR. In order to isolate *SUPERMAN*-like homologs from strawberry using PCR-based methods, heterologous reference sequences were used as a basis for primer design. Alignment of the translations of publicly available *SUPERMAN* sequences from Arabidopsis (NM_113214), petunia (AB117749), and tobacco (BE559491), and *SUPERMAN*-like sequences from Arabidopsis (the *RABBIT EARS* gene - NM_120689) and apple (EB113939) identified two conserved regions for design of degenerate primers. A conserved area found in the downstream portion of the zinc-finger region corresponding to the amino acid sequence QALGGHMN was used as the forward priming site. The C-terminal end of the protein contains a conserved area known as the transcriptional repression domain (Hiratsu et al., 2004), corresponding to the amino acid sequence LDLELR, and this was used as the reverse priming site. A semi-nested set of primers (Table 1) was designed consisting of two forward primers (QALGGH F1 and LGGHMN F2) in the QALGGHMN domain and a single reverse primer (LDLELR R1) in the LDLELR domain. In order to reduce primer degeneracy, inosine was incorporated at four-fold degenerate sites. Accounting for the variability in sequence length between these two conserved sites in the reference sequences, and given the absence of a known

intron in this region in any of the reference sequences, the strawberry PCR product was predicted to be approximately 450 bp to 600 bp.

Separation of L2 amplification products by agarose gel electrophoresis, yielded two bands within the predicted size range and a third band that was slightly larger (660 bp). Cloning and sequencing of these PCR products provided three *SUPERMAN*-like sequences that could be used as a basis for primer design for TVL-PCR. One clone turned out to be a false positive because although first 30 bases after the end of the primer sequence (comprising the zinc-finger) had homology to the *SUPERMAN* gene family, beyond that point there was no similarity between the clone sequence and the gene family. A BLAST to the VBI *F. vesca* genome showed that the majority of this clone had strong homology to a single contig. However, the zinc-finger region of the clone had no homology to the contig. Furthermore, a tBLASTx with the clone against the NCBI Viridiplantae database yielded no significant hits outside of the short zinc-finger region. However, at the time it was still believed to be a *SUPERMAN*-like gene (although possibly a pseudogene due to a high number of stop codons) so it was included in the alignment for further primer design. The two positive clones isolated from degenerate PCR were designated *FRASUP4* and *FRASUP5*.

TAIL-PCR for *FRASUP4* and *FRASUP5*

Cloning and sequencing of the products from TAIL-PCR using *FRASUP4* specific primers yielded four positive clones. One 900 bp product from the FRASUP4-5′ F3/AD3 primer combination was the upstream portion of the *FRASUP4* gene. Primer combinations FRASUP4-3′ F3/AD2 and FRASUP4-3′ F3/AD3 yielded 1.6kb fragments

of the downstream region of *FRASUP4*. The final combination FRASUP4-3′ F3/AD6 resulted in a 1.1kb fragment of the downstream region of *FRASUP4*.

Cloning and sequencing of third round TAIL-PCR products using the FRASUP5-5′ F3 and FRASUP5-3′ F3 primer combinations did not yield any desirable upstream or downstream fragments of the gene.

TVL-PCR for *FRASUP3*

Paralog Mining. To isolate more potential strawberry *SUPERMAN*-like genes, two sets of degenerate primers were designed based on the three previously isolated strawberry *SUPERMAN*-like gene fragments, *FRASUP4*, *FRASUP5* and the *SUPERMAN*-like "pseudogene". They were designed with the intent to amplify fragments of the 3′ end of novel *SUPERMAN*-like genes in the genetic background of L2. The primers, designed to anneal to the MNVHRRDRARL motif of the *SUPERMAN* zinc-finger (Table 1), were used in various combinations for TVL-PCR. The first round of TVL-PCR used primer FRASUPF1 as the G1 primer in combination with either the T3 or T7 vector primer, and produced a complex spectrum of bands (Figure 6A). The second round of TVL-PCR used primer C2H2 F2 as the nested G2 primer, in combination with the appropriate T_ primer, and resulted in a much-reduced spectrum of bands (Figure 6B). The second-round TVL-PCR products were shotgun cloned and colony PCR was used to identify clones with inserts of the targeted size range. A total of four clones of three different sizes were sequenced as indicated by the boxed gel bands in Figure 6B. The box with the star specifies clone *FRASUP3*, a novel *Fragaria SUPERMAN*-like sequence.

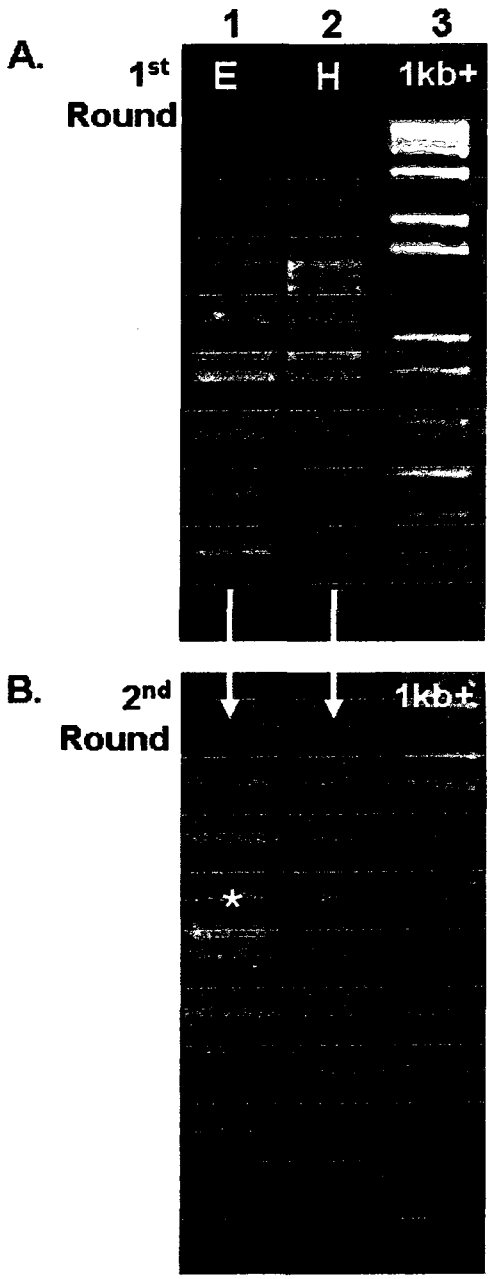


Figure 6. Isolation of *FRASUP3* via TVL-PCR paralog mining. **A.** First round TVL-PCR products generated from templates prepared using *Eco*RI (E) or *Hin*dIII (H) digestion of genomic DNA. **B.** Second round TVL-PCR products generated using the corresponding first round products as template. Inserts corresponding to the boxed product sizes in the second round were sequenced. Based on size, the box with a star is the *FRASUP3* fragment.

In another reaction, the first round of TVL-PCR used primer C2H2 F1 as the G1 primer in combination with either T3 or T7 vector primer. Template was *F. virginiana* L2 genomic DNA digested with either *Hin*dIII, *Eco*RI, or *Bam*HI. No amplification products could be visualized after the first round by agarose gel electrophoresis (data not shown). However, products from the second round of TVL-PCR using C2H2 F2 as the G1 primer and the appropriate T_ primer resulted in small number of, easily distinguishable bands (Figure 7). Total PCR products from lanes three, four, and five were cloned and the inserts corresponding to the sizes of the boxed fragments (Figure 7) were sequenced. The two boxed products in lane three both represent sequences at the 3′ end of the *FRASUP4* fragment. The differences in product length of the two *FRASUP4* fragments may be due to the amplification of two different alleles or to an incomplete restriction enzyme digestion. The single product generated from lane four turned out to be a 1.4kb fragment encompassing the 3′ end of *FRASUP5*. Sequencing of product from lane five did not yield any *SUPERMAN*-like genes.

Sequence Walking. Specific (non-degenerate) primers targeted to non-conserved sites in *F. virginana SUPERMAN* clone *FRASUP3* were designed and used to extend this gene fragment in both the 5′ and 3′ directions. For walking in the 5′ direction, G1 primer FRASUP3 5′F1 (Table 1) paired with the M13R vector primer produced a spectrum of products in the first-round reaction using *Hin*dII, *Eco*RI, or *Bam*HI digested and ligated genomic DNA as template (Figure 8A – top). The second round of TVL-PCR, using the first-round products as template with nested G2 primer FRASUP3 5′F2 (Table 1) and the nested T3 vector primer, also produced a spectrum of bands (Figure 8A - bottom). After shotgun cloning, colony PCR, and sequencing selected clones, a product corresponding

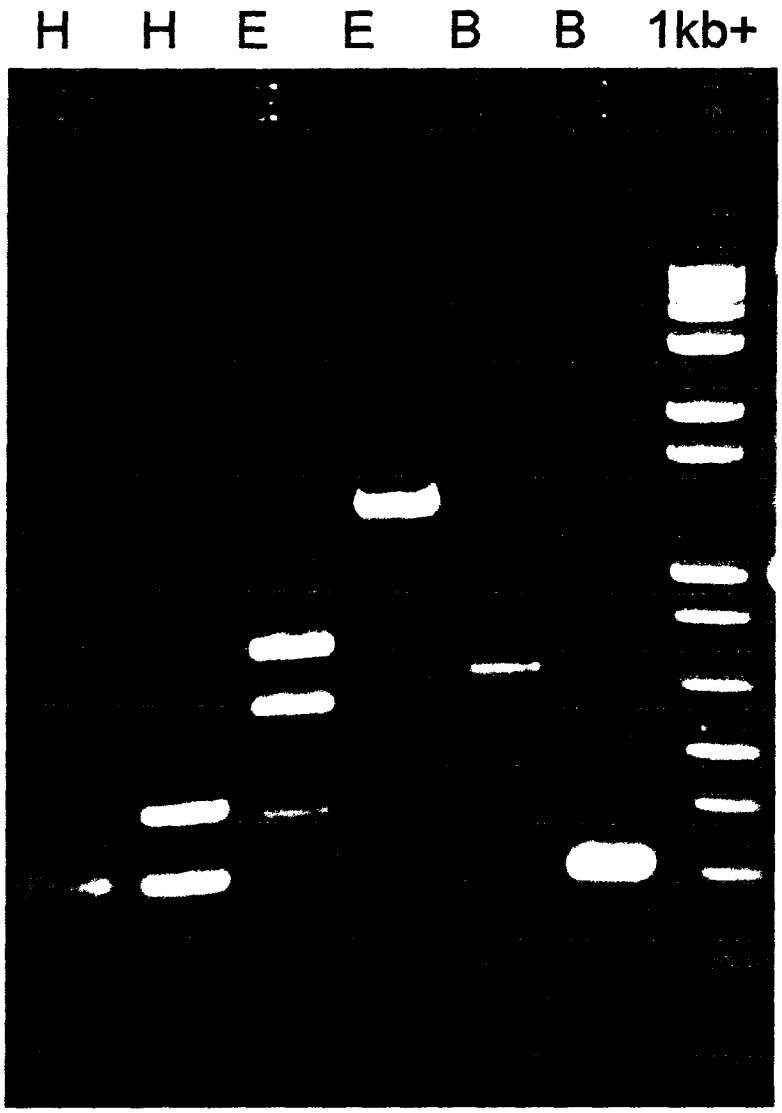


Figure 7. Isolation of the 3′ end of *FRASUP4* and *FRASUP5* via TVL-PCR paralog mining. Second round of PCR using degenerate primer C2H2 F2 and T7 (Lanes 1, 3, 5) or T3 (Lanes 2, 4, 6) using product from first round as templates. Original L2 genomic template DNA was cut with *Hin*dIII (H), *Eco*RI (E), or *Bam*HI (B) as indicated. Boxed fragments were cloned and sequenced.

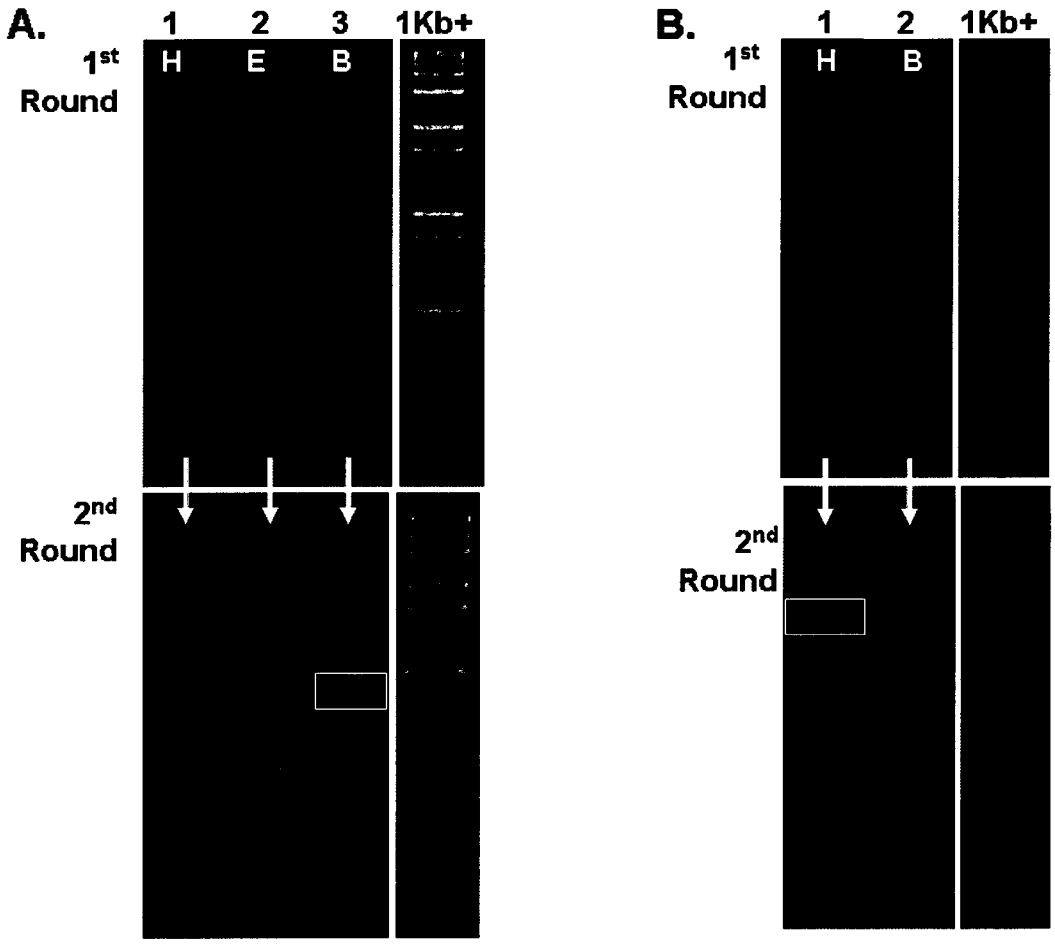


Figure 8. Extension of sequence from clone *FRASUP3* via TVL-PCR sequence walking. **A)** 5′ direction. **B)** 3′ direction. First round TVL-PCR products (top) were generated from templates prepared using *Hin*dIII (H), *Eco*RI (E), or *Bam*HI (B) digestion of L2 genomic DNA. First round products were used as templates to generate second round products (bottom). Boxes (bottom gels) indicate the locations of the second round TVL-PCR products that proved to be the desired products after cloning and sequencing.

in size to the boxed band in Figure 8A (bottom) was identified as the desired product. This sequence extended upstream from the known sequence to the putative start codon and beyond.

For walking in the 3′ direction, the first round of TVL-PCR employed G1 primer FRASUP3 3′F1 (Table 1) paired only with the T3 vector primer. The product from this reaction was used as template for the second round of TVL-PCR using the nested G2 primer FRASUP3 3′F2 (Table 1) in conjunction with the T3 vector primer. Shotgun cloning of the second round TVL-PCR products (Figure 8B - bottom), followed by colony PCR and sequencing of appropriately sized products (500 bp or larger), yielded one desired product of the size indicated by the boxed gel band (Figure 8B - bottom). This product was slightly less than 2kb in length, and extended well into the 3′ UTR.

Using the same methods described to isolate the upstream and downstream portions of the *FRASUP3* fragment, a pair of specific primers (Table 1) was used to capture the upstream sequence of the *FRASUP5* fragment beyond the putative start codon.

BLAST

A series of BLAST searches against the VBI *F. vesca* genome database using the full length *FRASUP3*, *FRASUP4*, and *FRASUP5* genes yielded two new strawberry *SUPERMAN*-like genes designated *FRASUP1* and *FRASUP2*. All three previously isolated genes also had matches with themselves in the database.

***Fragaria SUPERMAN*-like Alignments and Gene Phylogeny**

The results of the SUPERMAN-like alignments and phylogeny are presented in Figure 9, Figure 10 and Figure 11.

Figure 9. Alignment of the five predicted FRASUP proteins. Sequences were translated and then aligned by the ClustalW algorithm using the MegAlign sequence alignment software. Red bars indicate the highly conserved zinc-finger and transcriptional repression domain in the middle and carboxy-terminal end of the proteins, respectively.

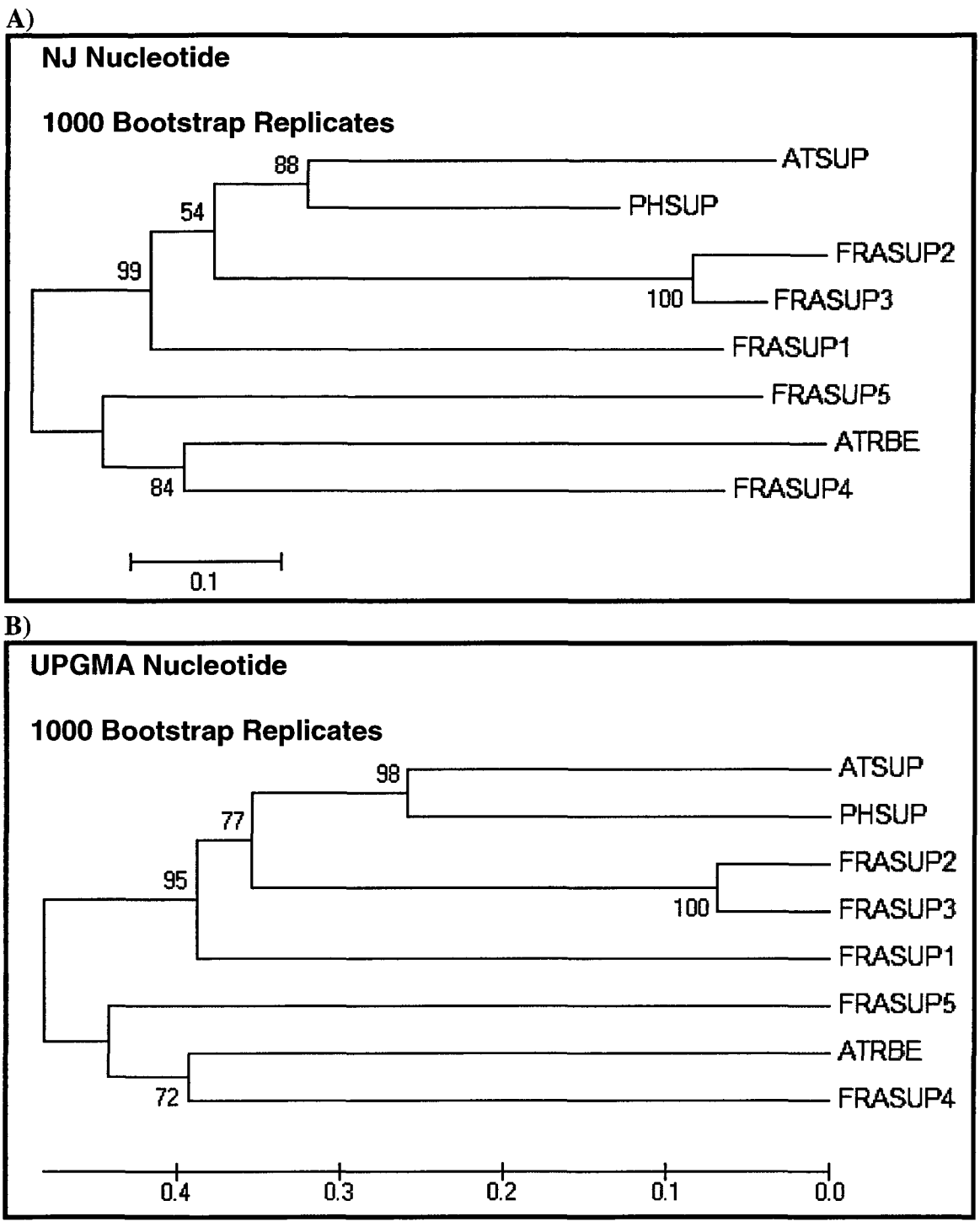


Figure 10. Nucleotide based phylogenetic trees. The five *FRASUP* genes as well as the Arabidopsis *SUP* and *RBE* and the Petunia *SUP* (*PhSUP*) genes were used to construct trees using the MEGA software. Trees were constructed from nucleotide sequences using both the **A)** Neighbor Joining (NJ) and **B)** Unweighted Pair Group Method with Arithmetic Mean (UPGMA) methods.

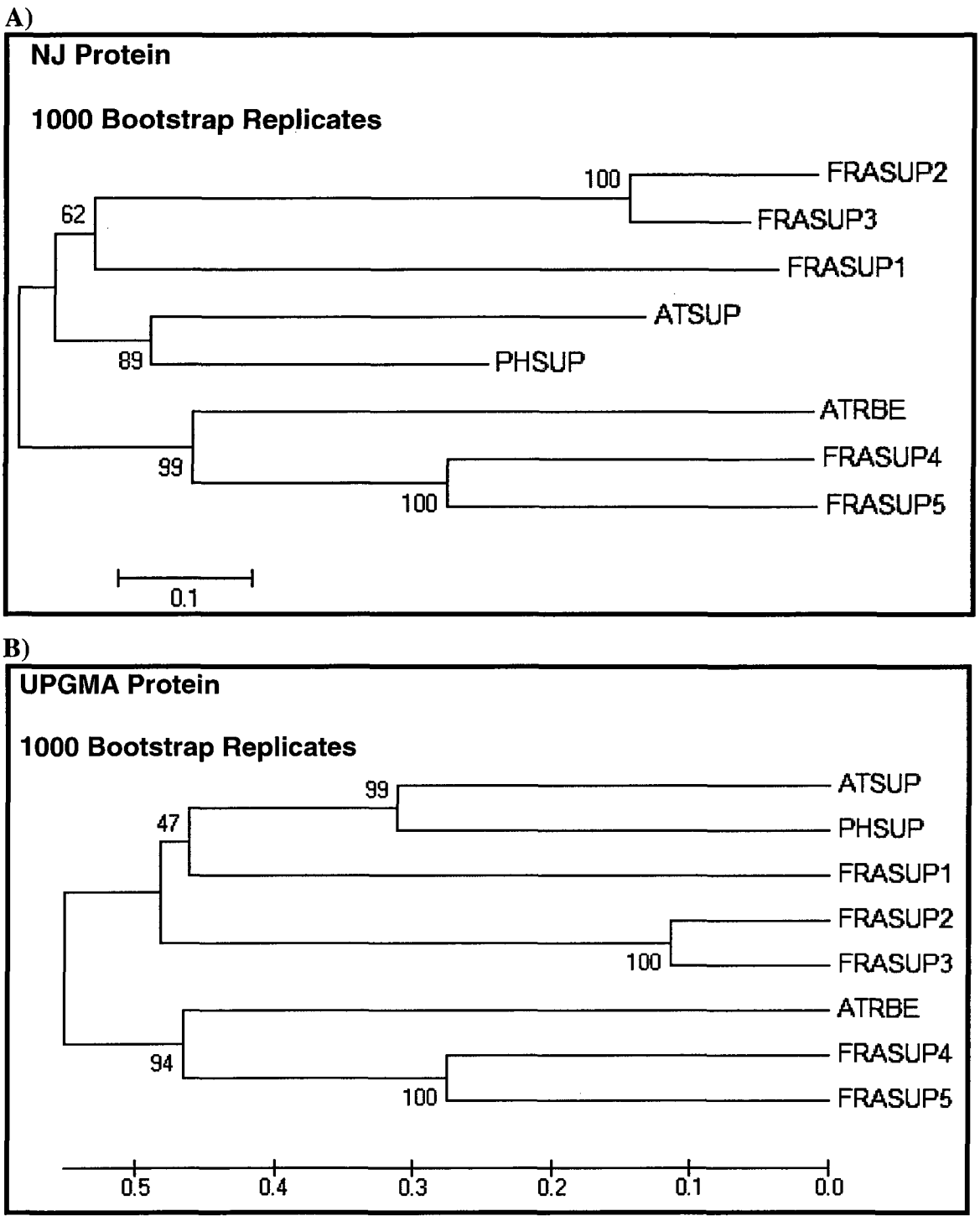


Figure 11. Protein based phylogenetic trees. The five FRASUP proteins as well as the Arabidopsis SUP and RBE and the Petunia SUP (PhSUP) proteins were used to construct trees using the MEGA software. Trees were constructed from protein sequences using both the **A)** Neighbor Joining (NJ) and **B)** Unweighted Pair Group Method with Arithmetic Mean (UPGMA) methods.

Preparation of *FRASUP* Genes for Alignments and Gene Phylogeny

Determining the full length (that is start to stop) coding sequence of the *FRASUP* genes was complicated by the fact that no cDNA was available to delineate the start and stop codons. To do this, ClustalW algorithm alignments of the *FRASUP* genes were performed using the MegAlign software (DNAStar, Madison, WI). Translations of all five genes were individually aligned to the translations of the Arabidopsis *SUPERMAN* (NCBI Accession U38946), *RABBIT EARS* (NCBI Accession AB107371) and Petunia *SUPERMAN* (NCBI Accession AB117749) genes. All previously characterized *SUPERMAN*-like genes lack introns, making the position of the putative stop codon easier to find, as there will be no false stop codon such as those within introns. Therefore, the first stop codon at the end of the open reading frame (ORF) of each *Fragaria SUPERMAN*-like gene was designated as the true stop codon. The putative start codons were more difficult to ascertain. As the *Fragaria SUPERMAN*-like genes presumably lack introns, the start codon should be somewhere inside the gene OR*F*. The start codons are found ~120 bp and ~150 bp away from the conserved zinc-finger region in Arabidopsis and *Petunia SUPERMAN* and Arabidopsis *RABBIT EARS* genes respectively. Therefore, start codons that were around this distance from the zinc-finger motif were designated the putative start codon. If two start codons were right next two each other, as in the case of *FRASUP1*, the first was designated as the true start codon.

Both the nucleotide and translated amino acid sequences of all five *Fragaria SUPERMAN*-like genes from *F. virginiana* accession L2 can be found in Appendix A. The sequences can also be viewed on the National Center for Biotechnology Information

(NCBI) GenBank website, using the genes' unique accession numbers. Both the method of initial isolation and subsequent isolation of the entire coding region for all seven candidate genes can be found in Table 4.

Discussion

Isolation of the candidate genes

The *AGAMOUS* and *ACS-7* genes were isolated with relative ease compared with the *FRASUP* genes. This is because of the availability of already existing strawberry gene sequence of these genes. Although the *Fragaria ACS-7* sequence was not annotated in GenBank it was of sufficient size and homology to other annotated *ACS* genes to be found in a BLAST search. Isolation of the *FRASUP* genes required a *much* greater commitment of resources simply due to the fact that no strawberry *SUPERMAN*-like sequences had been reported prior to this study. The effort put into this task is evident in the variety of PCR based sequence capture methods employed to isolate the five *FRASUP* genes.

Degenerate PCR

Degenerate PCR is a powerful tool if enough sequence information is available for primer design. Potential degeneracy can also be minimized by the use of inosine at the most highly degenerate primer sites. The effectiveness of degenerate primer PCR was demonstrated since fragments of two of the five strawberry *SUPERMAN*-like genes were isolated with this method. It is also possible that more than two gene fragments could have been isolated if more clones were sequenced. However, by itself, degenerate PCR cannot be used to obtain sequence outside of the priming sites. Therefore, to capture full

Table 4. Chronology of the isolation of the five *FRASUP* genes. A variety of strategies were employed to isolate the genes.

***Fragaria SUPERMAN*-like Gene**	**Initial Method of Isolation**	**Method for Isolation of Full Length Coding Sequence**
FRASUP1	*F. vesca* VBI BLAST Search	*F. vesca* VBI BLAST Search
FRASUP2	*F. vesca* VBI BLAST Search	*F. vesca* VBI BLAST Search
FRASUP3	TVL-PCR	TVL-PCR
FRASUP4	Degenerate Primer PCR	TAIL-PCR
FRASUP5	Degenerate Primer PCR	TVL-PCR

length gene sequence, degenerate PCR must be coupled with a chromosome walking technique.

TAIL-PCR

TAIL-PCR is one of the most utilized methods of sequence walking due both to its early development and to lack of need for linker sequence (vectors or adaptors) that other techniques require. Despite its widespread use, TAIL-PCR suffers from a host of problems that can limit its effectiveness. The most obvious drawback to this technique is the frequency of degenerate priming sites. If the degenerate primer cannot find a suitable annealing site close enough (and in the correct orientation) to the specific primer then no amplification can occur. One further drawback of TAIL-PCR is the possibility of a degenerate primer creating background amplification noise, even though this should be minimized by alternate high and low stringency PCR cycles (Table 2).

The application of TAIL-PCR in this study for sequence walking gave mixed results. Although it was initially successful in obtaining both the 5′ and 3′ ends of the *FRASUP4* fragment, it was unable to do so for the *FRASUP5* fragment. As a result, the technique was ultimately dropped in favor of TVL-PCR.

TVL-PCR

As demonstrated above, TVL-PCR is a powerful and flexible tool for both the isolation of novel gene family members (paralog mining) and sequence walking from a known portion of DNA. As opposed to TAIL-PCR, TVL-PCR takes advantage of the probability that a particular restriction site is much more likely to occur in genomic DNA than potential degenerate priming sites. This is because the majority of restriction enzymes recognize a four or six base sequence, while priming sites range from fifteen to

thirty bases. Therefore the odds of a certain restriction site occurring in a portion of DNA are much greater than a certain primer site, even with degeneracy as a factor. TVL-PCR has been applied to acquire partial or putatively full length coding sequences of *SUPERMAN*-like genes from genomic DNA of octoploid *F. virginiana* accession L2.

Two approaches to paralog mining were illustrated by the experiments described here. The initial approach employed conventional PCR using a pair of degenerate primers targeted to two conserved regions that were identified from an alignment of five heterologous *SUPERMAN*-like gene sequences obtained from GenBank. Although this commonly utilized approach enabled isolation of three *SUPERMAN*-like gene segments from strawberry, a limitation to this approach was that two suitable primer sites must be identified at a favorable distance from each other. Alignments of heterologous sequences do not always reveal two favorably located, conserved regions for forward and reverse priming within the genomic DNA. In many instances only a single conserved region can be found, precluding successful utilization of a conventional degenerate primer PCR approach.

In contrast, for paralog mining, TVL-PCR requires the identification of only a single conserved region, to which both the G1 and nested G2 forward primers can be targeted (perhaps even at partially overlapping sites), while the reverse priming site is supplied by the vector. Moreover, the conventional PCR approach yields only sequence fragments corresponding to a previously known region delimited by the two known genomic priming sites, while TVL-PCR enables isolation of sequences extending into previously unknown regions, delimited at only one end by a known genomic priming site and at the other end by a (potentially distant) site of restriction enzyme cleavage. In TVL-

PCR aimed at paralog mining, the use of degenerate genomic primers promotes the amplification of products from both orthologs and paralogs of the sequences used as a basis for primer design. Using this approach, a novel *SUPERMAN* clone *FRASUP3* from an octoploid strawberry was isolated.

A contrasting TVL-PCR strategy was used for sequence walking. This strategy employed specific (non-degenerate) primers targeted to non-conserved sites, with the goal of limiting amplification to orthologous sequences. Paralog amplification always remains a possibility, however, particularly in a polyploid species such as octoploid *F. virginiana*. The successful use of TVL-PCR has been demonstrated in five sequence walks, extending clone *FRASUP3* in both the 5′ and 3′ directions, extending clone *FRASUP5* in both the 5′ and 3′ directions, and extending the *Fragaria ACS-7* in the 3′ direction. These results clearly demonstrate the utility of TVL-PCR as a robust sequence walking technique, and it has become a routinely used tool in our lab.

TVL-PCR adds a useful new option to the toolkit of methodological choices for sequence walking outward from an initially known site and offers various advantages as compared with other ligation-dependent sequence walking methods. The method described herein relies on the ligation of the pCR4-TOPO vector to the unknown end of a target DNA molecule. The Invitrogen pCR4-TOPO vector and TOPO cloning technology takes advantage of the 3′ adenosine overhang created by *Taq* DNA polymerase. Because the entire pool of genomic DNA molecules generated by restriction digest and end-repair have adenosine overhangs, it is impossible for the adenylated genomic DNA fragments to re-ligate with each other. Additionally, genomic DNA fragments can only ligate with a

vector molecule having a 3′ thymidine overhang. Accordingly, chimeric ligation constructs comprised of multiple genomic fragments are not expected to occur.

In contrast to TVL-PCR, single specific primer PCR (SSP-PCR) (Novak and Novak, 1997) requires the genomic fragment to be cut with the same restriction enzyme used to cut the vector to allow for ligation. This limits the choice of enzymes that can be used for genomic digests, while also limiting sequential chromosome walking to the first flanking restriction site. The efficiency of SSP-PCR will be reduced by damage at the single-stranded cut site because ligation cannot occur and end-repair is not an available option. In addition to ssDNA damage, there is the potential for digested fragments to re-ligate to each other or to themselves. This may be overcome by having a high ratio of vector to genomic DNA (60-100 ng pUC19 vector and 100-200 ng genomic DNA), however with the cost of vector, this option can become expensive.

TVL-PCR utilizes the ligation product directly as the template for PCR: no transformation step is employed at this point. Some other ligation-based procedures such as that described by Boulter and Natarajan (1999) require that, after ligation of the fragment to both ends of the vector, the entire pool of recombinant plasmids be transformed into *E. coli*. After transformation, clones are collected, the plasmids isolated, and PCR is performed on the pool of plasmids to amplify the desired target. This is a very tedious and costly protocol, requiring an extra transformation step, and there are many variables that could cause it to falter. The first variable is the cloning step, which requires that the desired fragment must be ligated to both ends of the vector (circularized). If the desired fragment is very large as compared to the other digestion products, it is unlikely that ligation of both ends to the vector will occur. Secondly, the circularized plasmid

must be transformed into a bacterial cell. Entry of large plasmids into the cell is more difficult, meaning their transformation will not be as efficient as with smaller plasmids. TVL-PCR avoids these complications by eliminating the pre-PCR transformation step, and requiring that only one end of the desired DNA ligate to the vector to perform successful TVL-PCR.

One potential drawback of ligation-dependent methods is that ligation of the same kind of linker molecule to both ends of a genomic fragment would create opportunity for the linker primer to anneal to both ends of ligated fragments and to non-specifically amplify a multitude of ligated fragments via "single primer amplification". A ligation-dependent technique developed by Yuanxin et al. (2003) tends to eliminate the background noise created by single primer amplification. The technique is similar to other forms of ligation-dependent PCR in that a linker is ligated to digested DNA molecules. However, a subsequent step adds 3′ adenosine overhangs only to those fragments to which a site-specific genomic primer can anneal and prime a single cycle of replication. Thus, only a targeted subset of genomic fragments are available for ligation to a 3′ thymidine overhang linker, drastically narrowing the spectrum of ligation products that could be subject to single primer amplification.

The TVL-PCR procedure minimizes the problem of linker-dependent, single primer amplification by maintaining a high genomic DNA to vector ratio, thereby reducing the number of genomic fragments that acquire ligated vector molecules at both ends. For the ligation reaction 36 ng of genomic DNA, was used which, after complete digestion with a six-base restriction enzyme should result in 8.2 billion fragments present in the ligation reaction. At a concentration of 10 ng/μl, 0.5 μl of TOPO vector yields 1.2

billion molecules, and thus roughly a ratio of seven genomic DNA fragments to one vector molecule.

Methods presented by Cormack and Somssich (1997) and Yuanxin et al. (2003) provide only a limited length of linker sequence for external priming sites. These sites may not work well with primers designed to the known sequence because of differences in melting temperature, annealing temperature and the possibility of primer-dimer formation. The use of a large vector such as pCR4-TOPO offers the possibility of many priming sites if the usual vector primers do not match well with the genomic primers. The complete pCR4-TOPO vector sequence can be found at: http://www.addgene.org/pgvec1?f=c&cmd= viewvecseq&vectorid=6657.

Appropriate Restriction Enzymes

There are a number of factors to consider when choosing a restriction enzyme for TVL-PCR. First, the enzyme should not cut between the genomic primer annealing sites and the end of the initially known sequence, because a cut in this region would terminate the walk before it extended into unknown sequence. Second, after genomic DNA is digested, it must be amenable to end-repair to produce 3′ adenosine overhangs. Thus, the restriction enzyme must create ends that are blunt or that have recessed 3′ ends (5′ overhangs) that can be filled in by *Taq* DNA polymerase that also adds the requisite 3′ adenosine overhang. With these considerations, three restriction enzymes were chosen for digestion: *Eco*RI, *Bam*HI, and *Hin*dIII.

The average restriction fragment size is also an important consideration. To maximize the amount of unknown sequence amplified, an enzyme should be chosen that cuts infrequently. For this reason, six-base cutters were used, which produce fragments

with an average size of 4096 bp, thereby providing an average distance of 2048 bp between any given genomic priming site and the closest downstream restriction site. If desired, larger primer-to-cut-site distances could be obtained by employing partial digests.

Appropriate DNA Polymerase

To enhance the potential for large amplification products, the choice of DNA polymerase is also important. A high fidelity proofreading *Taq* polymerase mixture was chosen over standard *Taq* polymerase. This option was exploited because the size of the PCR product was unknown and high fidelity polymerase mixtures are known to have greater amplification range (Waggott, 1998).

Primer Design

A key factor to consider when designing genomic primers for vector ligation-dependent PCR is that the primers should not be too close to the end of the known sequence. It is essential to generate a TVL-PCR product that begins with at least 50-100 bases of known sequence before transitioning into unknown sequence, thus assuring that the homology between the newly generated product and the initially known sequence can be assessed.

In first round TVL-PCR, either the M13F/R or the T3/7 vector primers can be employed; however, use of an M13 primer has the advantage of allowing use of a nested T3/7 primer in the second round, perhaps increasing product specificity. Use of a T3/7 primer in the first round precludes primer nesting on the vector side during the second round of TVL-PCR, but can still lead to a successful walk (as demonstrated) and provides a helpful option if the M13F/R vector primers do not pair well with particular genomic primers. Importantly, an M13F/R primer cannot be used in the second round, because it

would generate a second round TVL-PCR product that would contain both an M13F/R and a T3/7 priming site. If such a product were subsequently cloned into the TOPO vector, difficulties could be encountered in sequencing the insert, due to the duplication of sequence primer sites in the vector and insert (Figure 5).

One further consideration when designing primers is their melting temperature. The vector primers used in these experiments have melting temperatures ranging from 47-50°C (Table 1), while the manufacturer's minimum recommended annealing temperature for use with AccuPrime HiFi *Taq* polymerase is 52°C. Therefore, genomic primers were designed with their melting temperatures limited to 50-55°C range so that the gene primers and vector primers would have similar melting temperatures, and still work with the designated thermocycler profile. Melting temperature values were determined by the Integrated DNA Technologies OligoAnalyzer application which can be found at: http://www.idtdna.com/analyzer/Applications/OligoAnalyzer.

The Consideration of Genome Size

The genus *Fragaria* has a relatively small basic genome size of ~200 Mb (Folta and Davis, 2006). Therefore, the 36 ng of restriction-digested genomic DNA used in the ligation reaction translates to 160,000 copies of the genome, and of the desired fragment (assuming a single copy gene). This quantity of genomic DNA was sufficient to support amplification of the desired templates in *Fragaria*. However, for organisms with significantly larger genomes, if the same amount of genomic DNA were to be used, the desired fragment copy number might be much lower. One way to overcome this potential obstacle would be to use more ligation reaction as template for the first round TVL-PCR. From the results presented herein, the ratio of genomic DNA to vector was optimal for

Fragaria. For much larger genomes, amplifying a target by TVL-PCR may require the use of larger amounts of both genomic DNA and vector, while maintaining the same 7:1 fragment to vector ratio.

BLAST

That a nearly complete genome sequence is worth its figurative weight in gold is realized in the discovery of strawberry *SUPERMAN*-like genes through BLAST searches. The fact that two novel *SUPERMAN*-like genes can be identified in a matter of seconds is testament to how invaluable a resource a sequenced genome can be. A BLAST search of the database using the Arabidopsis *SUPERMAN* and *RABBIT EARS* genes yielded hits to the zinc-finger region of all five *FRASUP* genes. Therefore, it appears that if a sequenced strawberry genome was available at the inception of this project, all five *FRASUP* gene sequences could have been obtained in seconds.

Since this time, periodic BLAST searches of the database with the five *FRASUP* genes have yielded no new *FRASUP* genes. Given that over 95% of the genome has been sequenced and assembled into contigs it seems highly unlikely that a new *FRASUP* gene will be discovered in *F. vesca* (Shulaev, unpublished).

***Fragaria SUPERMAN*-like Alignments and Gene Phylogeny** Alignments. As demonstrated in Figure 9, all of the *FRASUP* genes contain the zinc-finger with the conserved QALGGH motif except for *FRASUP2* which has the motif QSLGGH. The substitution of a serine for an alanine is not likely a sequencing error as there is also a serine in the *FRASUP2* gene on VBI database. At the carboxy-terminal end, all of the *FRASUP* genes contain the conserved DLELRL motif.

Phylogenetic Trees. The results of the phylogenetic trees revealed that the five *Fragaria SUPERMAN*-like genes cluster into two distinct clades. The first clade, composing *FRASUP1*, *FRASUP2* and *FRASUP3*, clusters with the two *SUPERMAN* genes from Arabidopsis and *Petunia*. The second clade is composed of *FRASUP4* and *FRASUP5* which clusters with the Arabidopsis *RABBIT EARS* gene. While the early branching that supports the two clades is maintained in all four trees, the internal branching differs slightly between nucleotide and protein based trees. The different branching patterns are most apparent in the *RABBIT EARS* clade. In the nucleotide based trees, *RBE* and *FRASUP4* cluster to the exclusion of *FRASUP5*. However in the protein based trees, FRASUP4 and FRASUP5 cluster together to the exclusion of RBE.

There is less agreement in terms of the branching pattern between trees in the *SUPERMAN* clade. In both the NJ and UPGMA nucleotide derived trees, early branching separates *FRASUP1* from the other members of this clade. The next branching point then separates into two groups with the *SUPERMAN* genes in one and *FRASUP2* and *FRASUP3* in the other. In the NJ protein derived tree, early branching creates two clusters, with the *SUPERMAN* genes in one and the three *FRASUP* genes in the other. In the UPGMA protein derived tree, early branching also creates two clusters. In this tree, the two *SUPERMAN* genes group with the *FRASUP1* gene while *FRASUP2* and *FRASUP3* form a separate group.

Bootstrap support for the clustering is quite strong throughout the trees in the *RBE* clade. The high bootstrap values strongly support the outcome of the trees regarding homology and possible orthology between the two *FRASUP* genes and the *RBE* gene. Interior branching in the *SUPERMAN* clade strongly supports the orthology of the

Arabidopsis and Petunia genes as well as the possible parology between *FRASUP2* and *FRASUP3*. Unfortunately, low bootstrap values between the *SUPERMAN* cluster and the *FRASUP2/3* group preclude any strong conclusions of orthology between the *SUPERMAN* and *SUPERMAN*-like genes. However, it is quite possible that one of the *FRASUP* genes in the *SUPERMAN* clade is orthologous to *SUPERMAN* but is too diverged to cluster. The second alternative may be that the true *Fragaria SUPERMAN* ortholog has not yet been isolated.

The slight differences in the organization of the nucleotide and protein phylogenetic trees are likely due to the algorithms used to create the trees. When generating a protein based tree, similarities in the behavior and structure of the functional groups add a layer of complexity to the trees. For instance leucine and isoleucine are different amino acids but have similar chemical properties. Therefore they will not generate as strong a penalty in the tree construction as for example a leucine and a lysine which are very different chemically. Information of this sort cannot be extrapolated from nucleotide based alignments.

Despite minor inconsistencies in the branching and bootstrap values between the four phylogenetic trees, overall the results support the hypothesis that the *FRASUP* genes do indeed belong to the *SUPERMAN* gene family. Furthermore the trees help to resolve the proposed orthology between the *FRASUP* genes and the *SUPERMAN* and *RABBIT EARS* genes. This is because the phylogenetic trees clearly define a *SUPERMAN* clade and a *RABBIT EARS* clade, with the *FRASUP* genes falling confidently into one or the other.

CHAPTER 3

MAPPING OF CANIDATE GENES AND PUTATIVE SEX LOCUS MARKERS WITH CONCURRENT ASSOCIATION STUDIES

Abstract

With the isolation and characterization of the candidate genes complete, it was of great interest to determine where in the strawberry genome the genes reside. More importantly, it was imperative to know if any of the genes were cosegregating with or at least syntenic with the sex determination locus. Addressing this question would not have been possible without an SSR marker named ARSFL7, which had been shown by Spigler et al. (2008; 2010) to cosegregate with sex in a separate population of *F. virginiana*. In the present study, ARSFL7 was also linked to sex in the L1L2 population, demonstrating that this marker is correlated with sex in two independent *F. virginiana* progeny populations. With a marker linked to sex in octoploid *F. virginiana*, it was now possible to perform a relevant linkage analysis using our own mapping population (designated YP) of diploid *F. vesca*. Segregation data in the YP population was collected not only for the candidate genes but also for ARSFL7 and additional markers surrounding the ARSFL7 locus as identified in two previously published diploid strawberry linkage maps developed by Sargent et al. (2006; 2007). The YP linkage map contained a total of 50 markers and was resolved into the seven predicted linkage groups. Furthermore, one candidate gene, *FRASUP5*, and all the additional Linkage Group 6 markers were on the same linkage group as ARSFL7, which is linked to sex in *F. virginiana*. Therefore, it is possible that *FRASUP5* may segregate with, or even be involved with, sex determination

in *Fragaria virginiana*. The incorporation of more markers near ARSFL7 permitted the use of comparative mapping between the YP linkage map and the two previously published diploid strawberry maps. These comparisons can be used to generate more markers that might be linked to sex determination in this area and lend support to the correct order of previously mapped markers.

Introduction

Plant Populations

The L1L2 population and its parents were used to gain more insight into the genetic basis of sex determination in *F. virginiana*.

ARSFL7 Genotyping

ARSFL is the prefix for a group of simple sequence repeat (SSR) markers developed by Lewers et al. (2005) for the purpose of creating linkage maps of various Rosaceous species. The markers have been widely used in the *Rosaceae* genetics community, particularly by Spigler et al. (2008; 2010) for the development of their *Fragaria virginiana* linkage map. Spigler et al. (2008) found one SSR marker, ARSFL7, which cosegregates with the sex determination locus. ARSFL7 is found on linkage group 41 (their map is as yet incompletely unresolved and presently contains 42 linkage groups) and is 6.8 cM away from the sex determination locus that causes male sterility (vestigial stamens) (Spigler et al., 2008). A follow up study using the same population but with additional markers showed that the SSR is only 2 cM from the sex determination locus (Spigler et al., 2010). In both studies, ARSFL7 was the molecular marker most closely linked to the sex determination locus (Figure 12). Given such a find, it was of great

interest to see if there was segregation of the ARSFL7 SSR in the L1L2 population and whether an allele would cosegregate with sex. According to Spigler et al. (2008) the size of the ARSFL7 marker allele cosegregating with male sterility was 277 bp, while the size of its alternate allele (male fertility) was 275 bp.

Mapping with the Diploid YP Population

Two highly saturated diploid strawberry linkage maps were recently published by Sargent et al. (2006; 2007). The original map, developed by Sargent et al. (2006) and updated in 2007 (Sargent et al., 2007) is based on an F2 population from an interspecific cross between two diploid species, *F. vesca* as the female parent and *F. bucharica* (this plant accession was previously misclassified as *F. nubicola*) as the male parent. The map, known as FVxFN, is well resolved, containing the seven predicted linkage groups which span 424 cM, and is very dense with an average distance of 2.3 cM per marker (Figure 13). This FVxFN linkage map contains markers based on both SSRs and genes, which were developed from genomic and EST sequences from *F.* ×*ananassa* and *F. bucharica*.

Materials and Methods

Features of Plant Materials

L1L2 Population. The sex of each plant in the L1L2 population was determined by looking at mature opened flowers in the greenhouse population. Hermaphrodite plants develop large protruding stamens with ample yellow pollen. Females develop tiny vestigial stamens that do not protrude out of the flower and make no pollen.

To see whether stamen number is a factor influenced by the sex determination locus, the average number of stamens from both parents was determined by counting the

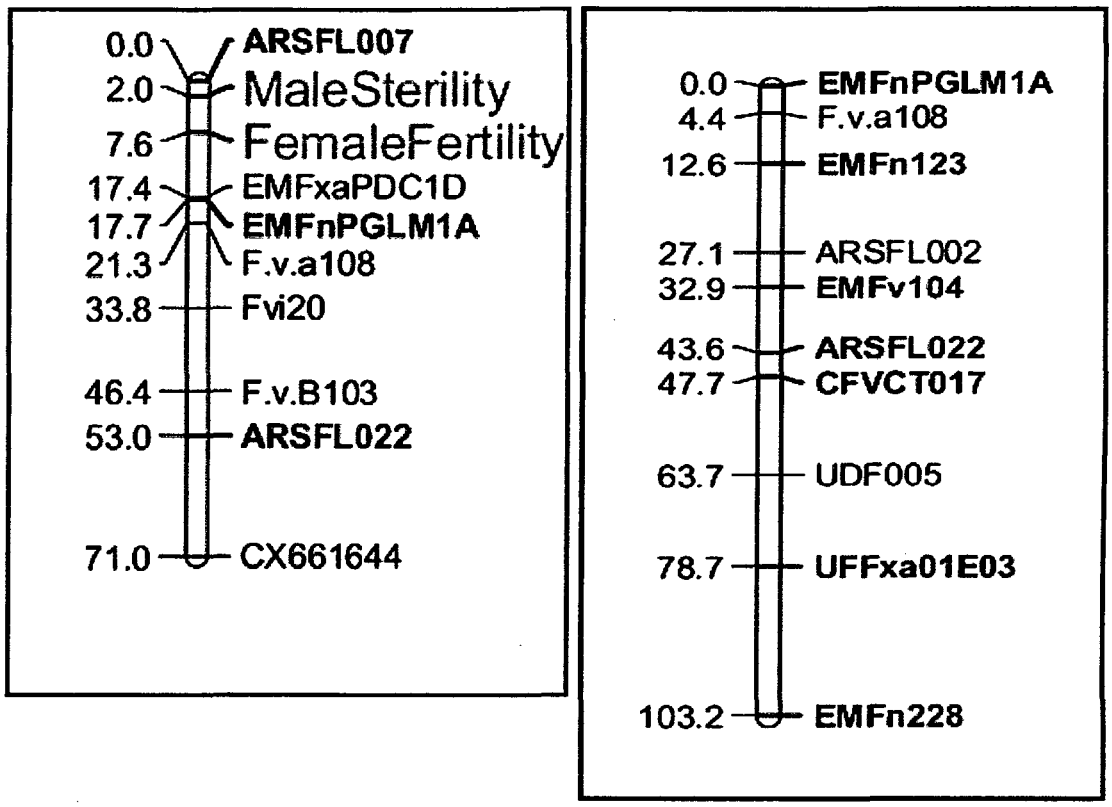


Figure 12. Previously published *F. virginiana* Linkage Group 6 maps. Maternal (VI.C-m) and paternal (VI.C-p) Linkage Group 6 maps from Spigler et al., 2010. These two maps represent one set of homeologous chromosomes in the *Fragaria virginiana* genome.

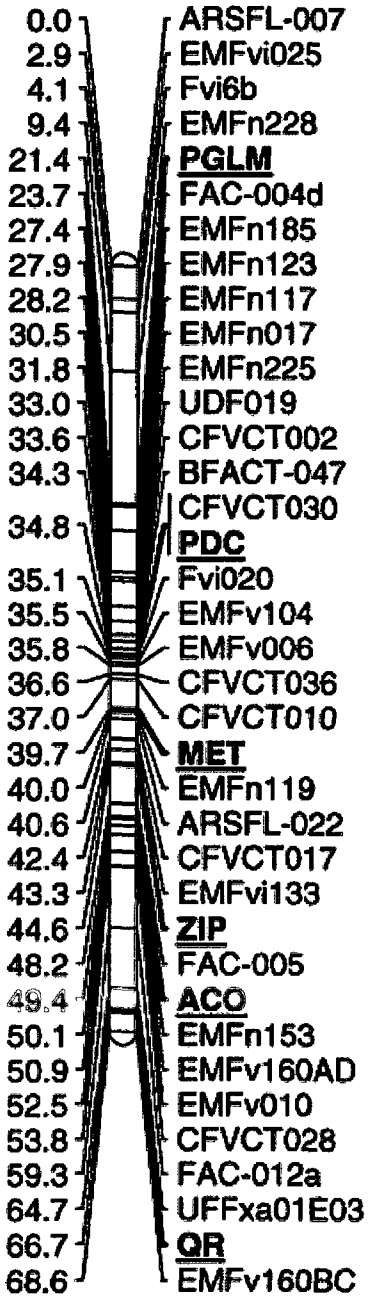


Figure 13. Previously published FVxFN diploid *Fragaria* Linkage Group 6 map. The FVxFN *F. vesca* Linkage Group 6 map from Sargent et al. (2007).

total number of stamens from six L1 flowers and six L2 flowers. Six flowers from each parent were dissected through the center of the flower. Each carpel half was removed with forceps. Stamens were counted under a dissecting microscope.

YP Population. The previously developed YP population as described in the general methods section was used exclusively for mapping in this study. The population is an F2 derived from a cross between two diploid *F. vesca* varieties, 'Yellow Wonder' (YW) and 'Pawtuckaway' (PAWT). A total of 72 F2 individuals were used for mapping the candidate genes and markers putatively linked to the sex determination locus.

ARSFL7 Genotyping of L1L2 Population. To perform genotyping, the ARSFL7 primers defined by Lewers et al. (2005) were acquired from Invitrogen with a fluorescent FAM label on the 5′ end of the forward primer. PCR was performed using 15 ng of each DNA with 0.4 mM dNTP mix, 0.4 μM of forward and reverse primers, 2.5 μl of 10X EconoTaq buffer (Lucigen, Middleton, Wisconsin) and 1 U EconoTaq DNA polymerase. The reaction was brought up to 25 μl with sterile deionized water. The thermocycler profile used was as follows: initial denaturation at 94°C for 1 min; 35 cycles of 94°C for 30 sec, 58°C for 30 sec, 72°C for 30 sec; and a final extension at 72°C for 10 minutes. Next, 10 μl of each reaction was separated on a 1% agarose gel at 100V for 70 minutes. The PCR products are expected to be within the range of 200-300 bp. A 1:10 dilution (1 μl PCR product and 9 μl dH_2O) was made and products were separated on an Applied Biosystems 3130 Genetic Analyzer. The fluorescent products were analyzed using the Peak Scanner version 1.0 software from Applied Biosystems.

Mapping with the Diploid YP Population

General Strategy for Length Polymorphism Analysis. PCR was performed using mapping primers to amplify products in both YW and PAWT according to the standard Invitrogen HiFi *Taq* protocol described above. After visualizing them by agarose gel electrophoresis, products were cloned using the standard Invitrogen TOPO Cloning kit protocol described above. Plasmids were isolated from a single clone and were sequenced bidirectionally on an Applied Biosystems 3130 Genetic Analyzer with M13 vector primers. The YW and PAWT contigs were then aligned using the MegAlign software package (DNAStar, Madison, WI) to determine if the size morphs amplified were allelic. The presence of parental alleles was checked in the F1 parent of the YP F2 population. Loci in the F2 population were then PCR amplified and scored. In all candidate genes and markers, both parental alleles were found in the YP F1. After separating the PCR products by agarose gel electrophoresis, the YP F2 progeny were scored (and coded) as either homozygous for the YW allele (a), homozygous for the PAWT allele (b), or heterozygous (h). The scoring results were then incorporated into an Excel spreadsheet of the YP F2 segregation data.

General strategy for CAPS Analysis

PCR was performed using mapping primers to amplify products in both 'Yellow Wonder' and 'Pawtuckaway' according to the standard Invitrogen HiFi *Taq* protocol described above. After visualization by agarose gel electrophoresis, products were cloned using the standard Invitrogen TOPO Cloning kit protocol as described above. Plasmids were isolated from a single clone and were sequenced bidirectionally on an Applied Biosystems 3130 Genetic Analyzer with M13 vector primers. The sequencing reads were trimmed of vector sequence, aligned using the NCBI sequence alignment software and

combined into a single contig. The YW and PAWT contigs were then aligned using the MegAlign software package (DNAStar, Madison, WI) to find restriction site polymorphisms. When a suitable polymorphism between YW and PAWT was found, CAPS analysis was performed on YW, PAWT, and the F1 to determine if the alleles of these products would be present in the F2 population. If the sequenced alleles were found in the F1 parent, then the locus was PCR amplified, restriction digested, and separated electrophoretically in the YP F2 population. The progeny were scored as described above.

Given the 3.2 kb size of the product amplified by the *AGAMOUS* mapping primers, the approach of identifying restriction site polymorphisms by sequencing was not considered viable. Instead a series of digests were performed on the YW and PAWT products with various restriction enzymes to find one that would yield a suitable polymorphism.

Generation of the YP linkage Map Using JoinMap

Once all of the segregation data for the candidate genes and the sex markers was collected, a linkage map could be constructed. The segregation data from the seven candidate genes and the six published markers were combined with data from 43 other markers from previous work. The data, comprised of 3 dominant and 47 codominant markers was then loaded into the mapping software program JoinMap version 4 (Plant Research International, Wageningen University, Netherlands). Using the default settings, the program was able to construct the predicted seven linkage groups from the data which can be seen in Figure 14.

Results

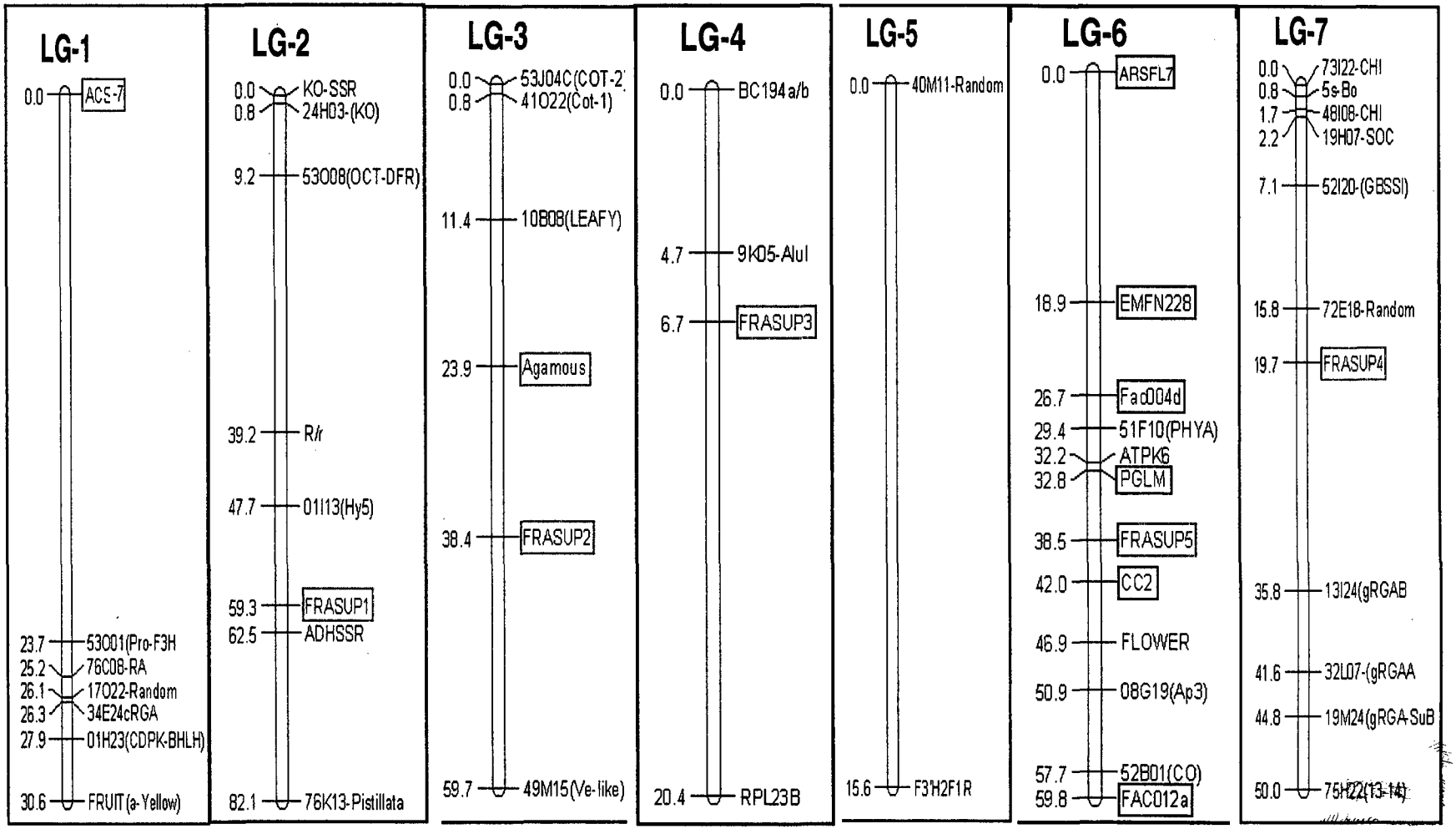


Figure 14. Candidate gene and sex marker mapping. The seven calculated linkage groups derived from the *F. vesca* YP mapping population. The marker name is on the right while the linkage group position (in centimorgans) is on the left. The candidate genes are boxed red, while the LG6 (sex determination chromosome) markers presented herein are boxed in blue. The total combined length of the map is 318.2 cM.

L1L2 Phenotypic and Segregation Data

It was determined that both the female (L1) and hermaphrodite (L2) parents contain an average of 23 stamens per flower; however the stamens in the female flowers were uniformly vestigial (Figure 3).

Out of a total of 57 F1 progeny, 25 were female and 32 were hermaphrodite. Once the sex segregation data was collected, it was used to determine if the L1L2 population was segregating for sex in the predicted 1:1 female to hermaphrodite ratio, as based on the model of Ahmadi and Bringhurst (1991). A χ^2 Goodness of Fit Test was then performed with the sex segregation data. In accordance with the model, the null hypothesis is that the L1L2 population does segregate in a 1:1 female to hermaphrodite fashion. As there are two classes (female or hermaphrodite) the degrees of freedom of the test are one. Under the Ahmadi-Bringhurst model, 28.5 F1 should be female and 28.5 should be hermaphrodite. The calculated χ^2 value for the test is 0.86. For one degree of freedom this value falls between a probability 0.5 and 0.3. These values are not statistically significant and so the null hypothesis was accepted. The sex of the L1 and L2 parents as well as all of the F1 progeny can be found in Table 5.

ARSFL7 Genotyping

Given the expected small size and possibly minute size variability of the SSR in the L1L2 population, it was determined that ARSFL7 would not be resolvable on an agarose gel. The best option was a genotyping gel, which is able to resolve single base pair differences. The entire L1L2 population including the parents L1 and L2 were genotyped for the ARSFL7 SSR. The Peak Scanner software showed a number of strong peaks ranging in size from 210 bp to 255 bp. Two alleles of the ARSFL7 SSR identified

Table 5. Segregation data for sex and the ARSFL7 locus in the L1L2 F1 population. The sex of both parents and all members of the F1 are scored as either female (F) or hermaphrodite (H). In addition, the genotyping results of parents and all F1 progeny at the ARSFL7 locus are given in base pairs (bp). Underlined-bold F1 represent recombinants.

Plant	**Sex**			**Plant**	**Sex**		
Parent		**Allele**	**Size (bp)**			**Allele Size (bp)**	
L1	F	243	249				
L2	H	253	255				
F1 Progeny							
L1L2				L1L2			
1	H	249	255	30	H	249	253
2	H	249	255	31	H	249	253
3	H	243	255	32	H	249	253
4	H	249	255	**33**	H	243	253
5	H	249	253	34	F	243	255
6	F	243	255	35	F	243	255
7	F	249	255	36	F	243	255
8	H	249	253	37	H	249	253
9	F	243	255	38	F	243	253
10	F	243	255	39	F	243	255
11	F	243	253	40	F	243	255
12	F	243	255	41	H	249	253
13	H	249	255	42	F	243	255
14	H	249	255	43	H	249	255
15	H	249	253	44	F	243	255
16	F	243	253	45	F	243	255
17	F	249	253	46	H	249	255
18	F	243	255	47	F	243	255
19	H	249	255	48	F	243	255
20	F	243	253	49	H	249	255
21	H	249	255	50	F	243	255
22	H	249	255	51	H	249	253
23	H	249	255	52	H	249	253
24	H	249	255	53	F	243	255
25	F	243	255	54	H	249	253
26	H	249	253	55	H	249	255
27	H	249	255	56	F	243	255
28	H	243	255	57	H	249	253
29	H	249	255				

in the female parent L1 appeared to be highly correlated with sex in the L1L2 population. Peaks at 243 bp and 249 bp appeared to segregate from one another, as they were never found together in the progeny. This would indicate that these alleles are in the same subgenome. In addition, peaks in the hermaphrodite L2 parent at 253 bp and 255 bp also appeared to be in the same subgenome, since they too were never found together in any of the F1. However, it is unclear without performing further crosses whether the ARSFL7 alleles in L2 are on chromosomes homologous to the sex chromosome in L1.

Allele 243 was highly correlated with sex, being found in 23 out of the 25 female F1 progeny. Conversely, allele 249 was found in 29 out of the 32 hermaphrodite F1 progeny. A visual representation of the linkage between the ARSFL7 alleles and the sex locus in the L1L2 population can be seen in Figure 15. Given that the 57 L1L2 plants represent 57 gametes from the female parent L1, and if peaks 243 and 249 are truly alternate alleles in the same subgenome, then there are a total of five recombinants in the 57 F1 progeny and a recombination frequency of 8.8% (8.8cM). The ARSFL7 locus segregation data in the F1 L1L2 population can be found in Table 5.

Mapping with the Diploid YP Population

Upon confirmation that the SSR marker ARSFL7 was closely linked to the sex determination locus in the L1L2 population, it was of great interest to develop more markers for finer mapping of this region and more importantly to determine if any candidate genes were associated with the sex determination locus. As luck would have it the ARSFL7 SSR was mapped in the FVxFN population and was found to be the first marker at the top of Linkage Group 6 (Figure 13). Now that the ARSFL7 marker was mapped to a particular chromosome it was of necessity to map markers flanking the SSR

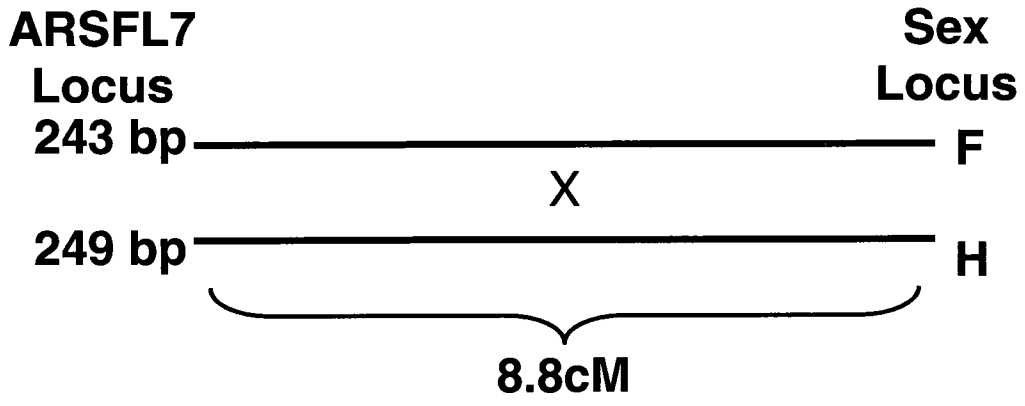


Figure 15. Diagrammatic representation of the map distance between the sex locus (F, H) and the ARSFL7 locus (243 bp, 249 bp). In the female parent, L1, of the L1L2 population, the 243 bp ARSFL7 allele is linked to the female (F) allele at the sex determining locus. The 249 bp allele in L1 is linked to the hermaphrodite (H) allele of the sex determining locus.

to determine where they mapped on the YP linkage map. This in turn could be compared to the placement of the candidate genes to determine if any were near ARSFL7. A total of six markers from the top of FVxFN Linkage Group 6 including ARSFL7 were mapped using the YP population. Three markers, EMFN228, FAC004d, PGLM, were from the most recent FVxFN map (Sargent et al., 2007) and can be seen in Figure 13, while markers FAC012a and CC2 were taken from two older unpublished versions of the FVxFN map (data courtesy of Dan Sargent). In one older map, ARSFL7 is still found in Linkage Group 6 but it is at the opposite end of the linkage group and surrounded by several different markers (including FAC012a). In the second unpublished map, ARSFL7 is in the now established order but is much nearer to marker CC2.

All six of the sex markers mapped to *F. vesca* Linkage Group 6 (LG6) in the YP population. This linkage group is homologous to the *F. virginiana* linkage group that contains the sex determination locus (Spigler et al., 2008, 2010). In addition one of the candidate genes, *FRASUP5*, also mapped to LG6. The linkage groups and map positions of all seven candidate genes are found in Table 6. It is of interest to note how dispersed the candidate genes are throughout the genome. The seven candidate genes are spread throughout six linkage groups and none of the *FRASUP* genes are found clustered together.

It was concluded that the best strategy to place both the markers and candidate genes onto the YP diploid map was to use a combined CAPS (Konieczny and Ausubel, 1993) and simple PCR product length polymorphism strategy, described below. Each candidate gene and published marker primer pair sequence was used in a BLAST search against the VBI *F. vesca* reference genome database. The contig from the best match was

Table 6. Linkage Group that each candidate gene is found on in the YP *F. vesca* mapping population.

Gene	Linkage Group	Distance From Zero (cM)
FRASUP1	2	59.3
FRASUP2	3	38.4
FRASUP3	4	6.7
FRASUP4	7	19.7
FRASUP5	6	38.5
AGAMOUS	3	23.9
ACS7	1	0

then used in a BLAST against the Viridiplantae database using the tBLASTx algorithm on the NCBI website. The tBLASTx algorithm translates the query sequence in all six reading frames and aligns the translations against the translated nucleotide database. The purpose of the BLAST was to find the location of possible protein coding genes near the markers. Coding sequences are undesirable for markers as they have a higher probability of being conserved and thus may have few polymorphic sites. The trade off of this strategy is that primer sites are more likely to be conserved in coding sequence as compared with non-coding sequence. However, given that 'Yellow Wonder' and 'Pawtuckaway' are the same species, they should have a high degree of sequence homology throughout the genome, coding or non-coding.

A region of the contig flanked by putative coding sequence was then imported into the PrimerSelect (DNAStar, Madison, WI) primer design software. Primers were designed to generate products in the size range of 500 bp to 2kb. The primer sequences (and their corresponding product lengths) used to map the Sargent map derived markers (putative sex linked markers) can be found in Table 7. The primer sequences (and their corresponding product lengths) used to map the candidate genes in the YP population can be found in Table 8, with the exception that the *AGAMOUS* gene was mapped based upon polymorphism detected by the large intron primer pair described in Chapter 2.

General Approach to Marker and Candidate Gene Mapping

A prerequisite for linkage mapping is finding an efficient and reliable way to distinguish alleles between parents of a mapping population. Presented here are the two strategies that were employed to map the candidate genes and published markers in the YP population. The procedure of primary use in the present study was Cleaved Amplified

Table 7. Primers used to map markers linked to the sex determination locus.

Marker Name	Marker Type and Source	Published Marker Primer Sequences	Size According to VBI
EMFN228	SSR	5′-TTGCTGAGGATTTGAAAATGG-3′	291 bp
	Sargent et al., 2006	5′-TGAAGTTTACTGGCGTTTGC-3′	
FAC004d	SSR	5′-GCCAATGTTCGATGTTTCACTA-3′	356 bp
	Lewers et al., 2005	5′-TCCTTGGGTCGATCACATAAAT-3′	
FAC012a	SSR	5′-TACACGTGTCCTAGGGTTTTCA-3′	169 bp
	Lewers et al., 2005	5′-AGCGGAGAATGAGTGACGATAG-3′	
CC2	RFLP	GenBank Accession: BI203070	GenBank:416 bp
	Vilanova et al., 2008		
PGLM	Phospho glyceromutase	5′-GTGATGGTGGAGTTCATTCTCG-3′	277 bp
	Sargent et al., 2007	5′-GGTCCTCTTCAAGTGTTTCAGC-3′	
ARSFL7	SSR	5′-GCGCGCATAAGGCAACAAAG-3′	214bp
	Lewers et al., 2005	5′-GCGAATGGCAATGACATCTTCTCT-3′	
Marker Name	**Mapping Method**	**YP Population Mapping Primer Sequences**	**Size According to VBI**
EMFN228	CAPS: *Taq*Ia Digest	5′-AGTGCAGGGCGTCTTCTTTGGTAT-3′	1.746kb
		5′-CATTGGGTGGGCTATCCTA-3′	
FAC004d	CAPS: *Dra*III Digest	5′-CACAACTGCCACCTAACTACCTC-3′	1.885kb
		5′-GTTAAAACGTGGGAATGTGAAAGA-3′	
FAC012a	AFLP	5′-GGTGGTGGCGGCAGCAGTAGG-3′	1.635kb
		5′-GTTTCGCGGTCGTCATCATTCGTC-3′	
CC2	CAPS: *Hin*fI Digest	5′-TTTCGGCTCTTCTCAACTCTCAG-3′	1.634kb
		5′-GTCGTCCCTCAGCACCTCACTTAT-3′	
PGLM	CAPS: *Ban*I Digest	5′-GCCGTTGATTTTATTGCGATTAGT-3′	1.709kb
		5′-GTGAAAAAGGGAGGCGGTGTTA-3′	
ARSFL7	CAPS: *Hae*III Digest	5′-CGCAAGACCTGCTGAAT-3′	1.547kb
		5′-GCGTATACAGTACTACTGGTG-3′	

Table 8. Primers used to map candidate genes.

Gene Name	Mapping Method	Mapping Primer Sequences	Product Size According to VBI
FRASUP1	CAPS: *Hpy*166II Digest	5′-GCAGCACCAAAAGTAGGATATG-3′	552 bp
		5′-GTGATTTCACTTTGCCCTG-3′	
FRASUP2	CAPS: *Bsr*I Digest	5′-AGGCTGATCAAGATTGGACGAGTA-3′	1.696kb
		5′-TCTGCGCATGCCTTTGGT T-3′	
FRASUP3	CAPS: *Mse*I Digest	5′-GCAGCGGGCATGGAGGTGTAT-3′	1.205kb
		5′-AAGGTGGCCAATGGGGGAAGTC-3′	
FRASUP4	CAPS: *Taq*1a Digest	5′-GACAGTTGCTTATATGGGACC-3′	2.014kb
		5′-CTAAGACTGAAACGAAAAGTGTTGG-3′	
FRASUP5	CAPS: *Msp*I Digest	5′-GGTGACCGTCCTAAGGTAAT-3′	1.501kb
		5′-CTACTACAGGTGGTGAGGG-3′	
ACS-7	CAPS: *Kpn*I Digest	5′-TCACCCCTCCGTCCATTCT-3′	1.404kb
		5′-GCACATATTCGGGGCTCAAG-3′	
AGAMOUS	CAPS: *Msp*I Digest	5′-TGA AAA CAA ACC AAA CAC TGA CCT-3′	3.225kb
		5′-TTCTTGTATCGTTCAATCGT-3′	

Polymorphic Sequence (CAPS) analysis. CAPS is analogous to RFLP analysis, but rather than digesting genomic DNA and detecting polymorphisms through probe hybridization, PCR products are digested and size separated by agarose gel electrophoresis. CAPS is used when the PCR products generated from the parents are the same size, and therefore cannot be mapped on the basis of PCR product length polymorphism. In this event, a sequence polymorphism that creates or eliminates a restriction site in one of the parents provides the basis for a restriction fragment size polymorphism that can be clearly discernible on an agarose gel.

The less frequently used strategy for mapping was simple PCR product length polymorphism analysis. This strategy directly uses differences in PCR product size (gel mobility) as polymorphisms for mapping. This form of mapping is obviously preferable to CAPS mapping as it forgoes the extra restriction enzyme digestion step that CAPS requires. However, given that readily distinguishable (agarose gel based) length polymorphisms were rarely observed, CAPS was used much more often.

When an obvious size difference between the PCR products of YW and PAWT was observed, mapping was performed by simply analyzing the PCR product size polymorphism. However, the majority of products from YW and PAWT were indistinguishable in size when run out on an agarose gel. Under these circumstances CAPS was employed to map these markers and candidate genes.

Discussion

L1L2 Phenotypic and Segregation Data

The equivalent average stamen counts from L1 and L2 support the claim from Ashman (1999) that females and hermaphrodites average the same number of stamens per flower. It also supports the hypothesis that stamen development (size) determines a plant's sex rather than the stamen number. It is clear that the number of stamen primordia in the females is equal to that of the hermaphrodites. It appears that females simply fail to follow through with the complete stamen developmental program.

According to the χ^2 Goodness of Fit Test, the null hypothesis was accepted, indicating that the L1L2 population is indeed segregating 1:1 for sex. This data further supports the Ahmadi-Bringhurst model (1991) that sex determination in *F. virginiana* is under the control of a single locus. This data also supports the Spigler et al. (2008; 2010) model that male fertility is under the control of a single locus whereby male fertility : sterility segregates 1:1.

ARSFL7 Genotyping

Spigler et al. (2008; 2010) used a total of 184 F1 individuals for their mapping study, for a maximum resolution of 0.54 cM. While the resolution of the L1L2 population (57 individuals) is only 1.75 cM, both populations gave very similar recombination frequencies between ARSFL7 and the sex locus, at least when comparing one of the studies. The recombination frequency of 8.8 cM between the sex determining locus and the ARSFL7 SSR in the L1L2 population is quite similar to the 6.8 cM published value found by Spigler et al. (2008) in their initial study. However, there is a notable discrepancy between both the aforementioned 8.8 cM and 6.8 cM distances and the 2 cM distance later reported by Spigler et al. (2010). It is difficult to account for the two disparate distances reported by Spigler et al. (2008; 2010), especially since both their

studies used the same mapping population and no new markers were added to the interval between ARSFL7 and the sex determination locus. Despite using the same population, it is clear that not all the same individuals were used in both studies. Given the dearth of markers available in the earlier study, Spigler et al. (2008) used the entire population of 184 individuals. However, in the later study (2010) the marker coverage for creating the map was increased nearly three fold. The drawback of this increased coverage was that five individuals were removed from the mapping dataset due to excessive genotypes at multiple loci which were inconsistent with Mendelian segregation patterns. Although not explicitly stated, it may be that these individuals constituted some recombinants between ARSFL7 and the sex locus which would inevitably lower the recombination frequency between these two loci when constructing the map. Despite inconsistencies with the three available map distances, it is still clear that the ARSFL7 SSR is closely linked to sex determination. This means that other markers linked to ARSFL7 in both the L1L2 and YP mapping populations have a high probability of also being linked to the sex determination locus.

Mapping with the Diploid YP Population

One of the most important results of the candidate gene mapping is that one of the *Fragaria SUPERMAN*-like genes is located on the sex determination chromosome. *FRASUP5* is located 38.5cM from ARSFL7 which is the marker on the top of YP LG6 and is also linked to sex in *F. virginiana* in both the L1L2 population and the Spigler population (Spigler et al., 2008; 2010). From the phylogenetic work done with the *FRASUP* genes, it is clear that *FRASUP5* belongs in the "*RABBIT EARS*" clade rather than the "*SUPERMAN*" clade. This is interesting in that if *FRASUP5* is indeed involved

in sex determination, it means that a gene more aptly defined as *RABBIT EARS*-like is controlling stamen development rather than the presumptive *SUPERMAN*-like gene. Still, it would be rash to jump to conclusions regarding the *FRASUP5* gene without knowing its function via transformation studies.

One other intriguing result of the mapping of the candidate genes in the YP population is the discovery that *ACS-7* is at the very end of Linkage Group 1. The significance of this map position derives from evidence presented by Davis et al. (1995b) who deduced that the putative differences in chromosome structure between diploid species *F. vesca* and *F. viridis* are due to translocations. This means that despite a great deal of genomic synteny between *Fragaria* species, there may still be large differences in chromosome structure. These putative translocations can serve as a driving force for speciation (especially if they are involved in sex determination) in the genus as well as being a powerful force for maintaining the sub-genomic (diploidized) composition of octoploid strawberry by preventing pairing of homeologous chromosomes.

Recently, Spigler et al. (2010) found direct evidence for translocations in the *F. virginiana* genome. By comparing the FVxFN diploid linkage map (Sargent et al., 2006;2007) to the *F. virginiana* linkage map they developed, they were able to find a number of markers which mapped to different linkage groups between the two species. One of these markers, ARSFL13 is quite intriguing in that there appears to be a translocation from LG1 in *F. vesca* to LG6 in *F. virginiana*. Furthermore, in the FVxFN map, ARSFL13 is at the end of Linkage Group 1. Unfortunately, the position of *ACS-7* relative to ARSFL13 is unknown at this time in either the FVxFN or *F. virginiana* maps. However, it is interesting to speculate whether *ACS-7* could have been part of the

translocation and may now reside on Linkage Group 6 in *F. virginiana*. If so, it may yet shown to be involved in sex determination in *F. virginiana*, especially given that the active gene yields a dominant phenotype (male sterility) in *Cucumis melo* and that male sterility in *F. virginiana* is dominant. Therefore, the *ACS-7* gene can be hemizygous in the *F. virginiana* genome due to a translocation and still impart a dominant phenotype.

With the generation of the YP linkage map, it is now possible to compare the three different *Fragaria* linkage maps: the YP map, the Spigler *F. virginiana* map (Spigler et al., 2010) and the Sargent integrated FVxFN map (Sargent et al., 2007). The Sargent FVxFN map is highly resolved with 37 markers mapped to LG6, while the YP and Spigler LG6 maps are wanting for LG6 markers with only 12 and 16, respectively. This is one impediment to comparative mapping; not all maps have the same markers. The population sizes of all three maps also vary markedly. In the YP population, 72 F2 individuals were used for mapping (representing 144 gametes), while Spigler et al. (2010) used 184 F1 progeny as a mapping population (representing 184 gametes from the female parent and 184 from the male). Despite the difference in size between the two populations they have very similar map resolution due to their population structure. Since the YP population is an F2 it has a resolution of 0.7 cM while the Spigler population is an F1 giving it a resolution of 0.5cM (for the female map). The integrated map developed by Sargent et al. (2007) combines the segregation data from the original FVxFN map with segregation from an interspecific backcross between *F. vesca* and *F. viridis* (FVxFV population). The total size of the mapping population is 187 individuals (94 from FVxFN F2 and 93 from FVxFV backcross) to give a resolution of 0.4cM (based on 281 gametes). Luckily, all the markers developed in the present study for mapping in the YP population

were codominant. This gives much better resolution to the map than using dominant markers, as it removes the ambiguity of whether a marker is heterozygous (presence/absence) or homozygous (presence/presence).

Despite the various differences in the derivation of the three maps, relationships can still be found among them which can be used to cement the structure and location of the sex determining locus. To begin with, all of the LG6 markers borrowed from the Sargent (2007) FVxFN map were able to be mapped to LG6 in the YP population. This is a good indication that there is large scale conservation of synteny between these two maps and may mean that marker loci borrowed from the FVxFN map have the potential to segregate with sex in the L1L2 population. In addition, the relative positions of markers between the FVxFN (Sargent et al., 2007) and YP maps are largely conserved. The only clear discrepancy is that the order of PGLM and FAC004d is reversed and that the most recent FVxFN map (Sargent et al., 2007) does not contains the CC2 marker. Of the six LG6 markers borrowed from the FVxFN map, only four (ARSFL7, PGLM, EMFn228, and FAC004d) markers were found in the Spigler *F. virginiana* map (Spigler et al., 2010). Of these four, ARSFL7, PGLM, and EMFn228 map to the LG6 homoeolog containing the sex determination locus. Aside from ARSFL7 which has already been shown to segregate with sex in the L1L2 population, PGLM, and EMFn228 should also be thoroughly studied to see if they do as well.

That the FAC004d marker did not map to the "sex homoeolog" in the Spigler (2010) *F. virginiana* map signifies one of the hazards of comparative mapping. Markers developed from diploid strawberry maps may not transfer easily to an octoploid *F. virginiana* map given that they are different species and vary markedly in ploidy. There is

always the chance that the marker will not amplify or that it is homozygous in all eight homeologous chromosomes. Furthermore even if a marker does segregate, the alleles may not be on the desired homoeolog, being homozygous or null in the other homoeologs.

Even when it is clear that a marker or gene belongs on a particular linkage group, there are still the issues of map distance and order relative to the other markers. In many parts of the genome, map distance can accurately reflect physical distances; however this is not always the case. Aside from underlying genetic components such as recombination suppression or recombination hot spots that can distort map distances, other factors can cause difficulty in translating map distances into physical distances. A small population size can adversely affect the reliability of the map. In small populations, a single recombinant may translate into a large recombination frequency while in a larger population, a recombination event has much less of an effect. Conversely, failure to find a recombinant between two loci in a small population may grossly underestimate the map distance (and consequently the physical distance) between two loci. The second issue that can increase the distance between two markers is an increase in number of markers on a linkage group. An increase in the number of markers between any two loci can increase the map distance between the two loci because double recombinants can be detected. Therefore the number of markers and size of the population play an important role in the size and reliability of the map.

One other challenge of comparative mapping is determining the correct order of the markers. Marker order between maps is not always maintained, especially between very closely linked markers or when there is segregation data missing for some population members. Indeed, the order of markers in the three maps can differ

significantly as in the case of EMFn228 which is located near ARSFL7 in the YP and Sargent FVxFN maps but is located at the opposite end of the chromosome in the Spigler (2010) *F. virginiana* map. In another case, the order of PGLM and FAC004d relative to the other markers in the YP and FVxFN maps is reversed.

There are a variety of reasons for the disagreement of marker order between maps, as with the case of PGLM and FAC004d. The first and most obvious is that the maps are built from populations of different species or accessions within a species. During evolution any number of chromosome deletions, duplications, translocations, and inversions could account for altered marker order. Another possibility could be missing marker data which makes it difficult for mapping software to predict the order. This problem can be compounded in highly saturated maps where the distance between two markers may be too close for the software to accurately assign order.

Maps are constantly evolving as new markers are added yet they serve as an invaluable compass for guiding answers to many genetic questions, especially when a candidate gene approach is taken. Although the overall map size and order of the markers may change over time, placing markers and genes on a map can give a strong indication of which genes to continue working with, especially if the associated phenotype has been mapped as well. Therefore knowing the actual map distance and order of *FRASUP5* is not as imperative as knowing that it is on the same linkage group as the sex determination locus. This however does not necessarily mean that *FRASUP5* is on the actual sex chromosome in *F. virginiana*. Even though *FRASUP5* and the sex determination locus are on LG6, it may be that they are on two separate homeologs in the octoploid strawberry. Even so, the focus of future work will now rest heavily on the study of this

gene. For instance, although not investigated in this study, it is of great interest to find whether *FRASUP5* is linked to sex in the L1L2 population.

CHAPTER 4

TRANSFORMATION OF CANDIDATE GENES

Abstract

Only after the seven candidate genes had been isolated and cloned could their functions be assessed. Gene function can be evaluated in either of two ways: blocking a gene's expression to create a knockout phenotype or over-expressing the gene. Both of these methods were employed on the seven candidate genes. To do this, the candidate genes were introduced into expression vectors using the Invitrogen Gateway cloning system (Invitrogen, Carlsbad, CA). The Gateway vectors utilized the constitutive 35S promoter from the cauliflower mosaic virus to transcribe the gene for either over-expression or knockout via RNA interference (RNAi). After using PCR and restriction enzyme digestion to ensure that the constructs contained the insert, the constructs were transformed into *Agrobacterium tumefaciens* for subsequent plant transformation. The ultimate goal is to introduce each of these constructs containing the native candidate gene back into *Fragaria vesca*, from which they were originally isolated. As an initial step toward this goal, the efficacy of the constructs was tested by using them to transform *Arabidopsis thaliana*. Of the fourteen constructs generated, twelve yielded positive Arabidopsis transformants harboring the candidate gene insert. Therefore it appears that at least these twelve constructs can be readily used for transformation into *F. vesca*.

Introduction

Transformation Overview

The isolation and characterization of the seven candidate genes was a significant step towards implicating one as the sex determination gene. Characterization thus far was only at the level of gene family and putative function. To truly understand the function of these genes we must observe the phenotype generated when the genes are silenced or over-expressed in vivo. To do this, candidate genes were cloned into vectors that allow the gene to be knocked out or over-expressed when introduced into a plant. The ultimate result of this work would be to see if one or more of the candidate genes generated a phenotype similar to the female flowers in *F. virginiana.* To accomplish this goal, all seven candidate genes were cloned into Gateway compatible vectors for use in plant transformation acquired from VIB-Research (University of Gent, Belgium). Each candidate gene was cloned into both an over-expression vector and an RNAi (gene silencing) vector.

The long term goal is to introduce the candidate gene containing vectors back into strawberry to determine the possible mutant phenotypes. However, in this study a preliminary transformation was performed with *Arabidopsis thaliana.* This was done because Arabidopsis has an extremely efficient transformation protocol and a rapid generation time. The strawberry community also has a transformation protocol (Oosumi et al., 2006) for diploid *F. vesca* which will be used eventually. However, for the present study it was found to be too laborious and time intensive within the confines of the project timeframe.

Gateway Cloning

In recent years, advances in gene cloning, vector development, and bacterial transfection have led to rapid and efficient transformation systems that can be applied to a

wide variety of plant species. One such breakthrough is the Gateway cloning technology introduced by Invitrogen which has allowed for the simple generation of clones for a variety of purpose such as over-expression, RNAi, and promoter analysis. The clones can be used for transformation of plant material through both *Agrobacterium tumefaciens* and biolistic methods. The overall strategy of the Gateway system is to use site specific recombinase enzymes to move the DNA fragment to be studied into the desired vector. To use this technology, two types of vectors are needed, each with its specific recognition sites (the att_ sequences) for the recombinase. In addition, a series of selectable markers are required to differentiate the desirable transformants of both bacteria and plant. The selectable markers for bacterial selection include a variety or antibiotic resistance genes in addition to the *ccdB* gene which is lethal to certain strains of *E. coli.*

One strong advantage that the Gateway system has over other cloning methods is control over the direction of the insert. Before the advent of the Gateway system, introduction of a DNA fragment into a vector was performed by digesting the vector with a restriction enzyme. The same enzyme was used to digest the fragment of interest so both the fragment and the vector had complementary ends. A DNA ligase was then added to the reaction and would ligate the vector to the DNA fragment. After transformation, clones had to be screened to ensure that the DNA fragment in the vector had been ligated in the proper orientation. A benefit of the Gateway system is that it provides directional cloning of the DNA fragment. Recombination between attB/attP and attL/attR sites is directional because sequence of the "att_" sites flanking either the DNA fragment of interest or the *ccdB* gene are not exactly the same and are designated attB1 and attB2 etc. So during the attB/attP recombination ("BP reaction"), attB1 sites will only

recombine with attP1 sites, and attB2 sites only recombine with attP2 sites. The same is true during the attL/attR recombination ("LR reaction") and thus directionality is maintained through the steps of Gateway cloning.

To begin, a PCR product of the DNA fragment to be studied is fitted with the site specific sequences at either end that will allow for recombination. This is performed using PCR with primers specific to the DNA fragment but with attB1 or attB2 specific sequence at primers' the 5′ end. The PCR product is then combined with a donor vector which contains the lethal *ccdB* gene flanked by the attP sequence, along with an antibiotic resistance gene. A recombinase known as BP Clonase is added to the reaction which initiates the recombination between the fragment of interest and the *ccdB* gene. The end result is that an entry clone is created with the fragment of interest but without the lethal *ccdB* gene. After recombination, the sequence flanking the insert is known as attL. The reaction is then transformed into *E. coli* TOP10 cells which are susceptible to the *ccdB* gene and the cells are plated out on the appropriate antibiotic. Cells that do not contain a plasmid die from the antibiotic while cells that received an unrecombined plasmid die from the *ccdB* toxicity. Positive transformants are then cultured, their plasmids isolated and confirmed by PCR or restriction enzyme digestion.

An entry clone is meant to propagate the DNA fragment of interest so it can be easily manipulated for the various types of transformation studies: over-expression, RNAi, or promoter analysis. The next step of the Gateway cloning system is to generate an expression clone which will be introduced into the plant cell through transformation. An expression clone contains all the elements needed to study the DNA fragment of interest in vivo. The expression clone begins with a destination vector which is analogous

to the donor vector. It also contains an antibiotic resistance marker (but not the same as the entry clone) and the lethal *ccdB* gene which is now flanked by attR sites.

To generate the expression clone, the entry clone is mixed with the destination vector. A second recombinase known as LR Clonase undergoes a recombination reaction between the attL sites of the entry clone and the attR sites of the destination vector. After recombination, the destination vector becomes the expression clone with the DNA fragment of interest now flanked by attB sites. The reaction is then transformed into *E. coli* TOP10 cells which are plated out on the appropriate antibiotic. Positive transformants can then be cultured and their expression clones isolated. Expression clones can then be transformed into *Agrobacterium* for subsequent plant transformation.

Elements of the Destination Vector

All destination vectors share a common backbone sequence which is needed for their proper function. The two vectors used in this study, pH7WG2D and pK7GWIWG2D(II) are based off the pPZP200 plasmid backbone (Hajdukiewicz et al., 1994). The critical components of the backbone include two origins of replication, one for *E. coli* and one for *Agrobacterium*, the right and left border sequence for transferring the insert into the plant (T-DNA), and a spectinomycin resistance gene for selection. Aside from these elements, each individual destination vector contains the unique elements for its specific function.

The pH7WG2D vector was used in this study to determine the phenotype caused by over-expression of the candidate genes in vivo. The purpose of this vector is to continually express the candidate gene in vivo using the 35S constitutive promoter. In addition, the other unique component of this vector is the hygromycin resistance gene

which is used to select for plants that have been transformed with the T-DNA. The components of the pH7WG2D vector can be seen in Figure 16a.

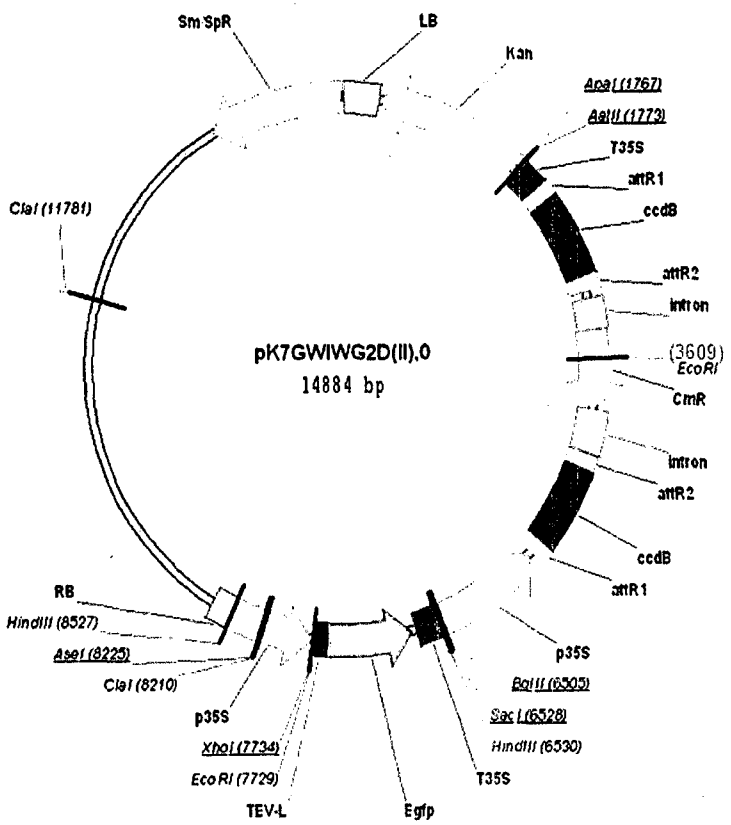
The pK7GWIWG2D(II) vector was used in this study to determine the phenotype caused by the silencing of the native candidate gene in vivo. This is done using RNAi, whereby a double-stranded RNA of the candidate gene is used to silence the native gene's expression. To do this, two copies of the candidate gene are inserted in opposite orientations separated by a short spacer sequence. The 35S promoter drives the transcription of the two candidate gene copies in a single transcript which forms a stem loop of dsRNA. The dsRNA is then processed by the RNAi pathway to silence the native gene. The pK7GWIWG2D(II) also contains a gene for kanamycin resistance used to select for plants that have been transformed with the T-DNA. The components of the pK7GWIWG2D(II) vector can be seen in Figure 16b.

Materials and Methods

Isolation of Full Length Genes from *F. vesca* var. Hawaii

The lack of any available cDNA clones for the candidate genes used in this study precluded the use of cDNA as a source for *Agrobacterium*-mediated transformation. As a result, full length genomic DNA was used as the gene source for cloning into the Gateway vectors and subsequent plant transformation. As the candidate genes were ultimately to be transformed into *F. vesca* var. Hawaii, it was concluded that 'Hawaii' would be the best source of genomic DNA for cloning the genes. This was easily accomplished using the Virginia Bioinformatics Institute (VBI) *F. vesca* reference genome database. Putative full length 'Hawaii' genomic sequence information (from start

a.



b.

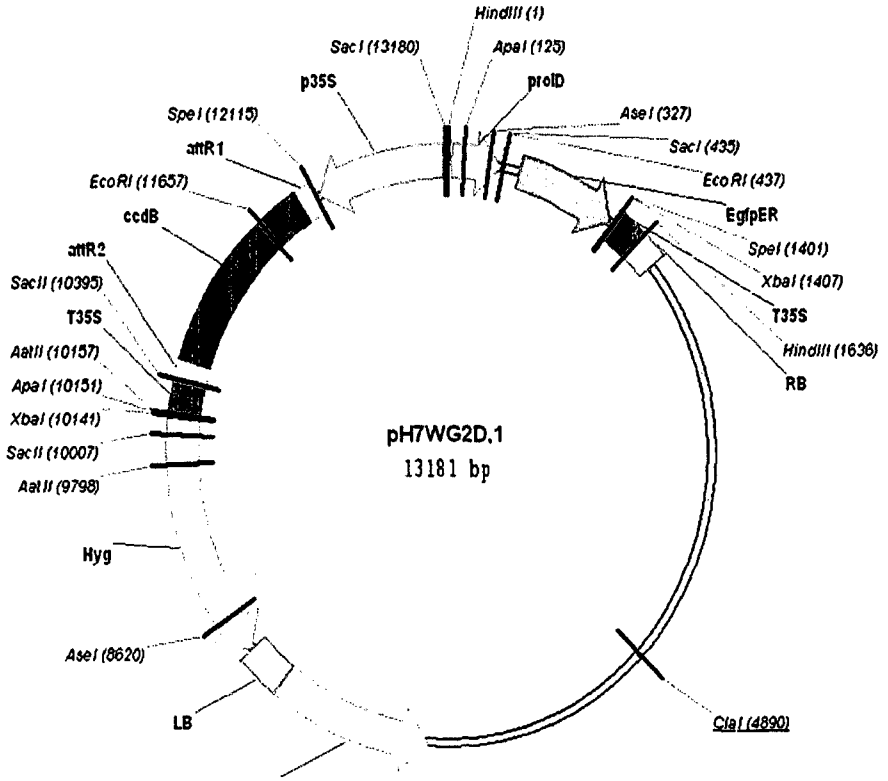


Figure 16. Plasmid maps of the vectors used for candidate gene transformation. Plasmids were obtained from VIB-Research (University of Gent, Belgium). **a.** The pHWG2D was used for over-expression studies. **b.** The pK7GWIWG2D(II) was used for RNAi studies.

codon to stop codon) was acquired by performing a BLAST of previously isolated full length or partial candidate gene sequence from *F. virginiana* against the *F. vesca* database. Contigs containing the candidate genes were then trimmed to span from the putative start codon to the putative stop codon. These edited sequences were used to design primers for the cloning of candidate genes into the Gateway vectors.

Transformation Primer Design

Primers spanning the complete coding region of the seven candidate genes were created in order to generate expression clones for plant transformation. Transformation primers were designed by using the first 20 bases of the gene as the forward primer and the last twenty bases of the gene as the reverse primer. In addition, 12 bp adaptors were added to the 5′ ends of the primers so all primers would have the following structure:

Forward: 5′ -AAAAAGCAGGCT-start codon plus 17 bases-3′

Reverse: 5′ -AGAAAGCTGGGT- stop codon plus 17 bases -3′

These chimeric primers were used to generate the first round PCR products needed for transformation. The purpose of the 12 bp adaptors is to provide a priming site for the attB1 and attB2 adaptor primers. The sequence of the attB adaptor primers serves as the site of recombination between the attB-containing PCR products and the attP-containing donor vector. All transformation primers and the attB1 and attB2 adaptor primers can be found in Table 9.

Generating PCR Products for Transformation

To generate the attB adaptor flanked PCR products, two rounds of PCR were performed with AccuPrime HiFi Taq and *F. vesca* var. Hawaii genomic DNA as the first round template. For the first round PCR, 25 µl reactions were prepared using 2.5 µl of

Table 9. Primer sequences used for the transformation of candidate genes. Products amplified by the primers span from the start to stop codon of each candidate gene and include attB adaptor sequence at the primers' 5′ end.

Transformation Primer	Primer Sequence
Universal attB1	5′-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3′
Universal attB2	5′-GGGGACCACTTTGTACAAGAAAGCTGGGT-3′
FRASUP1 F Trans	5′-AAAAAGCAGGCTATGATGGATTGTGTTAAGTT-3′
FRASUP1 R Trans	5′-AGAAAGCTGGGTCTAGAGATGCCCGAGTCGAA-3′
FRASUP2 F Trans	5′-AAAAAGCAGGCTATGAACACTGCTGGCTTCTG-3′
FRASUP2 R Trans	5′-AGAAAGCTGGGTTTAGCCTAATCGAAGCTCTA-3′
FRASUP3 F Trans	5′-AAAAAGCAGGCTATGCAGAGGAACAGTGGCTC-3′
FRASUP3 R Trans	5′-AGAAAGCTGGGTTCAGTAGCCTAATCGAAGCT-3′
FRASUP4 F Tra□s	5′-AAAAAGCAGGCTATGTTGAAGCAGTACTGCTA-3′
FRASUP4 R Trans	5′-AGAAAGCTGGGTTCATTTGATCTTGGGAGTTG-3′
FRASUP5 F Trans	5′-AAAAAGCAGGCTATGGAGCAAGCACGGTACTG-3′
FRASUP5 R Trans	5′-AGAAAGCTGGGTTTAATCTACTTTGATTACCT-3′
ACS-7 F Trans	5′-AAAAAGCAGGCTATGGCTATAGAGATTGAGCA-3′
ACS-7 R Trans	5′-AGAAAGCTGGGTCTAGTGCCTTCTCTCTCTTT-3′
AG F Trans	5′-AAAAAGCAGGCTATGGCCTATGAAAACAAACC-3′
AG R Trans	5′-AGAAAGCTGGGTTTACACTAACTGAAGGGAAA-3′

10X AccuPrime Buffer II, 0.4 μM of both forward and reverse gene primers, 0.5 U HiFi Taq DNA polymerase and 60 ng of template DNA. The PCR profile for the first round amplification was an initial denaturation at 94°C for 1 minute followed by 10 cycles at 94°C for 30 seconds, 52°C for 30 seconds and 68°C for 30 seconds per 500 bp, and ending with a final extension at 68°C for 10 minutes.

For the second round PCR, 25 μl reactions were prepared using 2.5 μl of 10X AccuPrime Buffer II, 0.12 μM of both forward and reverse attB adaptor primers, 0.5U AccuPrime HiFi Taq DNA polymerase and 6.2 μl of first round PCR product as template DNA. The PCR profile for the second round amplification was an initial denaturation at 94°C for 1 minute followed by 5 cycles at 94°C for 30 seconds 45°C for 30 seconds and 68°C for 30 seconds per 500 bp, and ending with a final extension at 68°C for 10 minutes. The thermocycler profile was repeated for another 30 cycles but with the annealing temperature at 55°C. 10 μl of second round PCR products were visualized on a 1% agarose 1X TBE gel run at 100V for 80 minutes. Finally, 10 μl of the second round PCR products were then column purified using the Promega SV PCR-Clean-Up System to remove unincorporated attB adaptor primers.

Performing the BP Reaction

For each candidate gene, a BP reaction was performed to recombine the candidate gene into the pDONR221 vector (Invitrogen, Carlsbad, CA), thus creating an entry clone. For each 10 μl BP reaction, 2 μl of purified attB flanked PCR product was combined with 150 ng of pDONR221, 2 μl BP-Clonase II Enzyme mix (Invitrogen, Carlsbad, CA) and brought up to 10 μl with TE buffer (pH 8.0). The reactions were incubated in the

dark at room temperature for one hour. At that time 1 µl of supplied proteinase K solution was added to each BP reaction to stop the reaction.

To transform the newly generated entry clones into *E. coli*, 2 µl of the BP reaction was added to 25 µl of TOP10 chemically competent cells. The cells were incubated for 10 minutes on ice then transferred to a water bath at 42°C for 30 seconds and returned to ice. 125 µl of supplied SOC medium was added to each tube of cells and the cells were shaken at 250 RPM at 37°C for an hour. Cells were then plated on LB agar (10 g/L tryptone, 10 g/L NaCl, 5 g/L hydrolyzed yeast, 15 g/L agar) containing 50 µg/mL kanamycin and grown at 37°C for 12-16 hours.

Confirmation of BP Clones

Individual colonies were screened for inserts using the colony screen method as described under the General Methods section of Chapter 1. Positive transformants were cultured in 2 ml of liquid LB medium containing 50 µg/mL kanamycin shaken at 250 RPM at 37°C for 14-16 hours. Plasmids were isolated and quantified on a NanoDrop spectrophotometer. To further confirm the presence of the insert, 100 ng of purified plasmid (and empty vector as a negative control) was digested with 2 U of the restriction enzyme *Bsr*GI (New England Biolabs, Ipswich MA) in a 20 µl reaction overnight at 37°C. The digest products were then run out on a 1% agarose 1X TBE gel run at 100 V for 80 minutes. *Bsr*GI was used to confirm the presence of inserts as it cuts within the attB adaptor primer sequence, thus cleaving the insert from the entry clone. After digestion confirmation, entry clones were sequenced on a Applied Biosystems 3130 Genetic Analyzer using the M13F and M13R vector primers as described above. In the case of *AGAMOUS*, only the coding region and smaller introns were sequenced, while the large 3

kb intron was not fully sequenced. Entry clone sequences were trimmed of vector and aligned to the putative full length VBI reference sequence to determine if there were any mutations in the entry clones.

Performing the LR Reaction

Once the integrity of the entry clones was confirmed, they were used to perform an LR reaction to generate the expression clone for plant transformation. To perform the LR recombination reaction, 150 ng of either the pH7WG2D over-expression destination vector or the pK7GWIWG2D(II) RNAi vector, 100 ng of the entry clone and 1 µl of LR Clonase II enzyme mix (Invitrogen, Carlsbad, CA) were combined and brought up to 10 µl with TE buffer (pH 8.0). The reactions were incubated in the dark at room temperature for two hours.

To transform the newly generated entry clones into *E. coli*, 2 µl of the LR reaction was added to 25 µl of TOP10 chemically competent cells. The cells were incubated for 10 minutes on ice then transferred to a water bath at 42°C for 30 seconds and returned to ice. 125 µl of supplied SOC medium was added to each tube of cells and the cells were shaken at 250 RPM at 37°C for an hour. Cells were then plated out on LB agar containing 50 µg/mL spectinomycin and grown at 37°C for 12-16 hours.

Confirmation of LR Clones

Individual colonies were screened for inserts as described above using the attB adaptor primers. Positive transformants were grown in 4 ml of LB liquid medium containing 50 µg/mL spectinomycin and shaken at 250 RPM at 37°C for 14-16 hours. Expression clones were isolated, quantified, and correct inserts were confirmed by *Bsr*GI digestion as described above. In addition, PCR was also performed on individual

expression clones using gene specific transformation primers to confirm the correct insert. This was done following the standard Lucigen EconoTaq protocol (General Methods).

Agrobacterium Transformation

After confirming the proper insert, both over-expression and RNAi expression clones were transformed into *Agrobacterium* strain GV3101. Cells were transformed with an Eppendorf 2510 Electroporator using 0.2 µl of expression clone and 40 µl of electrocompetent *Agrobacterium* cells. 250 µl of SOC was added to each transformation and cells were shaken at 250 RPM at 28°C for one hour. 40 µl of transformed cells were plated on LB agar containing 50 µg/mL spectinomycin and grown at 28°C for two days.

Confirmation of *Agrobacterium* Clones for Plant Transformation

Individual *Agrobacterium* colonies were screened for inserts by PCR following the standard Lucigen EconoTaq protocol (General Methods) using the attB adaptor primers. A single positive clone was then sub-cultured by streaking it on an LB agar plate containing 50 µg/mL spectinomycin. Plates were incubated at 28°C for 24 hours.

Transformation of Wild Type Arabidopsis

Transgenic plants were generated via the floral dip method (Clough and Bent, 1998) using wild type *Arabidopsis thaliana* ecotype Colombia-0 plants. To begin the transformation protocol, roughly 20 wild type seeds were sprinkled on growth media composed of a 1:1 mixture of Metromix 360 potting soil (SUN GRO Horticulture, Bellevue, WA) and perlite (Whittemore Company, Inc, Lawrence, MA). To ensure germination and continued growth, pots were placed in well-watered flats and moved to a growth chamber. The growth chamber conditions for the entirety of the experiment were: 21°C, a 16 hour photoperiod, and 100 µmol m^{-2} s^{-1} fluorescent light. When seedlings

developed 3-4 true leaves, each pot was thinned to contain between four and six plants. Aside from watering, plants were left undisturbed for 3-4 weeks until flower bolts emerged. At that time, the primary bolts were removed at the base of the plant with scissors to encourage growth of secondary bolts.

Once secondary bolts had developed, a starter culture of each *Agrobacterium* clone was grown using an isolated colony from each sub-cultured plate. The colony was used to inoculate 2 mL of liquid LB medium containing 50μg/mL spectinomycin and shaken at 250 RPM at 28°C for 14-16hours. The entire 2 ml starter culture was then used to inoculate 200 mL of liquid LB and grown under the same conditions for 12-14 hours. Cells from the 200 mL culture were harvested by centrifugation at 4000 x g for 20 minutes. The pelleted cells were resuspended in 5% sucrose (W/V) solution by vortexing and vigorous shaking. The cell suspension was diluted to an O.D. 600 of 0.800 ± 0.200 and Silwet L-77 (Lehle Seeds, Cat. No. VIS-02, Round Rock, TX) was added to a final concentration of ~0.05% (V/V).

To perform the Arabidopsis transformation, the entire *Agrobacterium* cell suspension was decanted into small Tupperware containers for dipping. The bolts of the plants were then submerged in the suspension for 15-20 seconds, being careful not to allow soil to get in the suspension. Four pots of plants were used for each individual transformation. The pots were laid on their sides overnight in a flat covered with a plastic dome to maintain humidity. The next day, plants were rinsed with water and grown at 21°C, 16 hour photoperiod, and 100 ± 10 μmol m^{-2} s^{-1} of fluorescent light. After 3-4 weeks, bolts containing mature siliques were removed from the plants and the seeds harvested and dried.

To select for transformants, 80mg of seeds were surface sterilized for each Petri plate containing selective media. A maximum of four plates for each construct were used although plants harboring some constructs produced less than the minimum amount of seed (320 mg) for this purpose. Each 80mg portion of seed was sterilized in 10 mL Falcon tubes (Falcon Plastics, Brookings, SD) as follows: 9 mL of 70% ethanol with three drops of 10% Triton, followed by agitation for five minutes. The 70% ethanol was decanted and replaced with 9 mL of 100% ethanol and 3 drops 10% Triton, followed by agitation for five minutes. The 100% ethanol was decanted for a final wash containing 9 mL of 100% ethanol. This was agitated for five minutes and then the ethanol was decanted. Residual ethanol was removed by laying open tubes horizontally in a laminar flow hood over night.

After sterilization and drying, seeds were sprinkled evenly across the selective plate. Selective medium was composed of 0.5X MS, pH 6, 0.8% PhytoBlend (Phytoblend, Caisson Labs., Inc., North Logan, UT) and either 60 μg/mL Hygromycin for the pH7WG2D (over-expression) containing constructs or 50 μg/mL Kanamycin for the pK7GWIWG2D(II) (RNAi) containing constructs. The plates were sealed with Micropore tape (3M, Minneapolis, Minnesota) and placed in the dark at 4°C for four days. After four days, plates were transferred to a growth chamber and grown at 21°C, 16 hour photoperiod, and 100 ± 10 μmol m^{-2} s^{-1} of fluorescent light. After two weeks under these conditions, plates were ready for the positive transformants to be removed.

Transformants were deemed "healthy" if they continued to grow on selective medium beyond initial germination (emergence of the radicle and cotyledon). Seedlings were chosen that either had (relatively) large cotyledons or had already developed true

leaves. Seedlings were removed from the agar substrate, being careful to remove attached un-transformed seedlings and planted in a 1:1 mixture of Metromix 360 potting soil (SUN GRO Horticulture, Bellevue, WA) and perlite (Whittemore Company, Inc, Lawrence, MA). Four seedlings per pot with a maximum of four pots per construct were sown. Pots were transferred to well-watered flats and grown at 21°C, 16 hour photoperiod, and 100 ± 10 μmol m^{-2} s^{-1} of fluorescent light. After 3-4 weeks under these growth conditions, plants had sufficient tissue for DNA extraction using the "smash prep" method.

Smash Prep DNA Extraction from Arabidopsis Positive Transformants

The "smash prep" is a method for rapid DNA extraction for short term use developed by Klimyuk et al. (1993). To perform a smash prep, a small piece of leaf tissue roughly 0.5 cm^2 was excised from each positive transformant and transferred to a 1.5 ml microfuge tube. Next, 40 μl of 0.25 M NaOH was added and the leaf piece was macerated with a pipette tip until the liquid turned green. Samples were then boiled for 30 seconds, after which 40 μl of 0.25 M HCl and 20 μl of 0.5 M Tris-HCl (pH 8.0) containing 0.25% (v/v) Nonidet P-40 were added to neutralize to sample. The sample was then boiled an additional 2 minutes and stored at 4°C.

Confirmation of Positive Arabidopsis Transformants

PCR was performed to determine whether the Arabidopsis plants that survived the selection process harbored the particular candidate gene. Following the standard protocol for Lucigen EconoTaq, half-sized reactions (12.5 μl) were made for each plant to be assayed. Each PCR contained the attB forward and reverse primers and 0.5 μl of smash prep DNA as template. The thermocycler profiles for amplifying both templates was as

follows: initial denaturation at 94°C for 1 min; then 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min; followed by a final extension at 72°C for 10 minutes. This was only done for plants transformed with a *FRASUP* candidate gene since all these genes are under 1kb. Since the smash prep template is of such poor quality, it was a concern that the larger *ACS-7* and *AGAMOUS* constructs would not amplify using a standard Taq polymerase. Rather, 0.5 µl of the *ACS-7* and *AGAMOUS* smash prep template DNA was amplified with AccuPrime HiFi Taq in half reaction volumes (12.5 µl) following the standard Invitrogen HiFi Taq protocol. The *ACS-7* gene was amplified with the attB forward and reverse primers while the *AGAMOUS* gene was amplified with gene specific primers (F: 5′-ACGATTGAACGATACAAGAAG-3′, R: 5′-CACTAACTGAAGGGAAACTTG-3′). The thermocycler profiles for amplifying both templates is as follows: initial denaturation at 94°C for 1 min; then 30 cycles of 94°C for 30 sec, 52°C for 30 sec, 68°C for 2 min; followed by a final extension at 68°C for 10 minutes. Total PCR product from all reactions was then separated on a 1% Agarose 1X TBE gel run at 100 V for 80 minutes.

In addition to testing for the presence of the insert, a series of positive and negative PCR controls were performed using wild type *Arabidopsis thaliana* (Colombia-0) as a template. Arabidopsis DNA was isolated by the smash prep method as described above. For the negative controls, attB primers were used to amplify potential loci in Arabidopsis both with the Lucigen EconoTaq and the Invitrogen HiFi taq protocols employing the same parameters for amplifying inserts. Furthermore, EconoTaq was also used to amplify potential products from Arabidopsis with all seven gene specific transformation primer pairs and the *AGAMOUS* gene specific primer pair (above). For

the positive control, the Hrabak Lab donated the 2AC4 primer pair (2AC43: 5′-CTACAGAGAAATGCGGACGAT-3′, 2AC44: 5′-TTGGGCAAAGACCTGAATGAA-3′), which is known to amplify a ~700 bp fragment in Arabidopsis. For both positive and negative controls, total PCR product was then separated on a 1% Agarose 1X TBE gel run at 100 V for 80 minutes.

Results

Confirmation of Gateway Clones

Sequencing of the candidate gene inserts within the entry clones yielded no discrepancy between the predicted amino acid sequence from the reference *F. vesca* 'Hawaii' database and the insert. Very rarely, there would be a sequence polymorphism between the chromatogram and the published *F. vesca* var. 'Hawaii' genome sequence. This occurrence was either due to a sequencing error or may be due to alternate alleles that were sequenced from the 'Hawaii' genome. Fortunately, these polymorphisms occurred within an intron or when in an exon would have resulted in silent mutations.

Digestion of both the entry clones and expression clones generated products displaying the desired insert lengths when separated on an agarose gel. The size of the insert digest product was the same size as the insert since *Bsr*GI did not cut in any of the candidate genes. PCR of the isolated expression clones with candidate gene specific primers also yielded product of the appropriate size as visualized by agarose gel electrophoresis. A colony screen for each of the *Agrobacterium* transformations was able to find at least one isolated positive clone for sub-culturing.

Transformation of Arabidopsis

Overall, the transformation of Arabidopsis with the fourteen different constructs was very successful. Positive selection was seen in twelve out of the fourteen transformations (Table 10). Wild type plants transformed with the RNAi construct of the *ACS-7* gene did not produce any positive transformants. Seeds from the *FRASUP3* RNAi construct transformation were accidentally plated on the wrong selective media. Aside from these two mishaps, all other plates produced ample positive transformants.

Confirmation of Arabidopsis Transformants

Inevitably, some positive transformants that were transferred to soil died, possibly due to damage during transfer or from some effect of the introduced construct. However, a total of 151 plants from 12 different constructs were able to grow well enough to provide a source of tissue for assay. Of these, a total of 90 plants tested positive for the presence of the insert using PCR (Table 10). This works out to be an overall regeneration efficiency of 60%. The total number of both plants screened and the number of transformants testing positive for the insert can be seen in Table 10.

Discussion

The overall goal of the Arabidopsis transformations was to determine with relative ease whether the over-expression and RNAi constructs can be used to transform *Fragaria vesca.* Although there is no way of knowing this for sure without actually undergoing the *F. vesca* transformations, it is now known that the construct-harboring *Agrobacterium* transformants contain all the necessary elements for an efficient plant transformation procedure, such that failure in the effort to transform *F. vesca* will not be attributable to a defect in the transformation vector.

Table 10. Transformation of *Arabidopsis thaliana* with seven candidate genes. All constructs yielded positive transformants except for the *FRASUP3* RNAi and *ACS-7* RNAi constructs.

Candidate Gene	Expression Vector	Number of Plants PCR Screened Following Positive Antibiotic Selection	Number of Plants with Candidate Gene
FRASUP1	Over-Expression	2	1
	RNAi	12	12
FRASUP2	Over-Expression	16	7
	RNAi	12	10
FRASUP3	Over-Expression	12	7
	RNAi	0	0
FRASUP4	Over-Expression	20	4
	RNAi	12	8
FRASUP5	Over-Expression	15	11
	RNAi	12	5
ACS-7	Over-Expression	12	4
	RNAi	0	0
AGAMOUS	Over-Expression	14	13
	RNAi	12	8
	Total	151	90

Despite the success with transformation, still two of the constructs, *FRASUP3* RNAi and *ACS-7* RNAi await confirmation. The *FRASUP3* RNAi construct may well have yielded positive transformants if the seeds had not been accidentally plated on the wrong selective medium. However, the *ACS-7* RNAi construct was plated on the correct medium yet produced no positive transformants. There are two likely explanations for this result. First, perhaps there were simply no transformants generated during the floral dip procedure. This could be caused by some key element missing from either the vector or the *Agrobacterium*. Even if all the proper elements for transformation were present in the *Agrobacterium*, since actual transformation is such a rare event, it may be that it just did not occur on the inflorescences. The second explanation for a lack of transformants is that knocking out the *ACS-7* gene by RNAi results in lethality. The *ACS* gene family is involved in the rate limiting step of ethylene synthesis and ethylene is a vital plant hormone involved in numerous regulatory processes. The consequence of knocking out *ACS* gene expression may wreak havoc on the plant's ability to function properly, killing it or preventing seed germination. It may be that one way to generate a mature positive transformant from this construct would be to provide exogenous ethylene during the selection phase of the transformation.

When it can be determined whether the *FRASUP3* RNAi and *ACS-7* RNAi constructs can generate positive transformants, the final step will be to transform the fourteen constructs into *F. vesca* var. Hawaii. This work will be performed using the protocol developed by Oosumi et al. (2006) for *Agrobacterium* mediated transformation of diploid strawberry. It will be of great interest to determine whether over-expression

and RNAi of the seven native genes will result in interesting and possibly informative floral phenotypes.

The actual number of Arabidopsis plants containing an insert is not as important as knowing that the constructs can be successfully employed to transform plant tissue. However, these Arabidopsis transformants are still an invaluable resource, especially as each one is an independent transformant. Once the plants are mature, seeds can be harvested from the positive transformants and can be germinated to yield T2 lines. Furthermore, these T2 lines can give valuable data in terms of both mutant phenotypes and expression via RT-PCR. Phenotypes may differ from phenotypes (or lack thereof) observed in the T1. This may either due to increased transgene copy number in homozygous T2 plants or that a transgene silenced in the T1 generation may have become un-silenced in the T2 generation. It is also of interest to see whether the T-DNA inserts are inherited stably in subsequent generations. The phenotypes generated in the T2 lines can then be compared with the phenotypes generated from the *F. vesca* transformation lines as a further line of confirmation for each gene's function. However there is always the possibility of generating phenotypes in the T1 and T2 not attributable to the transgenes. These phenotypes can be ascribed to the T-DNA inserting into a native gene (creating a knockout) or by activating an endogenous gene through the 35S promoter in the T-DNA. Insertional knockouts can only be seen in the T2 generation because the knockout needs to be homozygous to create a phenotype.

One lesson learned from the Arabidopsis transformations of candidate genes is to be aware of the high frequency of false positives, meaning those plants which survived the selection process but did not contain an insert. Based on PCR genotyping results, 40%

of the "transformants" were false positives. Two plausible reasons exist for getting false positives. First, the seedlings did not have enough contact with the selective media, which can happen when seeds are plated too thickly on the media. Second, the plants do actually contain an insert but the quality of the smash prep template is too poor to amplify the insert. Whatever the reason, a rigorous selection process should be taken during the strawberry transformation to minimize false positive transformants.

CONCLUSIONS

Study Summary

This study has been a great step forward in elucidating the genetic cause of sex determination in octoploid *Fragaria virginiana*. In total, seven genes, including five members of the strawberry *SUPERMAN*-like gene family have been isolated, characterized, and mapped. Although none of these genes has been directly implicated in sex determination in strawberry, it is well established that *AGAMOUS*, *SUPERMAN*, and *ACS-7* all play an important role in flower development in other species. However, this study has found tantalizing evidence that one of these *Fragaria SUPERMAN*-like genes may influence sex, as it has been mapped to the same chromosome in diploid strawberry that the sex determination locus is on in *Fragaria virginiana*.

The characterization and mapping of these genes yields valuable insight into whether any are involved in sex determination. However, it is through transformation of these genes that they can be truly implicated as the cause. The present study has made great strides in this avenue of work. All seven of the candidate genes have been cloned into transformation vectors to generate over-expression and knockout phenotypes. The phenotypes generated by over-expressing or knocking out these genes will give a better understanding of whether they are involved in flower development in strawberry. Genes whose phenotypes mimic those of female *F. virginiana* flowers can then undergo more rigorous study.

Future Directions

As is true with all of science, more answers inevitably lead to more questions. Despite all the progress that has been made, there is still much left to do in order to find the cause of sex determination in strawberry. In the short term it is imperative to introduce the candidate genes back into diploid strawberry to see if their floral phenotypes mimic those of female *F. virginiana*, namely vestigial stamens. Another short term goal is to develop additional markers more closely linked to sex than the ARSFL7 SSR. This will better characterize the sex determination locus, making it potentially easier to isolate.

We have now reached the age of genomics, whereby entire eukaryotic genomes can be sequenced, assembled, and annotated with relative ease and cost. The leap of progress made in this study given the access to the recently sequenced *F. vesca* genome database is testimony to how valuable a resource genomics has been to the life sciences. This resource was invaluable for isolating candidate genes and for marker development. Although the genome is over 95% sequenced, it remains in fragments of large contigs and is poorly annotated. As it is quickly becoming a cornerstone of research for the strawberry community it is imperative that the *F. vesca* genome be fully sequenced and assembled. A fully sequenced and assembled genome will allow future researchers to develop markers for fine mapping of the region associated with sex determination. In addition, with substantial annotation it should also be possible to find additional candidate genes in this area that can be further studied.

Although the *F. vesca* genome sequence has already proved its worth throughout the present study, it is still heterologous sequence and so is not as valuable a

resource as a sequenced, preferably female, *F. virginiana* genome would be. The sequencing of a genome from this species would reveal the genetic structure of the actual sex determination locus. Comparative genomics between the *F. vesca* and *F. virginiana* genomes will also give insight into possible translocations that may be important in the evolution of sex determination. Although it may be some years away, the sequencing of the *F. virginiana* genome and the answer to genetic basis of sex determination will be awaited with eager anticipation.

LIST OF REFERENCES

Ahmadi H, Bringhurst RS (1991) Genetics of Sex Expression in *Fragaria* Species. American Journal of Botany **78**: 504-514

Albani MC, Coupland G (2010) Comparative Analysis of Flowering in Annual and Perennial Plants. Current Topics in Developmental Biology **91**: 323-348

Alexander HM (1990) Epidemiology of Anther-Smut Infection of Silene Alba Caused by Ustilago Violacea: Patterns of Spore Deposition and Disease Incidence. Journal of Ecology **78**: 166-179

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. 215: 403-410

Ashley MV, Wilk JA, Styan SMN, Craft KJ, Jones KL, Feldheim KA, Lewers KS, Ashman TL (2003) High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill. (Rosaceae). Theoretical and Applied Genetics **107**: 1201-1207

Ashman TL (1999) Quantitative genetics of floral traits in a gynodioecious wild strawberry *Fragaria virginiana*: implications for the independent evolution of female and hermaphrodite floral phenotypes. Heredity 83: 733-741

Bereterbide A, Hernould M, Castera S, Mouras A (2001) Inhibition of cell proliferation, cell expansion and differentiation by the Arabidopsis SUPERMAN gene in transgenic tobacco plants. Planta **214**: 22-29

Bereterbide A, Hernould M, Farbos I, Glimelius K, Mouras A (2002) Restoration of stamen development and production of functional pollen in an alloplasmic CMS tobacco line by ectopic expression of the *Arabidopsis thaliana* SUPERMAN gene. The Plant Journal **29**: 607-615

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16: 1667-1678

Bors RH, Sullivan JA (2005) Interspecific hybridization of *Fragaria moschata* with two diploid species, F. nubicola and F. viridis. Euphytica **143**: 201-207

Boualem A, Fergany M, Fernandez R, Troadec C, Martin A, Morin H, Sari M-A, Collin F, Flowers JM, Pitrat M, Purugganan MD, Dogimont C, Bendahmane A (2008) A Conserved Mutation in an Ethylene Biosynthesis Enzyme Leads to Andromonoecy in Melons. Science **321**: 836-838

Boulter CA, Natarajan D (1999) A PCR-based method for isolation of genomic DNA flanking a known DNA sequence. Curr Issues Mol Biol **1**: 47-52

Bowman JL, Drews GN, Meyerowitz EM (1991a) Expression of the Arabidopsis floral homeotic gene AGAMOUS is restricted to specific cell types late in flower development. Plant Cell **3**: 749-758

Bowman JL, Sakai H, Jack T, Weigel D, Mayer U, Meyerowitz EM (1992) SUPERMAN, a regulator of floral homeotic genes in Arabidopsis. Development **114**: 599-615

Bowman JL, Smyth DR, Meyerowitz EM (1991b) Genetic interactions among floral homeotic genes of Arabidopsis. Development **112**: 1-20

Bringhurst RS (1990) Cytogenetics and Evolution in American *Fragaria*. Hort. Sci. **106**: 679–683

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

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature **353**: 31-37

Cormack RS, Somssich IE (1997) Rapid amplification of genomic ends (RAGE) as a simple method to clone flanking genomic DNA. Gene **194**: 273-276

Dathan N, Zaccaro L, Esposito S, Isernia C, Omichinski JG, Riccio A, Pedone C, Di Blasio B, Fattorusso R, Pedone PV (2002) The Arabidopsis SUPERMAN protein is able to specifically bind DNA through its single Cys2-His2 zinc-finger motif. Nucl. Acids Res. **30**: 4945-4951

Davis TM, Yu H, Haigis KM, McGowan PJ (1995a) Template mixing: a method of enhancing detection and interpretation of codominant RAPD markers. TAG Theoretical and Applied Genetics **91**: 582-588

Davis TM, Yu, H. and Haymes, K. M. (1995b) Genetic linkage group structure in the diploid strawberry: Evidence for chromosome differentiation between Fragaria vesca and F. viridis. Proceedings of the IV North American Strawberry Conference Orlando, FL, USA

Deyholos MK, Sieburth LE (2000) Separable Whorl-Specific Expression and Negative Regulation by Enhancer Elements within the AGAMOUS Second Intron. Plant Cell **12**: 1799-1810

Englbrecht CC, Schoof H, Bohm S (2004) Conservation, diversification and expansion of C2H2 zinc-finger proteins in the *Arabidopsis thaliana* genome. BMC Genomics **5**: 39

Fedorova NJ (1946) Crossability and phylogenetic relations in the main European species of Fragaria. Compte-rendu de l'académie des Sciences de l'URSS **52**: 545–547

Folta KM, Davis TM (2006) Strawberry Genes and Genomics. Critical Reviews in Plant Sciences **25**: 399 - 415

Gaiser JC, Robinson-Beers K, Gasser CS (1995) The Arabidopsis SUPERMAN Gene Mediates Asymmetric Growth of the Outer Integument of Ovules. Plant Cell **7**: 333-345

Hancock JF, Bringhurst RS (1979) Hermaphroditism in Predominately Dioecious Populations of Fragaria chiloensis (L.) Duchn. Bulletin of the Torrey Botanical Club **106**: 229-231

Hancock JF, Luby JJ (1993) Genetic Resources at Our Doorstep: The Wild Strawberries. BioScience **43**: 141-147

Hill TA, Day CD, Zondlo SC, Thackeray AG, Irish VF (1998) Discrete spatial and temporal cis-acting elements regulate transcription of the Arabidopsis floral homeotic gene APETALA3. Development **125**: 1711-1721

Hiratsu K, Mitsuda N, Matsui K, Ohme-Takagi M (2004) Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in Arabidopsis. Biochemical and Biophysical Research Communications **321**: 172-178

Hiratsu K, Ohta M, Matsui K, Ohme-Takagi M (2002) The SUPERMAN protein is an active repressor whose carboxy-terminal repression domain is required for the development of normal flowers. FEBS Letters **514**: 351-354

Huang H, Ma H (1997) FON1, an Arabidopsis Gene That Terminates Floral Meristem Activity and Controls Flower Organ Number. The Plant Cell **9**: 115-134

Irish VF, Sussex IM (1990) Function of the apetala-1 Gene during Arabidopsis Floral Development. Plant Cell **2**: 741-753

Jack T, Fox GL, Meyerowitz EM (1994) Arabidopsis homeotic gene APETALA3 ectopic expression: transcriptional and posttranscriptional regulation determine floral organ identity. Cell **76**: 703-716

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated SUPERMAN epigenetic alleles in arabidopsis. Science **277**: 1100-1103

Kater MM, Franken J, Van Aelst A, Angenent GC (2000) Suppression of cell expansion by ectopic expression of the Arabidopsis SUPERMAN gene in transgenic petunia and tobacco. The Plant Journal **23**: 407-413

Kazama Y, Fujiwara MT, Koizumi A, Nishihara K, Nishiyama R, Kifune E, Abe T, Kawano S (2009) A SUPERMAN-like Gene is Exclusively Expressed in Female Flowers of the Dioecious Plant Silene latifolia. Plant Cell Physiol. **50**: 1127-1141

Klimyuk VI, Carroll BJ, Thomas CM, Jones JDG (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis. Plant Journal **3**:493-494

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

Krizek BA, Lewis MW, Fletcher JC (2006) RABBIT EARS is a second-whorl repressor of AGAMOUS that maintains spatial boundaries in Arabidopsis flowers. The Plant Journal **45**: 369-383

Krizek BA, Meyerowitz EM (1996) The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. Development **122**: 11-22

Lewers KS, Styan SMN, Hokanson SC, Bassil NV (2005) Strawberry GenBank-derived and genomic simple sequence repeat (SSR) markers and their utility with strawberry, blackberry, and red and black raspberry. Journal of the American Society for Horticultural Science. **130**: 102-115

Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25**: 674-681

Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. The Plant Journal **8**: 457-463

Meyerowitz EM (1997) Genetic Control of Cell Division Patterns in Developing Plants. Cell **88**: 299-308

Meyerowitz EM, Bowman JL, Brockman LL, Drews GN, Jack T, Sieburth LE, Weigel D (1991) A genetic and molecular model for flower development in *Arabidopsis thaliana*. Dev Suppl **1**: 157-167

Mizukami Y, Huang H, Tudor M, Hu Y, Ma H (1996) Functional Domains of the Floral Regulator AGAMOUS: Characterization of the DNA Binding Domain and Analysis of Dominant Negative Mutations. Plant Cell **8**: 831-845

Mizukami Y, Ma H (1992) Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. Cell **71**: 119-131

Mullins ED, Chen X, Romaine P, Raina R, Geiser DM, Kang S (2001) Agrobacterium-Mediated Transformation of Fusarium oxysporum: An Efficient Tool for Insertional Mutagenesis and Gene Transfer. Phytopathology **91**: 173-180

Nakagawa H, Ferrario S, Angenent GC, Kobayashi A, Takatsuji H (2004) The Petunia Ortholog of Arabidopsis SUPERMAN Plays a Distinct Role in Floral Organ Morphogenesis. Plant Cell **16**: 920-932

Ng M, Yanofsky MF (2000) Three ways to learn the ABCs. Current Opinion in Plant Biology **3**: 47-52

Nibau C, Di Stilio VS, Wu H, Cheung AY (2010) Arabidopsis and Tobacco SUPERMAN regulate hormone signaling and mediate cell proliferation and differentiation. Journal of Experimental Botany **61**:1-13

Novak J, Novak L (1997) Amplification of flanking regions: New applications and performance optimization of single specific primer PCR and T-vector cloning. . Promega Notes Magazine **61**: 26-30

Ochman H, Ayala FJ, Hartl DL (1993) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. Meth. Enzymol. **218**: 309-321

Oosumi T, Gruszewski H, Blischak L, Baxter A, Wadl P, Shuman J, Veilleux R, Shulaev V (2006) High-efficiency transformation of the diploid strawberry (Fragaria vesca) for functional genomics. Planta **223**: 1219-1230

Orcheski BB, Davis TM (2010) An enhanced method for sequence walking and paralog mining: TOPO(R) Vector-Ligation PCR. BMC Res Notes **3**: 61

Parenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, Angenent GC, Colombo L (2003) Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in Arabidopsis: New Openings to the MADS World. Plant Cell **15**: 1538-1551

Rieger, M (2007) Strawberry (F. ×*ananassa*) In A Basra ed, Introduction to Fruit Crops, Ed 1 Vol 1. CRC Press, Boca Raton, pp 385-386

Rohde A, Grunau C, Beck DL, Montagu MV, Rosenthal A, Boerjan W (1999) carpel, a New Arabidopsis Epi-Mutant of the SUPERMAN Gene: Phenotypic Analysis and DNA Methylation Status. Plant Cell Physiol. **40**: 961-972

Rosin FM, Aharoni A, Salentijn EMJ, Schaart JG, Boone MJ, Hannapel DJ (2003) Expression patterns of a putative homolog of AGAMOUS, STAG1, from strawberry. Plant Science **165**: 959-968

Rousseau-Gueutin M, Lerceteau-Kohler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guerin G, Denoyes-Rothan B (2008) Comparative Genetic Mapping Between Octoploid and Diploid Fragaria Species Reveals a High Level of Colinearity Between Their Genomes and the Essentially Disomic Behavior of the Cultivated Octoploid Strawberry. Genetics **179**: 2045-2060

Sakai H, Krizek BA, Jacobsen SE, Meyerowitz EM (2000) Regulation of SUP expression identifies multiple regulators involved in Arabidopsis floral meristem development. Plant Cell **12**: 1607-1618

Sakai H, Medrano LJ, Meyerowitz EM (1995) Role of SUPERMAN in maintaining Arabidopsis floral whorl boundaries. Nature **378**: 199-203

Sargent D, Rys A, Nier S, Simpson D, Tobutt K (2007) The development and mapping of functional markers in Fragaria and their transferability and potential for mapping in other genera. Theoretical and Applied Genetics **114**: 373-384

Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arús P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH (2006) An enhanced microsatellite map of diploid Fragaria. Theoretical and Applied Genetics **112**: 1349-1359

Sather D, York A, Pobursky K, Golenberg E (2005) Sequence evolution and sex-specific expression patterns of the C class floral identity gene, SpAGAMOUS in dioecious Spinacia oleracea L. Planta **222**: 284-292

Schultz EA, Pickett FB, Haughn GW (1991) The FLO10 Gene Product Regulates the Expression Domain of Homeotic Genes AP3 and PI in Arabidopsis Flowers. The Plant Cell **3**: 1221-1237

Senanayake YDA, Bringhurst RS (1967) Origin of Fragaria Polyploids. I. Cytological Analysis. American Journal of Botany **54**: 221-228

Smith DR (1992) Ligation-mediated PCR of restriction fragments from large DNA molecules. Genome Research **2**: 21-27

Smyth DR, Bowman JL, Meyerowitz EM (1990) Early flower development in Arabidopsis. Plant Cell **2**: 755-767

Spigler RB, Lewers KS, Johnson AL, Ashman TL (2010) Comparative mapping reveals autosomal origin of sex chromosome in octoploid Fragaria virginiana. J Hered **101**: 107-117

Spigler RB, Lewers KS, Main DS, Ashman TL (2008) Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. Heredity **101**: 507-517

Staudt G (1962) Taxonomic studies in the genus *Fragaria*. Typification of Fragaria species known at the time of Linnaeus. Canadian Journal of Botany **40**: 881

Staudt G (1999) Systematics and geographical distribution of American strawberry species. Taxonomic studies in the genus *Fragaria* (Rosaceae: Potentilleae). Botany **81**: 100

Staudt G (2009) Strawberry Biogeography, Genetics and Systematics. Acta Hort. (ISHS) **842**: 71-84

Takatsuji H, Matsumoto T (1996) Target-Sequence Recognition by Separate-type Cys2/His2 Zinc-finger Proteins in Plants. Journal of Biological Chemistry **271**: 23368-23373

Takatsuji H, Nakamura N, Katsumoto Y (1994) A New Family of Zinc-finger Proteins in Petunia: Structure, DNA Sequence Recognition, and Floral Organ-Specific Expression. Plant Cell **6**: 947-958

Takeda S, Matsumoto N, Okada K (2004) RABBIT EARS, encoding a SUPERMAN-like zinc-finger protein, regulates petal development in *Arabidopsis thaliana*. Development **131**: 425-434

van Nocker S, Ek-Ramos J (2007) Control of flowering time. In KD Grasser ed, Plant Reviews: Regulation of Transcription in Plants, Ed 1 Vol 16. Blackwell Publishing, Hoboken, pp 225- 252

Vining KJ (2007) Studies of Verticillium wilt and characterization of candidate verticillium wilt resistance genes in the mint species *Mentha longifolia* (L.) Huds. PhD thesis Genetics. University of New Hampshire, Durham, p 118

Waggott W (1998) Clinical Applicaions of PCR. In Lo YM (ed) Methods in Molecular Medicine. Ed 1 Vol 16. Humana Press, Totowa, pp 81-91

Weigel D (1995) The APETALA2 Domain Is Related to a Novel Type of DNA Binding Domain. Plant Cell **7**: 388-389

Weigel D, Meyerowitz EM (1994) The ABCs of floral homeotic genes. Cell **78**: 203-209

White D, Chen W (2006) Genetic transformation of Ascochyta rabiei using Agrobacterium mediated transformation. Current Genetics **49**: 272-280
Wolfe SA, Nekludova L, Pabo CO (2000) DNA Recognition by Cys2His2 Zinc-finger Proteins. Annual Review of Biophysics and Biomolecular Structure **29**: 183-212

Yamagami T, Tsuchisaka A, Yamada K, Haddon WF, Harden LA, Theologis A (2003) Biochemical Diversity among the 1-Amino-cyclopropane-1-Carboxylate Synthase

Isozymes Encoded by the Arabidopsis Gene Family. Journal of Biological Chemistry **278**: 49102-49112

Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature **346**: 35-39

Yuanxin Y, Chengcai A, Li L, Jiayu G, Guihong T, Zhangliang C (2003) T-linker-specific ligation PCR (T-linker PCR): an advanced PCR technique for chromosome walking or for isolation of tagged DNA ends. Nucl. Acids Res. **31**: e68-

Yun JY, Weigel D, Lee I (2002) Ectopic Expression of SUPERMAN Suppresses Development of Petals and Stamens. Plant Cell Physiol. **43**: 52-57

APPENDICES

APPENDIX A

Putative full length nucleotide and predicted protein sequences of the five *FRASUP* genes. All five gene sequences have been deposited on the NCBI GenBank database under the indicated accession numbers.

FRASUP1 Predicted Protein, GenBank Protein Accession ADE34115.1

MMDCVKLRFGREEDLSDISPWTPKNFTCSFCKREFRSAQALGGHMNVHRRDRA

RLRLLPPNSVPSSSEYSPYPNPNPNPNPKSYFSSSPSSSSSLSGANKYLPHLVSPR

STYVSTPSSGTLDHENKKLMFNSGSHQHHAPPFLHPKTGLKRSSALDYEFGDQL

NGFAGFAQKHHLDDDFEVVKKDEKHINMIREELDMGLLIKDPNKVEVDLELRLG

HL*

FRASUP2 Predicted Protein, GenBank Protein Accession ADE34116.1

MNTSGFCKSLVKDPLHSTVMTEDVEVYMNGHPWPPRCYICDFCKREFKCAQSL

GGHMNVHRKERAILRIGQYTHSTILNLNLNPNPTFVSSLPPSLSSPFSTVSARQLSP

LTTSFISSVPSWISQSSHLPPPTFSSPSATCHTLDANLSGPSSNFMKFNDAKSQEED

ECMRLETGVLGGNSKDDIDLELRLG*

FRASUP3 Predicted Protein, GenBank Protein Accession ADE34117.1

MQRNTGSCKSLVMTKAIDGDSNTNNKKNTEDDEVCMNVYPWPPRCYICGFCKR

EFKSAQALGGHMNVHRKDRAKLRSSPPADWIGQYTHSTILNLNLKPNPNPTFVP

SPPPSPPSPFSTASTRQQLSPLTTSFSSVPSWISQSSHPPPTFSSPSATCHTENMKWL

VGGNLSGPSLNLKSFDLSSQMKINAESQEDGCMRLDLEIGVLGGNSEDEIDLELR

LGY*

FRASUP4 Predicted Protein, GenBank Protein Accession ADE34118.1

MLKQYCYSSSSSMNSVMSSSSATWEEKAFAEDAAGALGGGGCIWPPRSYSCSFC

MREFRSAQALGGHMNVHRRDRARLKQCLHNTSGTATPKLATDHVPHHLRNHS

RDFAAMTCPFSAPRAFISPYSSSKSCSTAAAAAADLSEKNIMVLGSMGSSSDEED

QSIISLMGGDQVGEYVETDLTVGLINTVDRGNKQLAGSCGEEASTCKRPKIASAM

KSAPFPFFLRPAEVSGLMAAANSMEDLDLELRLGDPTPKVK*

FRASUP5 Predicted Protein, GenBank Protein Accession ADE34119.1

MEQARYWMWAKRKHSSLMSASHEVQVPSSQDESSWEEQAFAEDAAGPLGGCI

WPPRSYSCSFCRREFRSAQALGGHMNVHRRDRARLKQSPNEQNVETTHHHHGL

PDQDHQDISVQINPFSSTLGTFHQYPSHHQVCALVYNPNNPNSNPLSYPSRVSAQ

LPGKDQICGQEQITLNIPTSRSSSVSYPPSRSILTGNRYYNNSVARNTEAAEKISRIV

ESGSCRANKGGDYVTTDLSVSLNLVVRCARPSVSAGCDEDEAIISCKRRRIMEEK

PSSKDSLPFFLKSEGFEVLSRSSSIEELDLELRLGDRPKVIKVD*

***FRASUP1* Nucleotide, GenBank Nucleotide Accession GU830919**

ATGATGGATTGTGTTAAGTTGAGGTTTGGAAGAGAAGAAGATTTAAGCGATA

TTTCTCCATGGACTCCAAAGAACTTTACATGTAGCTTCTGCAAGAGAGAATTC

AGGTCTGCTCAAGCTCTTGGGGGTCATATGAATGTCCACAGGAGAGACAGAG

CTAGGCTTCGGCTCCTACCACCTAACTCGGTCCCTTCTTCTTCCGAATATTCTC

CTTACCCTAATCCTAACCCTAACCCTAACCCTAAGTCTTACTTCTCTCCTTCGC

CATCATCATCATCATCTCTTTCAGGTGCTAACAAGTACTTACCACACTTGGTC

TCTCCTCGGAGCACTTATGTGTCCACACCTTCCTCAGGTACTTTAGACCATGA

AAACAAGAAACTGATGTTCAACAGCGGTTCTCATCAACATCATGCTCCTCCTT

TCCTTCATCCTAAGACAGGCCTCAAGAGAAGCAGTGCTCTTGATTATGAGTTT

GGAGATCAACTCAATGGATTTGCTGGGTTTGCACAAAAACATCATCTCGATG

ATGATTTTGAAGTTGTGAAGAAAGATGAAAAGCATATCAACATGATTAGGGA

AGAATTGGACATGGGGTTGCTGATTAAAGATCCTAATAAGGTGGACGTGGAT

TTGGAACTTCGACTCGGGCATCTCTAG

***FRASUP2* Nucleotide, GenBank Nucleotide Accession GU830920**

ATGAACACTGCTGGCTTCTGCAAAAGCTTAGTGAAAGATCCATTACACAGTA

CAGTCATGACTGAAGATGTTGAAGTCTATATGAATGGGCACCCATGGCCTCC

AAGGTGTTACATTTGTGACTTCTGCAAGAGAGAATTCAAATGTGCTCAATCTC

TGGGTGGCCATATGAATGTTCACAGGAAAGAACGTGCCATTCTCAGAATTGG

TCAGTACACTCACAGTACTATACTTAATCTCAACCTTAATCCTAACCCTACTT

TCGTATCCTCACTACCACCATCATTATCATCTCCCTTTTCTACTGTTTCAGCTA

GGCAGCTCTCACCATTAACTACTAGCTTTATAAGCTCAGTGCCTTCTTGGATA

TCACAATCCTCTCATCTTCCTCCTCCTACATTTTCATCGCCATCTGCTACTTGC

CATACTCTACGTGCTAACCTTTCGGGTCCTTCTTCAAACTTCATGAAGTTTAAT

GACGCAAAGTCTCAAGAAGAAGATGAGTGCATGTGGTTGGAGACTGGTGTTC

TTGGCGGTAACTCAAAGGATGACATCGATTTAGAGCTTCGATTAGGCTAA

***FRASUP3* Nucleotide, GenBank Nucleotide Accession GU830922**

ATGCAGAGGAACAGTGGCTCATGCAAAAGCTTAGTCATGACTAAGGCCATTG

ATGGAGATTCCAACACCAACATTAAGAAGAACACTGAAGATGATGAAGTCTG

TATGAATGTGTACCCATGGCCTCCGAGGTGTTACATTTGTGGCTTCTGCAAGA

GAGAATTCAAATCTGCTCAAGCTCTGGGTGGCCATATGAATGTTCACAGGAA

AGACCGTGCCAAGCTCAGAAGCTCACCCCCAGCTGATTGGATTGGTCAGTAC

ACTCACAGTACTATACTTAATCTCAACCTTAAACCTAACCCTAACCCTACTTT

CGTACCCTCACCACCACCATCACCACCATCTCCCTTTTCTACTGCTTCAACTA

GGCAGCAGCTCTCACCATTAACTACTAGCTTTAGCTCAGTACCTTCTTGGATA

TCTCAATCCTCTCATCCTCCTCCCACATTTTCATCGCCATCTGCTACTTGCCAT

ACTGAAAACATGAAATGGCTTGTGGGTGGTAACCTTTCCGGTCCTTCTTTAAA

TCTCAAGAGTTTCGATTTAAGCAGTCAGATGAAGATTAATGCAGAGTCTCAA

GAAGATGGGTGCATGAGGTTGGACTTGGAGATCGGTGTTCTTGGCGGTAACT

CAGAGGATGAAATCGATTTGGAGCTTCGATTAGGCTACTGA

***FRASUP4* Nucleotide, GenBank Nucleotide Accession GU830922**

ATGTTGAAGCAGTACTGCTACTCCTCGTCGTCGCGATCGTCGATGAACTCAGT

CATGAGTATCTCATCCGCGACATGGGAAGAGAAGGCTTTCGCTGAAGACGCA

GCTGGTGCGCTTGGTGGCGGTGGTTGCATATGGCCGCCGAGGTCTTATTCTTG

CAGTTTTTGTATGAGGGAGTTCAGGTCGGCTCAAGCTCTTGGGGGTCACATGA

ATGTTCACAGAAGAGACAGAGCTAGACTCAAGCAATGCCTCCACAACACTAG

TGGTACTGCTACTCCTAAACTAGCTACAGATCATGTCCCTCATCACCTTAGAA

ATCATAGCCATGGCTTTGGTGCGATGACTTGCCCTTTTTCAGCTCCTAGGGCT

TTTATCTCTCCTTATTCTTCTTCAAATTCTTGCTCAACTGCTGCTGCTGCTGCT

GCCGCTGCCGCTGATCGATCGGAGAAGAATATCATGGTGATGGGGTCATCAT

CAGACGAAGAAGACCAATCTGTTATTAGCCTGATGGGAGGTGATCAAGTTGG

GGAATATGTCGAAACCGATTTGACAGTGGGGTTGATTAACACAGTTGTTAGA

GGAAACAGACAATTAGCTGGGTCTTGTGGGGAAGAGGCAAGCACTTGTAAA

AGGCCTAAGATTGCTTCTGCTATGAACTCAGCACCATTTCCATTCTTTCTCCG

ACCTGCAGAGGTAAGTGGACTTATGGCTACAGCTAACTCCATGGAAGATTTG

GATCTTGAGCTCAGGCTTGGTGACCCAACTCCCAAGATCAAATGA

***FRASUP5* Nucleotide, GenBank Nucleotide Accession GU830923**

ATGGAGCAAGCACGGTACTGGATGTGGGCAAAGCGCAAGCACAGTAGTCTG

ATGAGTGCTTCTCATGAAGTTCAGGTGCCTTCTTCCCACGACGAATCGTCGTG

GGAAGAACAAGCTTTTGCAGAAGATGCAGCGGGGCCTCTCGGAGGTTGCATA

TGGCCACCGAGATCTTATTCCTGCAGCTTCTGCAGGAGAGAGTTTCGTTCGGC

TCAAGCTCTTGGCGGCCACATGAATGTTCACAGGAGAGACAGAGCTAGACTC

AAGCAGTCACCGAATGAACAAAATGTCGAAACTGCCCATCATCATCATGGCC

TTCCTGATCAAGATCATCAGGATATTTCTGTCCAAATCAATCCCTTCTCATCG

ACTCTGGGCACTTTTCATCAATACCCATCTCATCATCAAGTTTGTGCCTTGGTT

TATAACCCTAATAACCCTAGTTCGAATCCTCTCTCATCGCCATCTAGGGTTTC

GGCTCAATTACCAGGTAAAAATCAAATCTGTGGTCAAGAGCAAATTACCTTG

AATATCCCAACCTCTAGGTCATCCTCTGTATCAAATCCTCCGTCCAGGTCAAT

CTTGACTGGCAATAGATATTATAATAACTCTGTTGCAAGAAATACAGAAGCT

GCAGAGAAGATCTCAAGAATTGTGGAATCTGGGTCGTGTAGGGCAAATAAGG

GTGGTGACTATGTCACAACTGATTTGTCTGTGAGTCTGAATTTGGTCGTTCGC

TGTGCTCGTCCATCTGTGTCAGCTGGTAGTGATGAAGATGAAGCTATAATCAG

TTGCAAGAGAAGAAGAATAATGGAGGAGAAGCCATCATCCAAAGACTCTTTA

CCTTTCTTCCTAAAATCCGAGGGGTTTGAAGTACTAAGCCGCCCCAGCTCCAT

CGAAGAGTTGGATCTTGAGCTCAGGCTCGGTGACCGTCCTAAGGTAATCAAA

GTAGATTAA

APPENDIX B

The complete DNA alphabet including all of the degenerate bases to accompany the primers in Table 1.

Abbreviation	Base
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
I	Inosine
R	A,G
Y	C,T
M	A,C
K	G,T
S	C,G
W	A,T
H	A,C,T
B	C,G,T
V	A,C,G
D	A,G,T
N	A,C,G,T