NOTE TO USERS

Map-Based Cloning of a Regulator of the Vascular Cambium in *Arabidopsis thaliana*

Submitted by

Martha Rae Mullally

in partial fulfillment of the requirements

for the degree of Master of Science (Biology).

Carleton University

Your file Votre référence
ISBN: 0-494-13468-2
Our file Notre référence
ISBN: 0-494-13468-2

ABSTRACT

Wood is formed by the meristematic activity of the vascular cambium, a cylindrical meristem located between the xylem and phloem. Little is known about the regulation of the vascular cambium and there are no reports of mutants lacking a vascular cambium. An *Arabidopsis thaliana* mutant that lacks a cambium and does not undergo secondary growth was identified. This mutant was called *cambiumless* and it was postulated that the CAMBIUMLESS protein was a positive regulator of vascular cambium development. A map-based cloning strategy was used to identify the *CAMBIUMLESS* gene and it was found to be identical to *CURLY LEAF* (*CLF*), a gene that encodes a Polycomb-Group protein and known to be involved in epigenetic regulation of floral homeotic genes. This work identifies a novel role for *CLF* in the regulation of the vascular cambium.

ACKNOWLEDGEMENTS

There are many people I would like to thank for their assistance, friendship and support.

- My supervisor Sharon Regan, the most enthusiastic biologist I know. Thank you for your leadership, patience, mentoring, and teaching me how to become a scientist.

- My committee members, Doug Johnson, Sue Aitken, Shelley Hepworth and special thanks to Myron Smith for mapping conversations & agreeing to be my co-supervisor.

- Carleton University for being a great place to work and learn. Special thanks to Ed Bruggink for greenhouse space, Andrew Simons for use of his microscope, Tim Xing for use of space in his lab, and Lewis Ling in Earth Sciences for all of his help with the cryo-SEM.

- The great people in the Regan lab, Lisa McDonnell, Donna Yee, Jonathan Plett, Shuyou Han, Dan McPhee, Brendan O'Malley, Ed Harrison, Robin Vitez & Vijay Sharma (for help with the light microscope) who made it fun to come to work everyday. Special thanks to Graham Thurston, for always providing a thoughtful perspective, car-pooling to Kingston and being so darn funny.

- The biology department at Queen's, especially members of the Snedden lab, Wang lab, Plaxton lab, and our neighbours, the Moyes lab. Special thanks to Steph Dobney & Barb Vanderbeld for having so many answers to so many cloning questions. Enormous thanks to Sam Gennidakis for great chats, dinners and letting me stay at Rideau St.

- My enthusiastic volunteers, Michael Bush, Gabriel Lee & Nancy Truong. They made all of the tissue collection possible and reminded me that the fruits of labor cannot always be eaten.

- Scott Douglas at University of Toronto (UTSC), a beacon of light in the mapping darkness. Nick Provart at University of Toronto for designing the BAR and then answering lots of my questions about it.

- Friends and family who provided encouragement and support, including the Mills', Soren, Lingyun, Larissa; Genevieve, Peter, John, Maria and Laura Mullally; and Mary Jacques.

- My wonderful husband, Jason Jacques, who helped and supported me through the whole process. I am grateful for your presence in my life.

- Genome Canada, Natural Sciences and Engineering Research Council and the Council for Biotechnology Information for funding support.

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Full Name
	- A -
ABRC	Arabidopsis Biological Resource Centre
AFLP	Amplified Fragment Length Polymorphism
AG	*AGAMOUS*
AGI	Arabidopsis Genome Initiative
AGL17	*AGAMOUS LIKE17*
ANT-C	*Antennapedia* complex
AP1	*APETALA1*
AP2	*APETALA2*
AP3	*APETALA3*
ATP	Adenosine triphosphate
	- B -
BAC	Bacterial Artificial Chromosome
BAR	Botany Array Resource
bHLH	Basic-helix-loop-helix
bp	Basepair
BSA	Bovine serum albumin
BX-C	*Bithorax*-complex
	- C -
CAM	*CAMBIUMLESS*
CaMV 35S	Cauliflower Mosaic Virus 35S
CAPS	Cleaved Amplified Polymorphic Sequences
CLF	*CURLY LEAF*
CLV1	*CLAVATA1*
cM	centiMorgan
CNA	*CORONA/ATHB15*
Col-0	Columbia
Cryo-SEM	Cryogenic Scanning Electron Microscopy
	- D -
d	Day
dNTP	Deoxyribonucleotide
Dof	DNA-binding with one finger
	- E -
EDTA	Ethylenediaminetetraacetic acid
EMF	*EMBRYONIC FLOWER*
EMS	Ethyl methane sulfonate
En-2	Enkheim
e-Northern	electronic Northern
Esc	Extra Sex Combs
E[z]	Enhancer of zeste
	- F -
FIE	*FERTILSATION-INDEPENDENT ENDOSPERM*
FIS	*FERTILISATION INDEPENDENT SEED*

Abbreviation	Full Name
FIS2	*FERTILISATION INDEPENDENT SEED*
FLC	*FLOWERING LOCUS C*
FM	Floral Meristem
	- G -
GUS	β-Glucuronidase
	- H -
HD-ZIP III	Class III homeobox-leucine zipper
HMTase	Histone-methyltransferse
HOX	Homeobox
hr	Hour
	- I -
INDEL	Insertions or deletions
IPTG	Isopropyl-β -D-thioglactopyranoside
	- K -
kb	Kilobase
KNAT2	*KNOTTED-LIKE 2*
kV	kiloVolt
	- L -
L	Litre
LB	Luria broth
Ler	Landberg *erecta*
LFY	*LEAFY*
LRR	Leucine rich repeat
	- M -
MADS box	The first four MADS box genes: MCM1 AGAMOUS, DEFICIENS and SRF
Mb	Megabase
MBC	Map-based cloning
MEA	*MEDEA*
min	Minute
mL	Millilitre
M&S	Murashige and Skoog
MSI1	*MULTICOPY SUPPRESSOR OF IRA1*
	- O -
OD	Optical density
	- P -
Pc-G	Polycomb group
PCR	Polymerase chain reaction
PHB	*PHABULOSA/ATHB14*
PHE1	*PHERES1*
PHV	*PHAVOLUTA /ATHB9*
PI	*PISTILLATA*
PIPES	1,4-Piperazinediethanesulfonic acid
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2

Abbreviation	Full Name
PRE	Polycomb response element
	- Q -
Q-PCR	Quantitative polymerase chain reaction
	- R -
REV	*REVOLUTA / INTERFASCICULAR FIBRELESS*
RFLP	Restriction fragment length polymorphism
RIL	Recombinant Inbred Line
RNAi	RNA interference
rpm	Rotations per minute
RT	Room temperature
	- S -
s	Second
SAM	Shoot apical meristem
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscope
SeqView	Sequence Viewer
SNP	Single nucleotide polymorphisms
SSLP	Simple Sequence Length Polymorphism
STM	*SHOOTMERISTEMLESS*
Su[Z]	*Suppressor of zeste*
SWN	*SWINGER*
	- T -
TAE	Tris-Acetate-EDTA
TAIR	The Arabidopsis Information Resource
TBE	Tris-Boric Acid-EDTA
TBO	Toluidine Blue-O
T-DNA	Transfer DNA
TE	Tris-EDTA
TENS buffer	Tris, EDTA, NaOH, SDS
trx-G	Trithorax group
	- V -
V	Volt
VC	Vascular cambium
VP1	Viviparous-1
VRN2	*VERNALISATION2*
	- W, X, Y and Other -
WUS	*WUSCHEL*
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YEP	Yeast Extract Peptone
µL	Microlitre

LIST OF TABLES

LIST OF FIGURES

1.0 General Introduction

1.1 Wood Development

Trees are essential members of the ecosystem, comprising most of the terrestrial environment. Trees represent a vast percentage of the total biomass on Earth, are major carbon sinks, play a central role in climate control and are an important natural resource. Both the Earth's climate and biodiversity depend on forests. Trees are unique plants in terms of their long-life spans, large size, seasonal growth and dormancy and the formation of wood. Wood is vital for human communities and economies. Wood is one of the most valuable commodities in the industrialized world and is used in a wide range of industries, including construction, pulp and paper, and energy. In Canada, wood represents approximately 20 per cent of annual exports (Statistics Canada, 2003). In less developed countries, more than half of the world's annual wood harvest is used as fuel.

1.2 Biosynthesis of Wood

The process of wood development in trees is well characterized. Wood, or secondary xylem, is formed via xylogenesis: a primary xylem cell divides from the cambium, expands, a secondary cell wall is laid down and the cell undergoes programmed cell death to produce mature, dead, xylem cells that make up the wood tissue (Mellerowicz, *et al.*, 2001). Wood is made of cellulose, hemicellulose and lignin (Hu, *et al.*, 1999). The enzymes involved in biosynthesis of lignin and cellulose have been well characterized, the genes that encode these enzymes cloned, and their expression altered to examine the effect on wood development. There is much known about the developmental processes involved in wood development, the key enzymes involved and biosynthetic pathways have been proposed (Ahuja, 2001; Confalonieri, *et al.*, 2003). However, the underlying genetic mechanisms of wood

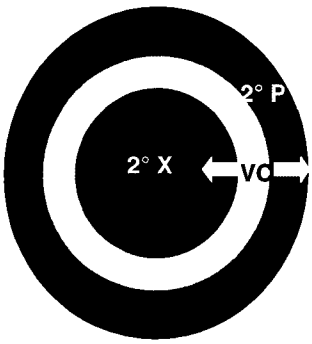
development and the genes that play key roles in the initiation of differentiation of the vascular cambium are less well characterized and are an area of research interest.

1.3 Development and Regulation of the Vascular Cambium

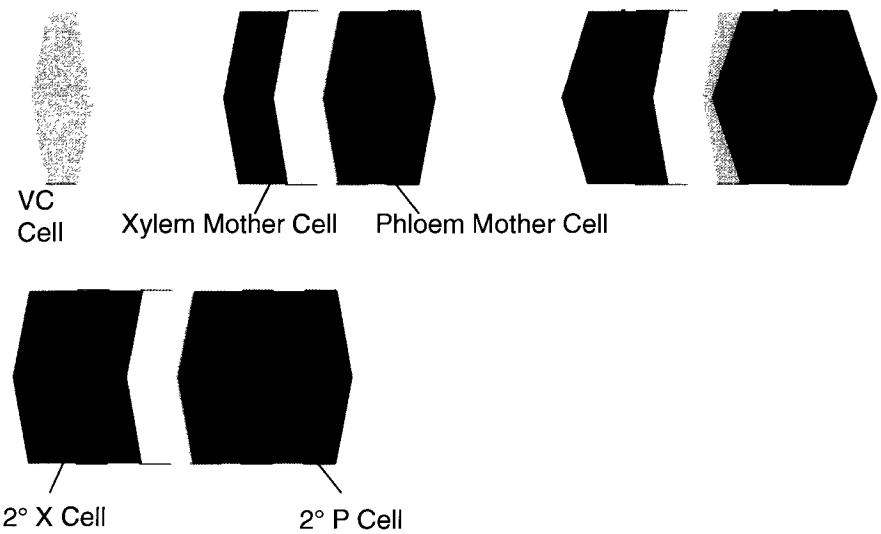
Development of primary vasculature is well characterized (Ye, *et al.*, 2002). Primary vasculature is made of xylem and phloem and is formed in bundles throughout the plant. Within the stem, the formation of secondary vasculature begins with cell proliferation in the primary bundles, or vascular cambium, initiation of xylem differentiation, regulation of cell expansion, deposition of a secondary cell wall and programmed cell death.

Meristems are groups of undifferentiated cells that serve as a source for new cells. Examples of meristems include the shoot apical meristem and root apical meristem, both of which are primary meristems. Wood is formed through the meristematic activity of a secondary and lateral meristem, the vascular cambium (VC). The VC is established in the secondary phase of growth and made of two types of meristematic cells: fusiform initials and ray initials. The fusiform initials produce long, vertically-oriented, vacuolated, secondary xylem mother cells on the inside of the VC and secondary phloem mother cells on the outside. Through periclinal divisions (*i.e.* divisions that are parallel to the outer surface), the mother cells produce secondary xylem and phloem cells that act in vertical transport in the plant. The ray initials are small cells whose derivatives include the radially oriented files of parenchyma cells within wood known as rays. A schematic representation of the development of the vascular cambium is presented in Figure 1.0. Secondary phloem cells, including sieve cells, companion cells and phloem fibers, conduct photosynthate from leaves to the rest of the plant. Secondary xylem cells in angiosperms, which include vessels and fibres, conduct gases and water from the

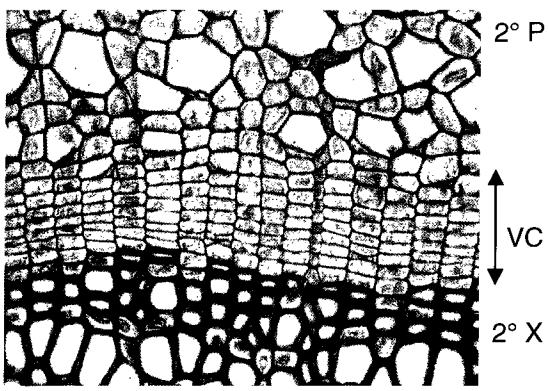
Figure 1.0. Schematic representation and micrograph of vascular cambium. (a) Schematic representation of secondarily-thickened Arabidopsis hypocotyl. The vascular cambium is a secondary meristem located between the secondary xylem and secondary phloem. Division of the vascular cambium gives rise to secondary xylem and secondary phloem, with secondary xylem predominating, resulting in increased girth of the hypocotyl. (b) Secondary xylem and secondary phloem are formed through the meristematic activity of the vascular cambium. Secondary xylem mother cells are formed on the inside of the vascular cambium and secondary phloem mother cells on the outside. Through periclinal divisions the mother cells produce secondary xylem and phloem cells. (c) Micrograph of vascular cambium from *Aristolochia* stem, stained with phloroglucinol. The vascular cambium is a thin layer of regularly shaped cells located between the secondary xylem and secondary phloem. The thick, lignified cell walls, stained read, of the secondary xylem are evident. *Aristolochia* micrograph obtained from http://botit.botany.wisc.edu. 2° P, secondary phloem, VC, vascular cambium, 2° X, secondary xylem.



(a)



(b)



(c)

root to the shoot and provide structural support. Secondary xylem cells develop thick secondary cell walls, undergo extensive lignification and programmed cell death to yield mature, dead, xylem cells. Mature phloem cells do not undergo lignification or programmed cell death. The cortex and bark is produced by the cork cambium and is deposited outside the layer of living phloem cells.

In recent years, more of the underlying molecular and genetic mechanisms that control the development of the VC and its production of xylem and phloem cells have been revealed, including the role of phytohormones such as auxins, cytokinins, and brassinosteriods, and the identification of a regulatory role for homeobox leucine-rich repeat (LRR) class III (HD-ZIP III) genes and microRNAs (Fukuda, 2004; Nieminen, *et al.*, 2004). Despite these developments, and extensive documentation at the structural level, relatively little is known about the underlying genetic mechanisms of secondary vascular development. The goal of this thesis was to identify genes that play a role in the regulation of the VC. In the research described, *Arabidopsis thaliana* was used as a model organism to examine genetic regulation of the VC.

1.4 Arabidopsis as a Model Organism for Wood Development

While extensive research has been published regarding identification of the enzymes involved in lignification and cellulose production in woody trees (Ahuja, 2001; Confalonieri, *et al.*, 2003), less is known about the initiation and differentiation of the VC. This absence of information reflects that trees are poorly suited to a laboratory setting due to their large size, long life span, and large genome size. However, the challenge of using trees as laboratory organisms has been tackled successfully, and the tree genus *Populus* (poplar) has emerged as a

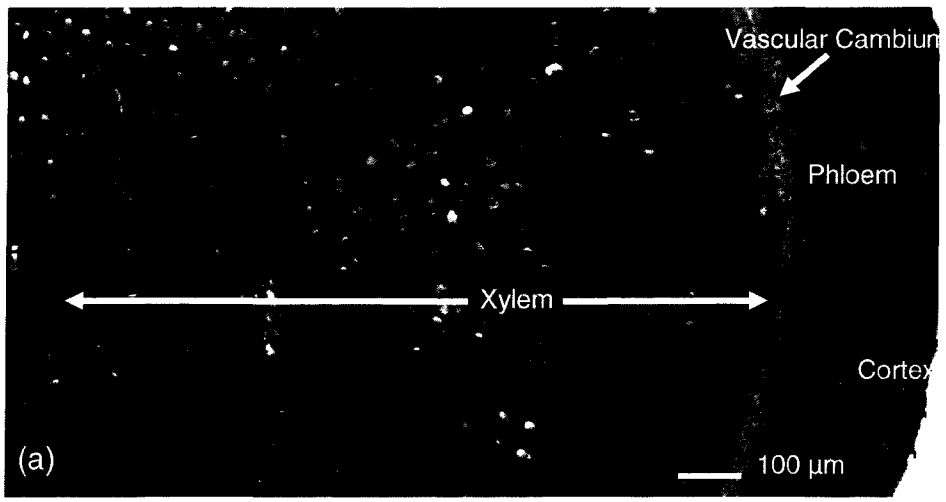
model organism for trees (Strauss & Martin, 2004). Effective protocols for genetic transformation of poplar have been developed (Han, *et al.*, 2000; Strauss & Martin, 2004), tools of molecular biology are increasingly applied, and the complete tree genome sequence, for *Populus trichocarpa*, was completed in 2004 (Joint Genome Institute, 2005). Despite these developments, there remains considerable interest in a smaller, model organism for wood development and Arabidopsis has been established as that model organism.

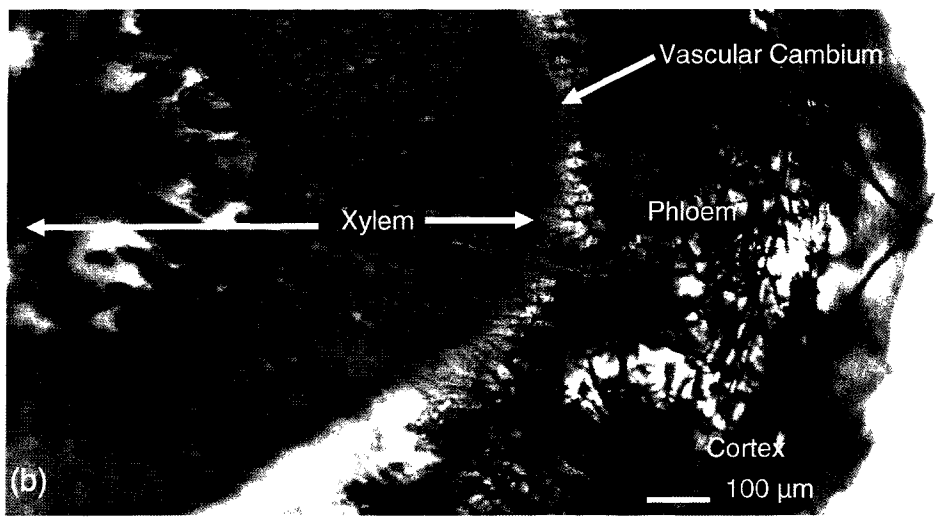
Several groups have demonstrated that *Arabidopsis* undergoes extensive secondary development in its hypocotyl and can be used to investigate development of secondary vasculature (Busse & Ray, 1999; Chaffey, *et al.*, 2002). Anatomical evidence indicates that cambial and wood cells, fibres and vessels elements, of Arabidopsis are morphologically and structurally similar to those in poplar (Chaffey, *et al.*, 2002). Figure 1.1 demonstrates the similarities in cellular anatomy and arrangement in the stem of the angiosperm tree *Populus* and the secondarily-thickened hypocotyl of Arabidopsis.

Extensive genetic homology has been demonstrated for genes from wood forming tissues in Arabidopsis and poplar (Hertzberg, *et al.*, 2001) as well as in Arabidopsis and Loblolly pine (Kirst, *et al.*, 2003). These anatomical and genetic similarities provide a compelling case for the use of Arabidopsis as a model organism in the examination of secondary vasculature development.

After prolonged secondary growth, it is possible to physically separate the secondary tissues, including xylem, phloem-enriched and nonvascular fractions from Arabidopsis (Zhao, *et al.*, 2000). Oh and colleagues (2003) have used this technique to analyse the transcriptome

Figure 1.1. Free-hand cross sections of *Populus tremula* x *P.alba* stem and Arabidopsis secondarily-thickened hypocotyl. (a) *Populus tremula* x *P.alba* stem, 6 months old, stained with phloroglucinol and (b) *Arabidopsis thaliana* secondarily-thickened hypocotyl, two months old, stained with Toluidine Blue O. The cross-sections show the similarities in cellular anatomy and arrangement of these tissues. Pictures courtesy of V. Sharma and S. Han, respectively.





profile in secondary xylem. Genes of unknown function make up the largest category of up-regulated genes, followed by transcription factors, including some members of the *R2R3-MYB* gene family. Some members of the pine *R2R3-MYB* gene family have also been implicated in lignin biosynthesis (Nieminen, *et al.*, 2004). Zhao and colleagues (2005) have used the same approach to analyse transcriptome profiles in the xylem, phloem-cambium and the non-vascular tissues. They identified genes specific to the xylem and phloem-cambium tissues and highlighted candidate regulatory roles for uncharacterized G2-like class III HD-ZIP genes, homologous to genes already linked to vascular development, as well as several members of large transcription factor families, such as *NAC*, *MYB*, *MADS-box*, *bHLH* and *WRKY* genes, for which roles in vascular differentiation have not previously been established (Zhao, *et al.*, 2005).

1.5 Gene Discovery Strategies

The goal of this thesis was to identify a regulatory gene of the VC. Forward and reverse genetics are the two main approaches to discovery of gene function. In both approaches, the normal function of a gene is deduced by analyzing the effect that a mutation of the gene has on a particular tissue, in this case the VC. Both forward and reverse genetic approaches are discussed with respect to discovery of gene function in Arabidopsis and poplar.

In a reverse genetics approach, activity of a known gene is disrupted and the corresponding effect on the target tissue is characterized. A reverse genetics approach has been previously used with success in the Regan lab to identify a new role for CLAVATA1 in regulation of the Arabidopsis VC. A known regulator of the size of the shoot apical meristem (SAM), CLAVATA1 (CLV1) was down-regulated by transformation of plants with an antisense

construct and the effect on the VC observed. CLV1 is a negative regulator of the SAM; *clv1* mutants have larger SAMs than wild-type plants. This increase in meristem size extends to the flowers; *clv1*plants display enlarged floral meristems and produce flowers with additional petals. The VC is derived in a linear fashion from the SAM. For this reason, it was hypothesized that down-regulation of *CLV1* expression would result in a larger SAM and a corresponding larger VC that differentiated to yield more secondary xylem. This was demonstrated to be correct; the transgenic *clv1* plants produced 50% more secondary xylem than their wild type counterparts (Han, 2004).

The second gene discovery strategy is forward genetics. In forward genetics, a mutant plant with altered vasculature is identified and the gene responsible for the phenotype is determined. This approach enables identification of novel genes with a role in the development of the VC and can identify novel roles for previously characterized genes. In this research project a forward genetics approach was used, a mutant plant lacking a vascular cambium was investigated and the gene responsible for the phenotype identified.

In poplar, both forward and reverse genetic strategies have been used in discovery of gene function. The techniques of reverse genetics are very similar to those used in Arabidopsis, both antisense and RNAi have been used to down-regulate target genes. In many cases poplars were transformed with Arabidopsis genes because the poplar genome was not available. The entire poplar genome has recently been sequenced and is publicly available on the Joint Genome Institute (JGI) website (http://genome.jgipsf.org/Poptr1.home.html). The availability of a complete poplar genome aids in the identification of target genes and assists in DNA construct design for down-regulation of gene expression.

Unlike Arabidopsis, there is no poplar mutant database where hundreds of different mutants with interesting phenotypes are listed and are available to be ordered. Therefore, reverse genetics in poplar starts with the generation of a poplar mutant. This has been done in the Regan lab; an activation-tagged line population of poplars has been generated. The poplars were transformed with a Transfer-DNA (T-DNA) construct consisting of three Cauliflower Mosaic Virus 35S (CaMV 35S) enhancers in tandem. The enhancers are inserted randomly into the genome, up-regulate genes in the vicinity and cause a mutant phenotype in the plants. The identity of the up-regulated genes can then be determined with genome-walking, a PCR-based approach that uses primers within the T-DNA and random primers to amplify the region where the insert is located. This approach has been used successfully to identify genes responsible for mutant phenotypes in the Regan population of activation-tagged line poplars.

Clearly, both forward and reverse genetics approaches are used in Arabidopsis and poplar with success. A common approach is for genes to be identified and characterized first in Arabidopsis and then poplar homologues or orthologues are identified in the poplar genome database and used as targets for up or down-regulation in poplar. This is a powerful approach because many genes have already been identified in Arabidopsis, it is easy to identify knock-out mutants in Arabidopsis, and there is considerable gene homology and protein function similarity between Arabidopsis and poplar. For this reason, the function of a gene in Arabidopsis provides clues about its function in poplar. A close similarity between trees and Arabidopsis means that knock-out mutants in Arabidopsis represent an important resource for wood formation research. As Chaffey has observed, the partnering of

Arabidopsis and poplar represents the way forward for the immediate future of wood research (Chaffey, 2002).

Finally, microarrays are increasingly used as a gene discovery strategy. In this approach the complete transcriptome of a mutant plant, or the complete transcriptome of a tissue, is compared to the transcriptome of a wild-type plant and altered gene expression identified. This is a data intensive approach to gene identification, where all genes whose expression profile changes, either are up-regulated or down-regulated, are identified. Microarrays have been used to identify xylem-specific genes in Arabidopsis (Oh, *et al.*, 2003; Zhao, *et al.*, 2005) and poplar (Hertzberg, *et al.*, 2001; Schrader, *et al.*, 2004).

1.6 Overview of Thesis

The following thesis is divided into two chapters. In the first chapter the use of map-based cloning (MBC) for gene discovery is discussed. The background work that led to MBC will be outlined and a detailed description of the experimental procedure to determine the identity a mutant gene presented. In the second chapter, the *CURLY LEAF* gene will be discussed. The previously uncharacterized role of *CULRY LEAF* in the regulation of the vascular cambium is outlined and further experiments to understand its role in vascular development discussed.

2.0 Chapter 2: Map-Based Cloning of the Gene Responsible for the Cambiumless Phenotype

2.1.1 *Arabidopsis thaliana* as a Model Organism in Plant Genetics

Arabidopsis thaliana is a herbaceous, dicotyledous weed, which belongs to the Brassicaceae (mustard) family and has a broad natural distribution across the Northern hemisphere. Originally proposed by George Redei in 1975 as a model organism for experiments in plant physiology (Meyerowitz, 2001), Arabidopsis has become a ubiquitous tool of plant molecular biology, physiology, biochemistry, genetics and genomics. Arabidopsis has all the characteristics of a good model organism, it is small, self-fertile, has a rapid lifecycle (from germination to mature seed in 6 weeks) produces many seeds (on average 5,000 per plant) and has a small genome (125 Mb) arranged on five chromosomes.

A wide array of tools have been developed for Arabidopsis. These include well established protocols for transformation, genetic and physical maps, comprehensive web-based databases, as well as stock centers with mutant, activation-tagged, and knock-out lines. In addition to the full genomic sequence completed in 2000 (The Arabidopsis Genome Initiative, 2000), these tools make Arabidopsis an excellent model for examining many aspects of plant biology.

2.1.2 Identification of the *cambiumless* Mutant

A selection of mutants was obtained from the Arabidopsis Biological Resource Center (ABRC, Ohio State University, Columbus, OH) and screened to identify candidates with altered vasculature. The screen consisted of macroscopic and microscopic observation, including plastic embedding and sectioning of stems and hypocotyls. Of the candidate pool,

ABRC stock number cs313 was deemed noteworthy as it had no vascular cambium and completely lacked secondary vasculature. For this reason, the mutant plant was referred to as *cambiumless* (*cam*) and the CAM protein postulated to be a positive regulator of the vascular cambium. The mutant was generated via ethyl methane sulfonate (EMS) mutagenesis and was part of a collection donated to the ABRC seed stock centre by Albert Kranz (Rhee, *et al.*, 2003). The purpose of the experiment described in this chapter was to identify the gene responsible for the *cam* phenotype using a MBC approach.

2.1.3 Mechanisms to Generate Mutants in Plants

Mutagenesis is the process where heritable alterations in the genome of an organism are produced. Mutant generation in genetics research has typically been done in two ways, via forward or reverse genetics. Forward genetics is an untargeted approach to mutant generation. An organism is exposed to a mutagen and its resulting offspring examined for a phenotypic effect of the mutation. The mutant population is screened to identify plants with interesting phenotypes. Reverse genetics is a targeted approach. Expression of a known gene is down-regulated, for example via antisense, RNAi, T-DNA insertion, *etc.*, and the phenotypic effect observed. Reverse genetics is a precise way to look at the function of one particular gene, its role in a pathway and phenotypic effect on the plants. However, the key drawback of reverse genetics is that it is not possible unless the target gene has been identified.

The classical approach to mutant generation is forward genetics. Forward genetics is an indirect, untargeted approach to mutagenesis, mutations are randomly introduced in the genome of an organism and the progeny screened to identify phenotypic effects. Common

mutagens used in Arabidopsis include ethyl methane sulfonate (EMS); fast neutron bombardment, and insertional elements, such as Transfer-DNA (T-DNA) from *Agrobacterium*, or Ac/Ds transposons. EMS mutagenesis causes point mutations, preferentially inducing G to A transitions, fast neutron bombardment causes small deletions, and insertional elements disrupt genes causing loss-of-function (Weigel & Glazebrook, 2002). The *cambiumless* mutant was generated via EMS-mutagenesis.

2.1.4 Mutagenesis with Ethyl Methane Sulfonate

To mutagenize seeds with EMS, they are imbibed in a low concentration (0.2 % v/w) solution of EMS for 16 hr. Following this, the seeds are thoroughly rinsed with sterile water, stratified at 4°C for 4-5 days and sown on soil (Jander, *et al.*, 2003). The individuals treated with the mutagen are referred to as the M_1 generation. EMS is typically used at concentrations that introduce multiple mutations, on the order of 500 mutations per genome, in the M_1 generation. The progeny derived from the self-fertilization of the M_1 generation are the M_2 generation. EMS predominantly induces recessive mutations, so phenotypes are only observed in M_2 homozygotes and therefore the M_2 generation is screened for mutants (Østergaard & Yanofsky, 2004). The high rate of mutagenesis usually permits phenotypic mutation identification in any gene by screening fewer than 5000 plants (Weigel & Glazebrook, 2002). Once a mutant is identified, it is backcrossed (*i.e.* crossed with a wild-type plant) for 5 – 10 generations and plants that retain the phenotype are selected to generate a true-breeding mutant line.

EMS mutagenesis affects gene function in a variety of ways, including complete loss of function, partially reduced function, qualitative alteration and constitutive expression. This

spectrum of expression changes occur because mutations in the gene, or its regulatory pathway, take different forms, including non-sense mutations, mis-sense mutations, splicing defects and disruption of regulatory sequences.

The wide variety of effects EMS mutagenesis can have on gene function is a strength of this approach and aids in the identification of genes which may not be identified by other methods. Further, the generation of a spectrum of alleles (*i.e.* from mild to strong) provides nuanced insight about the function of a gene. For example, strong alleles of *SHOOTMERISTEMLESS* (*STM*) result in plants with no shoot apical meristem, a lethal phenotype that does not provide much insight into the function of the gene beyond the fact that it is essential. However, weaker *STM* alleles result in a reduced shoot apical meristem, providing crucial information about the function of the gene (Carles & Fletcher, 2003). Insertional mutagenesis and reverse genetics only provide information about the effect of down regulating a gene, which may be lethal, or may have no visible effect on the phenotype. A benefit of insertional mutagenesis, however, particularly T-DNA, is that the T-DNA provides a "tag" that can be used later to identify the disrupted gene via PCR.

A disadvantage of EMS mutagenesis is that the mutated genes can only be identified with MBC. Until recently, MBC has been a laborious and time-consuming technique. However, with the advent of new genomic tools and complete Arabidopsis genomic sequence, MBC has become easier, and it has recently become possible for one person to undertake a mapping project and complete it within the time frame of a Master's degree.

2.1.5 Map-Based Cloning : A General Introduction

MBC is an iterative, indirect approach to assigning gene function. Mapping narrows the genetic interval containing the mutation by excluding regions of the genome that do not contain the mutation. Mapping resolution is determined by the size of the mapping population. A resolution of 10 – 40 kb can be obtained in populations as small as 1,000 plants (Lukowitz, *et al.*, 2000). Once a narrow region has been identified, clues about the mutant gene may be obtained by examining annotated genes within that region or by complementation testing.

Over 750 natural accessions of Arabidopsis have been collected from around the world and are available from the Arabidopsis seed stock center ABRC. The accessions vary in form, development and physiology and are referred to as ecotypes. MBC in Arabidopsis is greatly facilitated by the availability of complete genome sequences for two ecotypes, Landberg *erecta* and Columbia (Jander, *et al.*, 2002). Comparisons of the genomic sequences led to the identification of polymorphisms at known locations throughout the genome. The polymorphisms are genetic markers; known sequence differences at specified physical locations on the chromosomes.

A schematic representation of the steps involved in MBC is presented in Figures 2.0 and 2.1. MBC starts with identification of a mutant with an interesting phenotype in one of the ecotypes, for example in the L*er* ecotype. The mutant is then crossed with a Col-0 wild-type plant to generate the mapping population. The phenotypes of the F_1 plants are observed to determine if the mutation is dominant or recessive. Tissue is collected from the F_1 plants and they are left to self-fertilize. The seeds collected from the F_1 plants are planted to

generate the F_2 progeny. An initial mapping population of 600 F_2 plants is examined for segregation pattern of the mutant gene and tissue is collected from the F_2 homozygotes.

The first mapping reactions, referred to as first-pass mapping, are conducted with DNA from the F_1 parents and F_2 homozygous mutants. First-pass mapping is done with 22 markers, evenly distributed across the genome. For each marker, two bands are amplified in the F_1 parents, the *Ler* band and the Col-0 band, because the F_1 plants have an equal complement of both parental DNA types. In the F_2 homozygotes, there is also an equal complement of both parental DNA types at most marker loci and two bands are amplified. However, at the marker loci close to the mutant gene, the background DNA is from the parent where the mutant originated. In this example, the DNA close to the mutant gene is from the *Ler* ecotype, and only the *Ler* parental marker is amplified in this region. First-pass mapping determines the chromosome location of the gene and provides a narrow region within that chromosome where the gene is located.

Following first-pass mapping, a larger population of F_2 plants, 3400, is screened for recombination events within the narrow region and the region progressively narrowed. When a small area has been identified, the precise identity of the gene can be determined with a series of approaches, including complementation and examination of the sequencing database for candidate genes.

Figure 2.0. A step-by-step guide to map-based cloning. Schematic representation of the steps involved in map-based cloning. Adapted from Jander *et al.*, 2002.

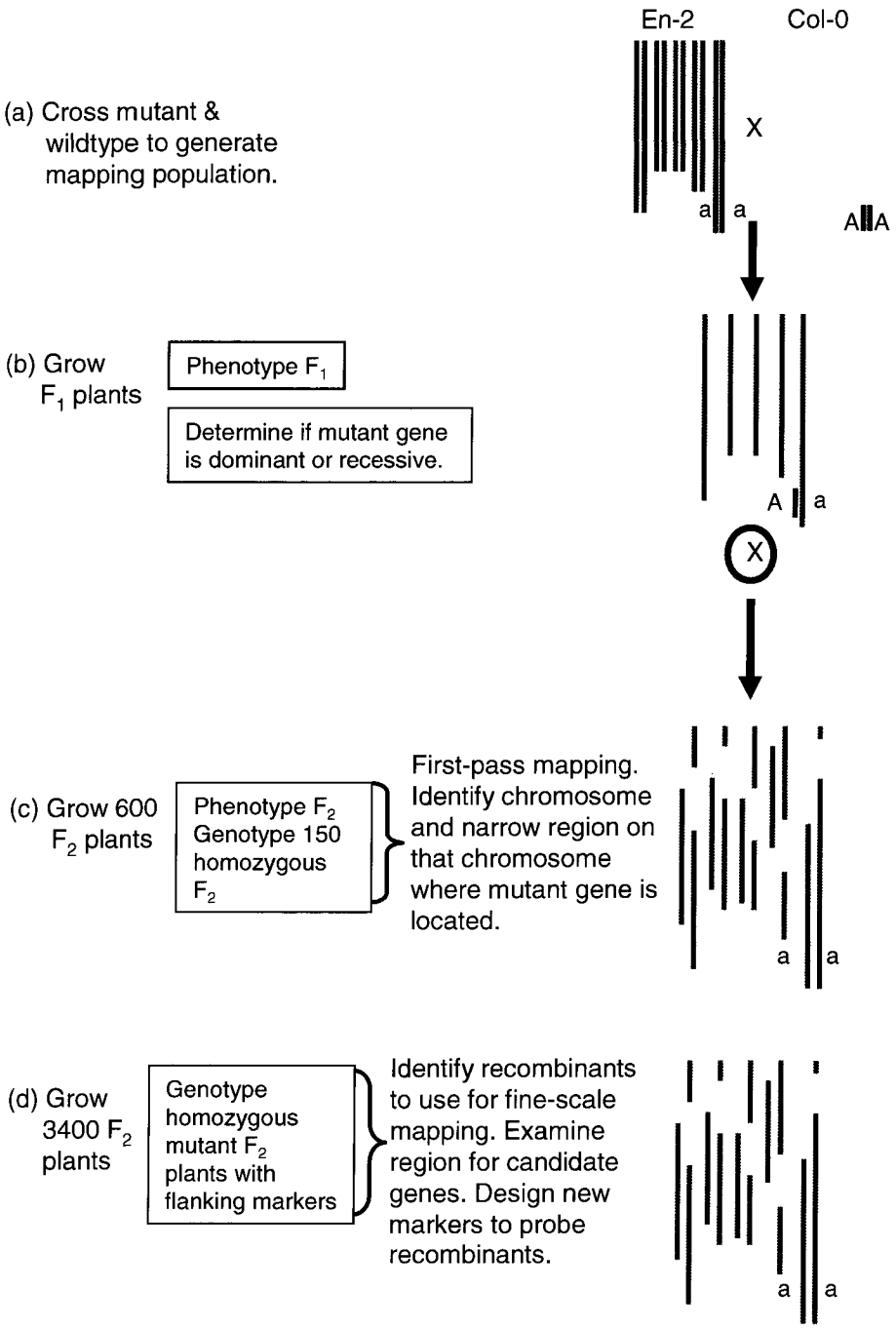
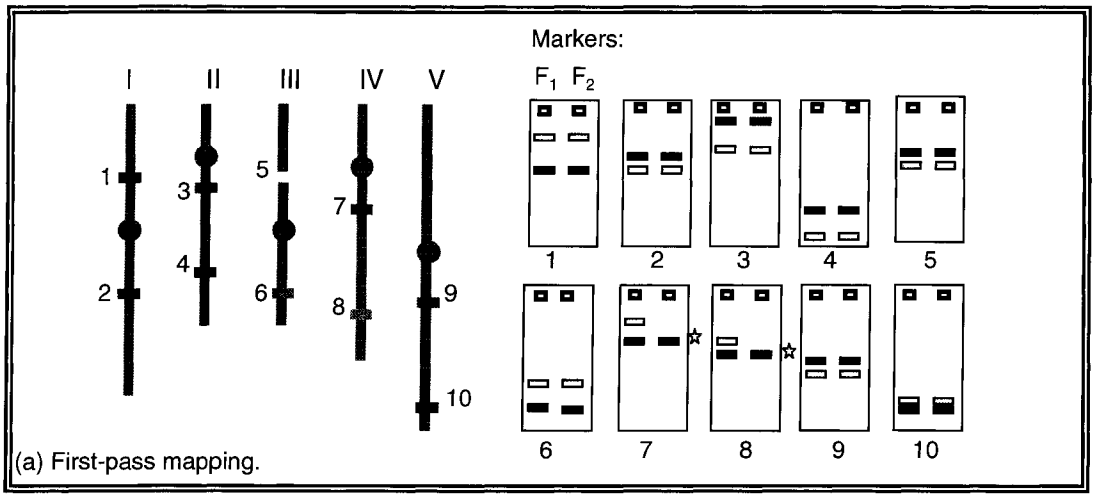
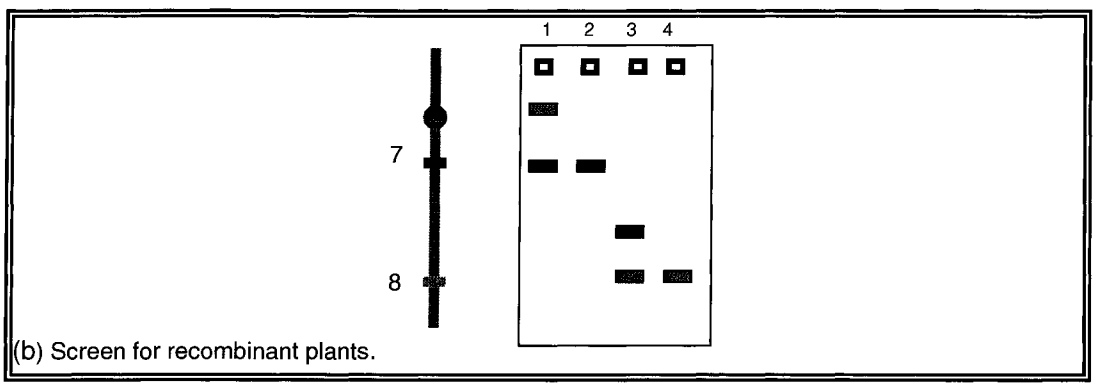
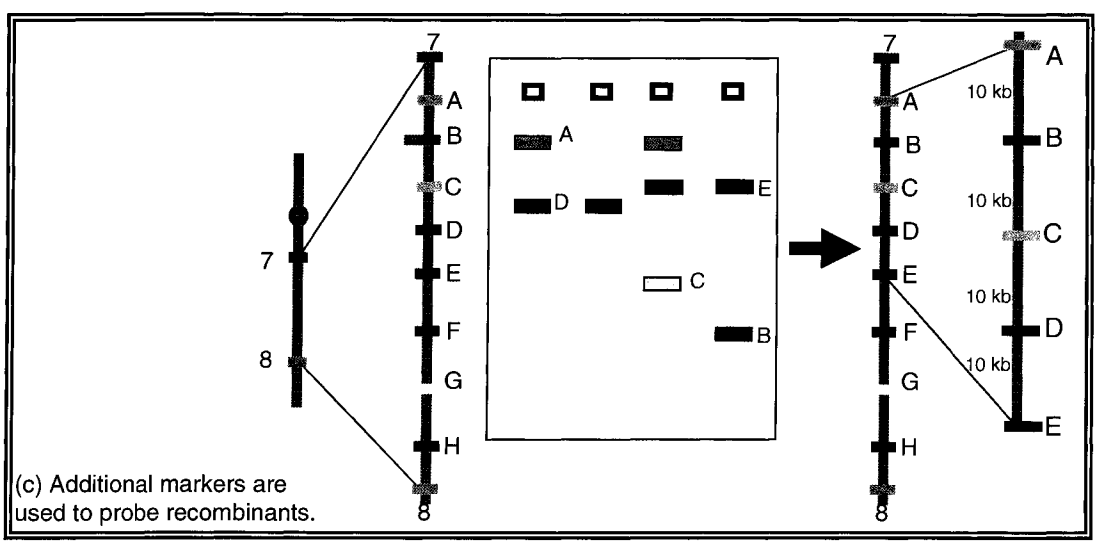


Figure 2.1. A step-by-step guide to first-pass mapping in map-based cloning. Schematic representation of the steps involved in first-pass mapping. Panel (a) F_1 and homozygous, bulked F_2 samples are screened with the first-pass markers to identify the chromosomal location of the mutant gene. Parental markers are amplified at all loci in the F_2 samples except markers 7 and 8, (starred) where only the En-2 parental band is amplified. This indicates that the gene is located on chromosome 4, linked to markers 7 and 8. Panel (b) Homozygous F_2 plants are screened to identify individuals with a recombination event in the area flanked by markers 7 and 8. Plants 1 and 3 have no recombination in this area, plants 2 and 4 have had a recombination event in this region. Only recombinants are screened further, plants without recombination discarded. Panel (c) Additional markers are chosen to probe the identified region. The presence of markers A, B, C, D and E indicates that these markers remained with the mutant gene after homologous recombination. The gene is located in the 40 kb region flanked by markers A and E. Annotated genes in this area can be examined for candidate genes and BACs used for complementation testing in the region.





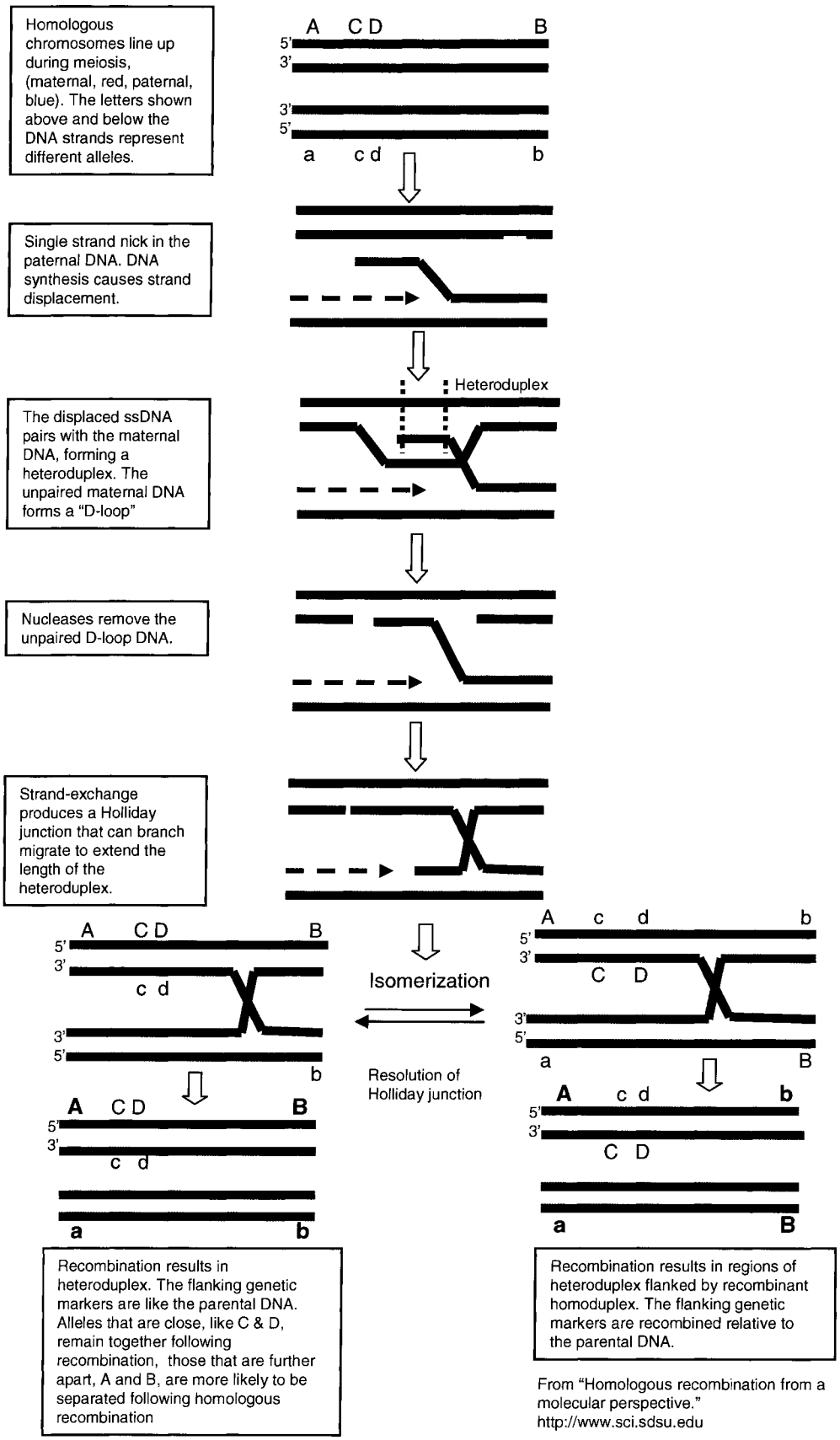


2.1.6 Recombination

The construction of genetic maps depends on recombination events between two non-sister chromatids of each pair of homologous chromosomes during meiosis. A general mechanism for homologous recombination is presented in Figure 2.2. Mapping takes advantage of the high frequency of genetic exchange events in meiosis. Meiotic recombination ensures proper chromosome disjunction and contributes to the genetic diversity of gametes. Each pair of homologous chromosomes undergoes at least one crossover to allow for correct segregation at the first meiotic division.

Recombination is not uniform across chromosomes, but varies in different regions of the chromosome. High levels of recombination occur in gene-rich zones and recombination is suppressed in centromeric regions. Frequency of recombination is not the same for all gene combinations; it is influenced by the proximity of one gene to another. If two genes are located close together on a chromosome, the likelihood that a recombination event will separate them is less than if the two genes are further apart. Mapping makes use of this phenomenon: markers located close to a target gene remain with the gene following recombination, whereas markers further away will assort independent from that gene following recombination. The frequency of recombination is measured as a percent, 1% recombination means that two genes are close together and recombine 1% of the time. The rates of recombination are converted to mapping units called centiMorgans (cM), where 1% recombination is equal to 1 cM. The first genetic maps were assembled by observing rates

Figure 2.2. Schematic representation of the general mechanism for homologous recombination.



From "Homologous recombination from a molecular perspective." http://www.sci.sdsu.edu

of recombination events, converting them to cMs and constructing genetic maps manually (Griffiths, *et al.*, 2000).

2.1.7 Genetic Maps and Physical Maps

Geneticists have a longstanding interest in determining the precise location of genes on chromosomes and consequently gene maps. There are two types of gene maps, genetic maps and physical maps. A genetic map, or linkage map, is built using genetic markers and linkage analysis to map a genetic region. The distances are based on co-segregation of alleles of nearby markers and depend upon recombination. These maps can be effective at determining gene order, but as recombination does not occur uniformly throughout the genome the genetic distance between two genes may not correspond with the physical distance. Despite this, genetic maps have been necessary resources for building physical maps, particularly sequence maps.

Sequence maps are the best example of physical maps and the most informative type of gene map available. These are built by cloning overlapping regions of the genome into a vector (*e.g.* a bacterial artificial chromosome, BAC), sequencing the BAC and then lining up the sequences to determine the precise base pair sequence of the entire genome. While these maps are time-consuming, expensive and difficult to produce, they also provide an important resource for molecular biologists.

The first linkage map of Arabidopsis was published in 1983 and consisted of 76 genes with phenotypes ranging from gibberellin dwarves, trichome morphology, male sterility and hypocotyl length (Koorneef, *et al.*, 1983). As mapping data was based on two-point crosses,

precise gene ordering was not determined. Despite this, the recognition that it was possible to map genes in Arabidopsis was a major development in plant biology and has been posited to be one of the reasons for the wide spread adoption of Arabidopsis as a model organism (Meyerowitz, 2001).

The first linkage map was followed by the first Restriction Fragment Length Polymorphism (RFLP) map in 1988 (Chang, *et al.*, 1988) and the first examples of MBC, or positional cloning, in the early 1990s (Arondel, *et al.*, 1992; Giraudat, *et al.*, 1992). From that point, genetic maps of molecular markers were rapidly developed and quickly expanded to include Random Amplified Polymorphic DNAs, Cleaved Amplified Polymorphic Sequences (CAPS), Simple Sequence Length Polymorphisms (SSLPs) and Amplified Fragment Length Polymorphisms (AFLPs). The Recombinant Inbred Lines (RILs) developed by Lister and Dean (Lister & Dean, 1993) became established as the standard for placement of molecular markers. Genes were placed on this map by determining recombination frequencies with linked markers. Consequently, two maps were developed in parallel, a classical genetic map and a recombinant inbred line map.

2.1.8 Map-Based Cloning is Facilitated by Genomic Tools

The major theoretical and technical breakthroughs in the intervening years have led to dense genetic maps and physical maps where genes and markers are located at precise sequence positions. Until relatively recently, MBC was a time-consuming activity. However, development of detailed genetic and physical maps, marker databases, as well as simple techniques to detect DNA polymorphisms has catalyzed MBC. Further details about the roles of these resources are provided below.

2.1.9 Molecular Markers

The markers in this mapping project included simple sequence length polymorphisms (SSLPs), single nucleotide polymorphisms (SNPs), and insertions or deletions (INDELs). All are PCR-based, abundant, and easy to analyse with agarose gels. SSLPs are loci where there is a stretch of dinucleotide repeats of varying length. SNPs are loci where the sequence difference between the two ecotypes varies by just one nucleotide. INDELs are loci where one ecotype has a region of the genome that has been inserted or deleted relative to the other. A schematic representation of these three marker types is presented in Figure 2.3.

2.1.10 Map-Based Cloning: Applicable in all Arabidopsis Ecotypes

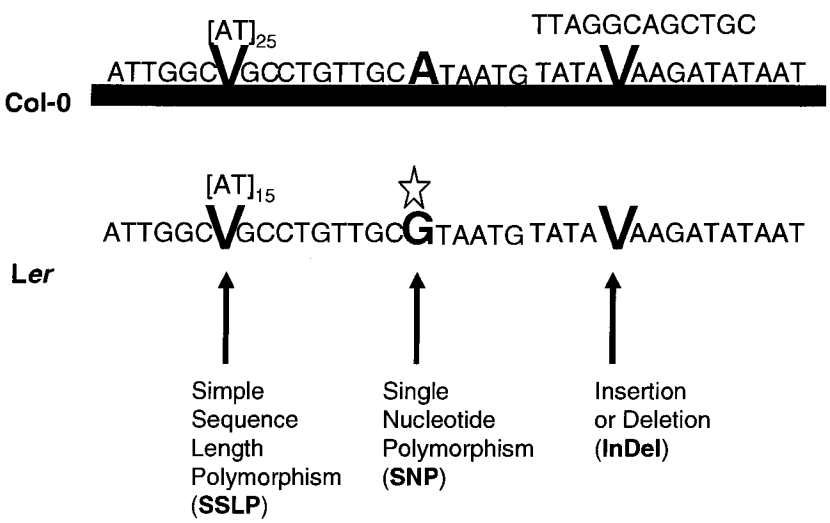
The majority of markers designed for mapping were generated through comparisons of Col-0 and L*er* genomic sequences. For this reason, it might seem that successful mapping projects can only be conducted using these two ecotypes. However, Jander contends that if there is random assortment of markers among Arabidopsis ecotypes, then 50% of the Col-0/L*er* polymorphisms should be useful to map with any pair of ecotypes (Jander, *et al.*, 2002). Mapping data obtained with AFLPs supports this assertion. AFLPs are generally due to underlying SNPs. It has been shown that 50% of the mapped Col-0/L*er* AFLP markers can be used for segregation analysis in a series of ecotype backgrounds, including Col-0/C24, Col-0/Wassilewskija and Col-0/Cape Verde Islands crosses (Peters, *et al.*, 2001). To use markers to map in alternate ecotypes, they need to be tested prior to mapping to determine if they amplify a PCR product and if that product can be distinguished from the one amplified with the other parental ecotype. However, as PCR markers are easy to use, this is a minor task. Further, even if a marker amplifies similar-sized pieces in both parents, there are many markers available so another one can be chosen.

2.1.11 Map-Based Cloning: A Step-by-Step Guide

2.1.12 Generating a Mapping Population

As Figure 2.0 indicates, after a mutant with an interesting phenotype has been identified, the first step of MBC is to generate a mapping population. This is done by crossing the mutant plant with a wild-type plant. MBC has to be done in F_2 plants with hybrid ecotype backgrounds. Thus, as depicted in Figure 2.0 (a), the original parent plants must be from two different ecotypes. Seeds from the parental cross are planted to generate the F_1 population. In this population the plant phenotypes are scored and it is determined if the mutant gene is recessive or dominant. For example, if all of the F_1 plants have the mutant phenotype then the mutant gene is dominant. Conversely, if all of the F_1 plants have the wild-type phenotype then the mutation is recessive. As EMS predominantly generates recessive mutations, plants in the F_1 population will not exhibit the mutant phenotype. F_1 plants are left to self-fertilize to generate the F_2 population. In the F_2 population the plants are scored for phenotype and the segregation pattern for the mutant gene determined.
If the mutant is distinct from the wild-type phenotype, recessive and easy to score, only homozygous mutant F_2 plants are used in mapping. The benefit of mapping just with homozygous F_2 mutants is that the phenotype indicates the mutant gene is present and fewer plants can be used. Mutations that result in co-dominant expression can be difficult to score, because it may not be possible to distinguish between a heterozygote and a homozygote. In this case, all of the F_2 plants are used in the mapping reactions. In either scenario, tissue from both F_1 and F_2 plants is collected and genomic DNA extracted. If the mutation demonstrates a 1:3 Mendelian segregation and homozygous mutants are easy to score (*i.e* is a recessive mutation), then, as indicated in Figure 2.0 (c), 600 plants are screened in first-pass mapping.

Figure 2.3. Schematic representation of three types of PCR-based molecular markers. The diagram depicts three types of PCR-based molecular markers: simple sequence length polymorphisms (SSLPs), single nucleotide polymorphisms (SNPs) and insertions or deletions (INDELs). These markers were identified by comparing the genomic sequences of the Columbia and Landsberg *erecta* ecotypes. All of these markers were identified in non-coding regions of the genomes.



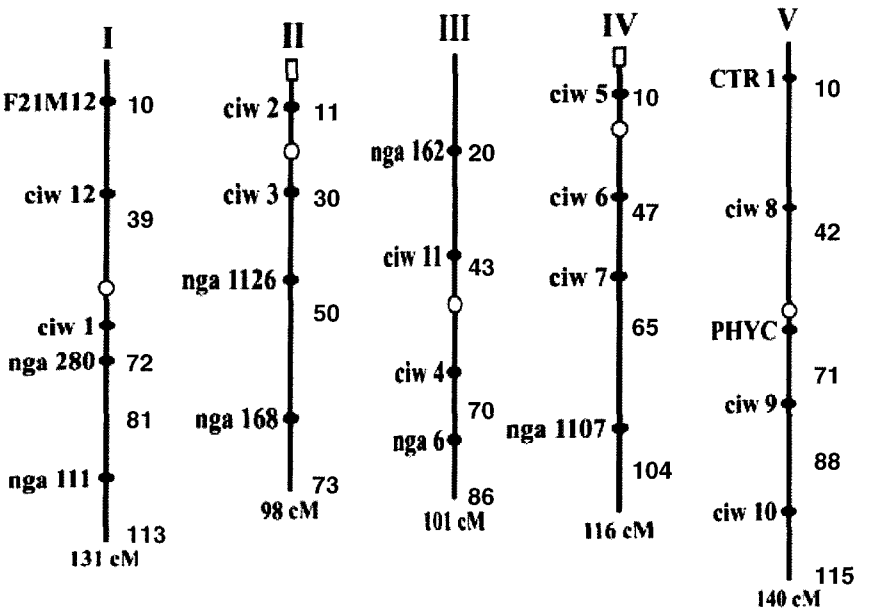
Of these, DNA is extracted from the homozygous mutants (expected to be 150 individuals). The extracted genomic DNA is pooled and screened with the first set of markers. For fine-scale mapping, 4,000 plants are grown and 1,000 homozygous mutants screened.

2.1.13 First-Pass Mapping

A schematic representation of first-pass mapping is presented in Figure 2.1 (a). The objective of first-pass mapping is to identify the chromosomal region where the mutant gene is located. This is done indirectly, by excluding regions that do not contain the mutant gene. In first-pass mapping the plants are examined with a group of 22 markers spaced evenly along each chromosome, roughly every 20 cM, as depicted in Figure 2.4.

Half of the DNA in the F_1 plants originated from one parent ecotype and half from the other parent ecotype. Therefore, in the F_1 plants both parental markers are present at each marker locus, as depicted in Figure 2.1 (a). In contrast, in the F_2 plants, both parental bands are present at the majority of the marker loci. However, for the markers close to the mutant gene only one parental band is present, assuming complete linkage. This is because in the region close to the mutation, the surrounding DNA originated from the mutant's parental ecotype, as depicted in Figures 2.0 (c). For example, if the mutant was generated in the L*er* ecotyoe, markers close to the mutant gene will be from L*er*. This is because genes, or a gene and a marker that are close together, do not assort independently during homologous recombination. In first-pass mapping, segregation of the markers in the F_1 plants serves as the control variable and is compared to the segregation pattern in the F_2 plants. The homozygous mutant F_2 plants are screened as a group, using a bulked segregant analysis approach (Michelmore, *et al.*, 1991). In bulked segregant analysis, DNA from the

Figure 2.4. Schematic representation of the five chromosomes of Arabidopsis, labeled with the 22 markers used in first-pass mapping. The cartoon depicts the five Arabidopsis chromosomes, labeled with the 22 PCR-based markers used in first-pas mapping. The markers are spaced evenly along the chromosomes, located roughly every 20 cM. Open rectangles represent telomeres and open circles centromeres. The distances between the markers, measured in centiMorgans (cM), are indicated in blue. Diagram from Lukowitz *et al.*, 2000.



150 homozygous mutant F_2 individuals is pooled and marker segregation analyzed in the entire population, permitting rapid detection of segregation patterns. The benefit of bulked segregant analysis is that it provides data about the overall segregation pattern, eliminating the need to screen each individual separately. First -pass mapping excludes regions where the mutant gene is not located, identifies the chromosome on which the gene is located and a narrow region within that chromosome. Consequently, first-pass mapping reduces the total area where the gene could be located from 5 chromosomes to a well defined candidate area on a single chromosome.

In the next step, only the narrow region is examined and the two markers that were shown to flank the region are used as the outer-most markers for fine-scale mapping.

2.1.14 Fine-scale mapping: Screening for Recombinants

The goal of fine-mapping is to narrow the area containing the target gene to 40 kb or less, which corresponds to roughly 0.16 cM. The number of F_2 plants that need to be analysed to have a 95% chance of recombination increases as the mapping region decreases. For this reason, more F_2 plants must be examined during this stage of mapping. There is no consensus about the total number of F_2 plants required to map a gene, with some recommending 3,000 – 4,000 (Jander, *et al.*, 2002) and others successfully mapping a gene with as few as 1,000 F_2 plants (Lukowitz, *et al.*, 2000). In general, a larger fine-scale mapping population of 3,000 – 4,000, increases the chances of mapping the gene to 40 kb and eliminates the time-consuming step of waiting for more plants to grow if the original population is too small.

Plants are genotyped individually with markers that are about 5% recombination, or 5 cM, apart and define the interval containing the gene (Lukowitz, *et al.*, 2000). Additional markers are selected from the The Arabidopsis Information Resource (TAIR) database (www.arabidopsis.org), which provides detailed information regarding markers, including type (*i.e.* SSLPs, SNPs, RFLPs, *etc.*) precise location, primer sequences, ecotypes they have been tested in and PCR product size. The Sequence Viewer (SeqView) tool on the TAIR website generates a map of marker locations relative to annotated genes in the same area which aids visualization. Additional markers in the region are used to probe the recombinant plants until the genetic region is narrowed down as much as possible. Figure 2.1 presents a schematic representation of fine-mapping, including diagrams of the expected results on agarose gels.

The benefit of mapping with markers developed from sequence data is that the precise location of the markers is known, assisting precise identification of the mutant locus. Further, the TAIR database SeqViewer tool provides detailed information about the Arabidopsis genome. This information assists in identification of new markers and examination of genes in a mapping region as possible candidate genes.

2.1.15 MBC: Enhanced by the Availability of Mutants in Stock Centres

After narrowing a genomic region containing the mutation, several approaches can be used to identify the gene. Primary among these are complementation experiments and sequencing. Complementation tests involve transforming overlapping pieces of wild-type DNA into the mutant to see which sequences are capable of restoring the wild-type phenotype. The entire Arabidopsis genome is available cloned into BACs in 10 kb pieces containing annotated genes or transcripts (Rhee, *et al.*, 2003). Once a genomic region is identified, the BACs that represent this region can be ordered and the complementation tests conducted by transforming the mutant plant with the BAC DNA. If transformation rescues the phenotype, then it can be concluded that the BAC contains the wild-type copy of the mutant gene.

If the region of interest is sufficiently small, it is also possible to sequence that region. Comparison of the mutant sequence to the wildtype makes it possible to identify the mutant gene and the location of the mutation within the gene. However, the definitive evidence should be provided by complementation experiments.

In this section of the thesis, I describe the MBC approach used to map the gene responsible for the *cam* phenotype. The details of the MBC process are described as well as the phenotypic characterization of the *cam* mutant plant. Finally, the gene discovered to be responsible for the *cam* phenotype is discussed.

2.2 Materials and Methods

2.2.1 Plant Strains

Columbia wildtype (Col-0), Landsberg *erecta* wildtype (L*er*), Enkheim (En-2) wildtype and *cam* (ABRC stock #cs313), created in the Enkheim ecotype, were used throughout the experiment. Seeds were obtained from the ABRC.

2.2.2 Growing Plants in Soil

Arabidopsis seeds were suspended in sterile 0.1% agarose (w/v) and stratified at 4°C for 4d to promote uniform germination. Suspension of the seeds in 0.1% agarose assists separation and sowing of individual seeds. Following stratification, seeds were direct-seeded onto moist soil (ProMix BX) in 48 or 72 cell-pak inserts via sterile Pasteur pipettes. The tray was covered in plastic wrap and two clear plastic humidity domes to create a warm, moist environment conducive to germination. Trays were placed in AC60 growth chambers (Enconair Ecological Chambers, Winnipeg, MB) and incubated under long day conditions (22°C, 130-190 $\mu E/m^2/s$; 16 hr light/8 hr dark). When the cotyledons were produced, the plastic wrap was removed and seedlings continued to grow under two humidity domes. The humidity domes were removed following production of first true leaves. Plants were watered every two days.

2.2.3 Phenotypic Characterization of the *cambiumless* Mutant

2.2.4 Flowering Time

The number of days after germination until the first appearance of a floral meristem was recorded for *cam* and En-2 wild-type plants. All plants were grown under long day conditions (16 hr light/8 hr dark).

2.2.5 Plant Morphology

Gross plant morphology was examined via dissecting microscope and cryo-scanning electron microscopy (cryo-SEM).

For dissecting microscopic examination, fresh tissues were removed from the plant and examined immediately. Images were recorded via a digital camera and iMac desktop computer attached to an Olympus SZX-12 dissecting microscope. Whole plant morphologies were compared, including leaves, early flowers, late flowers, abscission of floral organs and siliques.

For cryo-SEM examination, fresh tissues were frozen in liquid nitrogen slush at -190 °C and ice crystals were removed by sublimation at -90° C. Samples were placed on a cold stage fitted to the microscope, sputter-coated with aluminum and examined with a JEOL 6400 Scanning Electron Microscope, fitted with Oxford instruments, an Oxford CT 1500C Cryotransfer System and Cold Stage, using an accelerating voltage of 9 kV.

2.2.6 Microscopic Observations of Vasculature in *cam* and Wildtype

The morphology of the vasculature was examined via light microscopy and cryo-SEM.

For the light microscopy, samples were fixed in glutaraldehyde, dehydrated through an ethanol series and plastic-embedded in LR White. Sections 2 μm thick were cut with a microtome and stained with Toluidine Blue-O. Samples were observed with a Zeiss compound light microscope. This work was done by Sharon Regan.

For the cryo-SEM, fresh samples were frozen in liquid nitrogen at -196°C, embedded in tissue freezing medium (Electron Microscopy Sciences, Hatfield, PA), re-frozen in liquid nitrogen, at -196°C to solidify the tissue freezing medium, and planed with a cryo-microtome. Ice crystals were removed by sublimation at -90° C, samples were placed on a cold stage fitted to the microscope, sputter-coated with aluminum and examined with a JEOL 6400 Scanning Electron Microscope, fitted with Oxford instruments, an Oxford CT 1500C Cryotransfer System and Cold Stage, using an accelerating voltage of 9 kV.

2.2.7 Extraction of Genomic DNA

DNA was extracted from leaf tissue, using a rapid DNA extraction method (Weigel & Glazebrook, 2002). Leaves were flash frozen in liquid nitrogen, a leaf disc punched and was ground to a fine powder in 1.5 mL microtubes with the round end of a cold, flexible polyethylene stirring rod (Sarstedt, Germany). A 400 μL volume of DNA extraction buffer (200 mM Tris, pH 7.5, 250mM NaCl, 25 mM EDTA, 0.5% SDS) was added and the samples were vortexed. The samples were then directly centrifuged or left for a maximum of 1 hr while other samples were prepared. Samples were centrifuged at 13,000 rpm, at 20°C in an Fisher AccuSpin 3R benchtop centrifuge (Fisher Scientific, Ottawa, ON) for 1 min. The same centrifuge was used throughout the procedure. Following centrifugation, 300 μL of supernatant fluid was transferred to a new 1.5 mL microtube and 400 μL of isopropanol was added to precipitate the DNA. The tubes were vortexed, incubated at room temperature for 2 min and centrifuged at 13,000 rpm at 20°C for 10 min, to pellet the precipitated DNA. The supernatant was removed, 800 μL of 70% (v/v) ice-cold ethanol was added to wash the DNA pellet and the tubes were centrifuged at 13,000 rpm at 4°C for 5 min. The supernatant was decanted and the DNA pellets were dried by inverting tubes on a Kimwipe, on top of a

60°C heatblock, for 5– 10 min. The resulting DNA pellet was re-suspended in 100 μL sterile water or 100 μL 1X TE (10 mM Tris, pH 7.4, 0.1 mM EDTA) and stored at -20°C.

2.2.8 Amplification of Mapping Markers by PCR

The PCR amplifications were conducted with genomic DNA. Optimum $MgCl_2$ concentrations were pre-determined for each primer set (see Table 2.0). The PCR protocol was adapted from the one used by Lukowitz (Lukowitz, *et al.*, 2000). The PCR master mix was composed of PCR buffer (50 mM potasium chloride, 10 mM Tris-HCl pH 9.0 @ room temperature, 0.1% Triton X-100), 1 μM of each primer, 200 μM of each dNTP, 1.0 to 2.5 mM $MgCl_2$, 8.4 μL of template genomic DNA per reaction, resuspended in sterile H_2O or 1X TE, pH 7.4, and 0.5U of *Taq* polymerase (Invitrogen, Burlington, ON) for a final reaction volume of 20 μL. The cycling parameters were:

1. Initial denaturation: 1 min at 94°C
2. Denaturation: 30 s at 94°C
3. Primer annealing: 30 s at 55°C
4. Primer extension: 30 s at 72°C
5. Steps 2 – 4 repeated 39 times

For each mapping reaction, negative and positive controls were established. The negative control consisted of a PCR reaction with no DNA added and the positive control consisted of a PCR reaction set up with the F_1 DNA. All PCR reactions were conducted in an Eppendorf Mastercycler Gradient thermalcycler (Eppendorf, New York, NY).

Chromosome	Marker	Alias	Forward primer (5'->3')	Reverse primer (5'->3')	[$MgCl_2$] (mM)
I	F21M12	1.1	ggctttctcgaaatctgtcc	ttacttttgcctcttgtcattg	2.0
	ciw12	1.2	aggttttattgcttttcaca	ctttcaaaagcacatcaca	1.5
	ciw1	1.3	acattttctcaatccttactc	gagagcttctttatttgtgat	2.0
	nga280	1.4	ctgatctcacggacaatagtgc	ggctccataaaaagtgcacc	1.5
	nga111	1.5	tgtttttttaggacaaatggcg	ctccagttggaagctaaaggg	1.5
II	ciw2	2.1	cccaaaagttaattatactgt	ccgggttaataataaatgt	2.5
	ciw3	2.2	gaaactcaatgaaatccactt	tgaacttgttgtgagctttga	2.5
	nga1126	2.3	cgctacgcttttcggtaaag	gcacagtccaagtcacaacc	2.0
	nga168	2.4	tcgtctactgcactgccg	gaggacatgtataggagcctcg	2.0
III	nga162	3.1	catgcaatttgcatctgagg	ctctgtcactcttttcctctgg	1.0
	ciw11	3.2	ccccgagttgaggtatt	gaagaaattcctaaagcattc	2.5
	ciw4	3.3	gttcattaaacttgcgtgtgt	tacggtcagattgagtgattc	2.5
	nga6	3.4	tggatttcttcctctcttcac	atggagaagcttacactgatc	1.0
IV	ciw5	4.1	ggttaaaaattagggttacga	agatttacgtggaagcaat	2.0
	ciw6	4.2	ctcgtagtgcactttcatca	cacatggttagggaaacaata	2.0
	ciw7	4.3	aatttggagattagctggaat	ccatgttgatgataagcacaa	2.0
	nga1107	4.4	gcgaaaaaacaaaaaaatcca	cgacgaatcgacagaattagg	1.5
V	CTR1	5.1	ccacttgtttctctctctag	tatcaacagaaacgcaccgag	2.5
	ciw8	5.2	tagtgaaacctttctcagat	ttatgttttcttcaatcagtt	2.0
	PHYC	5.3	ctcagagaattcccagaaaaatct	aaactcgagagttttgtctagatc	2.0
	ciw9	5.4	cagacgtatcaaatgacaaatg	gactactgctcaaactattcgg	1.0

Table 2.0. The PCR-based markers used in the first-pass mapping reactions. A group of 22 markers were used in first-pass mapping. Each marker name and its alias is listed along with the primer sequences, forward and reverse, and the $MgCl_2$ concentration used for each primer pair in the PCR reactions. Obtained from Lukowitz *et al.* (2000).

2.2.9 Gel Electrophoresis of PCR Products

Products from the PCR amplification were analysed on a 4% TBE (1X Tris-Boric Acid-EDTA, pH 7.6) agarose gel prepared with 3:1 high resolution agarose (Amresco,Ohio,OH). Gels were pre-stained with 5 μL of 10 mg/mL ethidium bromide added to 100 mL of warm agarose solution. Orange G running buffer (30% glycerol in dH_20 with enough Orange G added to yield a bright orange solution) was added directly to the PCR products (4 μL into 20 μL PCR reaction) following amplification and the entire reaction electrophoresed at a rate of 8V/cm for 2 to 5 hr. Samples were electrophoresed alongside 10 μL of Norgen's Mininsizer DNA ladder (Norgen, St. Catherine's, ON) consisting of DNA pieces varying in size from 600 to 25 bp. All gels were visualized with the AlphaImager IS-2200 by Alpha Innotech (Alpha Innotech, San Leandro, CA)

2.2.10 Verifying Mapping Markers in the Enkheim-2 Ecotype

The markers used in this study were initially designed for use in mapping projects with Col x L*er* mapping populations. Before the markers could be used to map a gene with En-2 as parental ecotype, it was necessary to determine if the markers would (i) amplify products from En-2, and (ii) yield products of distinguishable from the Col-0 or L*er* products at the same loci. DNA was extracted from En-2 leaves and polymerase chain reactions set up using the first set of primers (see Table 2.0) to compare product size among each of the three ecotype backgrounds.

2.2.11 Crossing: Generation of F_1 and F_2 Mapping Populations

Two mapping populations were prepared, a Col-0 x *cam* and a L*er* x *cam* population. Crossing was initiated when all of the plants were flowering, approximately 4 weeks old. The *cam* plants served as pollen donors and the Col-0 and L*er* plants as pollen acceptors.

A stage 12 flower bud (Smyth, *et al.*, 1990) with closed sepals was used as the female for crossing. A small piece of laboratory tape was attached to the base of the chosen flower for later identification. The tape also provided a handle during crossing, decreasing the amount of mechanical stress on the stem. Using sharpened, fine-tipped forceps, the sepals, developing petals, and developing stamens were removed from the flower bud, exposing the carpel. The carpel was extended to its full length with stigmatic papillae on its surface. Pollen from a *cam* flower was gently dusted onto the stigmatic surface of the exposed carpel. To eliminate pollen contamination from wild-type plants, the carpel was covered with a piece of plastic wrap. In some cases the same carpel was pollinated several times over the course of 2 – 4 days to ensure successful pollination. When carpel extension was visible, the plastic wrap was removed from its surface. Siliques were left to develop on the plants. Siliques containing seeds from the crosses were harvested after the silique had fully extended and dried.

F_1 seeds from each mapping population, Col-0 x *cam* and L*er* x *cam*, were planted and grown under long-day conditions. The phenotypes of the F_1 plants were recorded. Rosette leaves were collected from the F_1 plants, flash frozen in liquid nitrogen and stored at -80°C. The F_1 plants were permitted to self-fertilize and the F_2 seeds were collected. F_2 seeds from each mapping population were sown directly on soil. A total of 600 F_2 seeds from each mapping

population (*i.e.* 1200 total plants) were grown under long-day conditions. The phenotypes of the F_2 plants were recorded. For each plant, rosette leaves were collected, flash frozen in liquid nitrogen and stored at -80°C. The DNA from each plant was extracted and used for first-pass mapping. Finally, F_2 plants were left to self-fertilize and F_3 seeds collected and cataloged from each individual F_2 plant.

2.2.12 First - Pass Mapping

2.2.13 Bulk Segregant Analysis of F_2 Mutants

First-pass mapping was conducted with plants from the Col-0 x *cam* mapping population. Two bulked DNA samples were prepared, one from the F_2 plants with the *cam* phenotype and one with the F_1 parental plants.

DNA was extracted from the rosette leaves of 96 F_2 plants with the *cam* phenotype. The DNA was divided into two aliquots of 50 μL, one of which was stored permanently at -20°C and the other used to prepare the bulked DNA pool (Michelmore, *et al.*, 1991). Half of each DNA sample, 25 μL, from 12 F_2 individuals was combined to make mini-pools of 300 μL. Each mini-pool was mixed gently by pipetting up and down. The mini-pools were then pooled, 25 μL of each pool into fresh microtubes, and gently mixed. The resulting pool included an equal amount of DNA from 96 individuals and was used for bulk-segregant mapping in the first-pass mapping.

The bulked DNA samples prepared from F_1 parental plants of the Col-0 x *cam* population contained DNA from 3 individuals.

2.2.14 First – Pass Mapping: Narrowing the Region of Interest

Once the mutation was narrowed to a region of a chromosome, additional markers were chosen in that region to further narrow the area of interest. PCR reactions were conducted with the F_1 and F_2 bulked DNA with primer sets flanking the region of interest.

2.2.15 Fine Mapping: Identifying Recombinants

For fine-scale mapping, a population of 3,400 F_2 Col-0 x *cam* plants was analysed. To rapidly screen the population for mutants, seeds were germinated on Murashige and Skoog (M&S) basal salt medium (Sigma - Aldrich, Oakville, ON), placed in AC60 growth chambers and incubated under long-day conditions. Plants which exhibited the *cam* phenotype in their true leaves were transferred to moist soil and returned to the growth chamber under long-day conditions.

Seeds were surface sterilized by incubating in 70% ethanol (v/v) for 2 min followed by an 8 min incubation in 30% bleach/0.2% Triton X-100 (v/v) solution. Sterilization solutions were removed with sterile Pasteur pipettes and seeds rinsed ten times with sterile, distilled water. Seeds were incubated in sterile, distilled water and placed in the refrigerator at 4°C for 4 d of stratification. Following stratification seeds were spotted onto the surface M&S basal salt medium (4.33 g M&S basal salts, 30 g sucrose, 0.8% agar (w/v), pH 5.7 – 5.8 with KOH) in 100 mm x 20 mm petri plates either directly with a sterile Pasteur pipette, or via a rapid spreading technique involving a spinning disc. On average, 100 – 250 seeds were spread on each plate.

Plants with the *cam* phenotype were carefully transferred from the petri plate with forceps and transplanted into moist soil. Plants were covered with two humidity chambers and returned to the growth chamber under long day conditions. The phenotypes of the F_2 plants were recorded and rosette tissue was collected, flash frozen in liquid nitrogen and stored at -80°C.

2.3 Results

2.3.1 Phenotypic Characterization of the *cambiumless* Mutant

2.3.2 Flowering Time

One of the first visible phenotypes of the *cam* mutant, besides overall stature, was the change in flowering time. Figure 2.5 presents the flowering time data for *cam* and wild-type plants grown under long-day conditions. The *cam* mutant flowered 18.5 +/- 3.5 days after germination as compared to the wild-type plant that flowered 27.3 +/- 2.6 days after germination. On average, the *cam* plants flowered 8.8 days before wildtype.

2.3.3 Plant Morphology

The plant morphology was investigated with dissecting microscopy and cryo-scanning electron microscopy (SEM). As Figure 2.6 (a) indicates, *cam* is smaller than the wild-type plant in all aspects of its biology. Images (b) and (c) show *cam* leaves, the leaves are small, partially folded with involute margins. The cells of the *cam* leaves are about 40 μm long, whereas the cells of the wild-type leaves are about 80 μm long, as depicted in (d) and (e). The images in (f) and (g) depict the development of a silique, as the silique develops the floral organs remain attached at the base. Images (h) and (j) show the *cam* flowers and image (i) depicts a wild-type flower, demonstrating that the *cam* flowers are smaller than wildtype and they have short, folded petals that do not extend beyond the stigmatic papillae at the end of the carpel. After about 5 weeks, the *cam* plant produced terminal flowers with homeotic conversions, specifically, unfused carpels as presented in (k) and (l).

Figure 2.5. Flowering time for *cam* mutant versus the En-2 wild-type plant. Average time to flower for En-2 wild-type plants versus *cam* mutant plants. On average, the *cam* plants flowered 18.5 +/- 3.5 days, (n = 50), after germination whereas the wild-type plants flowered 27.3 +/- 2.6, (n = 50), days after germination. All plants were grown under long-day conditions (16hr light/8 hr dark).

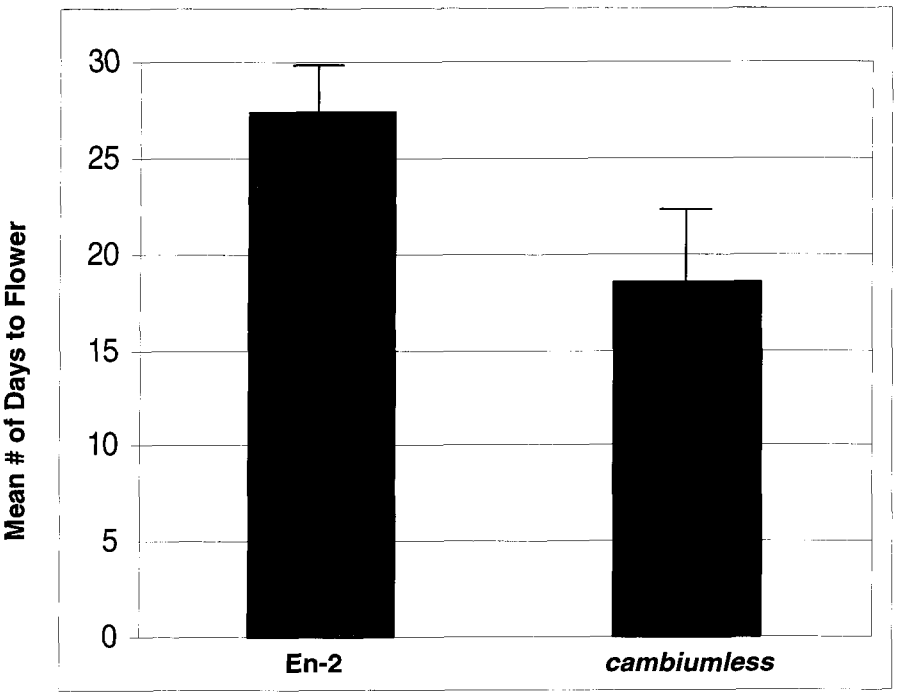
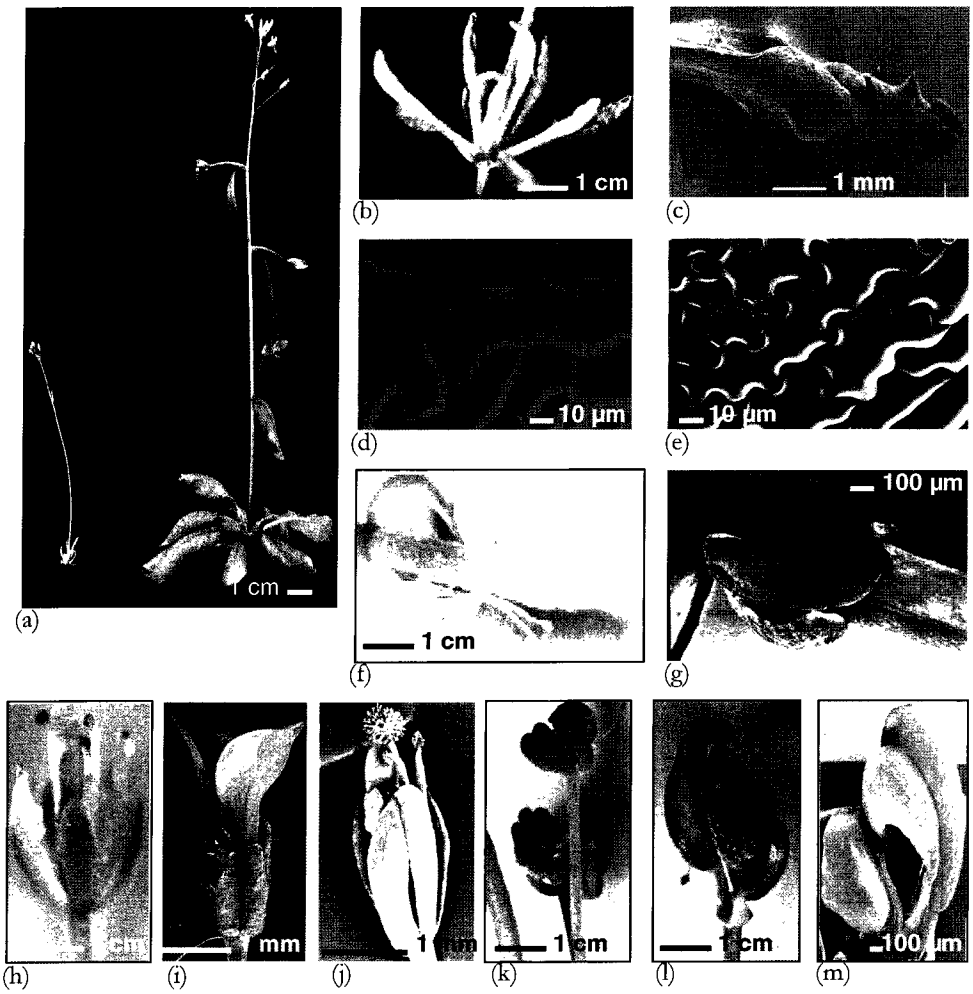


Figure 2.6. The *cam* mutant has several pleiotropic defects in morphology. Examination of the morphology of the *cam* mutant with dissecting and cryo-scanning electron microscopy. (a) The *cam* (left) and wild-type, En-2, plants, 4 weeks old; (b) *cam* rosette leaves, note that the rosette leaves are narrow with involute margins; (c) *cam* rosette leaf; (d) wild-type rosette leaf, abaxial, distal surface; (e) *cam* rosette leaf, abaxial, distal surface; (f – g) as the carpel develops into a silique in the *cam* plant, the floral organs don't abscise, but remain attached at the base of the silique, (f) *cam* developing silique; (g) *cam* developing silique; (h) early *cam* flower, note the folded, short petals that do not extend above the carpel; (i – j) comparison of wild-type flower and early *cam* flower; (k-m) late *cam* flowers with unfused carpels, (k), (l) interior of unfused carpels is visible, the arrow points to stigmatic papillae on the surface of the carpels; (l) SEM of unfused *cam* carpel.



2.3.4 Vasculature

It was initially observed that the *cam* mutant had considerably less rigidity in its stem and hypocotyl and hypothesized that it had altered secondary vasculature in its secondarily-thickened hypocotyl. This was investigated via light microscopy and clear differences were observed. Figure 2.7 displays the vasculature observed in a cross-section of a secondarily-thickened hypocotyl for a wild-type plant versus that observed for the *cam* mutant. The wild-type plant has secondary xylem and phloem, while the *cam* mutant plant has just primarily xylem and phloem. It is important to note that these two hypocotyls were observed at different magnifications, the wildtype was observed at 10X whereas *cam* was observed at 100X.

Cryo-SEMs of planed secondarily-thickened hypocotyls are presented in Figure 2.8. The diameter of the wild-type hypocotyl is greater than the diameter of the *cam* hypocotyl, 1.1 mm versus 0.47 mm. Despite the smaller diameter of the *cam* hypocotyl, the cells in the *cam* hypocotyl are larger than those in the wild-type hypocotyl, about 40 μm – 45 μm across whereas the wild-type cells are about 10 μm across. The *cam* hypocotyl image corresponds very well with the figures of a 29-day-old wild-type hypocotyl presented in Chaffey's work (Chaffey, *et al.*, 2002). The 29-day-old hypocotyl has a central stele containing the primary vasculature, encircled by a thin layer of the vascular cambium, the cork cambium and surrounded by large endodermis and epidermis cells. Further, the 29-day-old hypocotyl has an asymmetric shape, similar to that observed in Figure 2.8 (b) and unlike the symmetric shape observed in older secondarily-thickened hypocotyl, like that in Figures 2.7 (a) and 2.8 (a). For this reason, the large cells in Figure 2.8 (b) are labeled endodermis and epidermis cells. In a wild-type hypocotyl the periclinal divisions of the vascular cambium cause the

hypocotyl to expand, pushing against the endodermis and epidermis cells compressing them. In a secondarily-thickened hypocotyl, neither endodermis nor epidermis cells are visible. The presence of the large endodermis and epidermis cells in the *cam* hypocotyl provides further support for the observation that there is very little, or no secondary development in the *cam* plant. Considerable secondary thickening of the xylem cells in the wild-type hypocotyl and radial symmetry is visible. While some secondary thickening is visible in the centre of the *cam* hypocotyl, no radial symmetry is visible.

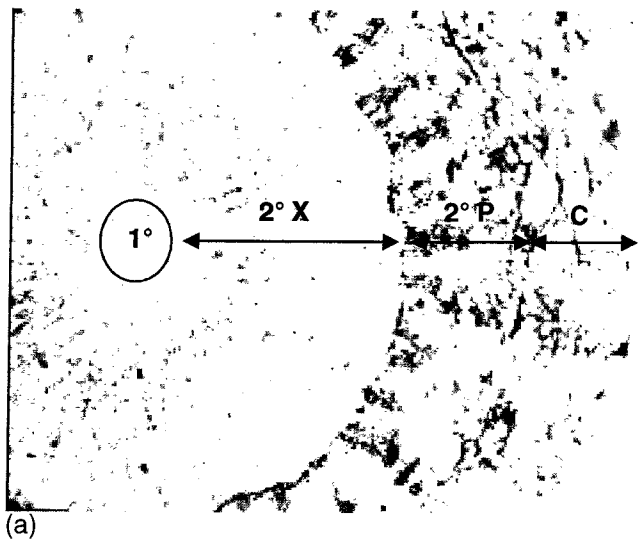
Cryo-planed SEM images of the *cam* stem and wild-type stem are presented in Figure 2.9. Vascular bundles are visible for both the *cam* stem, which has 6 bundles, and in the wild-type stem, with 14 bundles. The diameter of the wild-type stem is 0.82 mm and the diameter of the *cam* stem 0.35 mm. The cells in the centre of the *cam* stem, the pith cells, are about 20 μm in diameter whereas the pith cells in the wild-type stem are about 10 μm in diameter.

2.3.5 Map-Based Cloning of the *CAMBIUMLESS* Gene

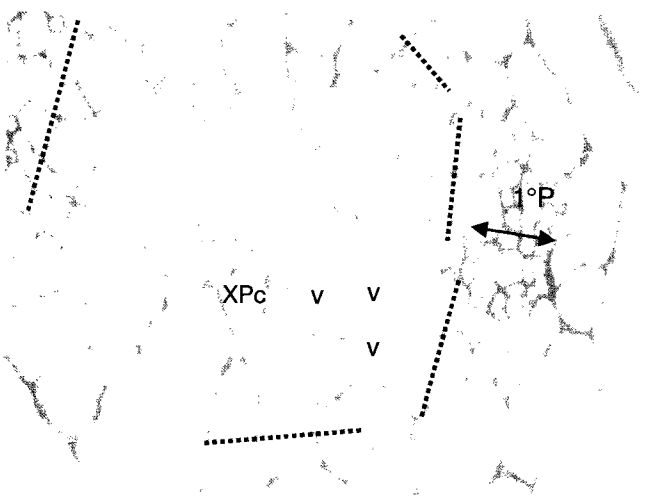
2.3.6 Crossing Data

All of the F_1 plants, in both mapping populations, displayed the wild-type phenotype, indicating that the *cam* mutation is recessive. In the F_2 population, 75.8% of plants exhibited the wildtype phenotype and 24.2% exhibited the *cam* phenotype. These percentages represent a 3:1 Mendelian inheritance pattern, and further indicate that *CAM* is a single gene mutation.

Figure 2.7. Light microscopy comparison of cross-sections of secondarily-thickened hypocotyls in wild-type plants versus *cam*. Plastic-embedded, 2 μm cross-sections of *cam* and wild-type hypocotyls stained with Toluidine Blue O. Plants were grown under long-day conditions (16 hr light/8 hr dark). (a) Two-month-old wild-type hypocotyl, observed at 10X magnification. (b) One-month-old *cam* hypocotyl, observed at 100X magnification. The dashed black lines outline cells that have recently undergone cell divisions, indicating that some cell division occurs in the hypocotyl of *cam*. C, cortex, 1°P, primary phloem, 2°P, secondary phloem, V, vessel, 2°X, secondary xylem, XPc, xylem parenchyma, 1°, primary vasculature.

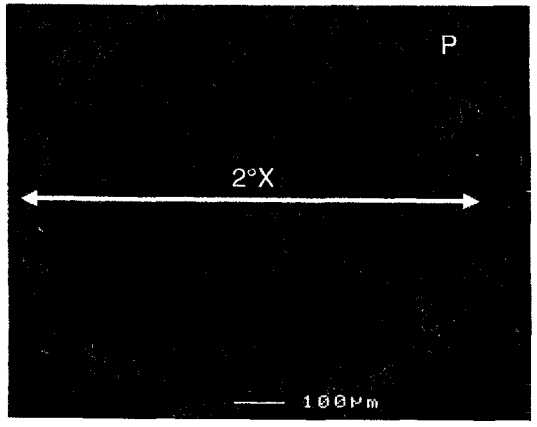


(a)

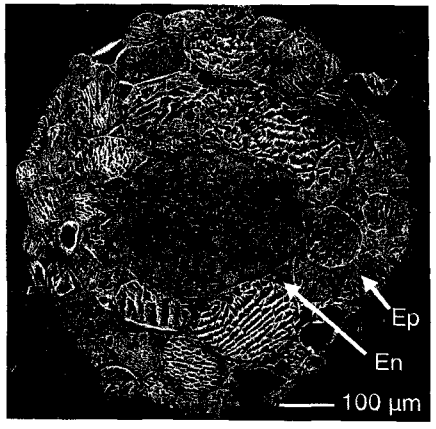


(b)

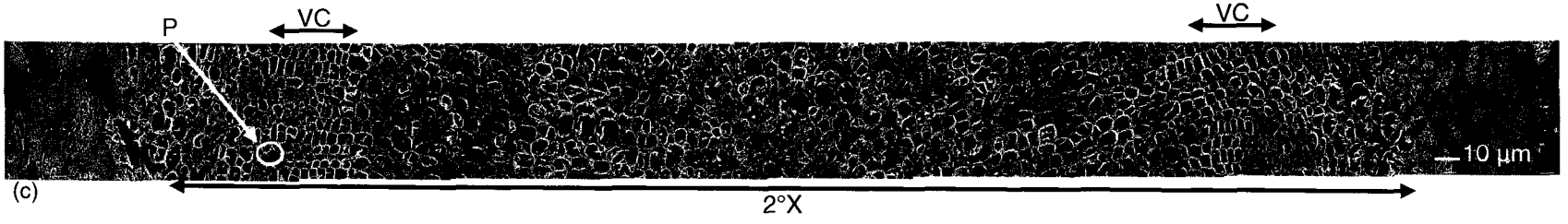
Figure 2.8. Scanning electron microscopy comparison of cross-sections of secondarily-thickened hypocotyls in wild-type plants versus *cam*. Cryo-planed SEM images of wild-type and *cam* hypocotyls. Samples were flash-frozen, embedded in tissue freezing medium, and planed with a microtome. Plants were grown under short day conditions (16 hr light/8 hr dark) to induce maximum secondary growth. (a) En-2 stem, 3 months old; (b) *cam* hypocotyl, 2 months old; (c) cross-section of wildtype hypocotyl; (d) cross-section of the *cam* hypocotyl. Ep, epidermis, En, endodermis, P, phloem, VC, vascular cambium, 2°X, secondary xylem.



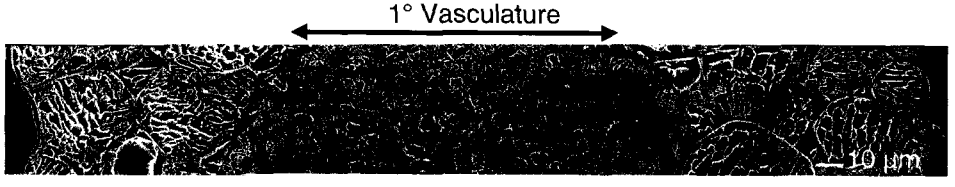
(a)



(b)

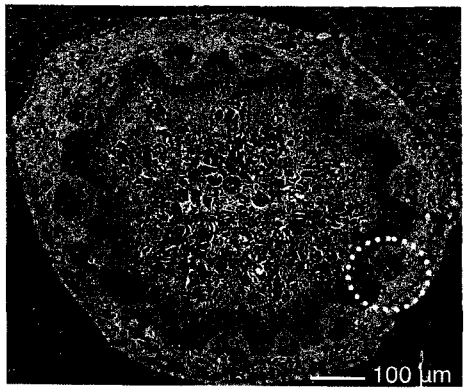


(c)

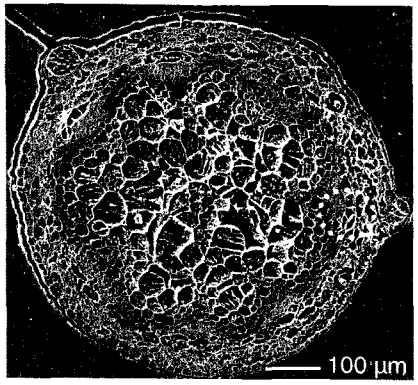


(d)

Figure 2.9. Scanning electron microscopy comparison of cross-sections of stems in wild-type plants versus *cam*. Cryo-planed SEM images of wild-type and *cam* stems. Samples were flash-frozen, embedded in tissue freezing medium, and planed with a microtome. Plants were grown under short day conditions (8hr light/16 hr dark) to induce maximum secondary growth. (a) En-2 stem, 3 months old, vascular bundle outlined in dashed circle; (b) *cam* stem, 2 months old, vascular bundle outlined in dashed circle; (c) cross-section of wildtype stem; (d) cross-section of the *cam* stem. A, arenchyma, P, phloem, PF, phloem fibres, X, xylem.

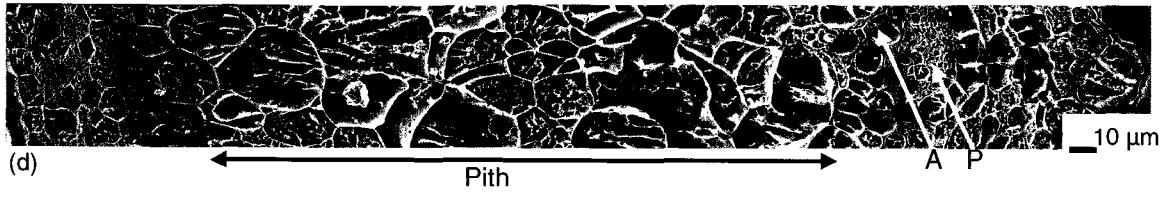


(a)



(b)

(c)



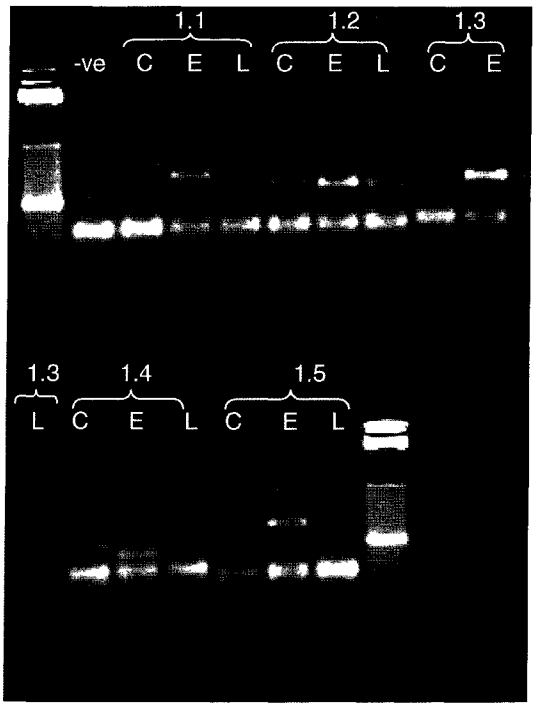
(d)

2.3.7 Verifying the Presence of Markers in the Enkheim Ecotype

Two ecotypes can be used to map a gene if they differ at the marker loci. To determine if the mapping markers were present in the En-2 genome, and if the amplification products from En-2 were a different size from those amplified from Col-0 and L*er*, the amplification products for each marker, with each ecotype, were compared. Figure 2.10 presents the PCR products amplified with the chromosome 1 markers for each ecotype. For markers 1.1, 1.2, 1.3 and 1.4 the products amplified for En-2 and L*er* are very similar in size whereas, the products amplified for Col-0 and En-2 are of different sizes.

The majority of the markers, 19 out of 22, amplified a PCR product in all three ecotypes, as presented in Table 2.1. The size of the En-2 bands was estimated by comparing them to the Col-0 and L*er* bands. In chromosomes 1 – 4, most of the markers amplified in the L*er* and En-2 generate bands of the same size, so they do not migrate differently on an agarose gel. Most of the markers amplified in the Col-0 and En-2 ecotypes generate bands of different sizes and migrate to different locations on an agarose gel. Mapping depends on separating individuals that are heterozygous for a marker versus those that are homozygous for a marker. For this reason, it is necessary that the parental markers amplify pieces of different sizes that are easily distinguished on an agarose gel. Therefore, the Col-0 x *cam* population was used to map the gene. As Table 2.1 indicates, some of the markers, 2.2, 3.3 and 5.5 did not amplify bands in the En-2 ecotype. However, the majority of them, 19 out of 22, did, indicating that En-2 can be used as a parental ecotype in mapping populations.

Figure 2.10. Polymerase chain reaction products for the five markers of chromosome 1 amplified with genomic DNA from three different ecotypes. The PCR products for the five markers of chromosome 1 were amplified from Columbia (C), Enkheim (E) and Landsberg *erecta* (L) genomic DNA. The chromosome 1 markers are labelled 1.1, 1.2, 1.3, 1.4 and 1.5. PCR products were analysed on a 4% TBE high-resolution agarose gel, electrophoresed at 8V/cm for 3hr, stained with ethidium bromide. The PCR products were analysed alongside a 100 bp DNA ladder, in the first and last wells, that did not resolve.



Marker	Size of PCR Product (bp)		
	Col-0	*Ler*	En-2
1.1	200	~160	~160
1.2	128	~115	<115
1.3	159	~135	>135
1.4	105	85	~85
1.5	128	162	>128
2.1	105	~90	~90
2.2	230	~200	-----
2.3	191	199	~199
2.4	151	135	~135
3.1	107	89	~89
3.2	179	~230	>230
3.3	190	~215	-----
3.4	143	123	~123
4.1	164	~144	~144
4.2	162	~148	>148
4.3	130	~123	<123
4.4	150	140	~140
5.1	159	143	~159
5.2	100	~135	~100
5.3	207	222	<207
5.4	165	~145	<165
5.5	140	~130	-----

Table 2.1. Size of PCR product amplified for each of the mapping markers in three ecotypes. The size of PCR product amplified for Col-0 and En-2 and estimated sizes for bands amplified from *Ler* are presented. Sizes for Col-0 and *Ler* bands are from Lukowitz *et al.* (2001). Sizes for En-2 bands were estimated by comparison with the Col-0 and *Ler* bands, values with ~ indicates that the En-2 band migrates to the same location as the Col-0 or *Ler* band, the other sizes are indicated based on whether or not the band is larger or smaller than the corresponding Col-0 or *Ler* band. On chromosomes 1 – 4 most of the PCR products for *Ler* and En-2 are a similar size, whereas the products for Col-0 and En-2 are of different sizes. On chromosome 5 most of the PCR products for Col-0 and En-2 are similar

in size whereas they differ between L*er* and En-2. Three of the marker reactions, 2.2, 3.3 & 5.5 did not amplify a product in En-2.

2.3.8 First-Pass Mapping: The Gene is on Chromosome 2

In first-pass mapping, each marker was amplified in the F_1 bulked-DNA sample and the F_2 bulked-DNA sample. In the F_1 individuals, for each marker, two parental loci are present and two bands are amplified. Among most of the F_2 samples, the markers of two parental loci are present and two bands are amplified. Only the marker close to the mutant gene has a parental bias at that locus and just one band is amplified. The results of the first-pass mapping are presented in Figure 2.11. The PCR reaction products for the markers on chromosomes 1 – 4 in the F_2 plants are presented in Figure 2.11. For each reaction, two bands are amplified from the F_1 DNA. Only one reaction with marker 2.3 with the F_2 tissue failed to amplify two bands. Amplification of two bands indicates the presence of both parental markers at all marker loci. The results in Figure 2.11 indicate that at each marker locus, with the exception of marker 2.3, both parental markers are present. For the reaction with marker 2.3, only the En-2 parental band is present in the F_2 plants, indicating that the mutated gene is on chromosome 2 and located close to marker 2.3.

2.3.9 First-Pass mapping: Narrowing the Region of Interest on Chromosome 2

First-pass mapping identified that the mutant gene was located on chromosome 2, close to marker 2.3. To narrow that region, additional chromosome 2 markers were chosen from the TAIR marker database. The region of chromosome 2 from 8.4 Mb to 13.2 Mb, between markers 2.2 and 2.4, was further examined to narrow the location of the gene. The markers in this region (PLS2, NB19, NB20, NB21, G009, G011, CZSODS, C4H and nga 361) are schematically depicted in Figure 2.12. These markers flank the area above (*i.e.* north of), and below (*i.e* .south of) marker 2.3. Each of these markers was tested to determine optimal amplifying conditions and the PCR product sizes for pieces amplified from Col-0 and En-2.

Figure 2.11. First-pass mapping results for markers on chromosomes 1 – 4. In the first-pass mapping PCR reactions, each reaction, with the exception of 2.3 amplified two bands. The starred band for marker 2.3 is the En-2 parental band amplified from the F_2 bulked DNA. No result was obtained for markers 2.2 or 3.3. PCR products were analysed on 4% high resolution agarose gels, electrophoresed at 8V/cm for 2 – 4 hr, stained with ethidium bromide.

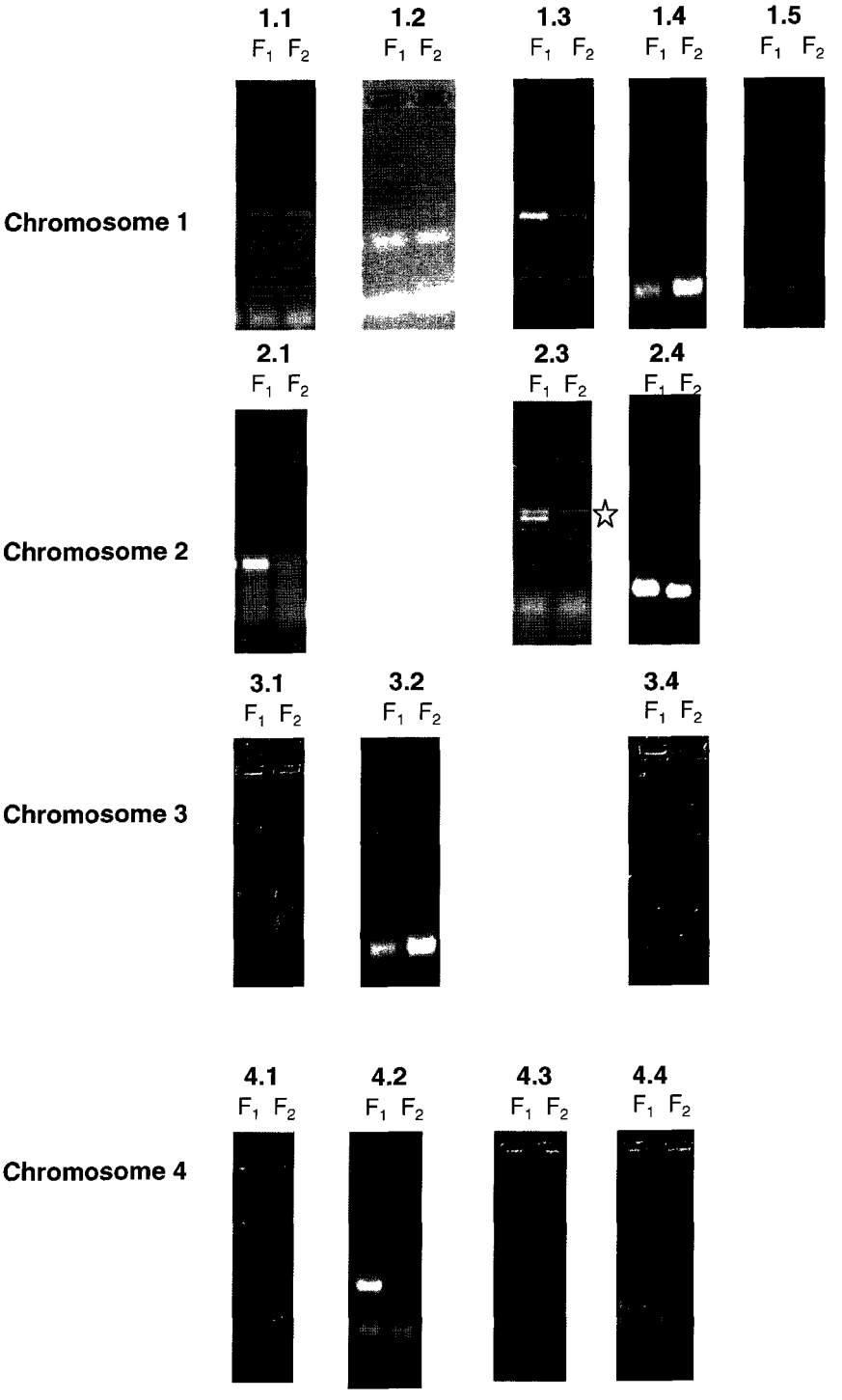
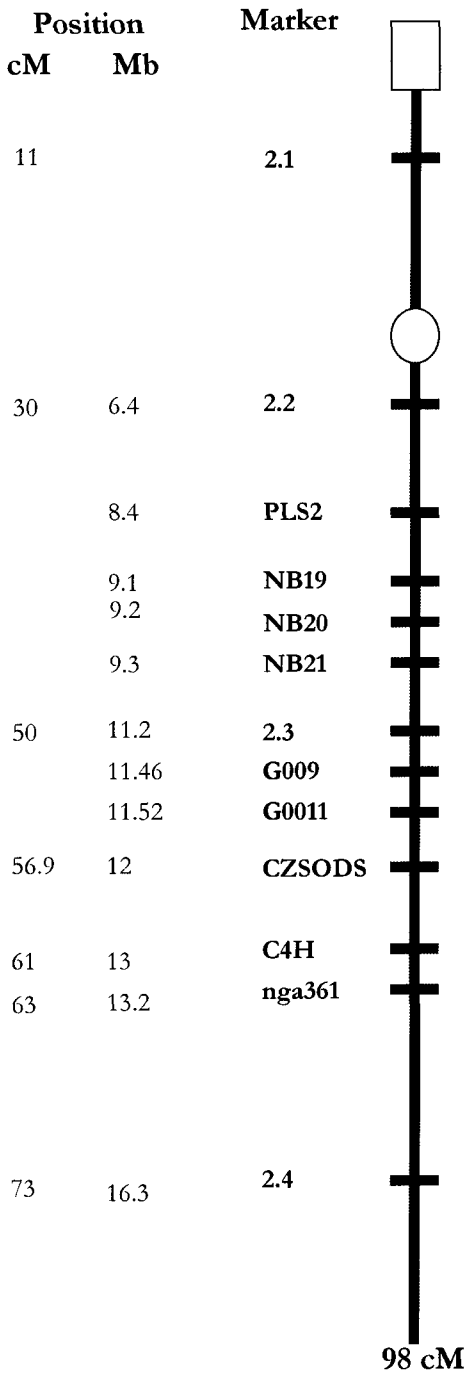


Figure 2.12. Schematic representation of chromosome 2 markers used to narrow the region of interest. Additional markers were chosen to probe the regions above (*i.e.* north of) and below (*i.e.* south of) marker 2.3 to narrow the location of the mutant gene. Distances are listed in centiMorgans (cM) and Megabases (Mb), (Mb distances were obtained from the TAIR marker database). The open rectangle represents the telomere region of the chromosome and the open circle represents the centromere. The total length of chromosome 2 is 98 cM.



Marker	Forward primer (5'->3')	Reverse primer (5'->3')	[MgCl] mM	PCR Product Size (bp)	
				Col-0	En-2
PLS2	tacgcgaattatttttaggaga	aatttattttgagtcggatgc	2.5	155	~155
NB19	aacaacctttgagcagagttacc	tgcagggaagagaaaaggaaatag	2.0	170	~170
NB20	tgcagggaagagaaaaggaaatag	accaagaaagcgagaaagacgag	2.5	1037	~1037
NB21	ctaaacccagagaagcaaaatc	acagcttcctccgtcaactac	2.5	246	~256
G009	aacttacattcttcaatccttcg	tgactagagtgtatttgatgtgg	1.5	201	~160
G011	aacaaagaacgaaactttagagg	gatcttctcttatgttgatccc	1.5	185	~100
CZSOD2	gaatctcaatatgtgtcaac	gcattactccggtgtcgtc	1.5	183	~180
C4H	gttcatggacggatgtgtatgc	ctagtggtggttaaaatatacgcg	2.5	75	~80
nga 361	acatatcaatatattaaagtagc	aaagagatgagaatttggac	2.0	114	~105

Table 2.2. Chromosome 2 markers used to narrow the region where the mutant gene could be located. Markers were chosen with the marker search tool available on the TAIR website. Optimal $MgCl_2$ concentrations determined for each primer pair are listed. PCR product sizes for Col-0 were obtained from the TAIR database, sizes for En-2 were estimated from gel electrophoresis analysis.

The primer sequences, determined $MgCl_2$ concentration and product sizes for Col-0 and En-2 are presented in Table 2.2.

Markers NB21, G009, G011, CZSOD2, C4H and nga 361 amplified products of different sizes for Col-0 and En-2 and were used in mapping reactions with the F_1 and F_2 bulked DNA samples. As Figure 2.13 indicates, reactions with NB21, G009 and CZOD2 amplified two parental bands in the F_1 DNA sample, but only the En-2 parental band in the F_2 DNA sample. Markers C4H and nga 361 amplified parental bands for both F_1 and F_2 samples. These results indicate that the mutant gene is on chromosome 2 and located in a 4.3 Mb area between NB21 and C4H. However, it is possible that the gene is centromere proximal, further north of marker NB21, but markers in this region did not amplify, potentially because they are close to the centromere. Figure 2.14 presents the TAIR map data for this region of chromosome 2.

2.3.10 Identification of *CURLY LEAF* as the Mutant Gene

During the preparation of tissue to identify plants with recombination events in the 4.3 Mb area of chromosome 2 to complete the mapping, it was discovered that this mutant had been previously identified. Goodrich and colleagues identified a gene called *CURLY LEAF* through Ac/Ds transposon mutagenesis, and designated the gene *curly leaf-2* (Goodrich, *et al.*, 1997). They observed a similar phenotype in a mutant generated via EMS-mutagenesis and listed in the Nottingham Arabidopsis Stock Centre (NASC) as N313. Through complementation they demonstrated that the mutant gene in N313 was *curly leaf-1* (*clf-1*). The N313 plant is available from the ABRC, listed with the catalog number cs313. The cs313 plant was the plant used in these experiments. The *CURLY LEAF* gene, At2g23380, is the

gene responsible for the "*cam*" phenotype; it is located on chromosome 2, 10 Mb south of the telomere, about 1Mb away from marker 2.3.

Figure 2.13. Results from chromosome 2 mapping reactions conducted to narrow the region of interest. The PCR products amplified with chromosome 2 markers are presented. Mapping reactions from markers NB21, 2.3, G009 and CZSOD2 amplified two parental bands in the F_1 samples and the En-2 parental band in the F_2 sample. Mapping reactions from markers C4H and nga361 amplified two parental bands in both the F_1 and F_2 samples. These results indicate that the mutant gene is located on chromosome 2 in the area between C4H and NB21, a 4.3 Mb region. No data was obtained for the G011 marker.

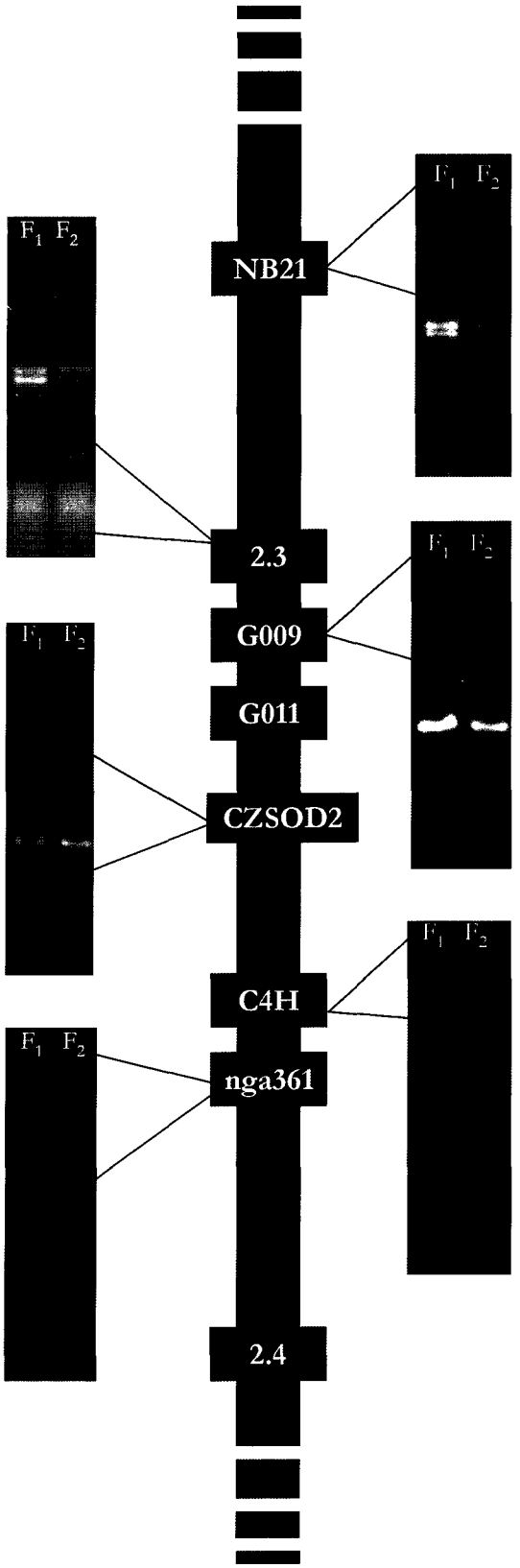
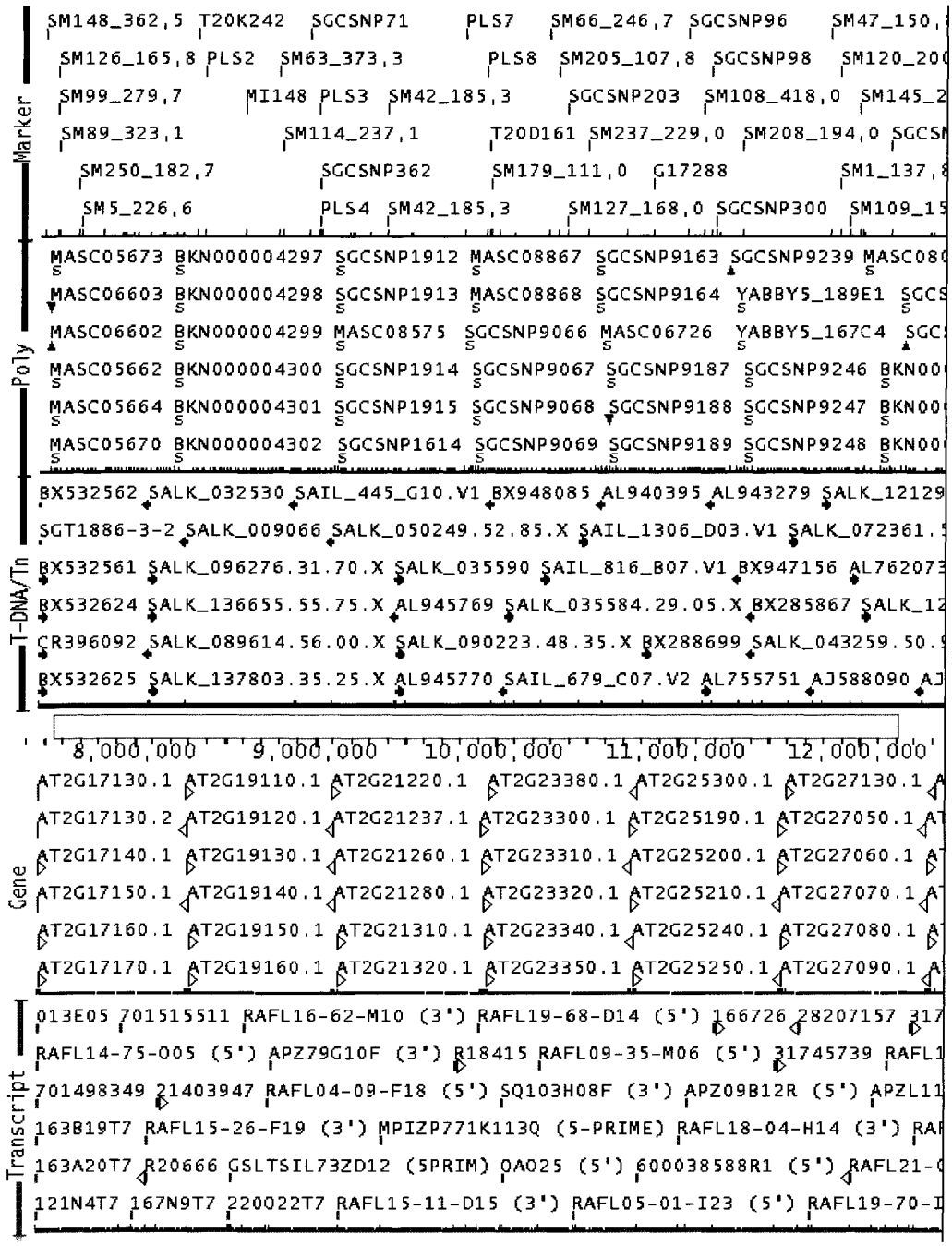


Figure 2.14. Genes located in the 4.3 Mb region identified as the location of the mutant gene. SeqViewer output, from the TAIR database (Rhee, *et al.*, 2003), listing the genes located in the 4.3 Mb region of chromosome 2 identified as the region where the mutant gene was located. The *CURLY LEAF* gene, At2g23380, is highlighted. The SeqViewer output includes detailed information about markers, polymorphisms, T-DNA insertions, gene models, transcripts and annotation units located in this area of the genome.



2.3.11 Why isn't there Data for Chromosome 5?

As indicated in Table 2.1, the markers for chromosome 5 amplified pieces of the same size for Col-0 and En-2 and pieces of different sizes for L*er* and En-2. For this reason, it was originally planned to use the Col x *cam* population for the mapping reactions on chromosomes 1 -4 and the L*er* x *cam* population for the mapping reactions on chromosome 5. However, because the region of interest was narrowed to chromosome 2, it was not necessary to proceed with chromosome 5 analysis.

2.4 Discussion

2.4.1 The *cam* Mutant Has Pleiotropic Phenotypes

2.4.2 Flowering Time

The *cam* plant flowers early, on average 8.8 days before the En-2 wild-type plant under long-day conditions. This result indicates that *CAM* has a role in flowering time.

2.4.3 Characterization of *cambiumless* Morphology

The mutation in the *CAM* gene has a pleiotropic effect on the morphology of the plant. The entire plant is smaller than the wild-type plant. The *cam* plant has small leaves with involute margins and the leaf cells are smaller than those of the wild-type plant. The cells of the *cam* leaves are smaller and the resulting organ (*i.e* the leaf) is smaller than a wild-type leaf. Perhaps the smaller size of the *cam* plant is due to a reduction in cell divisions throughout the plant. The flowers are smaller than wild-type flowers, with folded petals. Abscission of floral organs is also affected, as the carpel extends into a silique, the floral organs remain attached at the base of the silique. Finally, the later flowers exhibit homeotic conversions, with unfused carpels observed. Homeotic conversions in the late flowers suggest the mis-expression of one of the homeotic flowering genes. In Arabidopsis there are five homeotic genes that control floral organ identity in floral development, *APETALA1* (*AP1*), *APETALA2* (*AP2*), *APETALA3* (*AP3*), *PISTILLATA* (*PI*) and *AGAMOUS* (*AG*). Mis-expression of these genes results in altered floral morphology like that observed in the late flowers of the *cam* plant. These morphological observations taken together suggest that the *CAM* gene affects a wide range of processes, including flowering time, cell division, abscission, and floral morphology in late flowers.

2.4.4 Microscopic Observations of the Vasculature in *cam* and Wildtype

The examination of the vasculature reveals several differences between the *cam* plants and wildtype. The light microscopy images presented in Figure 2.6 demonstrate the differences between the cell types present in the *cam* hypocotyls, (i.e. only primary vascular cells), versus the secondary vascular cells observed in the wild-type hypocotyl. The cryo-SEM images of the *cam* and wild-type hypocotyls show very different structure of these two tissues. The *cam* hypocotyl has a smaller diameter than the wild-type hypocotyl, but the cells are, on average, 5X larger than the cells of the wild-type hypocotyl. The presence of the large endodermal and epidermal cells in the *cam* hypocotyl provides further support for the observation that there is very little, or no secondary development in the *cam* plant. The wild-type hypocotyl displays extensive secondary-thickening and radial symmetry analogous to the vasculature observed in an angiosperm, like that displayed in Figure 1.1. The stem of *cam* has a smaller diameter than the stem of the wild-type plant. Again a cell size difference is observed, specifically for the pith cells in the centre of the stem which are about 2X larger than the pith cells observed in the centre of the wild-type stem.

2.4.5 Mapping with Enkheim as a Parental Ecotype

Most of the genetic markers available for mapping were identified by examining the Col-0 and L*er* genomes for polymorphisms. Most mapping populations are generated by crossing Col-o and L*er* plants (Jander, *et al.*, 2002). In 19 of the initial 22 first-pass mapping markers, unique bands could be distinguished from the Col-0 bands that were amplified for En-2. Further, for the additional chromosome 2 markers used to narrow the region of interest, 5 out of 8 of the markers distinguished between Col-0 and En-2. This demonstrates that En-2 can be used successfully in mapping experiments and can be used as a parental strain when

generating mapping populations. The slight drawback to using En-2 as a parental strain is the necessity of pre-testing each marker to determine if it will amplify a product in En-2 and if the product size can be distinguished from the product generated for the other parent. However, the additional time required to do this is minimal, particularly when compared with the alternative of crossing a mutant into either Col-0 or L*er* and back-crossing it for 5 generations to obtain a true-breeding mutant. This provides support for the assertion that the Col-0/L*er* markers can be used to map in other ecotype backgrounds (Jander, *et al.*, 2002).

2.4.6 The Mutant Gene is Located on Chromosome 2

The first-pass mapping reactions, presented in Figure 2.11, clearly indicate that for all of the markers on chromosomes 1 – 4, with the exception of 2.3, two parental bands were amplified for both the F_1 parental DNA samples and the F_2 bulked homozygous samples. The PCR product for marker 2.3 in the F_2 bulked samples amplified just one product, the En-2 parental band. This indicates that the DNA in the area around the 2.3 marker is En-2 and that the mutant gene, generated in the En-2 background, is also located in this area. The success of the first-pass mapping indicates the efficacy of bulked segregant analysis. Instead of 96 separate PCR reactions for each of the 22 markers, 96 separate individuals were analysed with just one reaction per maker, reducing the total number of PCR amplifications from 96 to 22. Bulked-segregant analysis significantly reduces the number of PCR reactions required in the initial mapping stages and rapidly gives the researcher an indication of the location of the gene being mapped. Following the first-pass mapping, the additional markers chosen on chromosome 2 verified that the mutant gene was in the region and narrowed the area to a 4.3 Mb region.

While tissue was being prepared for the next stage of mapping, it was discovered that *CAM* is in fact the *CURLY LEAF* (*CLF*) gene, based on work done by Goodrich and colleagues (1997). *CLF* is located on chromosome 2, 10 Mb south of the top (*i.e.* northern-most) telomere, 1 Mb south of the 2.3 marker. Presumably, if the mapping had continued this locus would have been determined as the location of the *CAM* gene. The success of this project clearly illustrates the efficacy of MBC to identify a mutant gene in EMS mutagenized lines. The first-pass mapping delineated a narrow region on chromosome 2 and the gene was ultimately shown to be in that region. Further, it demonstrates that MBC is applicable within the time frame of a Master's degree, particularly with the assistance of the tools available on the TAIR database, and some able-bodied assistants to help with tissue collection. Finally, MBC is not limited to Col-0 x L*er* populations but can also be done using En-2 as a parental strain.

2.4.7 The *CURLY LEAF* Gene

The mutated gene in *cam* plants is *CLF*. *CLF* encodes a polycomb group (Pc-G) protein that acts in a multimeric protein complex to regulate expression of the homeotic floral genes *AG* and *AP3* (Goodrich, *et al.*, 1997). PcG proteins regulate gene expression through an epigenetic mechanism, specifically by methylating specific histone residues in chromatin and thereby making the chromatin inaccessible to the transcriptional machinery. CLF was originally described in terms of its effect on flower morphology. Similar to *ag* plants, *clf* plants flower early and the floral meristem is determinant, terminating in flowers with homeotic conversions similar to those observed in *ag* mutants. These defects in floral morphology suggested a role for *AG* in the *clf* phenotype and CLF has been shown to repress *AG* expression in vegetative tissues (Goodrich, *et al.*, 1997).

A role for CLF in vascular development has not previously been described. As such, this work suggests a novel function for the *CLF* gene. Precisely how CLF impacts development of the vascular cambium is not clear. However, because CLF is a global regulator of transcription it is likely to regulate another, unidentified gene with a direct role in secondary vascular development. A detailed discussion of the *CLF* gene and Pc-G proteins follows. Further experiments are described that will identify possible targets of CLF-mediated repression and determine the role it plays in cambial development.

3.0 Chapter 3: A Novel Role for *CLF* as a Regulator of the Vascular Cambium

3.1 Introduction

CURLY LEAF is a Polycomb-Group protein (Pc-G). Pc-G proteins are epigenetic regulators of gene expression. Pc-G proteins are usually considered to exclusively be repressors of homeotic genes, genes involved in developmental patterns and sequences, although recent evidence in Arabidopsis indicates that they regulate other kinds of genes. Accurate developmental patterning depends on the transcriptional state of early patterning events being fixed and inherited through cell division during somatic development. Regulation of development occurs at various levels, including DNA sequence, gene expression and epigenetic regulation. Epigenetic regulation of gene expression causes changes in gene expression that are inherited mitotically (or meiotically), but do not alter the underlying DNA sequence. For this reason, epigenetic regulation is often referred to as a form of "molecular memory" (Goodrich & Tweedie, 2002), because it is a mechanism through which the specified transcript profile is maintained or "remembered". One of the benefits of epigenetic control of gene expression is that it is reversible, particularly during meiosis, so that changes that occur during somatic development can be reset at the beginning of each new generation.

Although the precise mechanistic details of Pc-G-mediated repression are not known, it is hypothesized that the repression occurs through methylation of histones in the chromatin close to the target gene. Isolated chromatin has a "beads-on-string" conformation. The "beads" are nucleosomes, comprised of DNA wound around a group of histone proteins, and the "string" is additional DNA linking the nucleosomes. Methylation of a histone is posited to cause compaction of the surrounding chromatin. The compaction of the chromatin interferes with transcriptional machinery accessing the target gene and results in

repression of the gene. The mechanisms that maintain repression of the target gene through successive cell divisions are not known (Schubert, *et al.*, 2005).

3.1.1 Polycomb-Group Proteins

Epigenetic regulation via Pc-G proteins represents a highly conserved mechanism to control gene expression and is found in plants, animals and humans. In mammals, Pc-G proteins influence cell-cycle control, stem-cell self-renewal and have been implicated in several human cancers (Marx, 2005). Pc-G proteins were originally discovered and described in *Drosophila melanogaster* as regulators of developmental patterning through repression of homeobox (*HOX)* genes (Reyes & Grossniklaus, 2003). Specification of *HOX* genes maintains proper body patterning along the anterior-posterior axis and is mediated by members of the Pc-G proteins and the trithorax group (trx-G) proteins (Reyes & Grossniklaus, 2003). Pc-G proteins act antagonistically with trx-G proteins by maintaining transcriptional repression of homeobox genes while trx-G proteins maintain transcriptional activation of homeobox genes. Neither protein group initiates the transcriptional state of the target gene, but rather maintains it, thereby ensuring that early patterning events are stably inherited through somatic development.

Mutations in Pc-G proteins results in ectopic expression of homeotic genes. In *Drosophila* these mutations are phenotypically observed as posterior-anterior transformations of body segments, such as sex combs appearing on posterior legs rather than on the anterior set of legs. The *Drosophila* Pc-G proteins form large, multi-protein complexes and maintain repression of homeotic genes throughout development.

3.1.2 Pc-G Proteins in *Drosophila melanogaster*

Pc-G and trx-G proteins were originally described in *Drosophila melanogaster* with the discovery of a set of mutations that cause bizarre alterations in the body organization of the adult fly. In the *Antennapedia* mutant, for example, legs grow out of the head in place of antennae. These mutations, where body parts are transformed into structures that are supposed to be located in other positions are called homeotic mutations or homeotic conversions. The targets of Pc-G and trx-G protein regulation in *Drosophila* are *HOX* genes, members of a multigene family, known as the *bithorax* complex and the *Antennapedia* complex. These genes specify development along the posterior/anterior axis of the fly body, thus mutations in these genes or their regulators results in altered placement of body parts along this axis.

The details of Pc-G and trx-G protein regulation of gene expression have been well studied in *Drosophila* and there are many more mechanistic details available in this organism than others. Given the conservation of protein sequence of Pc-G and trx-G proteins across organisms, it is reasonable to assume that there may also be mechanistic similarities in how Pc-G proteins regulate gene expression in *Drosophila* and Arabidopsis. There are two groups of Pc-G protein complexes in *Drosophila*, the Polycomb Repressive Complex 2 (PRC2),that methylates a histone in the chromatin of the target gene, and another, the Polycomb Repressive Complex 1, (PRC1), that exerts repression of the target gene (Marx, 2005). Pc-G protein mediated silencing is directed by *cis*-acting DNA sequences called Polycomb response elements (PREs). These sequences are required for the repression of nearby genes by Pc-G proteins. PREs are thought to help "attract" the PRC2 to the appropriate genes.

Methylation of a histone by the PRC2 recruits PRC1 to the site of methylation and results in repression of target gene activity.

The mechanism for PRC1 repression is unclear, but chromatin compaction has been suggested. This view of the repression mechanism posits that chromatin compaction interferes with the ability of the transcriptional machinery to access and transcribe genes, resulting in repression. Chromosome observations via electron microscopy provide evidence for this mechanism. Francis and colleagues have demonstrated that PRC1 compacts nucleosomes, causing the beads to clump together and making them no longer individually distinguishable (Francis, *et al.*, 2004). The researchers further demonstrated that one of the components of the PRC1 required for repression is also required for chromatin compaction, linking these two outcomes. A second explanation for Pc-G protein mediated repression is that repression occurs because the binding of Pc-G proteins to the DNA physically blocks the transcriptional machinery interfering with transcription and resulting in repression (Levine, *et al.*, 2004), but this seems unlikely because Pc-G proteins lack the conserved DNA-binding domains usually associated with DNA-binding proteins

The PRC2 complex consists of Enhancer of zeste (E[Z]), Suppressor of zeste (Su[Z]), Extra Sex Combs (Esc) and P55 (Chanvivattana, *et al.*, 2004). The PRC2 modifies chromatin conformation through histone methyltransferase activity, specifically methylation of lysine 27 on the N-tail of histone H3 (Cao, *et al.*, 2002). HMTase activity of E(Z) relies on its being complexed with the other members of the PRC2 and the presence of the SET domain. HMTase activity measured *in vitro* is not detected when only E(Z) is present or when there are mutations in the SET domain (Czermin, *et al.*, 2002; Hsieh, *et al.*, 2003).

3.1.3 Pc-G Proteins in Arabidopsis

There are three groups of Pc-G proteins in Arabidopsis, each of which was identified based on: (1) their homology to *Drosophila* Pc-G proteins, (2) their plant Pc-G mutant phenotypes or (3) their effect on vernalization, the process by which prolonged exposure to cold temperatures promotes flowering. At this time, only Pc-G proteins with similarities to components of the PRC2 in *Drosophila* have been identified in Arabidopsis. Proteins with similarities to components of the PRC1 have not been identified, leaving the mechanism of Pc-G protein – mediated repression in Arabidopsis undetermined.

In Arabidopsis, double fertilization initiates seed development. One sperm cell fertilizes the egg cell and initiates embryogenesis and a second sperm cell fertilizes the central cell that develops into the endosperm. *FERTILISATION INDEPENDENT SEED (FIS)* genes were discovered by isolating mutants with the capacity to initiate seed development in the absence of fertilization. The *fis* mutants initiate seed development without fertilization, but the autonomous seeds do not contain an embryo, only an endosperm, resulting in seed abortion.

There are 4 *FIS* genes, all of which code for products with similarities to proteins of the *Drosophila* PRC2 complex. The *FIS* genes are *MEDEA* (*MEA*), *FIS2*, *FERTILISATION-INDEPENDENT ENDOSPERM* (*FIE*) and *MULTICOPY SUPPRESSOR OF IRA1* (*MSI1*). They bear protein sequence similarity to the *Drosophila* Pc-G proteins E(Z), Su(Z)12, Esc and P55, respectively. The *FIS* genes repress the expression of the type I MADS-box gene *PHERES1* (*PHE1*) in early seed development. *PHE1* was identified by microarray analysis and its role in seed development is unclear. In *fis* mutants *PHE1*

expression is highly up-regulated and expression levels remain high, whereas in wildtype expression levels decrease after 3 days. Currently it is not known why over-expression of *PHE1* leads to seed abortion.

Flower development in Arabidopsis occurs after formation of vegetative tissue. Flower development is a highly regulated part of Arabidopsis growth, largely controlled by five floral homeotic genes: *AP1*, *AP2*, *AP3*, *PI* and *AG*. Correct expression of these genes is required for proper development of the floral organs. The floral meristem (FM) initiates four different types of floral organs: sepals, petals, stamens and carpels. These organs are initiated in concentric circles, called whorls, in the centre of which is the FM. In a wild-type flower, the whorls are arranged as follows: the first whorl consists of four green sepals; the second whorl consists of four petals which are white at maturity; the third whorl consists of six stamens; and, the fourth whorl consists of the gynoecium or pistil, composed of an ovary with two fused carpels, that each contain ovules, and a short style capped with a stigma. The homeotic genes act together to specify the identity of the floral organs in each of the four whorls of the Arabidopsis flower.

Homeotic conversions in Arabidopsis are observed most easily in flowers and, like homeotic conversions in *Drosophila,* result in production of organs in the wrong locations. The second group of Arabidopsis Pc-G genes was identified based on common functionality to the Pc-G genes in *Drosophila*, their protein products repress the floral homeotic genes. The proteins are CURLY LEAF, with similarity to E(Z) and EMBRYONIC FLOWER (EMF), with similarity to Su(Z). Recently the *FIS* genes *FIE* and *MSI1* have also been shown to also

have a role in repression of homeotic floral genes during vegetative development (Hennig, *et al.*, 2003; Katz, *et al.*, 2004).

Vernalization is the process where flowering time is accelerated when plant species from temperate latitudes are exposed to periods of cold temperatures. Vernalization has several epigenetic features. First, there is a temporal separation between the stimulus (cold) and the response (flowering) that suggests a mitotic "memory", cells of the floral meristem, that were not present during the exposure to cold, retain a "memory" of the exposure and subsequently alter initiation of flowering. Second, vernalization is reset through meiosis, indicating that it involves a reversible genetic change (Goodrich & Tweedie, 2002). The Arabidopsis Pc-G protein VERNALISATION2 (VRN2) represses the floral repressor *FLOWERING LOCUS C* gene (*FLC*), thereby promoting early flowering in response to low temperatures. Table 3.0 presents a brief summary of the Pc-G proteins involved in seed development, flowering and vernalization.

3.1.4 CURLY LEAF is a Polycomb-Group Protein

CLF was the first Pc-G gene characterized in Arabidopsis and was described as a Pc-G gene based on its protein sequence similarity to the *Drosophila* protein E(Z). In addition to some other conserved domains, both the CLF and E(Z) proteins have a SET domain, which is associated with HMTase activity (Chanvivattana, *et al.*, 2004).

Pc-G Protein	Developmental Process Involved	Target (s)	Homology/Domains & Complex Affiliation	Function Predicted from Fly Counterparts
MEA	Repression of FIS	*PHE1*	E(Z), SET-domain	
FIS2	Repression of FIS	?	Su(Z)	
FIE	Repression of FIS, repression of floral homeotoc genes	*PHE1, AG, AP3, KNAT2, STM, BP, LFY, ANGL17*	ESC, PRC2	Increase of HMTase activity of E(Z)
MSI1	Repression of FIS, repression of floral homeotic genes	*PHE1, AG, AP3* & others	P55, PRC2	Nucleosome binding, increase HMTase activity of E(Z)
CLF	Repression of floral homeotic genes	*AG, AP3, , STM, KNAT2, AGL17*	E(Z), SET-domain, PRC2	HMTase towards K27H3
EMF2	Repression of floral homeotic genes	*AG, AP3*, & others	Su(Z)12, PRC2	Nucleosome binding, increase HMTase activity of E(Z)
SWN	Redundant with CLF	?	E(Z), PRC2	
VRN	Vernalisation	*FLC*	B3-domain transcription factors (plant-specific)	DNA binding? Interpretation of histone marks?

Table 3.0. Summary of the Arabidopsis Pc-G proteins involved in seed development, flowering and vernalization. The eight Pc-G proteins discussed in this thesis are listed as well as their role in Arabidopsis development, target genes, *Drosophila* Pc-G proteins they bear similarity to and predicted function. FIS, fertilization independent seed, refers to proliferation of endosperm in the absence of fertilization. Adapted from Schubert *et al.*, (2005).

3.1.5 CURLY LEAF is a Regulator of the Floral Homeotic Genes *AGAMOUS* and *APETALA3*

The CLF protein represses transcription of the floral homeotic gene *AG* in vegetative tissue and, to a lesser degree, *AP3* (Goodrich, *et al.*, 1997). *AG* encodes a type II MADS-box domain transcription factor. MADS-box domain transcription factors play important roles in developmental processes of plants and act as the "molecular architects" of flower morphogenesis. In wild-type plants, *AG* expression is confined to the third and fourth whorl of the flower and required for proper development of stamens and carpels. *AG* also specifies the termination of the floral meristem, through a suicide feedback loop with *WUSHCEL* (*WUS*) (Sharma, *et al.*, 2003). In this way, the indeterminate floral meristem becomes a determinate one, terminating in the carpel. In *ag* mutants, the floral meristem remains indeterminate, the flowers in the *ag* plant produce concentric whorls of sepals and petals followed by more sepals and petals. Over-expression of *AG* results termination of the floral meristem in determinate carpels and stamens (Mizukami & Ma, 1992).

The curled-leaf phenotype of the *clf* mutant resembles transgenic plants constitutively expressing *AG* (Mizukami & Ma, 1992). Transcript analysis of the *clf* mutant reveals ectopic expression of *AG*, *AG* is expressed throughout the plant, including in the early hypocotyl, rosette and cauline leaves & the stem (Goodrich, *et al.*, 1997). Further, the ectopic expression of *AG* appears to be responsible for many aspects of the vegetative and floral phenotypes, in *ag clf* double mutants, where *AG* activity is eliminated, leaf morphology is restored to near-wildtype levels (Goodrich, *et al.*, 1997; Serrano-Cartagena, *et al.*, 2000). Inappropriate expression of homeotic genes is further suggested by the partial homeotic

conversions of sepals and petals into carpels and stamens, respectively observed in the *clf* flowers (Goodrich, *et al.*, 1997).

Expression analysis suggests that CLF is required for the maintenance of *AG* repression during later stages of development, but not for the initial specification of the *AG* expression domain (Goodrich, *et al.*, 1997). This pattern of activity is consistent with Pc-G proteins in the *Drosophila* where the E(Z) protein is required to maintain, but not initiate, transcriptional repression of homeotic genes in the *Antennapedia* and *bithorax*-complexes (Reyes & Grossniklaus, 2003). This functional similarity represents a conservation of Pc-G protein function in plants and animals, despite the fact that homeotic genes in plants and animals are structurally unrelated, encoding homeobox and MADS-box genes, respectively.

Early investigations of *clf* plants focused on the floral phenotype. However, as discussed in Chapter 2, the mutation has a pleiotropic effect on the plant's morphology with one of the striking phenotypes being the small, curled leaves. Serrano-Cartagen and colleagues suggest that CLF restricts the expression of floral genes, *AP3* and *AG*, to flowers to protect the expression of leaf identity genes (Serrano-Cartagena, *et al.*, 2000). Kim and colleagues conducted a microscopic examination of cell size in the leaf, finding that the leaf cells of the *clf* plant are smaller than wild-type plants and there are fewer cells in a given area (Kim, *et al.*, 1998). They conclude that CLF plays a role in cell division and cell elongation throughout development of the leaf primordial. Recently, Katz and colleagues have used a RT-PCR approach to demonstrate an up-regulation of the MADS-box transcription factor *AGAMOUS LIKE 17* (*AGL17*), the homeobox genes *KNOTTED-LIKE 2* (*KNAT2*) and *SHOOTMERISTEMLESS* (*STM*) as well as the Pc-G gene *MEA* in rosette leaves of *clf*

plants (Katz, *et al.*, 2004). These studies suggest a wider role for CLF: as a regulator of floral homeotic genes; a role in proper development of leaf morphology; and a regulator of a broad range of developmental genes.

3.1.6 The Arabidopsis Genome Contains *CLF* Homologues

The Pc-G genes in *Drosophila* whose protein products are components of the PRC2 are single copy genes. In contrast, many of the Pc-G genes in Arabidopsis appear to consist of small gene families with similar protein structure, but possibly divergent protein function. For example, the CLF and MEA proteins are both similar to the *Drosophila* E(Z) protein and contain SET domains, but have functionally distinct roles: CLF represses floral homeotic genes whereas MEA represses endosperm proliferation (Chanvivattana, *et al.*, 2004). Another Arabidopsis protein similar to E(Z), SWINGER (SWN) (also known as EZA1), may have functional redundancy with CLF (Chanvivattana, *et al.*, 2004). The redundancy is inferred from the fact that the *clf* mutant phenotype was strongly enhanced by mutant *SWN* alleles, suggesting that the role of CLF is partially masked in *clf* mutants by the presence of SWN. However, there is incomplete redundancy because the *swn* mutants do not have a discernible phenotype and the *clf* phenotype is only rescued by transformation with 35S::*CLF* rather than 35S::*SWN*.

3.1.7 CLF is a Member of a PRC2-like Complex

Both genetic evidence and data from yeast-two-hybrid assays suggest that CLF physically interacts with several other Pc-G proteins, including EMF2, VRN and FIE (Chanvivattana, *et al.*, 2004). Based on this evidence, and the absence of PRC1 homologues in Arabidopsis, Chanvivattana and colleagues propose that in plants Pc-G proteins re-arrange forming a

series of different complexes, each of which has separate gene targets. Based on genetic, biochemical and physical interaction data (garnered from yeast-two-hybrid assays), they postulate that the Pc-G proteins re-combine to form three separate PRC2 complexes: a gametophyte development complex, consisting of the *FIS* gene protein products and targeting *PHE1;* a flowering development complex consisting of EMF2, CLF, FIE, SWN and MSI1, which targets *AG*, *AP2* and other genes; and a vernalization complex consisting of VRN2 and other Pc-G proteins and targeting *FLC*.

3.1.8 The Arabidopsis Genome May Contain PRE-like Sequences

Although a PRC1-like complex has not been found in plants, it has been suggested (Reyes & Grossniklaus, 2003) that Arabidopsis contains DNA sequences that act in a similar manner to the PRE sequences in *Drosophila* and act as a target for the PRC2 complex to initiate repression. For example, reporter gene fusions indicate that a second intron of *AG* is required for stable repression of *AG* by CLF (Sieburth & Meyerowitz, 1997). Sieburth transformed *clf* mutants with the upstream promoter region of *AG*, plus a 3.8 kb intragenic region, fused to the GUS reporter gene. In wild-type plants, transformation with this construct resulted in a predicted expression pattern for *AG*, GUS-staining was observed only in the stamens and carpels of the flowers. However, in *clf* mutants, ectopic expression of *AG* was observed, GUS-staining was visible in the leaves. This finding suggests that the intragenic region is required for stable repression of *AG* in wild-type plants, whereas in the *clf* mutants, the mutated CLF protein cannot bind the intragenic regulatory region and ectopic expression of *AG* results (Sieburth & Meyerowitz, 1997). Similarly, sequences required for stable repression by Pc-G proteins have been identified for the MADS-box

transcription factor *FLC* (Sheldon, *et al.*, 2002) and MEA and FIE have been shown to specifically bind to promoter regions of *PHE1* (Köhler, *et al.*, 2003).

3.1.9 Pc-G Proteins Have Distinct Functions in Plants and Animals

The functions of Pc-G proteins in Arabidopsis are distinct from those in *Drosophila.* In *Drosophila*, Pc-G proteins maintain expression states that are established during embryonic development thereby fixing developmental decisions. Pc-G repression is established when segmental differentiation starts and is maintained throughout the life of the fly. In contrast, Pc-G proteins in Arabidopsis are involved in regulating gene expression at transition points through the life cycle, from seedling to vegetative growth, to flowering and seed development. For this reason, Pc-G mediated repression has to be released at specific time points and in specific tissues as the life cycle progresses. For example, although CLF represses *AG* during vegetative growth, *AG* expression is required for proper flower development, the repression must be reversible as well as tissue and developmentally specific.

3.1.10 How Does CLF Affect the Vascular Cambium?

The *CLF* gene is not an obvious target as a regulator of the vascular cambium. Its discovery as the gene responsible for the "*cam*" phenotype raises more questions than answers, particularly given the absence of this observation in the literature. These questions include: (1) how does *CLF* expression change across development?; (2) when and where is CLF expressed *in vivo*?; and (3) how does CLF regulate the vascular cambium? The remainder of this introduction will focus on these questions and describe the experiments designed to answer them.

3.1.11 How does *CLF* Expression Change Across Development?

Initially this question will be investigated *in silico* with the Expression Browser tool from the Botany Array Resource (BAR) website (http://bbc.botany.utoronto.ca) (Toufighi, *et al.*, 2005). The data available in the BAR database was collected using the Affymetrix's GeneChip microarray, "ATH1" which consists of about 22,810 genes, representing most of the Arabidopsis genome (Toufighi, *et al.*, 2005). The data represents approximately 150 samples from the University of Toronto's microarray facility.

The Expression Browser makes it possible to perform an "electronic Northern" (e-Northern) and examine the expression profiles of a set of genes by inputting the gene AGI codes. The expression profiles generated by the Expression Browser are formatted by the Data Metaformatter and graphically depicted in a colour-coded table. Each row represents the gene of interest and each column represents a sample. Below the sample IDs there are four other categories including the research category, age, tissue type and whether the sample is a control sample. For this analysis the AtGeneExpress Developmental Series data set was probed because it consists of gene expression profiles in specific tissues across development (Schmid, *et al.*, 2005). The expression profile of *CLF*, its multi-protein partners in the PRC2, and their target genes will be examined.

Another tool available on the BAR website is the Expression Angler. The Expression Angler makes it possible to identify genes that respond similarly in terms of their gene expression levels or activation or repression levels as compared to a control. The Pearson correlation coefficient is used to identify co-regulated genes. This tool makes it possible to

identify novel genes that are co-regulated with the gene of interest and possibly identify previously unconsidered members of a pathway or signaling cascade.

Since the completion of the Arabidopsis genome sequencing project (The Arabidopsis Genome Initiative, 2000) and the development of microarray technologies, there is an enormous amount of data available to Arabidopsis researchers. The BAR assembles microarray data from a variety of sources and provides mechanisms to examine that data, enabling hypothesis testing *in silico* prior to *in vitro*. Toufighi and colleagues assert that the relationships between genes suggested by data mining with tools like the Expression Browser and Expression Angler recapitulate results obtained in the wet lab (Toufighi, *et al.*, 2005). This assertion provides support for the use of this database as a starting place to ask and generate valid hypotheses.

3.1.12 When and Where is *CLF* Expressed *in vivo*?

To investigate the spatial and temporal expression of *CLF*, constructs were made where the promoter region of the *CLF* gene was fused to the β-Glucuronidase (GUS) reporter gene. Two different constructs were made, one containing the complete upstream region of the *CLF* gene fused to the GUS reporter gene and another containing the complete upstream region of the *CLF* gene as well as 351 bp of the coding region, including the first intron, first exon and part of the second intron. Previous characterization of promoter regions indicates that they may be found both upstream of the coding region as well as within the intronic regions (Sieburth & Meyerowitz, 1997). This GUS fusion was designed to investigate when the *CLF* gene is expressed, where it is expressed and what regulatory regions are required for proper expression (i.e. regions upstream or within the gene). The GUS fusion constructs

were made in a pCAMBIA plant-expression vector and transformed into *Agrobacterium tumefaciens* and then transformed into wild-type Arabidopsis plants.

3.1.13 Do Physical Rearrangements in Chromatin Affect Genes Other than *AG*?

Recent studies suggest that in many organisms the genome is organized into chromatin domains whose expression patterns are coordinately regulated (Birnbaum, *et al.*, 2003; Spellman & Rubin, 2002). As such, genes with disparate functions and part of different pathways are co-regulated simply due to the fact that they are physically close together on a chromosome. Ma and colleagues found that chromatin domains of co-regulated genes in the Arabidopsis genome account for about 12% of the Arabidopsis genome (Ma, *et al.*,2005), similar to the 20% value reported for *Drosophila* (Spellman & Rubin, 2002). Given these observations and that Pc-G protein-mediated repression occurs via changes in chromatin structure, it is reasonable to hypothesize that the *clf* phenotype may be due to a change in the gene expression profile of a gene physically linked to *AG*. In the *clf* plant, ectopic expression of *AG* occurs because CLF does not repress its expression. It is possible that the ectopic expression is the result of a "loose" chromatin conformation that allows continuous transcription of *AG*. Perhaps another gene whose over expression is linked to altered vascular development is physically linked to *AG* and also ectopically expressed in the *clf* mutant.

Previous work demonstrating a relationship between co-regulation of genes and physical distance has relied on microarray data. In this case, data from the Gene Angler tool of the BAR will be used to examine genes that are co-regulated with *AG*. The extent to which

genes that are physically linked to *AG* are implicated in vascular development and genes that are physically linked to *AG* are significantly co-expressed with *AG* will be examined.

3.1.14 What Genes Regulate the Vascular Cambium?

Zhao and colleagues have recently examined the secondary tissues of the root-hypocotyl in Arabidopsis with a microarray and assembled a list of 39 genes with a role in vascular differentiation or function (Zhao, *et al.*, 2005). When preparing the tissue for analysis, the root hypocotyl was dissected and separated into three separate tissues, the xylem, the phloem/cambium and the nonvascular tissues. The transcript profile for each of these tissues was examined and compared to generate data about tissue-specific gene expression profiles in the root-hypocotyl.

Among the genes listed as having a role in vascular tissue differentiation or function are members of the class III homeobox-leucine zipper (HD-ZIP III) transcription factor family. Zhao and colleagues identified these genes as being specifically expressed in the xylem fraction of the root-hypocotyl (Zhao, *et al.*, 2005). In Arabidopsis, the HD-ZIP III gene family includes five members: *ATHB15/CORONA (CNA)*, *ATHB8*, *PHAVOLUTA (PHV)*, *PHABULOSA (PHB)*, and *REVOLUTA (REV)*. *PHV*, *PHB* and *REV* are expressed in various plant tissues, including vascular tissues, apical and floral meristems, and the adaxial region of lateral organs (Carlsbecker & Helariutta, 2005). In contrast, *ATHB15* and *ATHB8* are predominantly expressed in vascular tissue, suggesting a role in vascular development (Baima, *et al.*, 1995). Evidence suggests a link between HD-ZIP III gene expression and brassionosteriods, and that HD-ZIP III genes function in vascular differentiation in response to brassinosteriod signaling (Carlsbecker & Helariutta, 2005).

HD-ZIP III genes have also been shown to be regulated by microRNAs (Floyd & Bowman, 2004). Briefly, the various functions and roles of the HD-ZIP III genes in vascular development are: *ATHB8* is involved in vascular differentiation and cambial cell proliferation, *PHV* and *PHB* are involved in vascular bundle organization, *REV* is involved in vascular patterning, and although *ATHB15* is specifically expressed in the vasculature, its role is unknown.

A possible interaction between the Pc-G genes and the vascular genes identified by Zhao and colleagues (Zhao, *et al.*, 2005) will be examined with the Gene Angler tool, with a particular focus on the HD-ZIP III genes. These genes are of particular interest because of their tissue-specific expression patterns in the xylem and because of their association with patterning and differentiation in this tissue.

3.2 Materials & Methods

3.2.1 How Does *CLF* Expression Change Across Development?

Generate Expression profiles were generated *in silico* using the Expression Browser tool from the Botany Array Resource (BAR) website (Toufighi, *et al.*, 2005) with the AtGenExpress Developmental Tissue Series dataset (Schmid, *et al.*, 2005). The AtGenExpress Developmental Tissue Series dataset was used because it generates data that is specific to development (Toufighi, *et al.*, 2005). Expression profiles were generated for *CLF*, and the genes for the other members of the Arabidopsis Polycomb Repressive Complex 2, *FIE*, *EMF2*, *MSI1* and *SWN* as well as *AG* and several class III HD ZIP transcription factors known to play a role in vascular development *REV, PHB, PHV, ATHB8,* and *ATHB15 (CNA)*. The developmental baseline was chosen as the research area and all tissue types were chosen so that expression across development and throughout the plant was analysed. The data was examined in two ways. First, the average of replicate treatments was selected to view the absolute expression levels of the genes. Second, the average of replicate treatments relative to average of appropriate control was selected to view the log transformed clustered data. In both cases, the clustered data was selected to view graphical representation of the gene expressions.

3.2.2 When and Where is *CLF* Expressed *in vivo*? Construction of *CURLY LEAF* Promoter GUS Fusion Vectors

Two regions upstream of the *CLF* gene, the "short" and the "long" promoter were amplified via PCR and cloned into the TA-cloning vector pGEMT-easy. The inserts in pGEMT-easy were confirmed by sequencing and sub-cloned into pCAMBIA 1305.1. The pCAMBIA vectors are cloning vectors specifically designed for cloning into *Agrobacterium*

and subsequent transformation into Arabidopsis. The two constructs were sub-cloned into pCAMBIA 1305.1 and then transformed into Arabidopsis via the floral dip protocol (Clough & Bent, 1998). A schematic representation of the steps in the cloning strategy employed is presented in Figure 3.0. Detailed maps of both cloning vectors employed, pGEMT-easy and pCAMBIA 1305.1 are presented in Figure 3.1.

3.2.3 Extraction of Genomic DNA

Genomic DNA for PCR amplification of the *CLF* promoters was extracted with the rapid genomic DNA extraction protocol (Weigel & Glazebrook, 2002) described in Chapter 2 (see section 2.2.7).

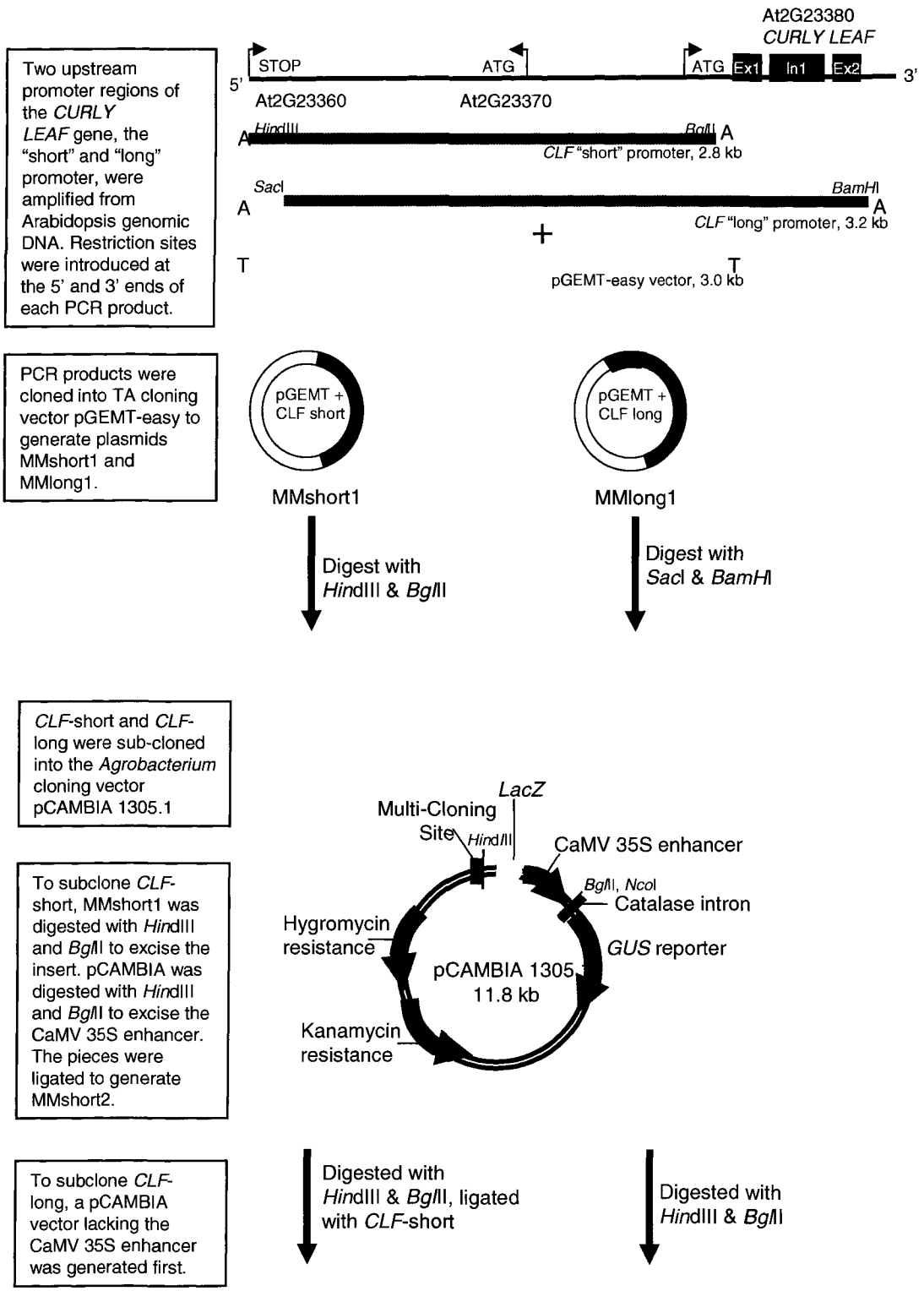
3.2.4 PCR Amplification of the *CLF* Promoter Regions

Two regions upstream of the *CLF* gene were cloned. Based on the entire genome sequence, a region from 9,959,808 to 9,962,650, corresponding to a 2,842 bp region upstream of the coding region, is referred to here as the short promoter. A region from the genomic sequence 9,959,808 to 9,963,001, is the same as the short promoter but also includes the first exon, first intron, and part of the second exon of the gene, a 3,193 bp region, referred to as the long promoter. The "CLF–short" promoter was amplified from the genomic DNA using the following PCR primers:

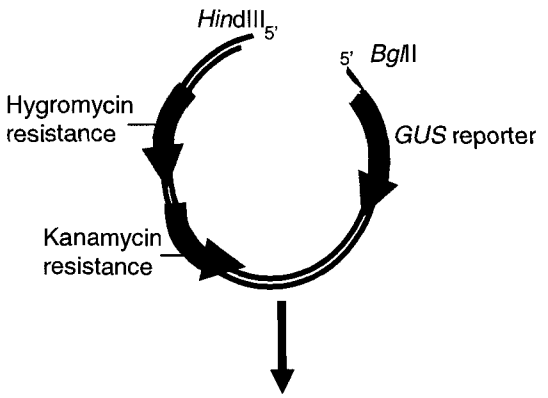
Forward: 5' - **AAGCTT**GAAGATGGTGGCATCTGTCATCTAGACT - 3'

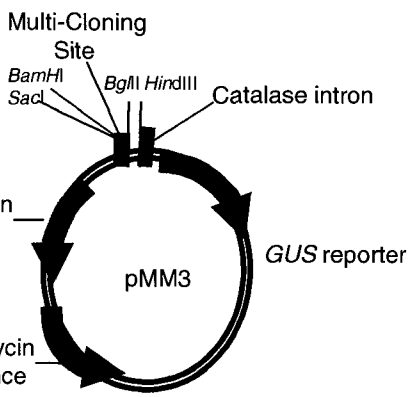
Reverse: 5' - **AGATCT**CATTGTCAAGAAACCAGATCGGAACCGA - 3'

Figure 3.0. Summary of the cloning strategy employed to clone the "short" and "long" promoter regions of CLF into the plant expression vector. Schematic representation of the cloning scheme used to clone the "short" and "long" promoter regions of *CLF*. The regions were amplified from genomic DNA, cloned into an *E.coli* TA cloning vector, pGEMT-easy, and then sub-cloned into a plant expression vector, pCAMBIA 1305.1.



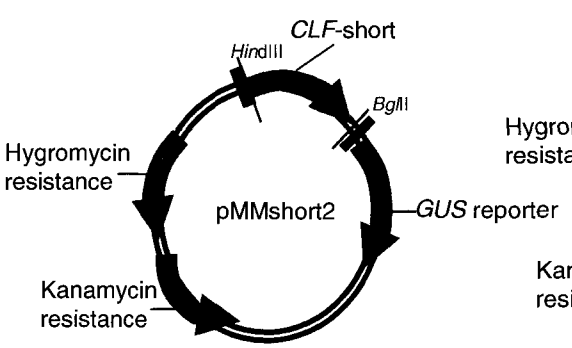
pCAMBIA was digested with *Hin*dIII and *Bgl*II to excise the CaMV 35S enhancer. Both *Hin*dIII and *Bgl*II generate 5' overhangs, these were filled in with the Klenow fragment of DNA polymerase & the plasmid closed via blunt end ligation to produce plasmid MM3. This was required because both restriction sites *Bgl*II and *Nco*I were present in the *CLF*-long piece.

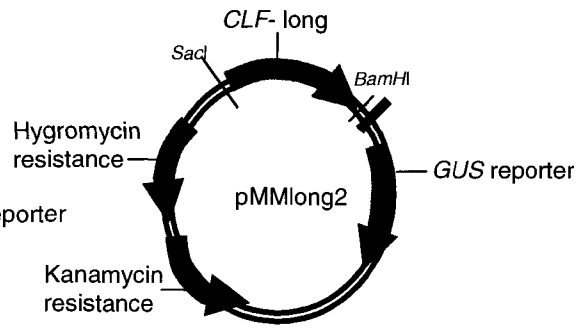




CLF-long was excised fom MMlong 1 by digestion with *Sac*I and *BamH*I. pMM3 was digested with *Sac*I and *BamH*I and ligated with *CLF*-long.

Digested with *Sac*I & *BamH*I, ligated with *CLF*-long

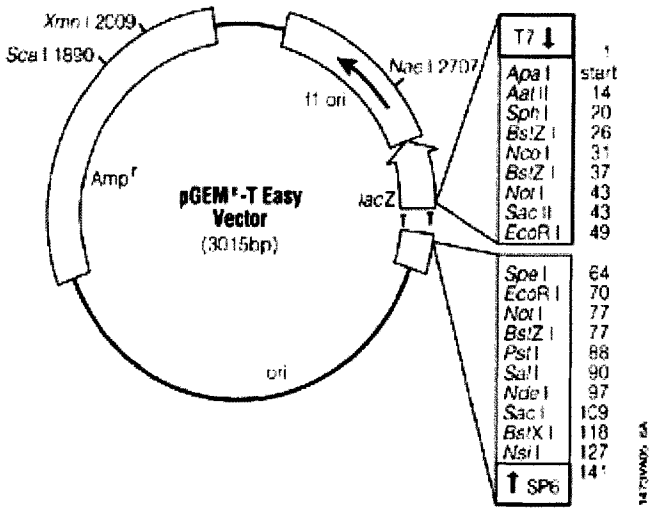




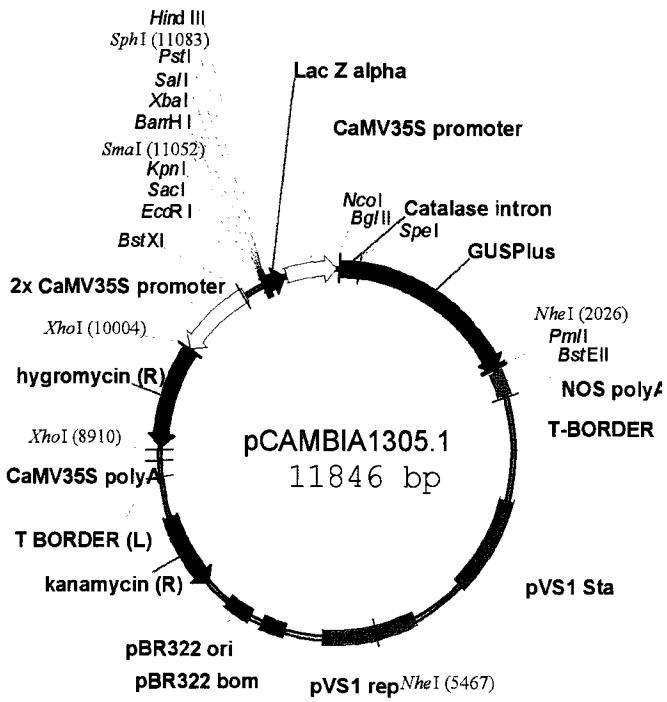
Each construct, MMshort2 and MMlong2 were transformed into *Agrobacterium* and then Arabidopsis, via the floral dip transformation protocol.

Figure 3.1. Maps for the two cloning vectors used, pGEMT-easy and pCAMBIA 1305.1. Cloning vectors used to clone the *CLF* promoter regions: (a) the *E.coli* TA cloning vector pGEMT-easy (Promega, Madison, WI), and (b) the plant expression vector pCAMBIA 1305.1 (Cambia, Canberra, Australia).

(a)



(b)



In the forward primer, the restriction site for *Hin*dIII, placed at the 5' end of the primer, is underlined. In the reverse primer, the restriction site for *Bgl*II, placed at the 5' end of the primer, is underlined.

The following PCR primers were used to amplify the "*CLF*–long" promoter:

Forward 5' – **GAGCTC**GAAGATGGTGGCATCTGTCATCTAGACT - 3'

Reverse: 5' - **GGATCC**CTCTCCTCCGCCTGAAACATCAAAATGA - 3'

In the forward primer, the restriction site for *Sac*I, placed at the 5' end of the primer, is underlined. In the reverse primer, the restriction site for *BamH*I, placed at the 5' end of the primer, is underlined.

The PCR was conducted with Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen, Burlington, ON). The PCR master mix was composed of 1X high fidelity buffer (600 mM Tris-SO_4, pH 8.9, 180 mM ammonium sulfate), 200 μM dNTPs, 200 μM $MgSO_4$, 2 mM forward primer, 0.2 μM reverse primer, 1 μL template DNA, 2.5 U Platinum *Taq* DNA Polymerase and 36.5 μL of sterile distilled water for a total reaction volume of 50 μL. The amplification cycling profile was:

1. Initial denaturation: 1 min, 30 s at 94°C
2. Denaturation: 30 s at 94°C
3. Annealing: 30 s at 55°C
4. Extension: 3 min at 68°C (1 min per kilobase)
5. Steps 2 – 4 repeated 39 times

6. Final Extension: 10 min at 72°C

All PCR reactions were conducted in an Eppendorf Mastercycler Gradient thermalcycler (Eppendorf, New York, NY).

3.2.5 Gel Electrophoresis of PCR Products

Products from the PCR amplification were analysed on a 1% TAE (1X Tris-Acetate-EDTA, pH 7.6) agarose gel, pre-stained ethidium bromide to a final concentration of 0.5 μg/μL. Orange G running buffer (30% glycerol in dH_20 with enough Orange G added to yield a bright orange solution) was added directly to the PCR products (10 μL into 50 μL PCR reaction) following amplification and the entire reaction electrophoresed for 45 min at 85 Volts. PCR reactions were electrophoresed alongside 10 μL of 0.1 μg/μL 1 kb DNA (Invitrogen, Burlington, ON) ladder used to determine the sizes of the PCR products.

3.2.6 Purification of PCR Products

PCR products were purified from the TAE agarose gels using a "Geneclean" kit from MP Biomedicals (MP Biomedicals, Irvine, CA). Following electrophoresis, the DNA bands were excised from the ethidium bromide stained agarose gel with a razor blade using long wave UV light. The excised pieces were weighed, the volume of the pieces (100 mg = 100 μL) determined, a 3X volume of 6M sodium iodide (NaI) was added and the pieces incubated at 55°C for 5 min with gentle mixing to melt the agarose.

Glassmilk was added to the DNA/NaI solution to bind the DNA. The volume of glassmilk was determined based on the volume of the DNA/NaI solution, larger total volumes required the addition of more glassmilk. In general, 5 – 10 μL of glassmilk was added to the

solution. To increase binding of the DNA to the glassmilk, the solution was incubated for 5 min at RT with periodic mixing. The glassmilk was then centrifuged at 14,000 x g for 5 s in a Fisher AccuSpin 3R benchtop centrifuge (Fisher Scientific, Ottawa, ON) and the supernatant discarded. All subsequent centrifugations were conducted with the same centrifuge. At this point the DNA is bound to the glassmilk and the bound DNA was washed with NEW Wash (50 mM NaCl, 10mM Tris-Cl pH 7.5, 2.5 mM EDTA, 50% (v/v) ethanol) by carefully re-suspending the pellet in 500 μL of NEW Wash and centrifuging at 14,000 x g for 5 s. The wash procedure was repeated once. The pellet was dried by removing residual ethanol and by inversion on a kimwipe on a 55°C heat block until no liquid was visible.

The DNA was eluted from the pellet by adding a volume of 1X TE (10 mM Tris, 0.1 mM EDTA, pH 7.4) equal to the volume of glassmilk added (5 – 10 μL) and resuspending the pellet by gently pipetting up and down. The solution was then centrifuged at 14,000 x g for 30 s to pellet. The supernatant containing the eluted DNA was carefully removed and placed in a clean, new tube. The elution step was repeated once and the supernatant combined with that form the first elution. All DNA samples were stored at -20 °C.

3.2.7 Cloning of Long and Short Promoter Fragments into pGEM-T Easy Vector

Following purification from the agarose gel, *CLF*-short and *CLF*-long were ligated into the pGEM-T Easy cloning vector (Promega, Madison, WI). The ligations were set up according to the manufacturer's instructions: 1X rapid ligation buffer (60mM Tris-HCl, pH 7.8, 20mM $MgCl_2$, 20 mM dithiothreitol (DTT), 2 mM adenosine triphosphate (ATP), 10% polyethylene glycol), 50 ng pGEM-T easy vector, 3 μL of purified PCR product and 3 Weiss units of

T4 DNA Ligase for a total reaction volume of 10 μL. The reactions were mixed by pipetting and incubated at 4°C overnight. For simplicity, the construct of pGEM-T-easy with *CLF*- short is referred to as "MMshort1" and the construct of pGEM-T-easy with *CLF*-long as "MMlong1".

3.2.8 Chemical Transformation of *E.coli* XL1-Blue Bacteria

3.2.9 Preparation of Chemically Competent *E.coli* XL1 – Blue

Chemically competent cells were prepared according to the protocol of Ausubel (Ausubel, *et al.*, 1997). A single colony of *E.coli* XL1-Blue was inoculated into 50 mL of liquid LB media, incubated overnight at 37°C with 200 rpm shaking. A 2L flask containing 400 mL of Luria Broth (10 g tryptone, 5 g yeast extract, 5 g NaCl, 1 mL 1M NaOH in 1 L dH_2O) was inoculated with 4 mL of the overnight culture and incubated at 37°C with 200 rpm shaking until an OD_{590} of 0.375. The cells were transferred into 50mL pre-chilled polypropylene tubes and incubated on ice for 5 – 10 min. Cells were harvested with centrifugation at 1600 x g for 7 min at 4°C. The supernatant was discarded and cells were reusupended in 10 mL of ice-cold $CaCl_2$ solution (60 mM $CaCl_2$, 15% glycerol, 10 mM PIPES pH 7.0). Cells were then centrifuged for 5 min at 1100 x g at 4°C, the supernatant decanted and cell pellet resuspended in 10 mL ice-cold $CaCl_2$ solution and incubated on ice for 30 min. Cells were centrifuged a final time at 1100 x g for 5 min at 4°C, resuspended in 2 mL of ice-cold $CaCl_2$ solution, aliquoted into pre-chilled 1.5 mL Eppendorf tubes, frozen at -69°C on dry ice and stored at -80°C until use.

3.2.10 Transformation of *E.coli* XL1-Blue Bacteria

Chemical transformation of *E.coli* XL1 – Blue was conducted with the protocol from Ausubel (Ausubel, *et al.*, 1997). Two μL of the ligation reaction was diluted with 23 μL of dH_2O and incubated on ice. Chemically competent *E.coli* XL1 – Blue cells were rapidly thawed, 100 μL was added to the DNA and incubated on ice for 10 min. The DNA and cell mixture was incubated at 42°C for 2 min, returned to ice for 2 min, 1 mL of LB was added and the tubes were incubated at 37°C for 1 hr with 200 rpm shaking. Following the incubation, 100 μL of cells were plated on solid LB media supplemented with 50 μg/mL ampicilin, 20μg/mL 5-bromo-4-chloro-3-indolyl-bD-galactoside (X-gal) and 0.1 mM isopropyl-beta-D-thioglactopyranoside (IPTG). Plates were incubated, inverted, for 12 – 18 hrs at 37°C.

3.2.11 Blue/White Colony Screening

E.coli transformed with MMshort1 and MMlong1 were screened on solid LB media supplemented with ampicillin, X-gal and IPTG. Colonies with an insert appear white on the plate, due to disruption of the *LacZ* gene. Colonies without an insert appear blue. Overnight cultures were prepared with white colonies. White colonies were plucked and placed in 5 mL of liquid LB supplemented with 50 μg/mL ampicillin and grown overnight at 37°C with 200 rpm shaking.

3.2.12 PCR Screening of Overnight Cultures

To confirm the presence of the insert, overnight cultures were screened using PCR to amplify the insert. The PCR reactions consisted of 10 μL of overnight culture, 1X PCR buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3 @ 25°C), 200 μM dNTPs, 1.5 mM $MgCl_2$, 1

μM forward primer, 1 μM reverse primer, 2.5 U *Taq* DNA polymerase and 28 μL of dH_2O for a final reaction volume of 50 μL. The cycling profile was the same as one used to amplify the piece from genomic DNA (see Section 3.2.2). PCR products were analysed on 1% TAE agarose gels, pre-stained ethidium bromide to a final concentration of 0.5 μg/μL. A 10 μL aliquot of Orange G loading buffer was added to each reaction and 25 μL was electrophoresed.

3.2.13 Plasmid Purification from Overnight Cultures

Presence of the insert in the MMshort1 and MMlong1 constructs was confirmed via PCR amplification of the inserts from overnight cultures. The plasmids were purified from the overnight cultures via the protocol from Aushbel (Ausubel, *et al.*, 1997). Cells were harvested from the overnight culture by centrifugation in the Fisher AccuSpin 3R benchtop centrifuge, 2.5 mL of culture was centrifuged at 2000 rpm for 5 min. Most of supernatant was decanted, the cell pellet resuspended in the remaining supernatant (50 – 100 μL) and placed on ice. A 300 μL aliquot of TENS buffer (10 mM Tris, 0.1 mM EDTA, 0.1 M NaOH, 0.5% SDS) was added to the cells and vortexed gently for 3-5 s until the mixture had a thick, sticky, consistency. A 150 μL aliquot of 3.0 M sodium acetate (NaOAc) was added, the mixture vortexed gently for 2 – 5 s and centrifuged at 12,000 x g for 2 min at 4°C. The supernatant, containing the plasmid DNA, was transferred to a new microfuge tube and 900 μL of ice cold, 100 % ethanol added to precipitate the DNA. The tubes were then centrifuged at 12,000 x g for 7 min at 4°C. The resulting pellet was washed three times with ice cold 70% ethanol by centrifuging at 12,000 rpm for 2 min. Following the third wash, the tubes were centrifuged to remove all residual ethanol. The tubes were dried by inverting on

a Kimwipe on a 55°C heatblock until no liquid was visible. The pellet, containing the plasmid DNA, was resuspended in 30 μL of 1X TE, pH 7.4 and stored at -20°C.

3.2.14 Confirmation of Insert via Restriction Digestion

Following extraction and purification of the plasmid construct, the presence of the insert was re-confirmed by restriction digestion. The restriction digest was conducted using the restriction sites originally incorporated into the PCR primers.

3.2.14.1 Confirmation of the *CLF*- short Insert: The MMshort 1 construct was digested with *Bgl*II and *Hin*dIII. The reaction consisted of 1X ReACT Buffer B (Promega, Madison, WI), 1μg /μL acetylated bovine serum albumin (BSA), 1 μL plasmid DNA, 5U *Bgl*II , 5 U *Hind*III and 14.8 μL dH_2O for a total reaction volume of 20 μL. The reaction was incubated at 37°C for 1 hr.

3.2.14.2 Confirmation of the *CLF*- long Insert: The MMlong1 construct was digested with *Sac*I and *BamH*I. The reaction consisted of 1X *BamH*1 buffer, 1 μL plasmid DNA, 12 U *Sac*I , 4 U *BamH*I and 16 μL for a total reaction volume of 20 μL. More *Sac*I enzyme was added to the reaction because of the manufacturer recommendation that a three-fold excess of *Sac*I enzyme be used when conducting a simultaneous double digest (Fermentas, Burlington, ON). The reactions were incubated at 37°C for 1 hr.

The restriction digests were analyzed on a 1% TAE agarose gel, pre-stained with ethidium bromide, to a final concentration of 0.5 μg/μL and electrophoresed at 85 V for 1 hr.

Further, the presence of the inserts in both constructs, MMshort1 and MMlong1 was confirmed by sequencing.

3.2.15 Sub-cloning of *CLF*-short into the Plant Transformation Vector pCAMBIA

To sub-clone the *CLF*-short promoter piece into pCAMBIA, both the MMshort1 construct and pCAMBIA 1305.1 were digested with the same restriction enzymes, the pieces were purified from agarose gels and then ligated together.

For MMshort1, a double restriction digestion was set up as described above (see section 3.2.14.1), with *Bgl*II and *Hin*dIII, the pieces electrophoresed, the 2.8 Kb piece excised from the gel and purified via a gene clean. The pCAMBIA 1305.1 vector was digested with *Bgl*II and *Hin*dIII in a reaction consisting of 1X ReACT Buffer B (Promega, Madison, WI), 2 μg/μL acetylated BSA, 1μL of plasmid DNA, 5U *Bgl*II , 5U *Hind*III and 14.8 μL dH_2O for a total reaction volume of 20 μL. The reaction was incubated at 37°C for 1 hr and electrophoresed. The appropriate piece (~11.1 Kb) was excised and purified from the agarose gel.

A ligation reaction was set up to ligate the *CLF*-short piece into the pCAMBIA 1305.1 vector. The ligation reaction included 1X rapid ligation buffer (60 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 20 mM dithiothreitol (DTT), 2 mM adenosine triphosphate (ATP), 10% polyethylene glycol) (Promega, Madison, WI), 1U T4 DNA Ligase, 1μL digested pCAMBIA and 3 μL *CLF*-short insert for a total reaction volume of 10 μL. The ligation reaction was incubated at 4°C overnight.

Following the overnight ligation, chemically competent *E.coli* XL1-Blue were transformed with the ligation reaction, as described (see section 3.2.7), and plated on solid LB/agar media supplemented with 30μg/mL kanamycin. Plates were incubated at 37°C overnight. The following morning, 5 mL overnight cultures in LB supplemented with 30 μg/mL kanamycin were prepared with colonies picked from the plate. The overnight cultures were screened with PCR for the presence of the insert, the construct was purified from the bacterial culture, and the plasmid was restriction digested to re-confirm the presence of the insert. Permanent glycerol stocks were prepared with a portion of the overnight culture. To do this, 500μL of overnight culture (OD_{600} <1.0) was added to 250 μL of glycerol solution (65% glycerol, 0.1 M $MgSO_4$, 0.025 M Tris-Cl, pH 8.0) in a screw-cap cryotube, flash frozen in liquid nitrogen and placed at -80°C for long term storage.

3.2.16 Generation of the pMM3 Plasmid

The purpose of the plasmid constructs is to investigate the activity of the *CLF* promoter by fusing it to the GUS reporter gene. The pCAMBIA 1305.1 vector contains the CaMV 35S enhancer. To successfully examine the normal activity of the CLF promoter the CaMV 35S enhancer must be removed. However, as Figure 3.0 shows, there are only two restriction sites immediately downstream of CaMV 35S, *Bgl*II and *Nco*I, that can be used to remove the enhancer form the plasmid. The *CLF*-long piece of DNA contains both *Bgl*II and *Nco*I sites, so these sites could not be used directly to sub-clone *CLF*-long into pCAMBIA 1305.1. Instead, the CaMV 35S enhancer was removed by digestion with *Hin*dIII and *Bgl*II, the resulting 5' overhangs filled with the Klenow DNA polymerase fragment to generate blunt ends and the plasmid ligated back together. The resulting plasmid, pCAMBIA 1305.1,

without the *LacZ* fragment or CaMV 35S enhancer, is referred to as pMM3 here. A schematic representation of production of pMM3 is presented in Figure 3.0.

A restriction digest of pCAMBIA 1305.1 with *Hind*III and *Bgl*II was conducted. The reaction consisted of 1X ReACT Buffer B (Promega, Madison, WI), 2 μg/μL BSA, 1 μL of purified plasmid DNA, 5U *Bgl*II, 5 U *Hin*dIII and 14.8 μL of dH_2O for a total reaction volume of 20 μL. The reaction was incubated at 37°C for 1 hr.

Restriction digestion with *Bgl*II and *Hin*dIII results in 5' overhangs. The 5' overhangs were filled in with the Klenow fragment of DNA polymerase. A 10 μL aliquot of the restriction digestion was added to 33 μM dNTPs, 5U Klenow fragment and 18 μL dH_2O for a total reaction volume of 30 μL. The 5'-fill-in reaction was incubated at 25°C for 15 min and the reaction terminated by adding EDTA, to a final concentration of 1.6 mM, and incubating at 75°C for 20 min. The DNA was analysed via gel electrophoresis, excised from the agarose gel and purified with a gene clean procedure. The resulting linear plasmid was blunt-end ligated to re-circularize the plasmid. The blunt ligation reaction consisted of 1X ligation buffer, 2 μL of linear pCAMBIA 1305.1, 10 U T4 DNA ligase, and 2 μL dH_2O for a total reaction volume of 10 μL. The reaction was incubated at 4°C overnight. The ligation reaction was used to transform chemically competent *E.coli* XL1-Blue as described previously (see section 3.2.10). The transformed bacteria were plated on LB plates supplemented with 30 μg/mL kanamycin and incubated at 37°C for 12 – 18 hrs. Overnight cultures were prepared by growing colonies in 5 mL liquid LB media with 30 μg/mL kanamycin and the plasmid was purified as previously described (see section 3.2.13). The resulting vector, pMM3, was missing both the *LacZ* gene and CaMV 35S enhancer, but

retained the multi-cloning site. The *CLF*- long fragment from the MMlong1 construct was sub-cloned into pMM3.

3.2.17 Sub-cloning of *CLF*-long into the Plant Transformation Vector pCAMBIA

At the time of thesis completion the *CLF*-long promoter region had not been successfully sub-cloned into pMM3. The sub-cloning will continue to be attempted until it is sub-cloned into pMM3 and successfully transformed into *Agrobacterium.*

3.2.18 Transformation of *Agrobacterium tumefaciens*

3.2.18.1 Preparation of Competent *Agrobacterium tumefaciens*

Competent *Agrobacterium tumefaciens* C58 pGV3850 was prepared according to the protocol of An and colleagues (An, *et al.*, 1988) with minor changes. *Agrobacterium tumefaciens* C58 pGV3850 was grown in 5 mL YEP (Yeast extract, Bacto-Peptone) supplemented with 150 μg/mL rifampicin overnight at 28°C. A 50 mL solution of YEP/ rifampicin (50 μg/mL) was inoculated with 2 mL of the overnight culture in a 250 mL baffled flask, shaken at 200 rpm and grown at 28°C to an OD_{600} of 0.5 to 1.0. The culture was chilled on ice, transferred to 50 mL polypropylene tubes and centrifuged at 4000 rpm for 20 min at 4°C. The supernatant was discarded and the cell pellet was resuspended in 500 μL of filter-sterilized, ice-cold 20 mM $CaCl_2$. The cells were aliquoted, 100 μL/tube into pre-chilled (on dry ice) 1.5 mL microfuge tubes, flash-frozen in liquid nitrogen and stored at -80°C.

3.2.18.2 *Agrobacterium* Transformation: Freeze-Thaw Transformation Method

To transform the *Agrobacterium* cells, about 1 ng of plasmid DNA was added directly to the frozen cells and the cells thawed by incubating for 5 min at 37°C. Following the incubation,

1 mL of YEP medium was added to the tubes and they were incubated for 2 – 4 hrs with 200 rpm shaking. Following this recovery period, the cells were centrifuged at 4,000 rpm for 30 s, the supernatant decanted and the cells resuspended in 100 μL of YEP media. The cells were spread on YEP/rifampicin (150μg/mL)/kanamycin (30μg/mL) plates and incubated at 28°C for 2 – 3 d before colonies appeared.

3.2.18.3 Preparation of Overnight Cultures from *Agrobacterium*

Overnight cultures of transformed *Agrobacterium* were prepared as described (see section 3.2.11) with the alteration that colonies were grown in 5 mL liquid YEP media supplemented with rifampicin (150μg/mL) and kanamycin (30μg/mL) and incubated at 28°C for 12 – 18 hr with 200 rpm shaking. The presence of the insert in transformed *Agrobacterium* was confirmed by PCR amplification with the overnight culture, as described previously (see section 3.2.12) and permanent glycerol stocks were prepared with overnight cultures as described previously (see section 3.2.15).

3.2.19 Floral Dip Transformation of Arabidopsis

The floral dip transformation of Arabidopsis was conducted with the protocol of Clough and Bent (Clough & Bent, 1998). Col-0 wild-type plants were grown under long-day conditions (18 hr light/6 hr dark) for 4 weeks. The primary shoots were removed and the plants transformed 4 – 6 days later, after the generation of several secondary shoots. A single *Agrobaterium* colony was used to inoculate in 5 mL of YEP liquid media supplemented with 150 μg/mL rifampicin and grown overnight at 28°C. The overnight culture was used to inoculate 495 mL of YEP/rifampicin and grown to an OD_{600} of approximately 0.8. Cells were harvested by centrifugation at 3345 x g at RT for 20 min. The cell pellet was

resuspended in an equal volume of infiltration media (5% sucrose, 0.05% Silwet L-77). For transformation plants, were inverted into the infiltration media and gently swirled for 3 – 5 s. The plants were then placed on their sides, in a humidity chamber and placed in the dark for 24 hrs. Following the recovery period, plants were returned to the growth chamber and allowed to grow to maturity under long- day conditions. Seeds were harvested after siliques had developed, extended and dried.

3.2.20 Use of Gene Angler to Identify Genes Co-expressed with *CLF*

The Gene Angler tool, available on the BAR website, was used to generate data regarding the degree to which other Arabidopsis genes are co-expressed with a target gene. The Affymetrix data set was used in this analysis. The correlation coefficient was set at -1 to generate a wide range of co-expression values, including those that indicate high levels of co-expression (values close to 1) and those that indicate opposite expression patterns (values close to -1). Gene Angler was used to examine genes co-expressed with *CLF* and *AG* to identify genes that may be regulated by these genes and directly implicated in vascular development. Genes co-expressed with each of the PRC2 components and the HD ZIP III genes were also examined.

3.2.21 Do Physical Rearrangements in Chromatin Affect Genes Other than *AG*?

To determine if genes that are physically linked to *AG* on chromosome 4 are also co-expressed with *AG*, all genes that are co-expressed with *AG* were examined with the BAR Gene Angler tool (Toufighi, *et al.*, 2005). A correlation coefficient was set at 0.6 to examine genes that show a 60% correlation of coexpression with *AG*. The AtGenExpress tissue series data set was used (Schmid, *et al.*, 2005). Meinke and colleagues state that the

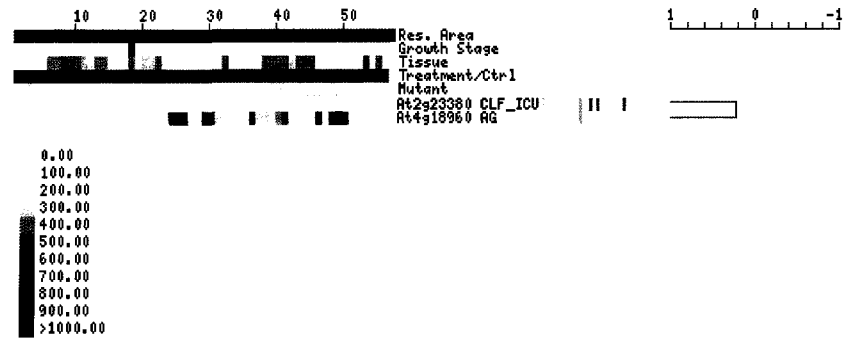
distribution of Arabidopsis genes across the 5 chromosomes is uneven, with 25.4% occurring on chromosome 1, 16.6% on chromosome 2, 20.6% occurring on chromosome 3, 14.8% occurring on chromosome 4 and 22.2% occurring on chromosome 5 (Meinke, *et al.*, 2003). For the analysis, it was assumed that the expressed genes would be similarly distributed across the chromosomes. Therefore, a significant shift away from this distribution, where proportionately more genes that are physically close to *AG*, on chromosome 4, are co-expressed with *AG*, would suggest that a physical change in the surrounding chromatin makes transcription more favourable and this change could be responsible for the observed biased co-expression rates.

3.3 Results

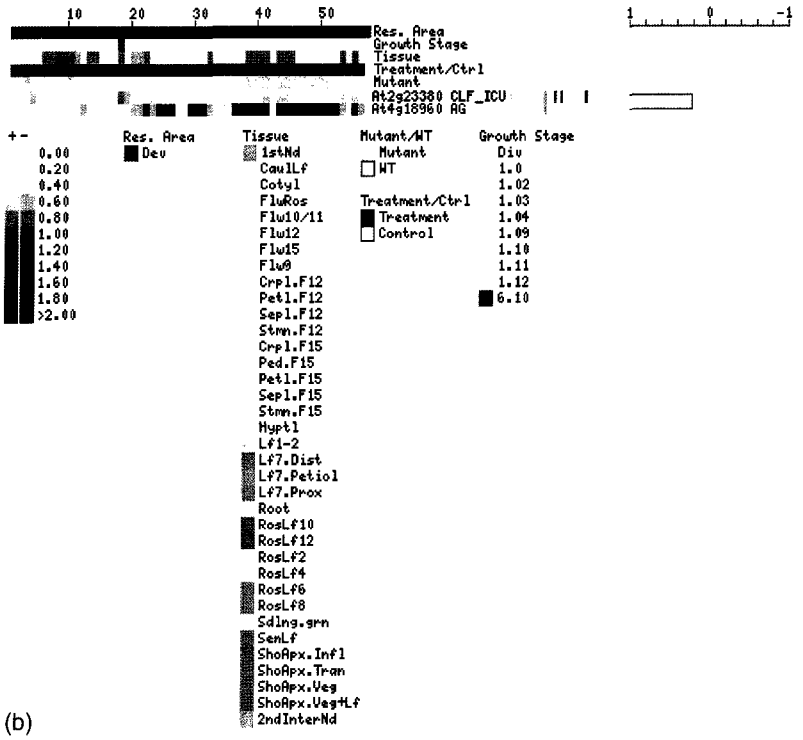
3.3.1 How does *CLF* Expression Change Across Development?

The expression profiles for *CLF* and *AG* across development were examined with the BAR e-Northern tool with the AtGenExpress Developmental Tissue Series dataset. The e-Northern gene expression profiles depict the gene expression levels of the *CLF* and *AG* genes throughout development in various tissue types. Two gene profiles were generated, one where the raw expression values were examined, Figure 3.2 (a), and another where the log-transformed expression values compared to mean values were examined Figure 3.2 (b). In the first gene expression profile, Figure 3.2 (a), the raw expression values for *CLF* across development show very little change, with an average expression level of 100. Low raw expression values are generated for *AG* in early vegetative tissue, with a range of 12 – 20, whereas high raw expression values are observed in stage 12 stamens and carpels, 1562.7 and 595.8, respectively. In the second gene expression profile, Figure 3.2 (b), the expression values indicates the degree of change in expression relative to the mean of the appropriate control. The expression level of *CLF* in the vegetative tissue shows very little change, from a 0.87- fold down-regulation, compared to the control, in the stage 1.02 rosette leaves, to a 1-fold up-regulation in the stage 1.06 rosette leaves. The largest increase observed for *CLF* expression across development is a 1.5- fold up-regulation in the late shoot apex. The expression profile of *AG* shows a slight down-regulation in vegetative tissue, with a 0.66-down-regulation in the second internode, and strong increases in the floral tissues ranging from 25 to 78 fold increases, with a 78- fold up-regulation in stamens of stage 12 flowers and a 30- fold up-regulation in carpels of stage 12 flowers.

Figure 3.2. Expression Browser data from the BAR depicting the expression profiles of *CLF* and *AG* across development. The expression profiles for *CLF*, At2g23380, and *AG*, At4g18960, were examined with an e-Northern. (a) Average raw gene expression values. (b) Average of replicate treatment relative to the average of appropriate control. The values are log2-transformed ratios, and can be thought of as 2 to the exponent of the given value. Thus 2 in the table indicates 2^2, which is 4-fold up relative to the appropriate control, while -2 indicates 2^{-2}, which is 0.25, or 4-fold down relative to the control. Log2-transformed ratios have an equal distribution in number space, aiding visualization.



(a)



(b)

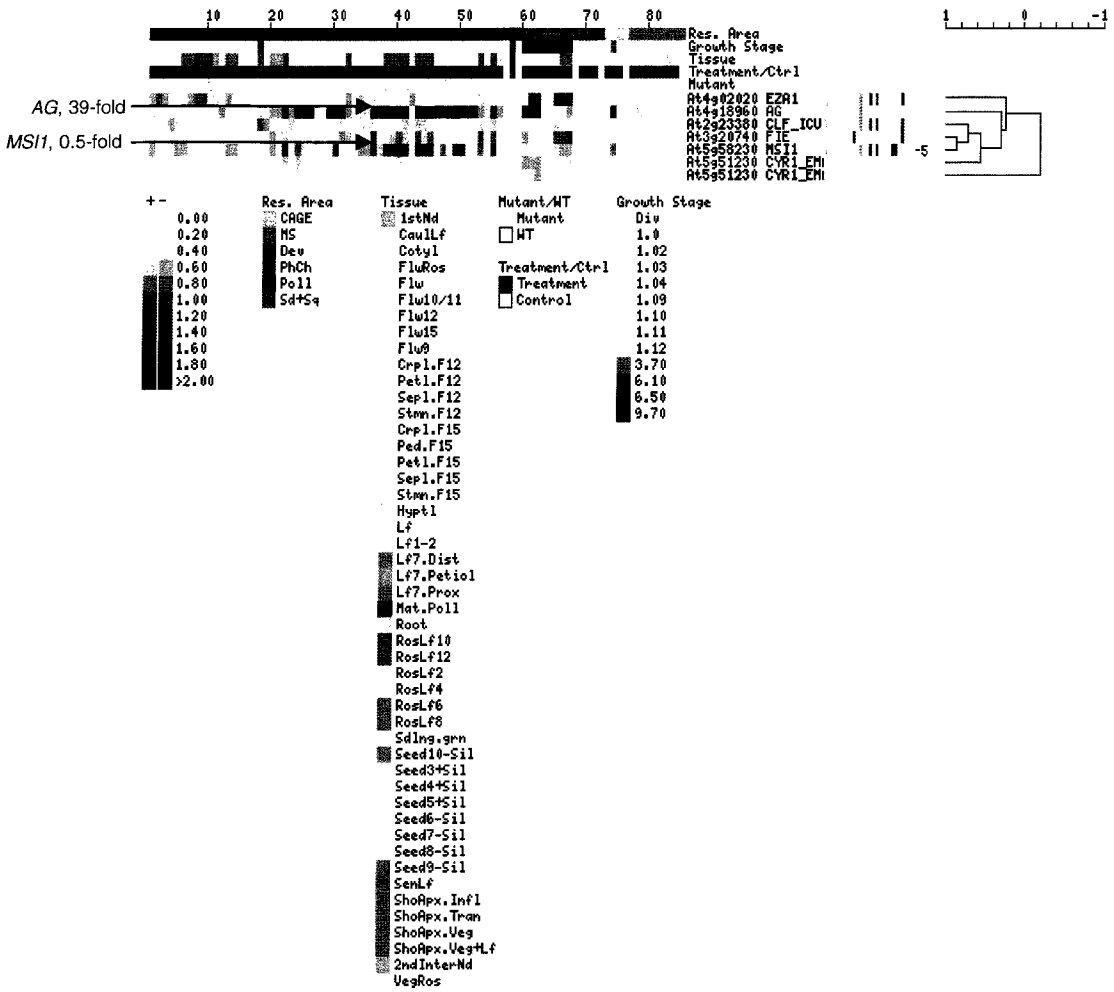
3.3.2. Gene Expression Profile for Pc-G genes and *AG* Across Development

The gene expression profiles for *AG* and the Arabidopsis Polycomb Repressive Complex 2 components, *SWN*, (listed as *EZA1)* At4g02020), *CLF, FIE,* At3g20740, *MSI1*, At5g58230, and *EMF2*, At5g51230 were examined throughout development in various tissue types. The expression profiles for the average replicate treatments compared to the average of an appropriate control were examined. The expression profiles are presented in Figure 3.3.

With the exception of *MSI1*, most of the PRC2 components exhibit little change in expression across development, with a maximum 1.74- fold up-regulation for *FIE,* as compared to the control, in the shoot apex transition and a 0.5- fold down-regulation of *SWN* in early rosette leaves. The expression profile for *AG* is similar to that observed in Figure 3.2, with up-regulation in general floral tissues, and strong up-regulation in the stamens, 97-fold, and carpels, 30-fold, of stage 12 flowers. *MSI1* shows a 0.5- fold down-regulation in stamens of stage 15 flowers that corresponds to a 40-fold increase in the expression of *AG*. This opposite expression profile in this tissue may suggest a regulatory interaction.

The clade diagram on the far right-hand side of the expression browser output indicates that four of the Pc-G genes, *CLF*, *FIE*, *MSI1* and *EMF2* exhibit 80% co-regulation.

Figure 3.3. Expression profile for the five Arabidopsis Polycomb Repressive Complex 2 components and *AG*. The expression profile for the Arabidopsis PRC2 components, *SWN*, (listed as *EZA1*), *CLF*, *FIE*, *MSI1*, *EMF2* and *AG*. The average expression values for the genes compared to appropriate controls were generated. The arrows point out expression levels for *AG* and *MSI1*in stage 15 stamens where *AG* exhibits a 39-fold up-regulation and *MSI1* a 0.5-fold down-regulation. The clade diagram to the right of the figure indicates the degree of co-expression for genes, measure on a scale from 1 (100% co-expression) to -1 (anti-correlation).



3.3.3 Co-expression Pattern for *AGAMOUS*

The Gene Angler tool of the BAR calculates the correlation coefficient between a target gene and the rest of the genes in the Arabidopsis genome. The correlation coefficients range in value from 1 to -1. A value of 1 indicates a perfect co-expression, a value of -1 indicates a perfect anti-expression, and a value of 0 indicates no correlation in the expression pattern of the two genes. To examine the extent to which the Gene Angler tool generates data about co-expression patterns that is consistent with existing genetic data, the data for genes co-expressed with *AG* was generated and is presented in Table 3.1. The Affymetix data set was used to "angle" *AG*. This analysis was done to test the assertion of Toufighi and colleagues that the data generated by the BAR tools re-capitulates laboratory findings (Toufighi, *et al.*, 2005). The data was examined to determine if genes that are known to be co-expressed with *AG* generated high correlation coefficients.

Genes Co-expressed with *AGAMOUS*	Correlation of Co-expression
APETALA 1 (AP1)	0.991
AGAMOUS LIKE 11 (AGL11)	0.983
AGAMOUS LIKE 16 (AGL6)	0.981
PISTILLATA (PI)	0.981
SEPALLATA (SEP1)	0.978
APETALA 3 (AP3)	0.976
SEPALLATA 2 (SEP2)	0.973
NO APICAL MERISTEM (NAM)	0.974
WUSHCEL (WUS)	0.932
AGAMOUS LIKE (AGL5)	0.922
AGAMOUS LIKE (AGL8)	0.864
SHATTERPROOF (SHP1 – AG like)	0.853
SHOOT MERISTEMLESS (STM)	0.850

Table 3.1. Genes co-expressed with *AGAMOUS*. Pearson correlation coefficients for genes demonstrated with Gene Angler to be co-expressed with *AG*.

The degree to which the Arabidopsis PRC2 component genes are co-expressed with *AG* was also investigated with the Gene Angler tool, using the Affymetix data set. This is of particular interest because the protein products of these genes form the multimer complex that regulates expression of *AG* (Chanvivattana, *et al.*, 2004). Table 3.2 presents correlation coefficients for the co-expression of *AG* and the PRC2 components.

3.3.4 When and Where is *CLF* expressed *in vivo*? Construction of *CURLY LEAF* Promoter::GUS Vectors

3.3.4.1 PCR Amplification of the *CLF* promoters and Cloning into Plant Expression Vector

Two upstream, regions of the *CLF* gene were amplified via PCR, a 2.8 kb piece, consisting of the entire 5' upstream region and a 3.3 kb piece, consisting of the 5' upstream region and part of the coding region of the gene. These two pieces were fused with the GUS reporter gene and cloned into a plant expression vector, pCAMBIA 1305.1. As Figure 3.4 demonstrates, the PCR reactions conducted to amplify the upstream regions of the *CLF* gene generated two bands, a 2.8 kb band, representing the "short" promoter and a 3.3 kb band, representing the "long" promoter. Both bands were excised from the agarose gel and purified. Both *CLF*-short and *CLF* -long were cloned into the *E.coli* p-GEMT-easy cloning vector and the resulting constructs dubbed MMshort1 and MMlong1. As indicated in Figure 3.5, digestion of the MMshort1 plasmid generated two bands, a 3 kb band representing the pGEMT-easy vector and a 2.8 kb band representing the cloned insert. The digestion pattern indicates that the short promoter was successfully cloned into the pGEMT-easy insert. As indicated in Figure 3.6, digestion of the MMlong1 plasmid with *Sac*I and *BamH*I, generated a

Arabidopsis PRC2 Components	Co-expression Value for *AGAMOUS*
FERTILISATION INDEPDENT ENDOSPERM (FIE)	-0.108
EMBRYONIC FLOWER (EMF2)	0.084
CURLY LEAF (CLF)	-0.028
SWINGER (SWN)	0.277
MULTISUPRESSOR OF IRA 1 (MSI1)	0.247

Table 3.2. Co-expression values for *AGAMOUS* and the Arabidopsis PRC2 components. Gene Angler co-expression values, expressed as correlation coefficients, for the co-expression of the Arabidopsis PRC2 components and the floral homeotic gene *AG*.

Figure 3.4. Gel electrophoresis of products from PCR amplification of the *CLF* promoter regions. The *CLF* promoter regions were amplified from genomic DNA. Lanes 1 and 5: DNA ladder with pieces ranging in size from 5000 bp – 300 bp, the arrow points to the 3 kb band, Lane 2: Negative control, PCR reaction with no DNA added, Lane 3: *CLF*-short, a 2.8 kb piece of DNA upstream of the start codon of *CLF*, Lane 4: *CLF*–long, a 3.19 kb piece of DNA consisting of the 2.8 kb 5' upstream region and 351 bp of the coding region of the gene .

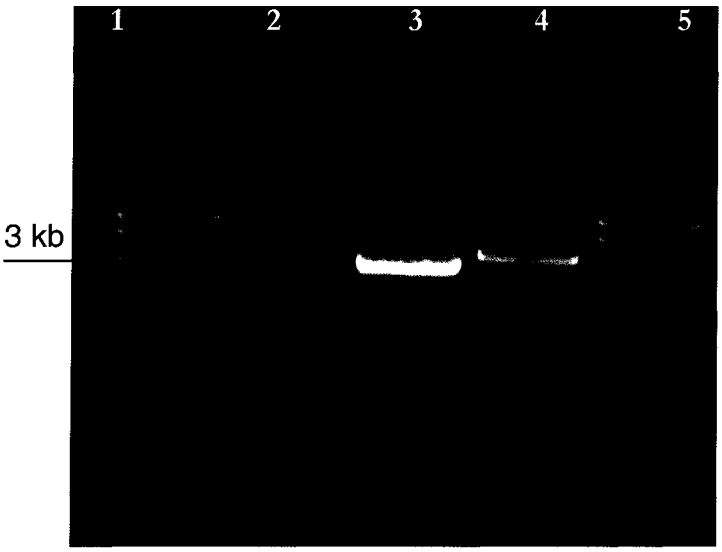


Figure 3.5. Confirmation of insert in MMshort1 via restriction digestion. The MMshort1 plasmid was digested with *Bgl*II and *Hin*dIII. Lanes1 and 4: 1 kb DNA ladder, Lane 2: Purified MMshort1 plasmid, uncut, Lane 3: MMshort1plasmid digested with *Bgl*II and *Hin*dIII, the arrow points to the 2.8 kb insert piece.

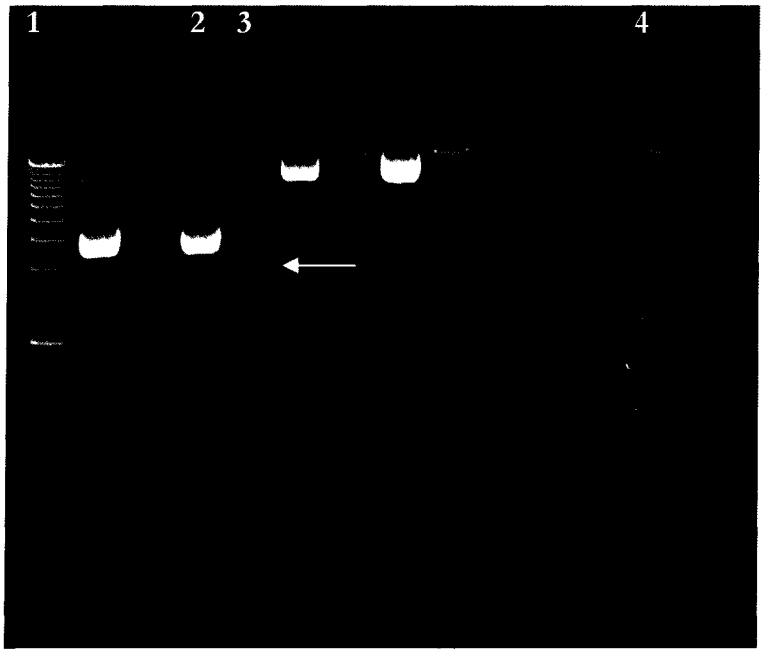
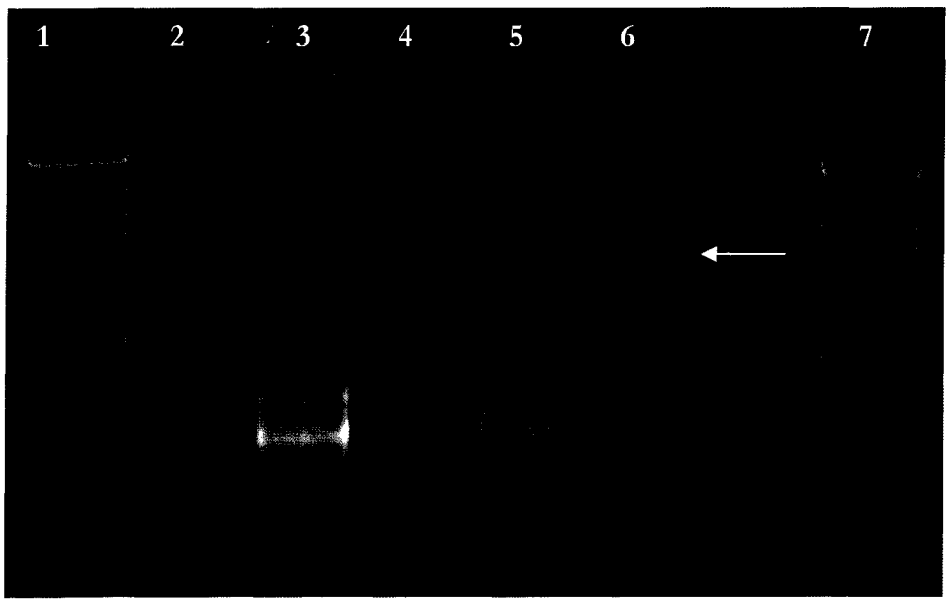


Figure 3.6. Confirmation of insert in MMlong1 via restriction digestion. The MMlong1 plasmid was digested with *Sac*I and *BamH*I. Lanes 1 and 7: 1 kb DNA ladder. Lane 2: Negative control, restriction digest with no DNA added. Lane 3: MMlong1 uncut plasmid. Lane 4: MMlong1 plasmid single digest with *Sac*I. Lane 5: MMlong1 plasmid single digest with *BamH*I. Lane 6: MM long plasmid digested with *Sac*I and *BamH*I, the arrow points to the 3.3 kb insert piece.



3.0 kb piece representing the pGEMT-easy vector, and a 3.3 kb piece representing the insert. The digestion pattern indicates that the long promoter was successfully cloned into the pGEMT-easy vector. Further, the presence of the correct genomic regions was confirmed by sequencing the MMshort1 and MMlong1 constructs.

Following cloning into pGEMT-easy, *CLF*-short was sub-cloned into the plant expression vector pCAMBIA 1305.1, and *CLF*-long into the altered pCAMBIA 1305.1vector, pMM3. The resulting constructs are referred to as MMshort2 and MMlong2. As shown in Figure 3.7, digestion of the MMshort2 plasmid with *Bgl*II and *Hin*dIII, generated an 11.8 kb piece representing the pCAMBIA 1305.1 vector, and a 2.8 kb piece representing the insert. The digestion pattern indicates that the short promoter was successfully cloned into the pCAMBIA 1305.1 vector. Unfortunately, the *CLF*-long piece has not yet been successfully sub-cloned into pMM3.

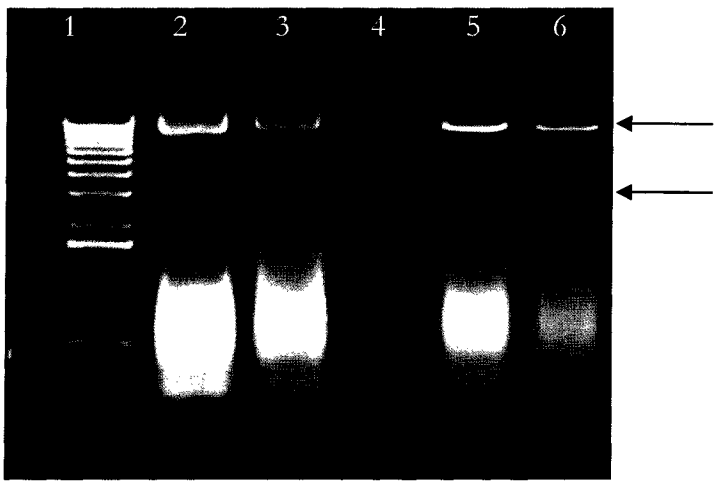
3.3.4.2 Transformation of *Agrobacterium* with MMshort2

Agrobacterium was transformed with the MMshort2 construct via the freeze-thaw transformation protocol (An, *et al.*, 1988). PCR amplification conducted with primers for the *CLF*-short promoter and an aliquot of overnight culture prepared from *Agrobacterium* transformed with the MMshort2 plasmid generated 2.8 kb pieces demonstrating that the *Agrobacterium* was successfully transformed with the MMshort2 construct.

3.3.4.3 Transformation of Arabidopsis with MMshort2

Generation of transgenic plants is currently underway. No preliminary data is yet available.

Figure 3.7. Confirmation of insert in MMshort2 via restriction digestion. The MMshort 2 plasmid was digested with *Bgl*II and *Hin*dIII. Lane 1: 1 kb DNA ladder, Lanes 2 and 3: MMshort 2 plasmid undigested; Lane 4: Negative control reaction, restriction digest with no DNA added. Lanes 5 and 6: MMshort2 plasmid digested with *Bgl*II and *Hin*dIII, the top arrow points to the 11.8 kb vector band and the bottom arrow to the 2.8 kb insert band.



3.3.5 How does *CLF* regulate the vascular cambium?

3.3.6 Could HD-ZIP III Genes be Targets of Pc-G-Mediated Regulation?

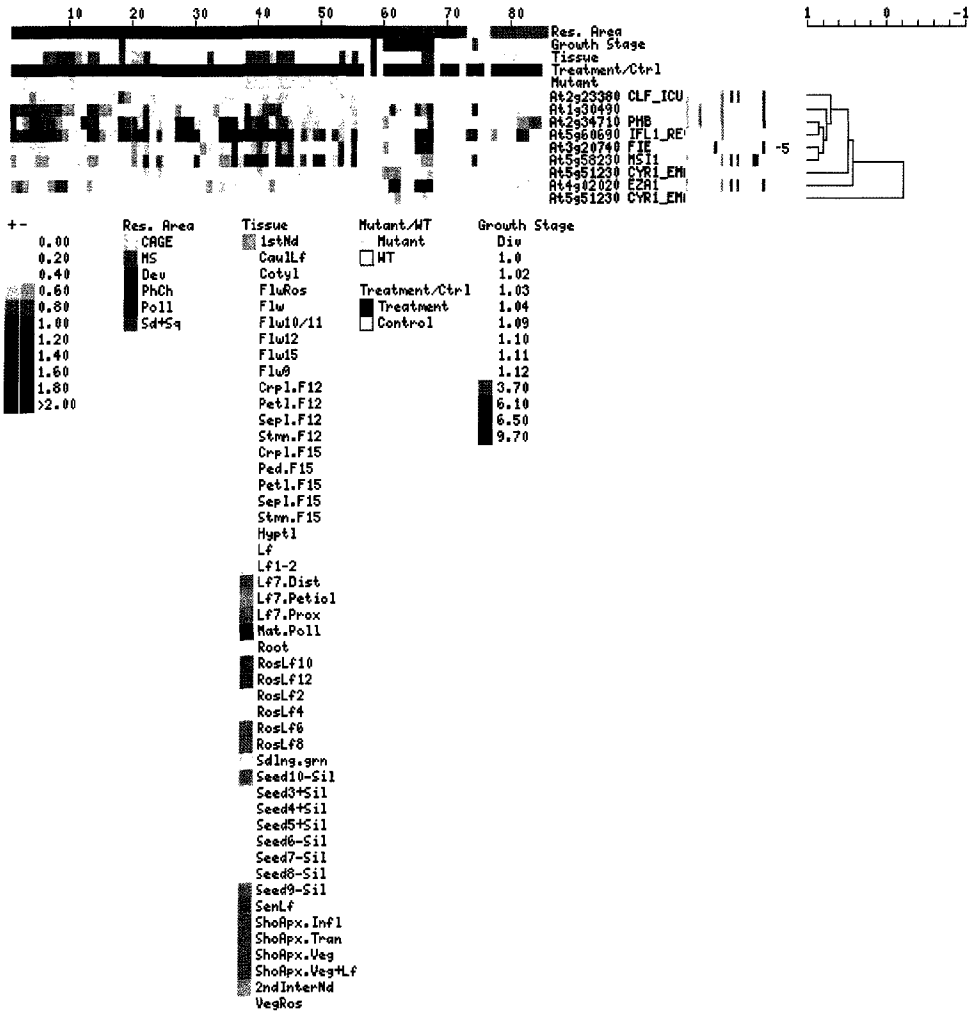
Figure 3.8 is the Expression Browser output for three of the HD-ZIP III genes, *REV, PHB, PHV,* known to play a role in vascular bundle organization and differentiation, and the five components of the Arabidopsis PRC2, *EMF2, FIE, CLF, SWN* and *MSI1.* The HD-ZIP III transcription factors show an increase in gene expression in the early hypocotyl and the shoot. *PHB*, in particular, is up-regulated 2-fold in the early hypocotyl whereas the PRC2 components are slightly down-regulated or show no change. In the first node and second internode, *PHV* displays a 1.2 and 1.4--fold up-regulation, respectively, *PHB* a 1.7 and 1.5 fold up-regulation and *REV* a 1.6 and 2- fold up-regulation. The gene expression profile for the PRC2 components in the shoot tissues shows no change in *CLF* expression, and slight down-regulation for *FIE, MSI10, SWN* and *EMF.*

The HD-ZIP III genes are weakly down-regulated in flower organs, specifically the sepals, petals and stamens, with slight up-regulation in the carpels. *PHV* shows down- regulation of 0.6, 0.9, and 0.9-fold; *PHB* shows 0.4, 0.5, and 0.3-fold down-regulation; and *REV* exhibits 0.5, 0.5, 0.2-fold down-regulation. All of the HD-ZIP III genes exhibit slight up-regulation in carpels, on the order of 1 to 2-fold.

3.3.7 Could Cambial-specific HD-ZIP III Genes be Targets of Pc-G-Mediated Regulation?

One hypothesis for the phenotype of the *clf* mutant, is that there is misexpression or down-regulation of cambial-specific HD-ZIP III genes. Two of HD-ZIP III genes are specifically expressed in the procambium, *ATHB8* and *ATHB15* and have recently been identified as

Figure 3.8. Expression profile for three HD-ZIP III transcription factors and the Arabidopsis PRC2 components. The expression profiles across development were generated for *REV*, *PHB* and *PHV* and the components of the Arabidopsis PRC2, *EMF2*, *FIE*, *CLF*, *SWN* and *MSI*.

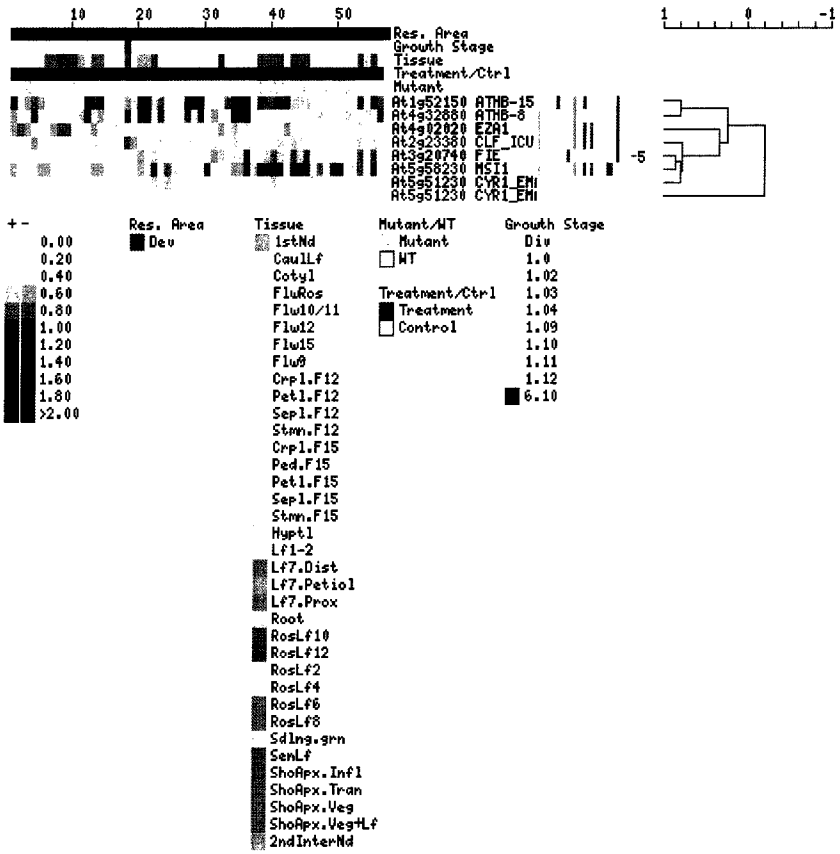


being specifically expressed in the xylem of the secondarily-thickened hypocotyl of Arabidopsis (Zhao, *et al.*, 2005). To examine the possibility that HD-ZIP III genes may be regulated by Pc-G proteins, the gene expression profiles of these genes were compared to the profiles for the PRC2 components.

The gene expression profiles for the two procambium specific HD-ZIP III genes, *ATHB8* and *ATHB15* and the five components if the Arabidopsis PRC2, *EMF2*, *FIE*, *CLF*, *SWN* and *MSI1* are presented in Figure 3.9. Both of the HD-ZIP III genes are up-regulated in the early hypocotyl, with 3.4 and 2.5-fold increases, respectively. The PRC2 components demonstrate very small changes in expression in the early hypocotyl, in the range of 1 to 0.8-fold changes. An increase in gene expression is observed for *ATHB8* and *ATHB15* in the shoot, specifically the first node and second internode, with 5 and 4.3-fold increases for *ATHB8,* respectively, and 2.8 and 2.6-fold increases for *ATHB15* respectively. In the first node and second internode, slight up-regulation of 1- fold is observed for *SWN* and *CLF*, whereas slight down-regulation is observed for the rest of the PRC2 components, 0.8 and 0.6- fold down-regulation for *FIE*, 0.8 and 0.6-fold down-regulation for *MSI1,* and 0.9 and 0.8-fold down-regulation for *EMF2.*

The clade diagram indicates that *ATHB8* and *ATHB15* are 80% co-regulated.

Figure 3.9. Expression profiles for two cambial-specific HD-ZIP III genes and the PRC2 components. The expression profile for *ATHB8* and *ATHB15* and the genes for the Arabidopsis PRC2 components, *EMF2*, *FIE*, *CLF*, *SWN* and *MSI1*, across development.



3.3.8 Are the PRC2 Components Co-expressed with Vascular Genes?

Zhao and colleagues have generated tissue-specific gene expression profiles for the secondarily-thickened root-hypocotyl of Arabidopsis (see section 1.5) (Zhao, *et al.*, 2005). They have identified 37 genes with a role in vascular tissue differentiation or function. To examine the possibility that the PRC2 component genes are co-expressed with these genes, the Pearson correlation coefficient was determined for the PRC2 component genes and the vascular genes identified by Zhao (Zhao, *et al.*, 2005), using the Botany Affymetrix Resource Data set. Gene pairs that exhibited a correlation value of 0.6 or higher with at least one Pc-G gene are presented in Table 3.3. Of the PRC2 component genes, only *FIE*, *CLF* and *MSI1* exhibit high correlation coefficients with vascular genes. *FIE* and *PHB* have a co-expression correlation of value of 0.67. *MSI1* and *XYP1* have a co-expression correlation of 0.77 and *MSI1* and *REV* have a co-expression correlation of 0.71. Of the 37 vascular-specific genes identified by Zhao and colleagues, *CLF* shows reasonably high co-expression correlations with 5 of them, a 0.6 correlation coefficient with *ATHB15,* 0.61 with *KAN2*, 0.63 with *GNOM/EMB30,* 0.67 with *REV* and 0.82 with *PHB*. Three of these genes, *PHB*, *ATHB15* and *REV* are members of the HD-ZIP III gene family.

3.3.9 Gene Expression Profiles for *CLF* and *AG* in Green Tissues

In the Gene Angler tool it is possible to examine co-expression of groups of genes in specific tissues. When all tissues are examined for genes co-expressed with a target gene, those with very high co-expression may have a sufficiently strong signal intensity that they "drown" out ones with lower signal intensities. For example, if the target gene is expressed in flowers, only other flower genes are generated because their coefficients of co-expression are sufficiently high. If, however, the floral tissue is eliminated from the analysis, genes with

Gene Locus	Gene Symbol	Description	Tissue Bias	Role in Vascular Tissue Differentiation	Pearson Correlation Coefficient for Co-expression with PRC2 Components				
					EMF2	FIE	CLF	SWN	MSI1
At4g32880	*ATHB-8*	Homeobox-Leu zipper family	Xylem	Vascular differentiation	0.252	0.178	0.511	0.291	0.287
At2g34710	*PHABULOSA/ ATHB14 (PHB)*	Homeobox-Leu zipper family	Xylem	Vascular bundle organization	0.218	**0.666**	**0.822**	0.064	0.590
At1g52150	*CORONA/ ATHB15*	Homeobox-Leu zipper family	Xylem	Vascular bundle organization	0.217	0.367	**0.601**	0.256	0.395
At5g60690	*REVOLUTA/ INTERFASCICULAR FIBERLESS (REV)*	Homeobox-Leu zipper family	Xylem	Vascular patterning	0.106	0.547	**0.665**	0.188	**0.709**
At1g32240	*KAN2*	G2-like transcription factor	PhloemCambium/ Nonvascular	Vascular bundle organization	0.020	0.507	**0.614**	0.068	0.529
At5g64080	*XYP1*	G2-like transcription factor	PhloemCambium/ Nonvascular	Vascular patterning	-0.182	0.460	0.350	-0.192	**0.766**
At1g13980	*GNOM/ EMB30*	Vascular patterning	No tissue bias	Vascular patterning	0.171	0.429	**0.634**	0.378	0.371

Table 3.3. Co-expression values for gene pairs of the PRC2 components and seven secondary vasculature-specific genes. The correlation coefficients for seven secondary vasculature-specific genes, identified by Zhao et al. (2005), and the PRC2 components were generated with the Gene Angler tool of the BAR. Genes for which the highest co-expression values were obtained are listed. Gene pairs with greater than a 60% co-expression correlation are highlighted. Description of the gene, tissue bias and its role in vascular tissue differentiation are from Zhao *et al.* (2005).

lower, but still significant coefficients of co-expression can be examined.

To examine genes co-expressed with *CLF* in green tissues, floral tissues were removed from the AtGenExpress dataset and genes with a coefficient of co-expression of 0.82 and above were examined in Gene Angler. The Gene Angler results are presented in Figure 3.10. Of particular interest is a cell division protein kinase gene (At1g18040) with a coefficient of co-expression of 0.891 and a trithorax group protein (At2g31650) with a coefficient of co-expression of 0.870. As the results in chapter 2 indicate, (see sections 2.2.5 and 2.2.6) it seems likely that there is an aberration in cell division in the *clf* mutant. The strong co-expression value for *CLF* and a cell division gene may indicate a regulatory function for *CLF* in cell division. The same analysis was conducted with *AG,* to examine genes co-expressed just in green tissues. The Gene Angler results are presented in Figure 3.11. In green tissues, *AG* is predominantly co-expressed with genes for transcription factors, proteins involved in transcription, such as a DNA-direct RNA polymerase II (At3g16980) and proteins of unknown function.

Figure 3.10. Gene Angler results for genes co-expressed with *CLF* in green tissues. An analysis of genes co-expressed with *CLF* just in green tissues, r = 0.82 or above, was done by removing the floral tissue from the AtGenExpress dataset and using the Gene Angler tool of the BAR. The arrows point to two genes of particular interest that show high co-expression with *CLF*, a cell divivision protein kinase gene (At1g18040) and a trithorax group gene (At2g31650).

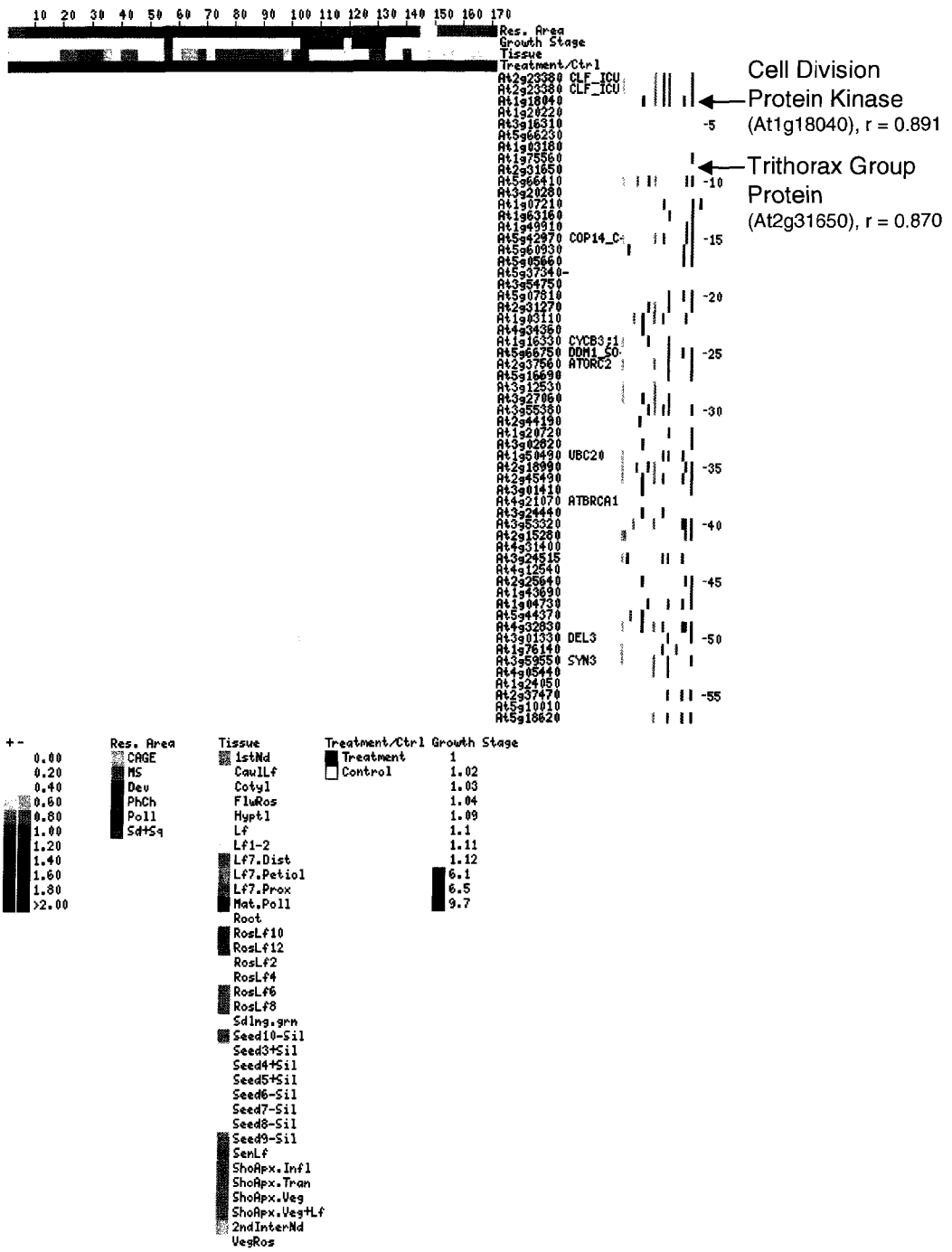
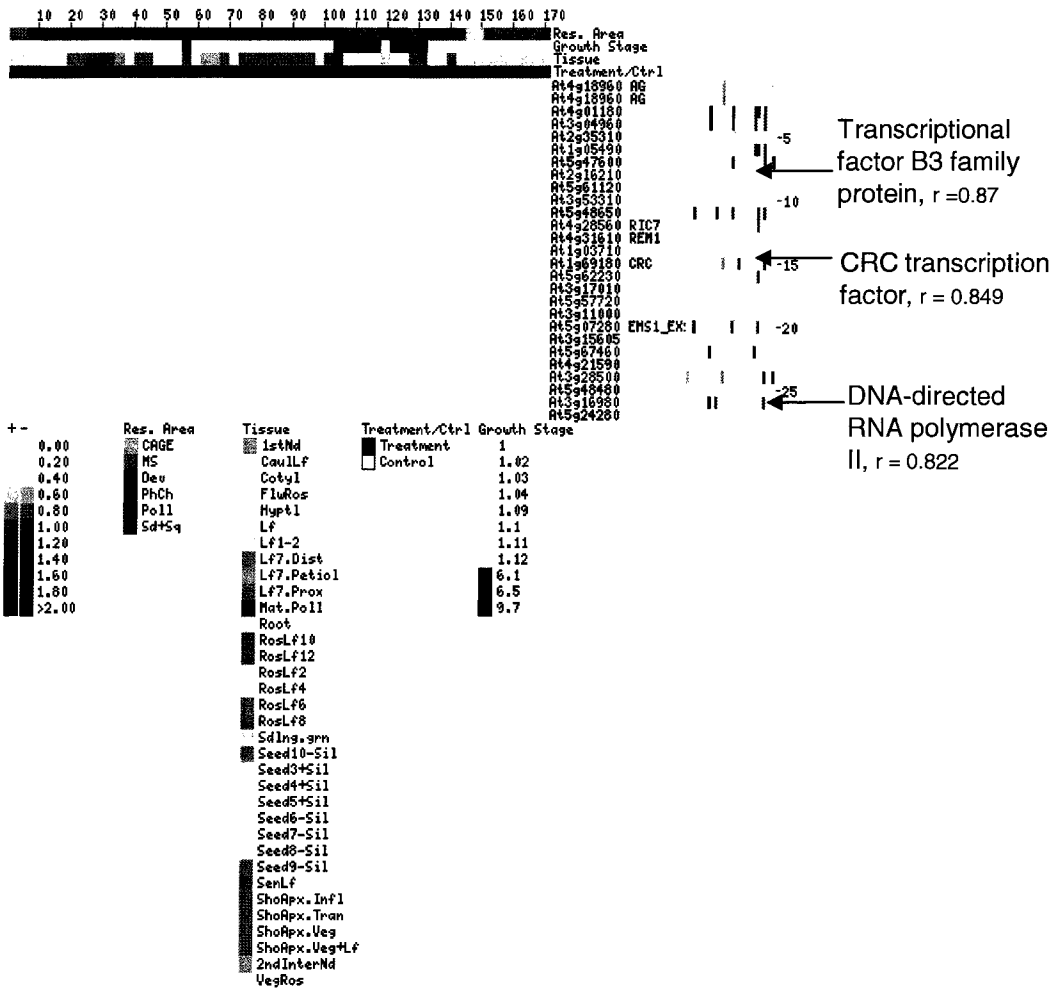


Figure 3.11. Gene Angler results for gene co-expressed with *AG* in green tissues. An analysis of genes co-expressed with *AG* in green tissues, r = 0.82 or above, was done by removing the floral tissue from the AtGenExpress dataset and using the Gene Angler tool of the BAR. In green tissues, *AG* is predominantly co-expressed with genes encoding transcription factors and other proteins involved in transcription, several of these are highlighted.



3.3.10 Do Physical Rearrangements in Chromatin Affect Genes Other than *AG*?

The full mechanistic details of Pc-G protein repression are unknown. Despite this, there is much evidence indicating that they maintain repression of gene expression by methylating specific histones in the chromatin of the target gene. The methylation causes chromatin compaction; transcriptional machinery is unable to access the target gene in the compacted chromatin, resulting in repression of the target gene.

One target gene of the CLF protein is *AG*. One hypothesis for the phenotype observed in *clf* mutants is that chromatin surrounding *AG* is not compacted, resulting in the over-expression of genes physically linked to *AG* (on chromosome 4) and causes the *clf* phenotype. In this scenario, it would be expected that when *AG* is expressed, *i.e.* when its surrounding chromatin is in a "loose" conformation, that other genes close to *AG* on chromosome 4 would also be expressed. To test this hypothesis, the following question was asked: is there a bias in the genes co-expressed with *AG*, *i.e* are more of the genes co-expressed with *AG* located on chromosome 4?

The number of genes that show a 60% rate of co-expression (Pearson Correlation Coefficient, r = 0.6) was examined with the Gene Angler tool of the BAR, using the AtGenExpress tissue series data set. 193 genes were shown to be co-regulated with *AG* with a Pearson Correlation Coefficient of 0.6 or higher. The distribution of these genes across the 5 chromosomes of Arabidopsis was compared to the expected distribution (Meinke, *et al.*, 2003) and is presented in Table 3.4. No significant difference was found between the chromosomal distribution of genes co-expressed with *AG* and the expected distribution of genes across the chromosomes.

Chromosome	% of Genes Distributed across the Genome	% of Genes Co-expressed with *AG*	t-value	Significance?
1	25.4	30.4	1.17	No
2	16.6	12.3	1.01	No
3	20.6	21.6	0.23	No
4	14.8	14.9	0.05	No
5	22.2	20.6	0.67	No

Table 3.4. Distribution of genes co-expressed with *AGAMOUS* across the five Arabidopsis chromosomes compared with the expected distribution of genes across the chromosomes. The percentage of genes co-expressed with *AG*, r = 0.6, across the five Arabidopsis chromosomes was compared to the expected distribution of the Arabidopsis genes across the genome. Genes on Chromosome 4 are not significantly more likely to be co-expressed with *AG* at the 95% level of significance.

3.4 Discussion

3.4.1 Gene Expression Profile for *CLF* Across Development : No Evidence of a Regulatory Relationship

The gene expression profiles for *CLF* and *AG* were compared across development, by generating an e-Northern with the tools available from the BAR website with the AtGenExpress Tissue Data set. This was done to determine if there was a clear difference in the expression patterns of *CLF* and *AG*. The CLF protein maintains repression of *AG* in vegetative tissues, so it was hypothesized that up-regulation of *CLF* might be observed in vegetative tissue with a corresponding down-regulation in *AG*. As the Expression Browser output displayed in Figure 3.2 indicates, the overall trend for *CLF* expression across development is small changes in expression. The raw expression value of *CLF* remains at about 100 in the vegetative tissue, which is a larger than raw expression values for *AG* in vegetative tissue, which are between 12 and 20. However, the raw expression values for *AG* increase considerably in the stamens and carpels of stage 12 flowers, to 1562.7 and 595.8, respectively.

When the average gene expression profile is compared to the average of a control, very little change in *CLF* expression levels are observed, no change greater than 1.6- fold is observed for either up or down-regulation. The expression profile for *CLF* does not suggest either a temporal or tissue-specific gene expression pattern for *CLF*. This is in contrast to the gene expression profile generated for *AG*, where an expected down-regulation is observed in vegetative tissue and expected strong up-regulation is observed in stamens and carpels. This is consistent with the genetic observations of *AG* expression.

The genetic data indicates that *CLF* is a negative regulator of *AG* expression (Goodrich, *et al.*, 1997). For this reason, it might be expected that a regulatory relationship between the two genes would be obvious when examining the e-northern (*i.e.* the target gene, *AG* and its negative regulator, *CLF*, would be expressed at opposite levels at the same time in the same tissue). However, a clear inverse relationship like this cannot be inferred based on the e-northern. The CLF protein contains a SET domain and acts as part of the PRC2 multi-protein complex. Proteins containing SET domains are not active unless complexed with their partner proteins (Czermin, *et al.*, 2002). For this reason, perhaps a clearer picture of CLF's effect on *AG* expression would be evident if the expression profiles for the PRC2 components were compared with the *AG* expression profile.

3.4.2 Gene Expression Profile for Pc-G genes and *AG* Across Development: : No Evidence of a Regulatory Relationship

The expression profiles in Figure 3.2 do not show a clearly inverted relationship between *CLF* and *AG* expression levels in vegetative tissues. Instead, the expression profile for *CLF* suggests a basal expression state that changes very little over time. Perhaps most of the Pc-G proteins, including CLF, are expressed at a low level and a change in expression level of just one component induces formation and activity of the Polycomb Repressive Complex 2. To investigate this hypothesis and examine the extent to which expression levels of the PRC2 components was inversely related to AG expression, the expresion profiles of *EMF2*, *FIE*, *CLF*, *SWN*, *MSI1* and *AG* were generated. It was hypothesized that a general trend for up-regulation of the PRC2 components would be observed in tissues where down-regulation of *AG* is observed, because the PRC2 negatively regulates *AG* expression.

As Figure 3.3 indicates, a similar trend is observed for *AG* and *CLF* expression as was previously noted, *CLF* expression does not change very much over development and *AG* expression is low in vegetative tissue, up-regulated in general floral tissue and strongly up-regulated in stamens and carpels. Almost no change is observed in *CLF* across development and changes in *AG* expression are consistent with it role in development. Most of the other Pc-G proteins show a similar expression pattern as *CLF*, with only small changes over the course of development.

The pattern of expression is different for *MSI1*. It exhibits greater changes than the other components of the PRC2, specifically with down-regulation in the stamen, where *AG* expression is highest. Perhaps the MSI1 protein acts to recruit the other PRC2 components, so that the PRC2 complex is formed in response to amount of MSI1 available. In this scenario, only a basal level of the other PRC2 components is required so limited changes in their expression profiles is observed across development. Instead MSI1 acts as an inducer for protein complex formation and activity, such that the changes in *MSI1* expression are indicative of changes in PRC2 complex activity. In support of this postulation, in the three Pc-G protein regulatory complexes proposed by Chanvivattana and colleagues, the MSI1 protein is a required member of each complex (Chanvivattana, *et al.*, 2004).

Interestingly, four of the PRC2 genes, *CLF*, *FIE*, *EMF2* and *MSI1*, show a high level of co-regulation, 80% co-regulation. This high level of co-regulation supports the proposal of Chanvivattana that these components are part of the same multi-protein complex that maintains transcriptional repression of their targets (Chanvivattana, *et al.*, 2004). The lower level of co-regulation observed for *SWN*, approximately 30% co-regulation, may further

support their assertion that there is only partial redundancy between the functions of the SWN and CLF proteins.

The *in silico* results do not definitely indicate either tissue or developmental specificity for expression of the PRC2 components. The expression profiles do not indicate one tissue or one point in development where there is up-regulation in the expression of *CLF* or its PRC2 partners. To examine tissue and developmental specificity of *CLF* expression, transgenic plants transformed with the *CLF* promoter fused to the GUS reporter gene were generated.

3.4.3 Floral Genes and Transcription Factors are Co-expressed with *AG*

To test the assertion of (Toufighi, *et al.*, 2005) that relationships suggested by the Expression Browser and Gene Angler are recapitulated in the lab, the genes co-expressed with *AG* were examined. As Table 3.1 indicates, many of the genes that are known to be co-expressed with *AG*, are listed and exhibit high correlation values for co-expression. For example, the floral genes *AP3* and *PI* which genetic data indicates are expressed in the third whorl of the flower with *AG*, have co-expression values of 0.976 and 0.991, respectively, with *AG*. These values indicate that the data obtained from the Gene Angler tool corroborate pre-existing genetic data about co-expression of floral genes and provide proof of concept to test biological questions *in silico* with the tools prior to designing experiments.

The Gene Angler results presented in Table 3.2 are different than from the *in vitro* results. As a negative regulator of *AG*, it might be expected that *CLF* expression would negatively correlated with *AG* expression. This is not the case for either *CLF*, which shows almost no correlation of expression, r =-0.028. Neither do any of the other PRC2 components show a

clear negative correlation in gene expression, all of the values are low, especially for *EMF2* and *FIE* which have correlation values of 0.084 and -0.108, respectively. These results suggest a slight gene expression relationship between *AG* and *SWN* and *MSI1*, with correlations of 0.277 and 0.247, suggesting that *AG* expression is correlated with expression of these Pc-G genes about 25% of the time. The genetic evidence suggests that *CLF* regulates *AG* and that is certainly the consensus in the literature (Goodrich & Tweedie, 2002; Schubert, *et al.*, 2005; Sieburth & Meyerowitz, 1997), however this relationship is not explicit in the results obtained with Gene Angler. This suggests one of two possibilities. First, the database does not generate accurate information about genetic relationships that have been previously observed. However, as the results for co-expression with *AG* indicate, many genetic interactions observed *in vitro* and *in vivo* are re-capitulated in the Gene Angler results. Another possibility is that the regulatory relationship between the PRC2 components and *AG* is not at the level of gene expression and is therefore obscured in gene expression data.

Finally, the ectopic expression of *AG* has been suggested to be responsible for the altered leaf morphology in the *clf* mutant, particularly in light of the observation that in *ag clf* double mutants, where *AG* activity is eliminated, leaf morphology is restored to near wildtype (Goodrich, *et al.*, 1997; Serrano-Cartagena, *et al.*, 2000). The AG protein is a transcription factor, so perhaps the over-expression of *AG* in *clf* plants up-regulates other genes that are causes the altered leaf morphology. Examination of Gene Angler results in all tissues (Table 3.1) and in just green tissues (Figure 3.11) does not yield an obvious target gene that could account for the leaf morphology. When co-expression with *AG* is examined across all tissues, only floral genes or regulators of *AG* exhibit co-expression. When genes co-

expressed with *AG* in green tissues is examined, genes encoding transcription factors, proteins involved in transcription, and proteins of unknown function exhibit co-expression. These results suggest that a gene regulated by *AG* is not responsible for the altered leaf morphology present in *clf* mutants.

3.4.4 *CLF* Promoter Region Cloning

Both the "short" and the "long" *CLF* promoter regions were successfully cloned into pGEMT-easy and sub-cloned into the *Agrobacterium*-specific pCAMBIA 1305.1 vector. The constructs were then transformed into *Agrobacterium*, and wild-type Col-0 plants transformed via the floral dip method (Clough & Bent, 1998). There are no preliminary data available at this time for the transgenic plants.

3.4.5 Gene Expression Profile for HD-ZIP III Genes and Pc-G genes across Development

Consistent with their requirement for vascular specification and patterning, the gene expression profiles for the HD-ZIP III genes *REV, PHB* and *PHV* exhibit up-regulation in the shoot with a corresponding down-regulation, or low level, of gene expression for the PRC2 components. This inverse expression profile suggests that there could be a regulatory relationship between the PRC2 components and the HD-ZIP III transcription factors.

The hypothesis that HD-ZIP III genes are regulated by Pc-G proteins was further investigated by looking at the expression profile of two transcription factors specifically expressed in the procambium, *ATHB8* and *ATHB15*. Consistent with the role of *ATHB8* in regulating cell proliferation and xylem differentiation, the greatest increase in expression is

in the early hypocotyl and the shoot, both tissues where procambial cell proliferation and xylem differentiation occur. *ATHB15* has a similar expression profile as *ATHB8*, with increased expression in the hypocotyl and shoot. There is corresponding down-regulation of the PRC2 component genes in the same tissue. This may suggest an inverse relationship in the expression levels of *ATHB8*, *ATHB15* and the Pc-G proteins, and the possibility that a regulatory interaction exists between these two groups of proteins.

3.4.6 What is The Mechanism for the Effect of CLF on Vascular Cambium?

It is notable that while much is known about Pc-G proteins, particularly with regard to their role in flowering, embryo development and vernalization, no previous link has been made between *CLF* and the vascular cambium. The CLF protein plays a role in the control of flowering and an undetermined role in cell cycling (Kim, *et al.*, 1998). The pleiotropic effect of the *clf* mutation suggests a role for the protein in a broad number of functions, or in specifying gene expression patterns early in development. Increasingly, Pc-G proteins are seen to play a broader role in specifying developmental patterns in plants (Hsieh, *et al.*, 2003; Reyes & Grossniklaus, 2003; Steimer, *et al.*, 2004). The expression patterns for the GUS constructs described in this thesis are expected to shed additional light on temporal and tissue-specific details regarding *CLF* expression.

3.4.7 Pc-G Proteins Are Responsible for Altered Vascular Development

Pc-G proteins are well described in the literature, specifically with respect to epigenetic control of developmental processes. Despite this interest, the examination of Pc-G proteins has been hindered by a limited examination of their phenotypic effects, specifically there has been a focus on flowering and reproductive effects of these mutations with limited

discussion of their role in other processes. The subsequent sections discuss the link between mutations in Pc-G proteins and altered vascular development.

3.4.8 Could Pc-G Proteins Regulate the Expression of HD-ZIP III Transcription Factors?

Pc-G proteins regulate homeobox genes in *Drosophila* and MADS-box genes in Arabidopsis. The latter have previously been identified as MADS-box transcription factors such as *AG*, *PHE1* and *FLC*. However, Katz and colleagues have recently demonstrated that both CLF and FIE also regulate the homeobox genes *STM* and *KNT2*. HD-ZIP III genes are also homeobox genes and are required for the proper development of the vasculature (Katz, *et al.*, 2004). Like flower development, the development of vasculature is a process dependant on specific and regulated patterning.

The results in Table 3.3 show correlations of co-expression for *CLF* and three xylem-specific HD-ZIP III genes, *ATHB14*, 0.82, *ATHN15*, 0.60, and *REV*, 0.67. Combined with evidence that CLF regulates other homeobox genes, this suggests that the HD-ZIP III transcription factors might also be regulated by CLF. However, CLF is a Pc-G protein, so the targets of its regulation should be repressed in wild-type plants. It therefore follows that in *clf-1* plants, the HD-ZIP III genes would be over-expressed and the mutant would produce excess vasculature. This is not the case in the *clf-1* mutant which has no secondary vasculature. If CLF regulates the expression of *ATHB14*, *ATHB15* and *REV*, the only type of regulation which would be consistent with the phenotype and the observed co-expression correlations is positive regulation.

Homeobox genes in *Drosophila* are regulated by two groups of proteins, Pc-G proteins and trx-G proteins. These two groups of proteins are typically considered to function antagonistically, with Pc-G proteins maintaining transcriptional repression of homeotic genes and trx-G genes maintaining their transcriptional activation. If HD-ZIP III genes were regulated by a similar mechanism, Pc-G proteins would be predicted to maintain transcriptional repression of HD-ZIP III genes and trx-G proteins to maintain activation of them. The CLF protein was described as a Pc-G based on its protein sequence similarity to E(Z), specifically the presence of a SET domain, the domain responsible for its HMTase activity (Kohler & Grossniklaus, 2002). The categorization of E(Z) as a Pc-G protein has come under scrutiny (LaJeunesse & Shearn, 1996). It has been demonstrated that mutations in the E(Z) protein can alter its function and it can act as either a repressor or activator of transcription (LaJeunesse & Shearn, 1996). It was concluded that E(Z) can act as both a Pc-G gene or trxG gene and that its functionality is tissue specific, so that E(Z) can result in activation of its target gene in one tissue and repression of the same gene in a different tissue.

There is some precedent in plant biology for proteins that act as both repressors and activators of transcription, including the VP1 protein in maize (Hoecker, *et al.*, 1995) as well as the Dof and WRKY proteins in Arabidopsis (Miao, *et al.*, 2004; Yanagisawa, 2004). Possibly CLF has similar bi-functionality and acts as a repressor of MADS-box transcription factors and an activator of homebox-leucine zipper genes. The phenotype of the *clf-1* mutant may be due to the altered CLF protein inducing ectopic expression of *AG* in vegetative tissues and late flowers, while reducing expression patterns of *ATHB8* and *ATHB15* in vascular tissues.

The gene expression profile presented in Figure 3.9 indicates that *CLF* is up-regulated in the same tissues where *ATHB8* and *ATHB15* are up-regulated. This could suggest that CLF promotes transcriptional activation of *ATHB8* and *ATHB15*. Further, *CLF* is highly co-expressed with a cell division protein kinase gene (At1g18040) in green tissues. The observation that the leaf cells of *clf* are smaller than those of wildtype suggests an aberration in cell division. Further, Kim and colleagues have suggested that CLF plays a role in cell division and cell elongation throughout the development of the leaf primordia (Kim, *et al.*, 1998). Perhaps the CLF protein is also a positive regulator of At1g18040 and the leaf phenotype in *clf* is due in part to down-regulation of this gene, resulting in small leaves composed of small cells. Finally, there may be a false separation between the Pc-G and trx-G proteins. Instead of acting antagonistically they may represent different possible combinations of proteins that have differential effects on transcription. The observation that *CLF* is highly co-expressed with a trx-G gene (At2g31650) suggests that there isn't an antagonistic relationship between *CLF* and this gene, and further supports the postulation that CLF can act like a trx-G protein and maintain transcriptional activity of target genes. The absence of a PRC1 in Arabidopsis suggests that Pc-G proteins function differently in plants than they do in animals (Kohler & Grossniklaus, 2002). This functional difference could include proteins that act to both activate and repress transcription.

The expression profiles generated *in silico* for the procambial-specific HD-ZIP III genes *ATHB8* and *ATHB15* and the components of the PRC2 suggest a small inverse expression relationship. An increase in expression in the hypocotyl and stem is observed for *ATHB8* and *ATHB15* with a corresponding decrease in expression for several of the Pc-G genes. However, this possible inverse relationship is not supported with the Gene Angler data. As

Table 3.3 indicates, there are no large negative values for co-expression correlations between PRC2 components and HD-ZIP III genes, while *FIE* shows a co-expression correlation of 0.67 with *ATHB15* and *MSI1* a co-expression correlation of 0.709 with *REV*. Further, *CLF* shows a high correlation of co-expression with two additional vascular genes, *KAN2*, r = 0.614, and *GNOM/EMB30*, r = 0.634. This suggests a new role for CLF in specifying the transcriptional state of vascular genes, in addition to regulating floral development, floral-organ abscission, flowering-time and cell-cycle control.

3.4.9 Do Physical Rearrangements in Chromatin Close to *AG* Result in Up-regulation of Chromosome 4 Genes?

In the *clf* mutant, *AG* is ectopically expressed. This is due to the fact that the mutation in *CLF* generates a mutant protein that cannot repress *AG*. Pc-G protein mediated repression alters chromatin structure such that the target gene cannot be accessed by the transcriptional machinery. It is possible that in the *clf* mutant, lack of chromatin compaction in the region surrounding *AG* results in over-expression of other genes physically linked to *AG*. Perhaps over-expression of these genes results in the *clf* phenotype. To examine this possibility, genes with a correlation coefficient with *AG* of 0.6 or higher were investigated. The 193 genes co-expressed with *AG* were examined to determine if there was a significant bias in favour of genes on chromosome 4. No significant difference was observed for the chromosome location of genes co-expressed with *AG* as compared to the normal distribution of genes across the five Arabidopsis chromosomes. This result suggests that the *clf* phenotype is not to the over-expression of a gene closely physically linked to *AG* on chromosome 4.

3.4.10 The Role of Data Mining in Arabidopsis Research

The use of the BAR in this paper highlights the importance and power of data mining in Arabidopsis research. The BAR tools make it possible to go beyond merely examining databases and assist in hypotheses testing prior to designing wet lab experiments. There is no obvious mechanistic link between epigenetic control of gene expression and the development of secondary vasculature. However, the use of the BAR in this analysis has permitted the examination of many genes and suggests connections amoung the genes that have not been previously tested. For example, the correlation coefficients for *CLF* and five vascular specific genes, *PHB*, *ATHB15*, *REV*, *Kan2*, and *GNOM/EMB30,* suggests a possible regulatory relationship not previously indicated in the literature. This novel observation will assist in the design of further experiments to elucidate the role of *CLF* in vascular development.

As a general caveat to this section, it must be pointed out that regulation of gene expression occurs in a variety of ways and is not limited to regulation of transcription. Gene expression can be regulated through mechanisms involved in RNA transport, transcript stability, translational initiation, post-translational modification, protein transport, protein stability and protein modification. These post-transcriptional types of regulation cannot be examined with the BAR tools. However, given that an interaction between *CLF* and the vascular tissue is a novel observation, for which there is no additional support in the literature, the transcriptional interactions discussed in this chapter are a reasonable starting place for the design of *in vitro* and *in vivo* experiments.

4.0 General Discussion

4.1 Map-Based Cloning is a Viable Strategy for Gene Discovery

The results of the MBC portion of this thesis, discussed in detail in Chapter 2, underline that MBC is an effective strategy for gene discovery in EMS-mutagenized lines. Mutagenesis with EMS creates point-mutations. Point mutations shed more light on a gene's function than knock-out mutations because they often produce alleles of various strengths, thereby providing nuanced information regarding gene function. For this reason, EMS mutants are helpful in understanding the full-spectrum of a gene's function. The disadvantage of EMS-induced point mutations is the difficulty of identifying the target gene and the precise location of the point mutation. However, MBC clearly is capable of identifying the mutated gene. In this example, a narrow region of chromosome 2 was identified as the mutant gene location and this region was ultimately determined to be the location of the *CLF* gene. Further, the success of this mapping project indicates that MBC can be done by just one person, especially if assisted with the tissue collection, within the limited time frame of a Master's degree.

4.2 Broad Developmental Role for Polycomb- Group Proteins

The discovery that the "*CAM*" gene is the *CLF* gene was surprising, particularly because *CLF* appears to be a well-characterized gene. This discovery provides evidence for a broader developmental role for Polycomb-group (Pc-G) proteins, and a previously uncharacterized role for *CLF* in the regulation of the vascular cambium. It remains unknown how the CLF protein exerts its effect on the vascular cambium, but the work described in this thesis clearly indicates that CLF regulates genes that are crucial for VC formation. Other Pc-G genes have previously been shown to control expression of

meristem identity genes, specifically of the root meristem. In *fie*, *clf swn* and *vrn2 emf2* double mutants, roots do not develop properly and are converted to hypocotyl – like tissue (Schubert, *et al.*, 2005).

Several experiments, described in Chapter 3, have been undertaken to examine how CLF exerts its effect on the vascular cambium, but preliminary data is not yet available. However, analysis of *CLF* co-expression with several HD ZIP-III genes suggests a possible regulatory relationship between these genes. This analysis was conducted with the Gene Angler tool of the Botany Array Resource (BAR). To lend credibility to this hypothesis both *in vitro* and *in vivo* experiments need to be performed. Specifically quantitative PCR (qPCR) could be carried out to determine if the expression profiles of the HD-ZIP III genes, specifically *PHB*, *ATHB15*, *REV* and *ATHB8* are altered in *clf* plants and other Pc-G mutants. It could also be useful to examine the expression profiles of the Pc-G genes in the HD- ZIP III mutants.

Chapter 3 introduced a hypothesis that the CLF protein has bi-functionality. It was posited that CLF acts in the maintenance of both transcriptional repression and activation. This hypothesis contrasts with the general consensus in the literature that CLF maintains repression of its target genes, but there is experimental evidence to support a bi-functionality of SET-domain proteins like CLF. Experimental examination of the bi-functionality of CLF would be interesting and perhaps shed light on how Pc-G proteins regulate gene expression in Arabidopsis. In the *Drosophila* experiments which found SET-domain proteins can act to both maintain activation and repression of their targets, mutants with knocked-out trx-G proteins were made and the gene expression levels of trx-G targets observed. It was

expected that maintenance of transcriptional activation would not be observed in the absence of the trx-G proteins, but instead the target genes remained active. A similar approach might be possible with Arabidopsis although fewer trx-G proteins have been characterized in Arabidopsis than in *Drosophila.* An initial comparison of protein sequences and structures might be helpful as other trx-G proteins also contain SET-domains. Perhaps there is a specific sequence within the SET-domain that permits a protein to act as both repressor and activator or specifies it as one or the other.

The observation that a mutation in *CLF* affects the vascular cambium suggests a broader role for Pc-G–mediated regulation of gene expression in developmental processes. As suggested by several groups (Kohler & Grossniklaus, 2002; Reyes & Grossniklaus, 2003), Pc-Gs play a role at transitional stages in plant development, maintain repression of flowering genes in vegetative tissue, regulate flowering time, play a role in floral morphology, effect the transition from primary to secondary vascular development, effect seed development and germination. These transitions and the role of various Pc-G proteins in regulating gene expression at these transitions are presented in Figure 4.0. In *Drosophila*, Pc-G proteins maintain expression states that are established during embryonic development and through this mechanism fix developmental decisions. Pc-G-mediated repression is established when segmental differentiation starts and once established is maintained throughout the life of the fly. Further, in *Drosophila,* genetic and biochemical data has demonstrated that there are two functionally distinct Pc-G protein complexes, PRC1 and PRC2 and many of the mechanistic details of Pc-G mediated repression in *Drosophila* have been determined.

In contrast, in Arabidopsis, proteins with similarities to PRC1 components have not been identified. Only proteins with similarity to members of the PRC2 complex have been identified, and the mechanistic details of Pc-G mediated repression remain unclear. Further, it has been argued that a novel difference between Pc-G proteins in *Drosophila* and Arabidopsis is that Pc-G-mediated repression in Arabidopsis is reversible (Reyes & Grossniklaus, 2003). Reyes and Grossniklaus argue that Pc-G proteins repress developmental pathways that would be active by default in the absence of Pc-G proteins, such as the endosperm proliferation observed in *fis* mutants. Many Pc-G proteins may act early in development and as a consequence their repression must be released in response to stimuli that allow progression through the life cycle. For example, *AG*, the target of CLF-repression, is repressed in vegetative tissue. With the induction of the floral meristem, this repression is reversed and *AG* is expressed in the stamens and carpels of flowers. This major difference in the activity of *Drosophila* and Arabidopsis Pc-G proteins suggests that Pc-G-mediated repression may be the default state and that it has to be overcome by some unknown mechanism.

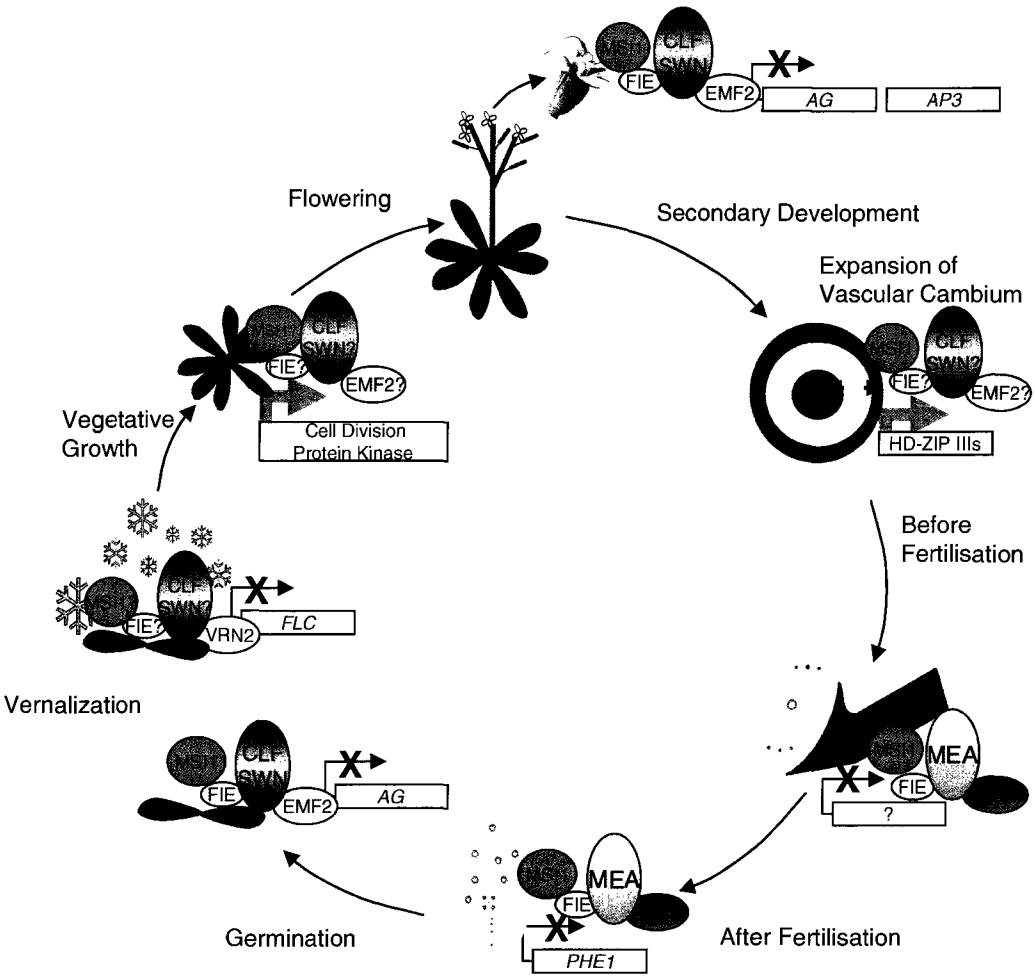
4.3 Future Directions

In light of the finding described in this thesis that CLF is a novel regulatory gene of the vascular cambium, two specific experiments are underway. The first one is a microarray analysis of the transcript profiles of Enkheim wild-type plants and *clf-1* mutants. In this microarray, total RNA profiles of leaves and stems in wildtype and mutant plants will be compared. These two separate tissues were chosen to determine if the affected genes would be the same in both or if a tissue-specific profile would be generated. Microarrays are an effective strategy for the identification of novel genes. For example, Zhao and colleagues

have used this approach to identify several novel genes with tissue-specific expression in secondary vascular tissues (Zhao, *et al.*, 2005). Further, this approach has been used to identify the target of Pc-G-mediated repression in the *fis* mutants *mea*, *fie* and *fis2*. Each of these mutants was shown to exhibit over-expression of a novel type I MADS-box gene, *PHE1*, previously unknown to play a role in seed development (Köhler, *et al.*, 2003). This approach has also been used in *emf2* mutants, but direct targets have not yet been identified (Schubert, *et al.*, 2005).

The unique finding that a Pc-G protein, CLF, plays a role in vascular development provides greater insight regarding wood development. This finding will be applied to poplar. A poplar homologue of *CLF* has been identified. The poplar *CLF* will be knocked-out and the effect on the vascular cambium observed. Further, entire transcript profile analysis will be used to identify other affected genes in the knock-out poplar. An advantage of using poplar is that it is relatively easy to separate the xylem from the phloem-cambium. Thus, the effect of down-regulation of *CLF* will be examined specifically in these tissues and compared to generate tissue-specific profiles similar to those generated by Zhao and colleagues for wood-forming tissues in Arabidopsis (Zhao, *et al.*, 2005). Identification and examination of Pc-G protein function in trees will provide additional insight into both the function of Pc-G proteins in plants, their impact on wood development and their target genes that are key genetic players in wood development.

Figure 4.0. Proposed roles for the Pc-G protein complexes throughout the Arabidopsis lifecycle. In Arabidopsis flowers the putative Polycomb Repressive Complex 2 (PRC2) consisting of EMF2, FIE, CLF, SWN and MSI1 regulate flower organ development by repressing *AG* and *AP3*. In the secondarily vasculature of the plant, the CLF protein, possibly complexed with the same PRC2 components, may act as a positive regulator of the HD-ZIP III vasculature genes, *ATHB8, ATHB15, PHB*, and *REV* as well as two secondary-vasculature specific genes, *Kan2*, and *GNOM/EMB30*. In the female gametophyte, MEA, FIE, FIS2 and MSI1 inhibit precocious proliferation of the central cell by repressing an unknown target gene. These same proteins are postulated to act together in seed development as well, specifically by regulating embryo and endosperm proliferation. At the seedling stage, CLF, FIE and EMF2 repress the switch from vegetative to generative development. Mutants in all three genes express *AG* at the seedling stage and demonstrate precocious flowering. VRN2 may complex with FIE, MSI1, CLF and SWN to repress the floral inhibitor *FLC* following vernalization to regulate flowering time. Finally, in the vegetative stage of development, CLF may complex with the PRC2 components to maintain repression of *AG* in vegetative tissue and may also maintain activation of a cell division protein kinase gene (At1g18040) to maintain cell division in the leaves.



LITERATURE CITED

Ahuja MR (2001) Recent advances in molecular genetics of forest trees. Euphytica **121:** 173-95

An G, Ebert P, Mitra A, Ha S (1988) Plant Molecular Biology Manual. **A3:** 1-19

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in Arabidopsis. Science **258:** 1353-4

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1997) Current Protocols in Molecular Biology. **1:** 1.1 – 1.10.

Baima S, Nobili F, Sessa G, Lucchetti S, Ruberti I, Morelli G (1995) The expression of the *Athb-8* homeobox gene is restricted to provascular cells in *Arabidopsis thaliana.* Development **121:** 4171-82

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DM, Benfey PN (2003) A gene expression map of the Arabidopsis root. Science **302:** 1956-60

Busse JS, Ray FE (1999) Vascular differentiation and transition in the seedling of *Arabidopsis thaliana* (Brassicaceae). Int J Plant Sci **160:** 241-51

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science **298:** 1039-43

Carles CC, Fletcher JC (2003) Shoot apical meristem maintenance: the art of a dynamic balance. Trends Plant Sci **8:** 394-401

Carlsbecker A, Helariutta Y (2005) Phloem and xylem specification: pieces of the puzzle emerge. Curr Opin Plant Biol **8:** 512-7

Chaffey N (2002) Why is there so little research into the cell biology of the secondary vascular system of trees? New Phytologist **153:** 213-23

Chaffey N, Cholewa E, Regan S, Sundberg B (2002) Secondary xylem development in Arabidopsis: a model for wood formation. Physiol Plant **114:** 594-600

Chang C, Bowman JL, DeJohn AW, Lander ES, Meyerowitz EM (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana.* Proc Natl Acad Sci **85:** 6856-60

Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon Y, Sung Z, Goodrich J (2004) Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development **131:** 5263-76

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J **16:** 735-43

Confalonieri M, Balestrazzi A, Bisoffi S, Carbonera D (2003) *In vitro* culture and genetic engineering of *Populus* spp.: synergy for forest tree improvement. Plant Cell Tiss Org Cult **72:** 109-38

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V (2002) *Drosophila* Enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell **111:** 185-96

Floyd SK, Bowman JL (2004) Gene regulation: Ancient microRNA target sequences in plants. Nature **428:** 485-86

Francis N, Kingston R, Woodcock C (2004) Chromatin compaction by a Polycomb-group protein complex. Science **26:** 1574-77

Fukuda H (2004) Signals that control plant vascular cell differentiation. Nat Rev Mol Cell Biol **5:** 379-91

Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F, Goodman HM (1992) Isolation of the Arabidopsis *ABI3* gene by positional cloning. Plant Cell **4:** 1251-61

Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G (1997) A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature **386:** 44-51

Goodrich J, Tweedie S (2002) Remembrance of things past: Chromatin remodeling in plant development. Annu Rev Cell Dev Biol **18:** 707-46

Griffiths, A. J. F., Miller JH, Suzuki DT, Lewontin RC, Gelbart WM (2000) An introduction to genetic analysis., Ed 7th. W. H. Freeman, New York, N.Y.

Han K, Meilan R, Ma C, Strauss SH (2000) An *Agrobacterium tumefaciens* protocol effective on a variety of cottonwood hybrids (genus Populus). Plant Cell Reports **19:** 315-20

Han S (2004) The growth of the vascular meristem is negatively regulated by CLV1 in Arabidopsis. MSc thesis. Ottawa-Carleton Institute of Biology, Ottawa, ON.

Hennig L, Taranto P, Walser M, Schönrock N, Gruissem W (2003) Arabidopsis *MSI1* is required for epigenetic maintenance of reproductive development. Development **130:** 2555-65

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Bhalerao R, Uhlen M, Teeri TT, Lundeberg J, Sundberg B, Nilsson P, Sandberg G (2001) A transcriptional roadmap to wood formation. Proc Natl Acad Sci **98:** 14732-7

Hoecker U, Vasil IK, McCarty DR (1995) Integrated control of seed maturation and germination programs by activator and repressor functions of *VIVIPAROUS-1* of maize. Genes Dev **9:** 2459-69

Hsieh TF, Hakim O, Ohad N, Fischer RL (2003) From flour to flower: how Polycomb group proteins influence multiple aspects of plant development. Trends Plant Sci **8:** 439-45

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang V (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotech **17:** 808-12

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. Plant Physiol **129:** 440-50

Jander G, Baerson SR, Hudak JA, Gonzalez KA, Gruys KJ, Last RL (2003) Ethyl methane sulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant Physiol **131:** 139-46

Joint Genome Institute (2005) JGI P. trichocarpa v1.0 Home. **2005:** 11

Katz A, MOran O, Mosquna A, Hakim O, Ohad N (2004) FIE and CURLY LEAF Polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. Plant J **37:** 707-19

Kim G, Tsukaya H, Uchimiya H (1998) The *CURLY LEAF* gene controls both division and elongation of cells during expansion of the leaf blade in *Arabidopsis thaliana.* Planta **206:** 175-83

Kirst M, Johnson AF, Baucom C, Ulrich E, Hubbard K, Staggs R, Paule C, Retzel E, Whetten R, Sederoff R (2003) Apparent homology of expressed genes from wood-forming tissues of loblolly pine (*Pinus taeda* L.) with *Arabidopsis thaliana.* PNAS **100:** 7383-88

Köhler C, Hennig L, Spillane C, Pien S, Gruissem W, Grossniklaus U (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1.* Genes Dev **17:** 1540-53

Kohler C, Grossniklaus U (2002) Epigenetic inheritance of expression states in plant development: the role of Polycomb-group proteins. Curr Opin Plant Bio **14:** 773-9

Koorneef M, van Eden J, Hanhart CJ, Stam P, Braaksma FJ, Feenstra WJ (1983) Linkage map of Arabidopsis thaliana. J Hered **74:** 265-72

LaJeunesse D, Shearn A (1996) E(Z): A Polycomb group gene or a trithorax group gene? Development **122:** 2189-97

Levine SS, King IF, Kingston RE (2004) Division of labor in Polycomb group repression. Trends Biochem Sci **29:** 478-85

Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana.* Plant J **4:** 745-50

Lukowitz W, Gillmor SC, Scheible W (2000) Positional Cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiol **123:** 795-806

Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW (2005) Organ-Specific expression of Arabidopsis genome during development. Plant Physiol **138:** 80--91

Marx J (2005) Combing over the Polycomb group proteins. Science **308:** 624-6

Meinke DW, Meinke LK, Showalter TC, Schissel AM, Mueller LA, Tzafrir I (2003) A sequence-based map of Arabidopsis genes with mutant phenotypes. Plant Physiol **131:** 409-18

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W (2001) Unravelling cell wall formation in the woody dicot stem. Plant Mol Biol **47:** 239-74

Meyerowitz EM (2001) Prehistory and history of Arabidopsis research. Plant Physiol **125:** 15-9

Miao Y, Laun T, Zimmermann P, Zentgraf U (2004) Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. Plant Mol Biol **55:** 853-67

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci **88:** 9828-32

Mizukami Y, Ma H (1992) Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic Arabidopsis plants alters floral organ identity. Cell **71:** 119-31

Nieminen K, Kauppinen L, Helariutta Y (2004) A weed for wood? Arabidopsis as a genetic model for xylem development. Plant Physiol **135:** 653-59

Oh S, Park S, Han K (2003) Transcriptional regulation of secondary growth in *Arabidopsis thaliana.* J Exp Bot **54:** 2709-22

Østergaard L, Yanofsky M (2004) Establishing gene function by mutagenesis in *Arabidopsis thaliana.* The Plant Journal **39:** 682-96

Peters JL, Constandt H, Neyt P, Cnops G, Zethof J, Zabeau M, Gerats T (2001) A physical amplified fragment-length polymorphism map of Arabidopsis. Plant Physiol **127:** 1579-89

Reyes JC, Grossniklaus U (2003) Diverse functions of Polycomb- group proteins during plant development. Semin Cell Dev Biol **14:** 77-84

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

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet **37:** 501-06

Schrader J, Nilssona J, Mellerowicza E, Berglundb A, Nilssonc P, Hertzberg M, Sandberg G (2004) A high-resolution transcript profile across the wood-forming meristem of Poplar identifies potential regulators of cambial stem cell identity. Plant Cell **16:** 2278-92

Schubert D, Clarenz O, Goodrich J (2005) Epigenetic control of plant development by Polycomb-group proteins. Curr Opin Plant Bio **8:** 553-61

Serrano-Cartagena J, Candela H, Robles P, Ponce MR, Perez-Perez JM, Piqueras P, Micol JL (2000) Genetic analysis of *incurvata* mutants reveals three independent genetic operations at work in Arabidopsis leaf morphogenesis. Genetics **156:** 1363-77

Sharma VK, Carles C, Fletcher JC (2003) Maintenance of stem cell populations in plants. Proc Natl Acad Sci **100:** 11823-29

Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. Plant Cell **14:** 2527-37

Sieburth LE, Meyerowitz EM (1997) Molecular dissection of the *AGAMOUS* control region shows that *cis* elements for spatial regulation are located intragenically. Plant Cell **9:** 355-65

Smyth D, Bowman J, Meyerowitz E (1990) Early flower development in Arabidopsis. Plant Cell **2:** 755-67

Spellman PT, Rubin GM (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. J Biology **1:** 5-10

Statistics Canada (2003) Statistics Canada. **2005:** 11

Steimer A, Schob H, Grossniklaus U (2004) Epigenetic control of plant development: new layers of complexity. Curr Opin Plant Bio **7:** 11-19

Strauss SH, Martin FM (2004) Poplar genomics comes of age. New Phytol **164:** 1-4

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

Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The Botany Array Resource: e-Northerns, expression angling, and promoter analyses. Plant J **43:** 153-63

Weigel D, Glazebrook J (2002) Arabidopsis: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Yanagisawa S (2004) Dof domain proteins: Plant-specific transcription factors associated with diverse phenomena unique to plants. Plant Cell Physiol **45:** 386-91

Ye ZH, Freshour G, Hahn MG, Burk DH, Zhong R (2002) Vascular development in Arabidopsis. Int Rev Cytol **220:** 225-56

Zhao C, Craig JC, Petzold HE, Dickerman AW, Beers EP (2005) The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-hypocotyl. Plant Physiol **138:** 803-18

Zhao C, Johnson BJ, Kositsup B, Beers EP (2000) Exploiting secondary growth in Arabidopsis. Construction of xylem and bark cDNA libraries and cloning of three xylem endopeptidases. Plant Physiol **123:** 1185-96